Nonparametric statistics for scalar ergodic diffusion processes

Inauguraldissertation
zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
der Universität Mannheim

vorgelegt von

Cathrine Aeckerle-Willems
aus Saarlouis

Mannheim, 2019

Dekan: Dr. Bernd Lübcke, Universität Mannheim
Referent: Prof. Dr. Claudia Strauch, Universität Mannheim
Korreferent: Prof. Dr. Enno Mammen, Universität Heidelberg
Korreferent: Prof. Dr. Mark Podolskij, Aarhus University

Tag der mündlichen Prüfung: 10. Dezember 2019

Acknowledgments

Mein größter Dank gilt meiner Betreuerin Frau Professor Dr. Claudia Strauch. Ihre Begeisterung und Leidenschaft für die Forschung waren von Beginn an mitreißend und die wichtigste Grundlage für die Erstellung dieser Arbeit. Ich danke ihr zutiefst für die hervorragende fachliche Begleitung, die eine sprudelnde Quelle für Motivation und Inspiration war. Sie ließ mich an ihrem Wissen teilhaben und ich konnte so die Zeit bestmöglich nutzen, um viel zu lernen. Neben der Betreuung dieser Arbeit war Frau Professor Dr. Claudia Strauch eine allzeit ansprechbare Mentorin, die mich in die Welt der Wissenschaft einführte und Perspektiven aufzeigte, die mir Mut machen, den Weg in der Wissenschaft weiter zu verfolgen.

Mein besonderer Dank gilt auch meinem Zweitbetreuer und -gutachter Herrn Professor Dr. Enno Mammen, der mir initial die Möglichkeit eröffnete, mein Promotionsvorhaben zu beginnen. Seine Bereitschaft, sein großes Wissen und seinen Erfahrungsschatz zu teilen sowie stets für Rat und Tat verfügbar zu sein, weiß ich sehr zu schätzen.

Ich möchte mich zudem bei Herrn Professor Dr. Mark Podolskij für seine Bereitschaft, diese Dissertation zu begutachten, bedanken.

Herrn Professor Dr. Richard Nickl danke ich für sein Interesse an meiner Arbeit sowie insbesondere für die Denkanstöße und hilfreichen Diskussionen, die zur Entstehung eines Teiles dieser Dissertation führten. Zudem war seine Einladung an die University of Cambridge und die Gelegenheit, meine Forschung dort vorzustellen, eine tolle Erfahrung.

Ich war Mitglied des Graduiertenkollegs 1953 der Universitäten Mannheim und Heidelberg und bedanke mich bei allen Beteiligten für die finanzielle Unterstützung und den akademischen Rahmen mit vielen Workshops, Konferenzen und dem Raum für fachlichen Austausch. Ich habe dort nette Kolleginnen und Kollegen gefunden und möchte insbesondere Alexander Kreiß und Lena Reichmann erwähnen. Mit beiden hatte ich viel Freude und sie waren jederzeit sehr hilfsbereit.

Während der gesamten Zeit war mir mein Mann, Marc Willems, eine große Stütze und dafür danke ich ihm von ganzem Herzen. Durch seine tatkräftige und mentale Unterstützung konnte ich die Zeit, in der unsere Tochter Malia in unser Leben getreten ist, genießen und zugleich meine Forschung fortführen. Tausend Dank gilt auch meiner Familie, die meinen Weg begleitet, ermöglicht und dabei stets an mich geglaubt hat.

Zusammenfassung

Diese Dissertation widmet sich zweierlei Zielen, die ein statistisches Problem und seine probabilistischen Grundlagen betreffen. Forschungsgegenstand ist die Schätzung des Drifts und der invarianten Dichte für eine große Klasse skalarer, ergodischer Diffusionsprozesse, basierend auf stetigen Beobachtungen, unter Berücksichtigung des Verlustes in Supremumsnorm. Es werden Konzentrationsungleichungen und Momentabschätzungen für Analoga von klassischen empirischen Prozessen in stetiger Zeit bereitgestellt. Diese stellen das zentrale probabilistische Hilfsmittel für die Analyse des Schätzrisikos in sup-Norm dar.

Es wird angenommen, der unbekannte Drift gehöre zu einer nichtparametrischen Klasse glatter Funktionen unbekannter Ordnung. Ein adaptiver Ansatz zur Konstruktion datengetriebener Driftschätzer, die minimax optimale Konvergenzraten in sup-Norm erzielen, wird vorgeschlagen. Außerdem wird ein Donsker-Theorem für den klassischen Kernschätzer der invarianten Dichte und dessen semiparametrische Effizienz gezeigt. Schließlich werden beide Resultate zusammengeführt und ein rein datengetriebener Selektionsmechanismus für die Bandweite entwickelt, der simultan sowohl einen ratenoptimalen Driftschätzer als auch einen asymptotisch effizienten Schätzer für die invariante Dichte liefert.

Essentielle Werkzeuge für die Untersuchung sind uniforme Exponentialungleichungen für empirische Prozesse und verwandte stochastische Integrale skalarer, ergodischer Diffusionsprozesse. Mit der Entwicklung dieser probabilistischen Hilfsmittel wird die Grundlage, welche üblicherweise für die Untersuchung der sup-Norm-Eigenschaften von Schätzverfahren für eine reiche Klasse von Diffusionsprozessen benötigt wird, geschaffen. Die Idee hat ihren Ursprung im klassischen i.i.d. Kontext, in dem Konzentrationsungleichungen vom Talagrand-Typus ein Schlüssel für die statistische Analyse in sup-Norm sind. Mit der Zielsetzung einen entsprechenden Ersatz im Diffusionskontext zu entwickeln, wird ein systematischer, in sich geschlossener Ansatz für solche uniformen Konzentrationsungleichungen präsentiert. Dieser beruht auf einer Martingalapproximation und Momentabschätzungen, die durch Anwendung der Generic Chaining-Methode hergeleitet werden.

Abstract

This thesis is directed towards a twofold aim concerning a statistical problem and its probabilistic foundations. We consider the question of estimating the drift and the invariant density for a large class of scalar, ergodic diffusion processes based on continuous observations in supremum-norm loss. Concentration inequalities and moment bounds for continuous time analogues of classical empirical processes driven by diffusions are provided. These serve as the central probabilistic device for the statistical analysis of the sup-norm risk.

The unknown drift is supposed to belong to a nonparametric class of smooth functions of unknown order. We suggest an adaptive approach which allows to construct data-driven drift estimators attaining minimax optimal sup-norm rates of convergence. In addition, we prove a Donsker theorem for the classical kernel estimator of the invariant density and establish its semiparametric efficiency. Finally, both results are combined to propose a fully data-driven bandwidth selection procedure which simultaneously yields both a rate-optimal drift estimator and an asymptotically efficient estimator of the invariant density of the diffusion.

Crucial tool for our investigation are uniform exponential inequalities for empirical processes and related stochastic integrals driven by scalar, ergodic diffusion processes. Providing these probabilistic tools, we lay the foundation typically required for the study of sup-norm properties of estimation procedures for a large class of diffusion processes. The idea originates in the classical i.i.d. context where Talagrand-type concentration inequalities are a key device for the statistical sup-norm analysis. Aiming for a parallel substitute in the diffusion framework, we present a systematic, self-contained approach to such uniform concentration inequalities via martingale approximation and moment bounds obtained by the generic chaining method.

Contents

1 Introduction

Nonparametric statistics for continuous time processes is a recent and active field of research. Diffusion processes constitute a large class of these stochastic processes exhibiting a Markovian structure and continuity of trajectories. They have gained increasing attention in mathematical statistics during the last decades. The book Kutoyants (2004) provides an overview of inference for scalar diffusion processes based on continuous observations. Spokoiny (2000) and Dalalyan (2005) establish adaptive, i.e., data-driven methods for pointwise and integrated risk measures. Multivariate models are investigated in Dalalyan and Reiß (2007), Strauch (2015) and Strauch (2016). Nonparametric estimation from discrete data is of interest, as well (see Hoffmann (1999), Gobet *et al.* (2004), Söhl and Trabs (2016)), and the recent literature also comprises Bayesian approaches for diffusion models (e.g., van der Meulen *et al.* (2006), van Waaij and van Zanten (2016), Nickl and Ray (2018)). For a more detailed literature review, we refer to the introduction of Chapter 4.

The statistical analysis of diffusion processes differs from parallel investigations of classical models such as independent and identically distributed (i.i.d.) observations. This is mainly due to the very distinct underlying probabilistic structure characterised by a certain kind of dynamics under dependency governed by a stochastic differential equation. Beyond the theoretical interest, the need for statistical investigations based on recent nonparametric, data-driven methods becomes apparent in view of the broad practical relevance of diffusion process models. There are numerous applications of these models in various areas of science among them applications in mathematical finance, physics, and genetics, just to name very few relevant fields. Popular examples within these fields are diffusion models for stock prices, the Ornstein-Uhlenbeck process, or the Wright-Fisher model in population genetics.

1.1 Sup-norm adaptive estimation of diffusion characteristics in dimension one

The diffusion processes under consideration are given as solutions $X = (X_t)_{t\geq 0}$ of a stochastic differential equation of the form

$$\mathrm{d}X_t \;=\; b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t,\; t \geq 0, \tag{1.1.1}$$

where $W = (W_t)_{t\geq 0}$ is a standard Brownian motion. Assuming that we can observe the whole trajectory of the process for a finite period of time, there exists a continuous record of observations $(X_s)_{s=0}^t,\ t > 0$. The maps $b : \mathbb{R} \to \mathbb{R}$ and $\sigma^2 : \mathbb{R} \to \mathbb{R}$ denote the drift coefficient and the diffusion coefficient, respectively. A diffusion process can be interpreted as a continuous process that follows some trend and at the same time is subject to a certain volatility around this trend. The latter is represented by the integrated drift term whereas the volatility is induced by the stochastic integral which involves the diffusion coefficient. From a continuous record of observations σ^2 is identifiable as far as possible which is due to the convergence property

$$\sum_{i=0}^{n-1}(X_{t_{i+1}} - X_{t_i})^2 \longrightarrow \int_0^t \sigma^2(X_s)\mathrm{d}s, \quad \text{as } \Delta \to 0, \quad \text{almost surely,}$$

where $0 = t_0 < \dots < t_n = t$, $\Delta := \max_{i=0,\dots,n-1} |t_{i+1} - t_i|$. Therefore, the focus is on estimation of the drift coefficient b.

We assume that the diffusion process has ergodic properties. In this case, the process admits an invariant measure μ_b with invariant density ρ_b given via the relation

$$\rho_b(x) \;=\; \frac{1}{C_{b,\sigma}\sigma^2(x)} \;\exp\left(\int_0^x \frac{2b(y)}{\sigma^2(y)}\mathrm{d}y\right), \quad x \in \mathbb{R},$$

with

$$C_{b,\sigma} := \int_{\mathbb{R}} \frac{1}{\sigma^2(u)} \;\exp\left(\int_0^u \frac{2b(y)}{\sigma^2(y)}\mathrm{d}y\right)\mathrm{d}u$$

denoting the normalising constant. This explicit relation between the invariant density and the drift coefficient accounts for the strong connectivity of the problems of drift estimation and estimation of the invariant density. Our aim is suggesting purely data-driven estimators for these characteristics, which behave optimally in a sense that will be explained. To this intent, we propose using kernel-type estimators. For determining the invariant density ρ_b, we adopt a continuous time analogue of the classical kernel density estimator defined as

$$\rho_{t,K}(h)(x) \;:=\; \frac{1}{th}\int_0^t K\left(\frac{x - X_u}{h}\right)\mathrm{d}u, \quad x \in \mathbb{R}, \tag{1.1.2}$$

for a positive bandwidth h and some smooth kernel function $K\colon \mathbb{R} \to \mathbb{R}$ with compact support. A natural estimator of the drift coefficient b, which relies on the analogy between the drift estimation problem and the model of regression with random design, is given by a Nadaraya–Watson-type estimator of the form

$$b_{t,K}(h)(x) := \frac{\overline{\rho}_{t,K}(h)(x)}{\rho^+_{t,K}(h)(x) + r_t}, \quad \text{where} \tag{1.1.3}$$

$$\overline{\rho}_{t,K}(h)(x) := \frac{1}{th}\int_0^t K\left(\frac{x - X_s}{h}\right)\mathrm{d}X_s, \quad \rho^+_{t,K}(h)(x) := \max\{0, \rho_{t,K}(h)(x)\}, \tag{1.1.4}$$

and $(r_t)_{t\geq 0}$ is a suitable, strictly positive sequence such that $\lim_{t\to\infty} r_t = 0$.

As a measure for the quality of the estimation procedures, we consider the expected maximal error, i.e., for an estimator $\widehat{b}$ of b we analyse the risk $\mathbb{E}_b\left[\|\widehat{b} - b\|_\infty\right]$ where

$\|f\|_\infty := \sup_{x\in\mathbb{R}} |f(x)|$, for any function $f : \mathbb{R} \to \mathbb{R}$. This perspective is a worst-case consideration which is highly relevant in practical applications for several reasons. Beyond the obvious desire to analyse worst cases in the presence of uncertainty, measuring the quality of an estimator in sup-norm risk is insightful because it takes into account the performance for estimation of the whole function instead of assessing the risk only pointwise. Furthermore, it has a straightforward interpretation in contrast to other global risk measures such as the mean integrated squared risk.

Parametric statistics rely on the fundamental assumption that the target can be described by a finite number of parameters which means that the search for a good estimator is restricted to a finite dimensional set. In contrast, nonparametric procedures aim at finding an estimator for a target that is contained in an infinite dimensional function class imposing some regularity on the target.

Nonparametric estimators typically depend on a tuning parameter such as the bandwidth for kernel estimators. The choice of the tuning parameter has a significant impact on the quality of the procedure, for instance in terms of convergence rates for the risk which tends to zero for a reasonable estimator. The convergence rate determines how fast the risk vanishes. Another phenomenon frequently occurring is the dependence of these rates on the assumed order of regularity. For estimation of the drift coefficient in sup-norm loss we obtain the rate $\left(\frac{\log t}{t}\right)^{\frac{\beta}{2\beta+1}}$ where $\beta > 0$ denotes the order of regularity. Loosely speaking, in our context, a bandwidth choice is considered to be a good one if it yields an estimator which achieves this fast convergence rate. Such a bandwidth can be obtained from balancing the opposite effects of the bias and the variance of the estimator and is given as $h^* \simeq \left(\frac{\log t}{t}\right)^{\frac{1}{2\beta+1}}$. However, this choice again depends on the assumed regularity β. Hence, the associated estimator is not data-driven and requires the a priori knowledge of the regularity. Our objective are adaptive methods that can overcome this circumstance. An adaptive estimator with the associated adaptive bandwidth choice is supposed to be purely data-driven and its risk is supposed to vanish as fast as possible, ideally as it would if the regularity was known in advance. This means that it achieves or adapts to the optimal convergence rate which depends on the unknown regularity.

Beyond convergence rates for the risk of an estimator, it is a natural aspiration to investigate the asymptotic distribution of an estimator. This is an important first step for testing or for the construction of confidence bands. We analyse the asymptotic distribution of the kernel estimator for the invariant density in a functional sense. This means that we do not only investigate the asymptotic distribution pointwise. Instead, we show a Donsker-type result which is a weak convergence result for the estimator as a random map in the space of bounded functions. Such a result is charming since it immediately yields asymptotic convergence results for continuous functionals of the invariant density.

Another goal of this thesis is to show that the performance of the suggested estimation procedures for the drift coefficient and the invariant density of a scalar diffusion process is optimal. Corresponding to the different nature of the results we look at two different optimality concepts. The first one is minimax optimality which refers to the optimality of the convergence rate for the sup-norm risk of the drift estimator. Secondly, semiparametric efficiency of the invariant density estimator is addressed. We deal with the question what an optimal limit distribution is and show that it is achieved by our estimator.

1.2 Concentration of empirical processes driven by diffusion processes

Beyond the statistical question, this thesis is also concerned with more probabilistic problems that naturally arise from the statistics. The investigation of the sup-norm risk requires deep probabilistic results from the theory of empirical processes in the continuous diffusion framework. The probabilistic toolbox that we aim for contains moment bounds and concentration inequalities for suprema of additive functionals and stochastic integrals over infinite dimensional function classes. These results parallel Talagrand-type inequalities known from the classical i.i.d. context. Our statistical investigation based on continuous observations exploits the probabilistic features of diffusion processes and aims at results which reflect the very nature of these processes. Reducing the problem to a discrete or even i.i.d. context in order to use available Talagrand-type inequalities would lead to a loss of information on the probabilistic structure to a large extent which contradicts our objective. Therefore, we develop the probabilistic tools within our diffusion context proposing at the same time a general machinery for the derivation of concentration inequalities for empirical processes driven by other processes, as well. Our results constitute the probabilistic foundation for solving the statistical problem treated in this thesis, and moreover they play an essential role for other recent questions in statistics for diffusion processes. Furthermore, they are clearly of independent probabilistic interest. The analysis of parametric, high-dimensional diffusion models as well as the nonparametric, adaptive estimation in sup-norm within discrete observation schemes or based on diffusion models with multivariate state variables are examples for other relevant statistical problems.

Given real-valued i.i.d. observations $X_1, ..., X_n \sim P$, $n \in \mathbb{N}$, the law of large numbers motivates the estimation of expected values $P(f) := \int f \mathrm{d}P$ where $f : \mathbb{R} \to \mathbb{R}$, $f \in L^1(P)$, by sample means $\mathbb{P}_n(f) := \frac{1}{n}\sum_{i=1}^n f(X_i)$, i.e., by additive functionals of the observations. In case the interest is not only in exploring the behaviour of this estimator with respect to one specific function f but with respect to a whole class of functions $\mathcal{F}$, the need to investigate the empirical process $(G_n(f))_{f\in\mathcal{F}}$ arises. It is defined as $G_n(f) := \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n f(X_i) - P(f)\right)$ and $\mathcal{F}$ denotes a possibly infinite dimensional class of functions. Typical questions then address consistency, i.e., $\sup_{f\in\mathcal{F}} n^{-1/2}|G_n(f)| \to 0$, in mean, as $n \to \infty$, or concentration inequalities of the form

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}} |G_n(f)| \geq \phi(u)\right) \leq \exp(-u), \quad u > 0, \tag{1.2.5}$$

for a function $\phi : \mathbb{R}_+ \to \mathbb{R}_+$. A famous result in that spirit which is the central ingredient for a number of statistical sup-norm investigations is a concentration inequality by Talagrand (1996b). We cite a version by Bousquet (2003) as follows. It refers to a countable class of real-valued functions $\mathcal{F}$ such that $\mathcal{F}$ is uniformly bounded by a positive constant U and such that $\sup_{f\in\mathcal{F}} P(f^2) \leq \nu^2$ for another positive constant ν^2. Then, setting $V := n\nu^2 + 2U\mathbb{E}\left[\sup_{f\in\mathcal{F}} \left|\sum_{i=1}^n f(X_i)\right|\right]$ it holds, for any $x \geq 0$,

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n f(X_i)\right| \geq \mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n f(X_i)\right|\right] + \sqrt{2Vx} + Ux/3\right) \leq 2\exp(-x). \tag{1.2.6}$$

Studying ergodic diffusion processes, there exists a law of large numbers and apparently, it is natural to look at empirical processes in this context, as well. Beyond this very straightforward motivation, our specific statistical problem of drift estimation requires developing Talagrand-type concentration inequalities for continuous time analogues of empirical processes driven by diffusion processes.

The additive functional analogue is given via

$$\mathbb{G}_t(f) := \sqrt{t}\left(\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s - \mu_b(f)\right) \tag{1.2.7}$$

where $(X_s)_{s\geq 0}$ is a scalar, ergodic diffusion process given as a solution of (1.1.1) and $\mu_b(g) := \int g\mathrm{d}\mu_b$, for any $g : \mathbb{R} \to \mathbb{R}$, $g \in L^1(\mu_b)$. Moreover, we analyse the stochastic integrals

$$\mathbb{H}_t(f) := \sqrt{t}\left(\frac{1}{t}\int_0^t f(X_s)\mathrm{d}X_s - \mu_b(b\cdot f)\right). \tag{1.2.8}$$

Our approach targets upper bounds for all moments $\left[\mathbb{E}\sup_{f\in\mathcal{F}}|\mathbb{G}_t(f)|^p\right]^{\frac{1}{p}}$ and $\left[\mathbb{E}\sup_{f\in\mathcal{F}}|\mathbb{H}_t(f)|^p\right]^{\frac{1}{p}}$ specifying very detailed the explicit dependence on p and other constants involved which is crucial for the statistical applications. From these upper bounds, (nonadaptive) convergence rates for the sup-norm risk of estimators for diffusion characteristics can be deduced immediately. Charmingly, another direct implication are the corresponding concentration inequalities of the form (1.2.5) which are required for the adaptive procedure in this thesis.

These results can be shown under typical assumptions on the class of functions $\mathcal{F}$ which are the same as in the i.i.d. context. Moreover, we even allow for a certain unboundedness which is remarkable since results for additive functionals of unbounded functions are scarce even for the nonuniform consideration.

The investigation of stochastic integrals as in (1.2.8) relies on an analysis of the local time $(L_t^x(X), x \in \mathbb{R}, t \geq 0)$ associated to the diffusion process X. Roughly speaking, the local time at t and x measures how often x is visited by the process X until time t. The scaled local time serves as a natural estimator for the invariant density if a continuous record of observations is available.

Remarkably, the local time is a phenomenon only existing in dimension one which illustrates that the scalar case is very different from the multivariate setting and deserves special attention. As a consequence, the statistical procedures for scalar diffusions exhibit an exceptional behaviour which cannot immediately be deduced from a general multivariate analysis.

1.3 Outline of the thesis and main contributions

In this section, the structure of the thesis is outlined including a concise summary of the main contributions. We begin with an introduction on diffusion processes in Chapter 2 and present basic facts that are used without further notice throughout the subsequent chapters.

The main part of this thesis consists of Chapter 3 and Chapter 4 which are both presented in a self-contained manner and can be read independently of each other. These chapters are based on joint work with Prof. Dr. Claudia Strauch and correspond to a large extent to the preprints Aeckerle-Willems and Strauch (2018a) and Aeckerle-Willems and Strauch (2018b), respectively.

In Chapter 3, we examine concentration inequalities for empirical processes and related concepts driven by diffusion processes thereby laying the probabilistic foundations for the statistical investigation which is subject of Chapter 4.

We derive moment bounds for the sup-norm of the local time of a continuous semimartingale under assumptions on the moments of the martingale and the finite variation part of the process. These upper bounds are a central ingredient for the investigation of the stochastic integrals of the form (1.2.8). Interestingly, proving upper bounds on the sup-norm of the local time serves as a blueprint for the further analysis of empirical processes. In Section 3.3, we propose an approach for the derivation of moment bounds for continuous time analogues of empirical processes driven by continuous semimartingales via generic chaining and localisation based on a given martingale approximation. The machinery developed for this purpose is transferable to other classes of processes. The concentration equalities we aim for come for free, once upper bounds for all pth moments, $p \geq 1$, revealing the explicit dependence on p, are available.

We continue restricting to a broad class of diffusion processes in Section 3.4 and prove the existence of a martingale approximation for additive functionals. The general result for continuous semimartingales from the previous section then implies moment bounds and concentration inequalities for empirical processes of the form (1.2.7). Exploiting the moment bounds on the sup-norm of the local time, we can finally prove parallel results for stochastic integrals as in (1.2.8).

In Section 3.5, we turn to some first statistical applications that follow from the obtained concentration inequalities for empirical processes. Given a continuous record of observations $(X_s)_{s=0}^t$ of a diffusion process X, we analyse the kernel density estimator $\rho_{t,K}(h)$ and the local time estimator for estimation of the invariant density ρ_b and establish upper bounds on the sup-norm risks $\mathbb{E}_b\left(\|\rho_{t,K}(h) - \rho_b\|_\infty^p\right)$ and $\mathbb{E}_b\left(\|t^{-1}L_t^\bullet(X) - \rho_b\|_\infty^p\right)$, for any $p \geq 1$. Additionally, we provide an exponential inequality for the sup-norm distance between these two estimators.

In Chapter 4, we continue the statistical analysis of the kernel density estimator. Based on the results of the previous chapter, we prove a Donsker-type result for the kernel density estimator establishing its asymptotic normality in a functional sense. More precisely, the law of $\sqrt{t}(\rho_{t,K}(h) - \rho_b)$ converges to a Gaussian limit for a certain range of bandwidths $h = h_t$ where the convergence takes places in the space of bounded functions. Following the general conceptual framework proposed in van der Vaart and Wellner (1996), we establish semiparametric optimality of the limit distribution.

Section 4.4 concerns the adaptive estimation of the drift coefficient of a scalar, ergodic diffusion process. Conceptually, the results are based on the analog situation of i.i.d. observations examined by Giné and Nickl (2009) where estimation of the distribution function and the density with respect to the sup-norm risk is considered. We suggest an adaptive Nadaraya-Watson-type estimator for the drift coefficient b of a scalar, ergodic diffusion process. Providing upper and lower bounds for the sup-norm risk, we conclude that the suggested estimator behaves optimally in a minimax sense uniformly over non-parametric Hölder classes of unknown order. In particular, it turns out that in contrast to other situations such as pointwise estimation for example, there is no price for adaptation in terms of the convergence rate. Finally, we suggest another adaptive, i.e., purely data-driven, bandwidth which simultaneously yields an asymptotically normal estimator for the invariant density and at the same time a minimax rate-optimal estimator for the drift coefficient in Section 4.5.

2 An introduction to diffusion processes

The central objects of interest in this thesis are diffusion processes. The notion of a diffusion process is closely related to the theory of stochastic differential equations. In fact, a diffusion process is a solution to a certain kind of stochastic differential equation (SDE) as will be discussed shortly. These stochastic processes are widely used to model the evolution of various systems in many areas of science, for instance as models for stock prices in finance. Another example is the diffusion approximation of the Wright-Fisher or the Moran model in population genetics.

From the theoretical perspective diffusion processes constitute a special class of continuous Markov processes. There are different approaches, all of them leading to the same notion of a diffusion process. Firstly, we can describe a diffusion as a Markov process characterised by a certain type of infinitesimal generator. Such processes can secondly be obtained from a solution to a martingale problem. These martingale problems correspond to stochastic differential equations which consequently leads to another third approach already mentioned before. In this chapter, we give a short overview of basics on diffusion processes thereby setting the scene for the main part of the thesis.

2.1 Markov processes

Studying diffusions, we are working within the larger class of continuous Markov processes. Therefore, we collect some very basic facts about continuous time Markov processes in $\mathbb{R}^d$ and fix associated notation. For a detailed account, we refer to Kallenberg (1997).

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space equipped with a filtration $(\mathscr{F}_t)_{t\geq 0}$. Note that during all subsequent chapters $(\Omega, \mathscr{F}, \mathbb{P})$ will always denote the underlying probability space even if it is not explicitly mentioned.

Definition 1. An $\mathbb{R}^d$-valued stochastic process $X = (X_t)_{t\geq 0}$ is called Markov process if X is $(\mathscr{F}_t)_{t\geq 0}$ adapted and satisfies the Markov property:

$$\mathbb{P}(X_t \in B|\mathscr{F}_s) = \mathbb{P}(X_t \in B|X_s),$$

for any $0 \leq s \leq t$, and $B \in \mathcal{B}(\mathbb{R}^d)$ where $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel σ-algebra of $\mathbb{R}^d$.

The Markov kernels defined via

$$\mu_{s,t} : \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \to [0, 1], (x, B) \mapsto \mu_{s,t}(x, B) := \mathbb{P}(X_t \in B|X_s = x)$$

for $0 \leq s \leq t$ are called transition kernels. Given Markov kernels μ and ν on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, we define Markov kernels from $\mathbb{R}^d$ to $\mathbb{R}^d$ by

$$\mu\nu(x, A) \ := \ \int \nu(y, A)\mu(x, \mathrm{d}y) \quad x \in \mathbb{R}^d, A \in \mathcal{B}(\mathbb{R}^d).$$

It is straightforward to check that the transition kernels of a Markov process satisfy:

$$\mu_{s,t} = \mu_{s,u}\mu_{u,t}, \quad \text{for any } s \leq u \leq t. \tag{2.1.1}$$

Furthermore, the finite-dimensional distributions of a Markov process X with initial distribution $\mathbb{P}^{X_0} =: \mu_0$ are given via

$$\begin{aligned} &\mathbb{P}_{\mu_0}(X_{t_1} \in B_1, ..., X_{t_n} \in B_n) \qquad (2.1.2)\\ &= \int \int_{B_1} \cdots \int_{B_{n-1}} \mu_{t_{n-1},t_n}(x_{n-1}, B_n)\mu_{t_{n-2},t_{n-1}}(x_{n-2}, \mathrm{d}x_{n-1}) \cdots \mu_{t_0,t_1}(x, \mathrm{d}x_1)\mu_0(\mathrm{d}x) \end{aligned}$$

for any $0 = t_0 \leq t_1 \leq ... \leq t_n$, $B_1, ..., B_n \in \mathcal{B}(\mathbb{R}^d)$.

If X has continuous paths, these finite dimensional distributions uniquely determine the measure $\mathbb{P}_{\mu_0}$ of a continuous Markov process on the measure space $\left(C(\mathbb{R}_+, \mathbb{R}^d), \mathcal{C}\left(C(\mathbb{R}_+, \mathbb{R}^d)\right)\right)$ where $\mathcal{C}\left(C(\mathbb{R}_+, \mathbb{R}^d)\right)$ is the natural $\sigma-$algebra generated by the cylindrical sets.

Conversely, given a family of Markov kernels $(\mu_{s,t})$ that satisfy (2.1.1) and a probability measure μ_0 on $\mathcal{B}(\mathbb{R}^d)$, it can be shown that there exists an $\mathbb{R}^d$-valued Markov process X with initial distribution μ_0 and transition kernels $(\mu_{s,t})$ (see Theorem 7.4, Chapter 7 in Kallenberg (1997)).

We consider time-homogeneous Markov processes with transition kernels satisfying

$$\mu_{s,t} = \mu_{0,t-s} =: \mu_{t-s} \quad \text{for all } \ 0 \leq s < t.$$

The Markov property and (2.1.1) then read

$$\mathbb{P}(X_t \in B | \mathscr{F}_s) = \mu_{t-s}(X_s, B)$$

for any $0 \leq s < t$, $B \in \mathcal{B}(\mathbb{R}^d)$, and

$$\mu_{s+t} = \mu_s\mu_t$$

for all $s, t \in \mathbb{R}_+$, respectively. If the latter holds, a family of transition kernels is called transition semigroup.

For a probability measure μ_0, $\mathbb{P}_{\mu_0}$ denotes the distribution of a Markov process with transition kernels (μ_t) and initial distribution μ_0. Assuming that these processes have continuous versions, the measure $\mathbb{P}_{\mu_0}$ is a measure on $\mathcal{C}(C(\mathbb{R}_+, \mathbb{R}^d))$. Set $\mathbb{P}_x := \mathbb{P}_{\delta_x}$ for the dirac measure δ_x at x, $x \in \mathbb{R}^d$. Then, the measure $\mathbb{P}_{\mu_0}$ is given via

$$\mathbb{P}_{\mu_0}(A) = \int \mathbb{P}_x(A)\mu_0(\mathrm{d}x), \tag{2.1.3}$$

for all A in $\mathcal{C}(C(\mathbb{R}_+, \mathbb{R}^d))$ and any initial distribution μ_0 (see Lemma 7.7, Chapter 7 in Kallenberg (1997)). We also write $\mathbb{E}_{\mu_0}$ for the expected value with respect to the measure $\mathbb{P}_{\mu_0}$.

2.1.1 Invariant measures

Our statistical analysis of time-homogeneous diffusion processes, a special class of Markov processes, will rely on the assumption of stationarity of the process. The existence of such a stationary process is closely related to the existence of a so-called invariant measure. One of the objects we want to estimate is the density of this stationary measure with respect to the Lebesgue measure.

If μ_0 is the stationary measure, the distribution of X_0 is equal to the distribution of X_s, for any $s \geq 0$, which is given via $\mu_0 \mu_s$, so that μ_0 must be invariant under the transition kernels $(\mu_t)_{t \geq 0}$. It can thus be interpreted as an equilibrium with respect to the transition of the distribution of the process over time. For a general Markov process such a measure is called invariant measure.

Definition 2. We call a measure μ on $\mathcal{B}(\mathbb{R}^d)$ an invariant measure of the family $(\mu_t)_{t \geq 0}$ of transition kernels on $\mathbb{R}^d$ if it satisfies:

$$\mu(A) = \int \mu_t(x, A) \mathrm{d}\mu(x), \quad \text{for any } A \in \mathcal{B}(\mathbb{R}^d) \text{ and } t \geq 0.$$

If the family of transition kernels of a time-homogeneous Markov process admits an invariant measure μ_0, it is straightforward to deduce from (2.1.2) that the Markov process with initial distribution μ_0 is stationary. In particular, this applies to diffusion processes with an initial value X_0 distributed according to μ_0.

2.2 Overview of different approaches to diffusion processes

Shortly speaking, a diffusion process is a continuous Markov process characterised by a certain type of generator. Depending on context and intention different approaches to a rigorous introduction of this class of processes can be found in the literature. The varying notions entail different probabilistic techniques. Therefore, one or another approach might be beneficial for different objectives.

In this thesis, the class of scalar diffusion processes given as solutions to time-homogeneous stochastic differential equations (SDEs) are considered. References for this approach are Kallenberg (1997), Durrett (1996), or Kutoyants (2004) among others.

Definition 3. Given measurable, locally bounded functions $b : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}$ such that the stochastic differential equation

$$X_t = X_0 + \int_0^t b(X_s) \mathrm{d}s + \int_0^t \sigma(X_s) \mathrm{d}W_s, \quad t \geq 0, \tag{2.2.4}$$

where $(W_s)_{s \geq 0}$ denotes a Brownian motion, admits a weak solution $X = (X_t)_{t \geq 0}$, which is unique in law, we call X a (homogeneous) diffusion process driven by the drift coefficient b and the diffusion coefficient σ^2 with initial distribution $\mathbb{P}^{X_0}$.

More precisely, $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \geq 0}, \mathbb{P}, X, W)$ is referred to as a weak solution of (2.2.4) if W is an $(\mathscr{F}_t)$-Brownian motion and X is a continuous, $(\mathscr{F}_t)$-adapted process such that

(2.2.4) is satisfied for X and W. Existence and uniqueness of (even strong) solutions to the stochastic differential equation will be implied by our basic assumptions

(i) $b, \sigma \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R})$, (local Lipschitz-continuity)

(ii) $|b(x)| + |\sigma(x)| \le C(1+|x|)$, (at-most linear growth)

for some $C > 0$, and any $x \in \mathbb{R}$. Note that in the literature, it is common to simply define diffusion processes as solutions to (2.2.4). We include weak existence and uniqueness in law in Definition 3 because these requirements are very natural and assumptions to meet these usually have to be imposed anyway. Moreover, weak existence and uniqueness in law ensure the (strong) Markov property to hold (see Theorem 18.11 in Kallenberg (1997) or Theorem 8.6 and Corollary 8.8 in Le Gall (2016)) and therefore the definition is more in line with another approach we are about to discuss.

This approach to diffusion processes can be considered as even more classical and starts from a Markov process with a certain type of generator. The degree of generality as well as assumptions vary across the relevant literature, so it is advisable to pay some regard to the context. We give a short overview of the approach via the generator and via martingale problems as treated e.g. in Revuz and Yor (1999) or Rogers and Williams (2000) for (multivariate) diffusion processes in $\mathbb{R}^d$.

We begin with a definition based on the generator of a diffusion process in $\mathbb{R}^d$.

Definition 4 (Definition 2.1, Chapter VII in Revuz and Yor (1999))**.** Let $b : \mathbb{R}^d \to \mathbb{R}^d$, $a : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be measurable, locally bounded maps and let $a(x)$ be symmetric and semi positive-definite, for any $x \in \mathbb{R}^d$. Define a second order differential operator

$$L = \frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(\cdot)\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{d} b_i(\cdot)\frac{\partial}{\partial x_i}. \tag{2.2.5}$$

Then, a Markov process $(\Omega, \mathscr{F}, \mathscr{F}_t, X_t, \mathbb{P}_x)$ with state space $\mathbb{R}^d$ is a diffusion process with generator L if it has continuous paths, and for any $x \in \mathbb{R}^d$, and any $f \in C_c^\infty$,

$$\mathbb{E}_x\left(f(X_t)\right) = f(x) + \mathbb{E}_x\left(\int_0^t Lf(X_s)\mathrm{d}s\right), \tag{2.2.6}$$

where C_c^∞ denotes the space of infinitely differentiable functions with compact support. The maps a and b are called the diffusion coefficient and the drift, respectively.

The connection to the martingale problem formulation is immediate, given a diffusion process as in Definition 4. In this case, for any $f \in C_c^\infty$, $M_t^f := f(X_t) - f(X_0) - \int_0^t Lf(X_s)\mathrm{d}s$ is a martingale for $\mathbb{P}_x$ (see Proposition 2.2, Chapter VII in Revuz and Yor (1999)). Thus, for any $x \in \mathbb{R}^d$, the measure $\mathbb{P}_x$ on the measure space $\mathcal{W} = (\ C(\mathbb{R}_+, \mathbb{R}^d),$ $\mathcal{C}\,(\ C(\mathbb{R}_+, \mathbb{R}^d)\ )$ is a solution to the martingale problem $\pi(x, a, b)$ defined in the following way:

Definition 5 (Definition 2.3, Chapter VII in Revuz and Yor (1999)). A probability measure $\mathcal{Q}$ on $\mathcal{W}$ is a solution to the martingale problem $\pi(x,a,b)$ if $\mathcal{Q}(X_0 = x) = 1$ and, for any $f \in C_c^\infty$, the process $M_t^f = f(X_t) - f(X_0) - \int_0^t Lf(X_s)\mathrm{d}s$ is a $\mathcal{Q}$-martingale with respect to the filtration $(\sigma(X_s, s \le t))$ where (X_t) denotes the process of coordinate projections.

The converse direction starting from a solution to the martingale problem is more involved. The central task is to establish the Markov property whereas Dynkin's formula (2.2.6) follows immediately. The martingale problem is said to be well-posed if there exists a unique solution of $\pi(x,a,b)$, for any $x \in \mathbb{R}$. Under this uniqueness assumption, solutions can be shown to be Markov processes:

Theorem 6 (Theorem 1.9, Chapter IX in Revuz and Yor (1999)). *If for every $x \in \mathbb{R}^d$, there is one and only one solution $\mathbb{P}_x$ to the martingale problem $\pi(x,a,b)$ and if for every $A \in \mathcal{B}(\mathbb{R}^d)$ and $t \ge 0$ the map $x \mapsto \mathbb{P}_x(X_t \in A)$ is measurable, then $(X_t, \mathbb{P}_x, x \in \mathbb{R}^d)$ is a Markov process with transition function $P_t(x,A) = \mathbb{P}_x(X_t \in A)$.*

There is a one-to-one correspondence between solutions of martingale problems and stochastic differential equations which accounts for the link to the SDE approach as in Definition 3. This correspondence is treated in Revuz and Yor (1999) in a more general, non-homogeneous setting. They introduce Itô processes as solutions of martingale problems with arbitrary initial distribution and a non-homogeneous generator L (see Definition 2.5, Chapter VII). This class includes the diffusion processes, i.e., the homogeneous case. Their Proposition 2.6 then shows that the solution of an SDE of the form (2.2.4) is indeed an Itô process, and in particular, it is the solution of a martingale problem with b and $a = \sigma\sigma^\intercal$.

The converse direction is content of Theorem 2.7 of Chapter VII in Revuz and Yor (1999)(see also Theorem 20.1, Chapter V in Rogers and Williams (2000)). It says that given an Itô process, and in particular a diffusion process X with drift and diffusion coefficient b and a, there is a Brownian motion W in $\mathbb{R}^n$ on an enlargement of the probability space and a map $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times n}$ such that $a = \sigma\sigma^\intercal$ and

$$X_t = X_0 + \int_0^t b(X_s)\mathrm{d}s + \int_0^t \sigma(X_s)\mathrm{d}W_s.$$

2.3 Properties of scalar diffusion processes

In this section, we consider a diffusion process on $\mathbb{R}$ as introduced in Definition 3 which solves the SDE

$$X_t = X_0 + \int_0^t b(X_s)\mathrm{d}s + \int_0^t \sigma(X_s)\mathrm{d}W_s.$$

Recall that such a solution is indeed a continuous strong Markov process.

2.3.1 Scale function and speed measure

We call a diffusion process regular if, for every $y \in \mathbb{R}$, the hitting time $\tau_y := \inf_{t>0}\{X_t = y\}$ is finite with positive probability, for all initial values $x \in \mathbb{R}$; i.e.:

$$\mathbb{P}_x(\tau_y < \infty) > 0.$$

For $-\infty < a < x < b < \infty$ one can show that the diffusion starting in x exits the interval (a, b) with probability one, thus

$$\mathbb{P}_x(\tau_a < \tau_b) + \mathbb{P}_x(\tau_b < \tau_a) = 1.$$

Moreover, the strong Markov property and the regularity yield the existence of a so-called scale function $s : \mathbb{R} \to \mathbb{R}$ which is continuous, strictly increasing and such that, for any $-\infty < a < x < b < \infty$,

$$\frac{s(x) - s(a)}{s(b) - s(a)} = \mathbb{P}_x(\tau_b < \tau_a) \tag{2.3.7}$$

(see Proposition 3.2, Chapter VII in Revuz and Yor (1999)). The scale function of X is unique up to an affine transformation. Considering the transformation $Y = s(X)$, we obtain another regular, continuous strong Markov process Y. Looking at the defining equation (2.3.7), it is straightforward that the scale function of Y is the identity $s_Y(x) = x$. Such a Markov process is said to be on natural scale. Let $I := (a, b)$ be a bounded interval in $\mathbb{R}$ and denote by σ_I the exit time of I, i.e.

$$\sigma_I = \tau_a \wedge \tau_b, \ \mathbb{P}_x \text{ a.s., for } x \in I, \quad \text{and} \quad \sigma_I = 0, \ \mathbb{P}_x \text{ a.s., for } x \notin I.$$

Revuz and Yor (1999) prove that the function $m_I(x) := \mathbb{E}_x(\sigma_I)$ is bounded on I, and consequently σ_I is finite with probability one. Intuitively, m_I gives an idea how fast the diffusion moves through an interval I. This explains the term speed measure that we are about to define. The notion of the speed measure relates the function m_I to the scale function s. Set

$$G_I(x, y) := \begin{cases} \frac{(s(x)-s(a))(s(b)-s(y))}{s(b)-s(a)} & \text{if } a \le x \le y \le b, \\ \frac{(s(y)-s(a))(s(b)-s(x))}{s(b)-s(a)} & \text{if } a \le y \le x \le b, \\ 0 & \text{otherwise.} \end{cases}$$

Then, there is a unique Radon measure $\widetilde{m}$ on $\mathbb{R}$ such that for any bounded, open interval $I = (a, b)$

$$m_I(x) = \int G_I(x, y)\widetilde{m}(\mathrm{d}y), \tag{2.3.8}$$

for any $x \in I$ (Theorem 3.6, Chapter VII in Revuz and Yor (1999)). Note that we have $\widetilde{m} = \widetilde{m}^Y \circ s$ where $\widetilde{m}^Y$ denotes the speed measure of $Y = s(X)$. Therefore, we can draw conclusions concerning $\widetilde{m}$ from results available for speed measures of Markov processes which are on natural scale.

We call a diffusion X on $\mathbb{R}$ recurrent if, for any $x, y \in \mathbb{R}$, it holds $\mathbb{P}_x(\tau_y < \infty) = 1$. In particular, a recurrent diffusion is regular. For our investigation it is central that given a recurrent diffusion X, the finiteness of the speed measure $\widetilde{m}$ implies that X is ergodic with unique invariant measure $\mu := \frac{\widetilde{m}}{\widetilde{m}(\mathbb{R})}$ (cf. Kallenberg (1997) or van der Vaart and van Zanten (2005)). Ergodicity (wrt. to $\mathbb{P}_\mu$) means that the shift invariant sigma-field is trivial. This property entails a kind of law of large numbers for the Markov processes under consideration with initial distribution μ (cf. Cattiaux *et al.* (2012), Theorem 9.8 in Kallenberg (1997)): For any $f \in L^1(\mu)$,

$$\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s \overset{t\to\infty}{\longrightarrow} \int f\mathrm{d}\mu \quad \mathbb{P}_\mu\, a.s. \text{ and in } L^1(\mu). \tag{2.3.9}$$

Connections of the scale function and speed measure to ergodic properties can be studied in a quite general setting of regular, strong Markov processes. However, we want to focus on diffusions given as a solution of the SDE

$$X_t = X_0 + \int_0^t b(X_s)\mathrm{d}s + \int_0^t \sigma(X_s)\mathrm{d}W_s.$$

We present a standard set of conditions that grant existence and ergodic properties of the diffusion process:

Assumption 1.

(i) $b : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}$ are locally Lipschitz continuous and satisfy the at-most linear growth condition

$$|b(x)| + |\sigma(x)| \leq C(1 + |x|), \qquad \text{for all } x \in \mathbb{R},$$

and some constant $C > 0$,

(ii) $|\sigma(x)| > 0$, for all $x \in \mathbb{R}$,

(iii) letting

$$s(x) := \int_0^x \exp\left(-2\int_0^y \frac{b(z)}{\sigma^2(z)}\mathrm{d}z\right)\mathrm{d}y,$$

assume that $\lim_{x\to-\infty} s(x) = -\infty$ and $\lim_{x\to\infty} s(x) = \infty$ and

$$D := \int_{\mathbb{R}} \frac{1}{\sigma^2(x)s'(x)}\mathrm{d}x < \infty.$$

Assumption (i) yields the existence of a strong solution to the SDE on the whole real line. Together with assumption (ii), it can be shown that the scale function of the solution X is given by s and the speed measure $\widetilde{m}$ has density $2/(s'\sigma^2)$ with respect to the Lebesgue measure on $\mathbb{R}$, i.e.,

$$\widetilde{m}(\mathrm{d}x) = \frac{2}{s'(x)\sigma^2(x)}\mathrm{d}x.$$

A derivation of the scale function and the speed measure through the characterisation (2.3.8) can be found in (Karatzas and Shreve, 1988, p. 343 ff.). The additional assumption (iii) yields recurrence (Proposition 5.22 in Karatzas and Shreve (1988)) and finiteness of the speed measure. This in turn implies that X is ergodic with unique invariant measure $\mu = \frac{\widetilde{m}}{\widetilde{m}(\mathbb{R})}$.

The following assumption that we will encounter in Chapter 3 and 4 is sufficient for (iii) to hold.

(iv) For some $\nu \geq 0$, $\mathcal{C} > 0$, the diffusion coefficient σ^2 satisfies

$$\left|\sigma^{-1}(x)\right| \leq \mathcal{C}(1 + |x|^{\nu}), \quad \text{for all } x \in \mathbb{R}, \tag{2.3.10}$$

and there exist constants $A > 0, \gamma > 0$ such that

$$\frac{b(x)}{\sigma^2(x)}\operatorname{sgn}(x) \leq -\gamma, \quad \text{for all } |x| > A. \tag{2.3.11}$$

In the statistical literature addressing diffusion models, σ is very often assumed to equal one and in this case (2.3.10) is clearly satisfied. In this thesis, central ingredients for the statistical analysis provided in Chapter 3 are shown for bounded (from above and below), not constant, diffusion coefficients thereby maintaining a remarkable degree of generality. In order to ease the exposition in Chapter 4, we will restrict to $\sigma \equiv 1$ therein. Condition (2.3.11) to ensure ergodic properties and the existence of the invariant measure can be considered as a standard assumption in statistics for scalar, ergodic diffusion processes.

2.3.2 Local time

The local time of diffusion processes or more general continuous semimartingales is a phenomenon which exists in dimension one only. Concerning our statistical investigation, we will also discover a behaviour which occurs for scalar, not for multivariate, diffusion processes. And thus, not surprisingly the peculiarities of the local time are found to be located at the core of the analysis. In this section, we give a short overview of basic definitions and characteristics of the local time of continuous semimartingales.

The local time is related to the occupation time of a continuous semimartingale X which is a measure for the time spent in a Borel set. It is common to measure occupation time of a diffusion X with respect to the quadratic variation. Formally, the occupation time $T_t(A)$ of a Borel set $A \in \mathcal{B}(\mathbb{R})$ is given by

$$T_t(A) := \int_0^t \mathbb{1}_A(X_s)\mathrm{d}\langle X\rangle_s, \quad t \geq 0.$$

If, for any $t \geq 0$, the occupation time $T_t(\cdot)$ is absolutely continuous with respect to the Lebesgue measure, $T_t(\mathrm{d}x) << \lambda(\mathrm{d}x)$, we call the density

$$L_t^x(X) := \frac{T_t(\mathrm{d}x)}{\lambda(\mathrm{d}x)}, \quad x \in \mathbb{R}, t \geq 0,$$

the local time of X. If it is clear which process the local time belongs to, we also write $L_t^\bullet$ for the local time of X. It can be shown that local time exists for continuous semimartingales. A common approach to rigorously introduce the local time process is via Tanaka's formulas (cf. Revuz and Yor (1999) or Le Gall (2016)). We fix some notation and set, for any $x \in \mathbb{R}$, $x^+ := \max\{x, 0\}$, $x^- := -\min\{x, 0\}$, as well as $\operatorname{sgn}(x) := 1$ if $x > 0$ and $\operatorname{sgn}(x) := -1$ if $x \leq 0$.

Proposition 7 (Proposition 9.2 in Le Gall (2016)). *Let X be a continuous semimartingale and $x \in \mathbb{R}$. There exists an increasing process $(L_t^x(X))_{t\geq 0}$ such that the following three identities hold:*

$$\begin{aligned}
|X_t - x| &= |X_0 - x| + \int_0^t \operatorname{sgn}(X_s - x)\mathrm{d}X_s + L_t^x, \\
(X_t - x)^+ &= (X_0 - x)^+ + \int_0^t \mathbb{1}\{X_s > x\}\mathrm{d}X_s + \frac{1}{2}L_t^x, \\
(X_t - x)^- &= (X_0 - x)^- - \int_0^t \mathbb{1}\{X_s \leq x\}\mathrm{d}X_s + \frac{1}{2}L_t^x.
\end{aligned}$$

The increasing process $(L_t^x(X))_{t\geq 0}$ is called the local time of X at level x. Furthermore, for every stopping time τ, we have $L_t^x(X^\tau) = L_{t\wedge\tau}^x(X)$.

Moreover, for a fixed $x \in \mathbb{R}$, the process $(L_t^x(X))_{t\geq 0}$ is continuous in t and the process $(L_\bullet^x)_{x\in\mathbb{R}}$ with values in $C(\mathbb{R}_+, \mathbb{R}_+)$ has a modification which is càdlàg in x (see Theorem 9.4 in Le Gall (2016)). We will always consider this modification without further notice. The next proposition addresses the occupation times formula which accounts for the heuristic introduction of the local time as the density of the occupation measure at the beginning of this section.

Proposition 8 (cf. Corollary 1.9, Chapter VI in Revuz and Yor (1999)). *Let X be a continuous semimartingale with local time $(L_t^\bullet(X))_{t\geq 0}$. Then, for any bounded, measurable function $f : \mathbb{R} \to \mathbb{R}$, it holds*

$$\int_0^t f(X_s)\mathrm{d}\langle X\rangle_s = \int f(x)L_t^x(X)\mathrm{d}x.$$

A straightforward consequence of the occupation times formula is the approximation of the local time via

$$L_t^x(X) = \lim_{\epsilon\downarrow 0}\frac{1}{\epsilon}\int_0^t \mathbb{1}_{[a,a+\epsilon)}(X_s)\mathrm{d}\langle X\rangle_s, \quad t \geq 0,\, x \in \mathbb{R}$$

(see Proposition 9.9 in Le Gall (2016)). This representation emphasises the meaningful interpretation of the local time $L_t^x(X)$ as a measure for how often x is visited by the process X until time t.

2.4 Limit theorems

Analysing the kernel invariant density estimator $\rho_{t,K}(h)$ (see (1.1.2)) viewed as a random element in the space of bounded functions $\ell^\infty(\mathbb{R})$ leads to the study of additive functionals $\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s$ for f from a class of functions which is given as translations of the kernel function K in the specific case. In particular, we are interested in the asymptotic distribution of $\sqrt{t}\,(\rho_{t,K}(h)-\rho_b)$ and the optimality of the obtained limit distribution.

In this section, we recap the tools we will resort to and start with an introduction of a pointwise law of large numbers and a central limit theorem for the process

$$\mathbb{G}_t(f)=\sqrt{t}\left(\frac{1}{t}\int_0^t f(X_u)\mathrm{d}u-\int f\mathrm{d}\mu_b\right)$$

for an ergodic, scalar diffusion process X with invariant measure μ_b. These results will be used in the subsequent chapters without further notice. In addition, we recall some basics on uniform weak convergence and state a known result on uniform weak convergence of the local time process.

2.4.1 Law of large numbers and central limit theorem for scalar ergodic diffusion processes

Let X be a scalar, ergodic diffusion process given as a solution of the following SDE

$$X_t=X_0+\int_0^t b(X_s)\mathrm{d}s+\int_0^t \sigma(X_s)\mathrm{d}W_s$$

for coefficients b,σ satisfying Assumption 1 with invariant density

$$\rho_b(x)=C_{b,\sigma}^{-1}\sigma^{-2}(x)\exp\left(2\int_0^x\frac{b(z)}{\sigma^2(z)}\mathrm{d}z\right),\ x\in\mathbb{R},$$

where $C_{b,\sigma}:=\int_{-\infty}^{\infty}\sigma^{-2}(x)\exp\left(2\int_0^x\frac{b(z)}{\sigma^2(z)}\mathrm{d}z\right)\mathrm{d}x$. In the sequel, the corresponding invariant measure and its distribution function are denoted by μ_b and F_b, respectively, and we assume that the initial value X_0 is distributed according to μ_b.

We repeat that the process having ergodic properties means that a law of large numbers holds (see also (2.3.9)), and furthermore a central limit theorem for additive functionals can be shown:

Proposition 9 (Law of large numbers (Theorem 1.16 in Kutoyants (2004))).
The diffusion process X satisfies a law of large numbers in the sense that for any measurable function $f:\mathbb{R}\to\mathbb{R}$ with $\int|f(x)|\rho_b(x)dx<\infty$, almost surely

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s=\int f(x)\rho_b(x)\mathrm{d}x.$$

Proposition 10 (Central limit theorem (Proposition 1.22 and 1.23 in Kutoyants (2004))).

For any measurable function $f : \mathbb{R} \to \mathbb{R}$ *with* $\int f^2(x)\rho_b(x)\mathrm{d}x < \infty$,

$$\frac{1}{\sqrt{t}} \int_0^t f(X_s)\mathrm{d}Ws \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0, \int f^2(x)\rho_b(x)\mathrm{d}x\right).$$

Furthermore, for any measurable function $f : \mathbb{R} \to \mathbb{R}$ *with* $\int |f(x)|\rho_b(x)\mathrm{d}x < \infty$, $\int f(x)\rho_b(x)dx = 0$, *and such that*

$$\delta^2 := 4 \int_{-\infty}^{\infty} \left[\int_{-\infty}^{x} \frac{f(v)\rho_b(v)}{\sigma(x)\rho_b(x)}\mathrm{d}v\right]^2 \rho_b(x)\mathrm{d}x < \infty,$$

it holds

$$\frac{1}{\sqrt{t}} \int_0^t f(X_s)\mathrm{d}s \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0, \delta^2\right). \tag{2.4.12}$$

Here, $\stackrel{\mathcal{L}}{\Longrightarrow}$ *denotes weak convergence.*

An important consequence is a pointwise central limit theorem (see Proposition 1.25 in Kutoyants (2004)) for the local time process $(L_t^x(X), t \geq 0, x \in \mathbb{R})$ or stated differently the asymptotic normality of the local time estimator at a point x for the invariant density $\rho_b(x)$ defined as $\rho_t^\circ(x) := \frac{L_t^x(X)}{t\sigma^2(x)}$:

For every $x \in \mathbb{R}$, assuming that $\int \sigma^2(x)\rho_b(x)\mathrm{d}x < \infty$ and

$$\delta^2(x) := 4\rho_b^2(x) \int_{\mathbb{R}} \left[\frac{\mathbb{1}\{u > x\} - F_b(u)}{\sigma(u)\rho_b(u)}\right]^2 \rho_b(u)\mathrm{d}u < \infty,$$

the central limit theorem implies

$$\sqrt{t}\,(\rho_t^\circ(x) - \rho_b(x)) \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, \delta^2(x)).$$

A statistical objective of this thesis is the maximal error between estimator and estimation target, and therefore we are not only interested in pointwise, but also in uniform limit theorems. This is dealt with in Negri (2001) for the local time estimator as well as in a more general setting in van der Vaart and van Zanten (2005). They show that under certain conditions

$$\sqrt{t}\,(\rho_t^\circ - \rho_b) \stackrel{\mathcal{L}}{\Longrightarrow} \mathbb{H} \tag{2.4.13}$$

in $\ell^\infty(\mathbb{R})$ where the limit $(\mathbb{H}(x), x \in \mathbb{R})$ is a Gaussian random map with covariance structure

$$\mathbb{E}\mathbb{H}(x)\mathbb{H}(y) \;=\; 2\widetilde{m}(\mathbb{R})\rho_b(x)\rho_b(y) \int_I (\mathbb{1}\{[x,\infty)\} - F_b)(\mathbb{1}\{[y,\infty)\} - F_b)\mathrm{d}s$$

where s and $\widetilde{m}$ denote the scale function and speed measure introduced in Section 2.3, respectively. The notion of weak convergence in $\ell^\infty(\mathbb{R})$ is introduced rigorously in the subsequent section.

Note that the speed measure m which appears in van der Vaart and van Zanten (2005) is equal to $\frac{1}{2}\widetilde{m}$ for $\widetilde{m}$ as introduced in Section 2.3. Taking into account that $\rho_b \mathrm{d}x = \widetilde{m}(\mathbb{R})^{-1}\widetilde{m}(\mathrm{d}x) = \frac{2}{s'\sigma^2\widetilde{m}(\mathbb{R})}\mathrm{d}x$, it is clear that the limit coincides with the pointwise, one dimensional limit in (2.4.12).

2.4.2 Uniform weak convergence

We have seen a uniform weak convergence result (2.4.13) for the local time density estimator of a scalar, ergodic diffusion process X in the previous section which we also refer to as a Donsker-type theorem. We proceed with providing some background knowledge on the subject of weak convergence of bounded processes indexed by an arbitrary set $\mathcal{S}$ interpreted as random elements in $\ell^\infty(\mathcal{S})$ also referred to as uniform weak convergence. Note that $\ell^\infty(\mathcal{S})$ is complete but not separable except in the case of finite $\mathcal{S}$ and so the cylindrical $\sigma-$algebra on $\ell^\infty(\mathcal{S})$ does not equal the Borel σ-algebra with respect to the sup-norm. In particular, stochastic processes indexed by $\mathcal{S}$ do not necessarily induce a Borel measure on $\ell^\infty(\mathcal{S})$, and the law of such a process does not need to be tight. As tightness is closely related to weak convergence, this is an issue that has to be dealt with. A minor concern, on the other hand, is that looking at continuous functions on $\ell^\infty(\mathcal{S})$ of the process the resulting random variable does not even need to be measurable. This circumstance can be solved via outer expectations. Details can be found in (Giné and Nickl, 2016, Section 3.7.1). Measurability issues will be mostly omitted in the sequel.

We introduce a definition of uniform weak convergence for bounded processes indexed by an index set $\mathcal{S}$.

Definition 11 (Definition 3.7.22 in Giné and Nickl (2016))**.** Let $\mathbb{H} = (\mathbb{H}(s))_{s\in\mathcal{S}}$ be a bounded process whose finite-dimensional distributions equal the laws of the finite-dimensional projections under a tight Borel probability measure on $\ell^\infty(\mathcal{S})$. Furthermore, let $\widetilde{\mathbb{H}}$ denote a Borel measurable version of $\mathbb{H}$ with separable range. Let $\mathbb{H}_t = (\mathbb{H}_t(s))_{s\in\mathcal{S}}, t \geq 0$, be a sequence of bounded processes. Then, we say that the sequence $(\mathbb{H}_t)_{t\geq 0}$ converges in law (converges weakly) to $\mathbb{H}$ in $\ell^\infty(\mathcal{S})$, or uniformly in $s \in \mathcal{S}$ if

$$\mathbb{E}^*[G(\mathbb{H}_t)] \underset{t\to\infty}{\longrightarrow} \mathbb{E}[G(\widetilde{\mathbb{H}})],$$

for any bounded and continuous function $G : \ell^\infty(\mathcal{S}) \to \mathbb{R}$, where $\mathbb{E}^*$ denotes the outer expectation. In this case, we write

$$\mathbb{H}_t \overset{\mathcal{L}}{\Longrightarrow} \mathbb{H} \text{ in } \ell^\infty(\mathcal{S}).$$

An important result is the following characterisation of uniform weak convergence which says that it is equivalent to finite-dimensional convergence and asymptotic equicontinuity:

Theorem 12 (Theorem 3.7.23 in Giné and Nickl (2016))**.** *Let $\mathbb{H}_t = (\mathbb{H}_t(s))_{s\in\mathcal{S}}, t \geq 0$, be bounded processes indexed by $\mathcal{S}$. Then, the following statements are equivalent:*

(a) *The finite-dimensional distributions of the processes* $\mathbb{H}_t$ *converge in law, and there exists a pseudo-metric* d *on* $\mathcal{S}$ *such that* $(\mathcal{S}, d)$ *is totally bounded, and*

$$\lim_{\delta \to 0} \limsup_{t \to \infty} \mathbb{P}^* \left(\sup_{d(r,s) \le \delta} |\mathbb{H}_t(r) - \mathbb{H}_t(s)| > \epsilon \right) = 0, \quad \forall \epsilon > 0. \tag{2.4.14}$$

(b) *There exists a process* $\mathbb{H}$ *whose law is a tight Borel probability measure on* $\ell^\infty(\mathcal{S})$ *and such that*

$$\mathbb{H}_t \xrightarrow{\mathcal{L}} \mathbb{H} \text{ in } \ell^\infty(\mathcal{S}).$$

Note that the results in Giné and Nickl (2016) are formulated for sequences indexed by natural numbers but they still hold for sequences indexed by positive real numbers, and even more generally for nets (see van der Vaart and Wellner (1996)).

Asymptotic equicontinuity is related to asymptotic tightness (see Theorem 1.5.7 in van der Vaart and Wellner (1996)) which gives another useful characterisation of uniform weak convergence. A bounded process $\mathbb{H}_t = (\mathbb{H}_t(s))_{s \in \mathcal{S}}, t \ge 0$, is called asymptotically tight if, for every $\epsilon > 0$, there exists a compact set $K \subseteq \ell^\infty(\mathcal{S})$ such that, for any $\delta > 0$,

$$\liminf_{t \to \infty} \mathbb{P}_* \left(\mathbb{H}_t \in K^\delta \right) \ \ge \ 1 - \epsilon,$$

where K^δ denotes the δ-enlargement of K, and $\mathbb{P}_*$ stands for the inner probability.

Theorem 13 (cf. Theorem 1.5.4. in van der Vaart and Wellner (1996))**.** *Let* $\mathbb{H}_t = (\mathbb{H}_t(s))_{s \in \mathcal{S}}$, $t \ge 0$, *be bounded processes. Then, the following statements are equivalent:*

(a) *The finite-dimensional distributions of the processes* $\mathbb{H}_t$ *converge in law and* $(\mathbb{H}_t)_{t \ge 0}$ *is asymptotically tight.*

(b) *There exists a process* $\mathbb{H}$ *whose law is a tight Borel probability measure on* $\ell^\infty(\mathcal{S})$ *and such that*

$$\mathbb{H}_t \xrightarrow{\mathcal{L}} \mathbb{H} \text{ in } \ell^\infty(\mathcal{S}).$$

Furthermore, if $\mathbb{H}_t = (\mathbb{H}_t(s))_{s \in \mathcal{S}}$, $t \ge 0$, *is asymptotically tight and the finite-dimensional distributions converge weakly to the marginals of a process* $\mathbb{H} = (\mathbb{H}(s))_{s \in \mathcal{S}}$, *then there is a version* $\widetilde{\mathbb{H}}$ *of* $\mathbb{H}$ *with uniformly bounded paths and* $\mathbb{H}_t \xrightarrow{\mathcal{L}} \widetilde{\mathbb{H}}$ *in* $\ell^\infty(\mathcal{S})$.

The theory of uniform weak convergence allows the investigation of asymptotic properties of empirical processes in continuous time. Let $\mathcal{F}$ be a class of uniformly bounded functions, i.e., with $\sup_{f \in \mathcal{F}} \|f\|_\infty < \infty$. We introduce the empirical process

$$\mathbb{G}_t(f) := \sqrt{t} \left(\frac{1}{t} \int_0^t f(X_s) - \int f(x) \rho_b(x) \mathrm{d}x \mathrm{d}s \right)$$

where X is a scalar, ergodic diffusion with invariant density ρ_b as in Section 2.4.1. The process $\mathbb{G}_t$ indexed by $\mathcal{F}$ is an element of $\ell^\infty(\mathcal{F}) = \{z : \mathcal{F} \to \mathbb{R} : \|z\|_\infty < \infty\}$. Following van der Vaart and van Zanten (2005) we call $\mathcal{F}$ a Donsker class if $\mathbb{G}_t$ converges weakly in $\ell^\infty(\mathcal{F})$ to a tight, Borel measurable random map $\mathbf{G}$.

A goal of this thesis is to establish a Donsker-type result for the kernel invariant density estimator, which means that we look at a specific class $\mathcal{F}$ of translations of a kernel function. The investigation relies on a Donsker theorem for the local time of a diffusion process. The latter turns out to be a special case of Donsker theorems for empirical processes provided in van der Vaart and van Zanten (2005). There, a more general framework of scalar diffusions with finite speed measure is considered.

Due to the Cramér-Wold and the central limit theorem we already know that the finite dimensional distributions of the empirical process $\mathbb{G}_t$ converge weakly. More precisely, for any finite number $f_1, ..., f_n \in \mathcal{F}$, $n \in \mathbb{N}$, we have

$$(\mathbb{G}_t(f_1), ..., \mathbb{G}_t(f_n))^\intercal \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0_n, \Gamma)$$

with $\Gamma \in \mathbb{R}^{n\times n}$ and

$$\Gamma_{i,j} = \Gamma(f_i, f_j) = 4\int_{-\infty}^{\infty} \frac{1}{\sigma^2(x)\rho_b(x)} J(f_i, x) J(f_j, x) \mathrm{d}x$$

where $J(f, x) := \int_{-\infty}^{x} f(v)\rho_b(v)\mathrm{d}v - \mathbb{E}(f(X_0))F_b(x)$, for any $x \in \mathbb{R}$, $f \in \mathcal{F}$. Therefore, the existence of a tight Gaussian random map $\mathbf{G} \in \ell^\infty(\mathcal{F})$ with covariance structure Γ is necessary for the uniform weak convergence of $\mathbb{G}_t$. In contrast to the classical i.i.d. setting where additional entropy conditions, as for example in Theorem 3.7.36 in Giné and Nickl (2016), are needed to satisfy the asymptotic equicontinuity criterion in part (a) of Theorem 12, van der Vaart and van Zanten (2005) prove that for a large class of diffusion processes the existence of the tight Gaussian limit is already sufficient for $\mathcal{F}$ to be Donsker. As a corollary, they obtain conditions under which the local time density estimator converges uniformly (cf. (2.4.13)). We will come back to this result when translating it into a Donsker theorem for the kernel density estimator of the invariant density.

3 Concentration of scalar ergodic diffusions and some statistical implications

3.1 Introduction

With regard to the very basic idea of estimating expected values via sample means as motivated by the law of large numbers, the relevance of concentration inequalities which quantify the deviation behaviour of more general additive functionals from their mean is pretty obvious. It is thus natural that they can be identified as being a central device in many statistical investigations, both from a frequentist and a Bayesian point of view. From an applied perspective, expected maximal errors describing worst case scenarios are of particular interest for quantifying the quality of estimators. The analysis of sup-norm risk criteria when estimating densities, regression functions or other characteristics thus is of immense relevance. Nevertheless, even in classical situations like density estimation from i.i.d. observations, the sup-norm case is a delicate issue and usually not treated as exhaustively as L^p or pointwise risk measures. Analysing the sup-norm risk often requires to resort to empirical process theory. More precisely, it leads to the need of finding moment bounds and concentration inequalities for the supremum of empirical processes, i.e., the supremum of additive functionals, over possibly infinite-dimensional function classes. This turns out to be a probabilistic challenge. In case of diffusion processes with unbounded state space, estimation of diffusion characteristics in sup-norm risk is a mostly open question even in the most basic setting of continuous observation of a scalar process. The current work aims at providing the fundamental probabilistic tool box, including uniform concentration inequalities for empirical processes and related concepts, in the continuous scalar diffusion context as they are essential for further statistical research on the sup-norm risk.

Since they are taken as a standard model for a number of random phenomena arising in various applications, statistical inference for ergodic diffusion processes, based on different observation schemes, has been widely developed during the past decades. While observation data as the central ingredient of any estimation procedure in practice are always discrete, it is insightful to start the statistical analysis in the framework of continuous observations, thereby providing both benchmark results and a starting point for estimation schemes based on discrete data. Within this framework, we demonstrate that our approach to concentration results can be specified as needed for proving sharp upper bounds on sup-norm risks. Moreover, we introduce a machinery for obtaining uniform concentration inequalities for empirical processes based on martingale approximation and the generic chaining device that allows for the analogue treatment of more general classes of Markov processes as well. In particular, with regard to the diffusion process set-up, our approach could also be adapted for sup-norm risk investigations based on discrete observations or multivariate state variables. While the basic idea of martingale approximation is applied at several places in the statistical literature, we are not aware of *any* systematic attempts to exploit the approach for deriving concentration results.

3.1.1 Basic framework and main results

Given a continuous-time Markov process $X = (X_t)_{t\geq 0}$ with invariant measure μ, the counterpart to the empirical process $\sqrt{n}\left(n^{-1}\sum_{i=0}^{n-1} f(Y_i) - \mathbb{E}\left[f(Y_0)\right]\right)$, $f \in \mathcal{F}$, based on i.i.d. observations $Y_0, ..., Y_{n-1}$, is given as

$$\sqrt{t}\left(\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s - \int f(x)\mathrm{d}\mu(x)\right), \quad f \in \mathcal{F}, \tag{3.1.1}$$

$\mathcal{F}$ denoting a class of functions, typically satisfying suitable entropy conditions. This continuous-time version of the classical empirical process is our first object of interest. For the goal at hand, we will focus on diffusion processes given as a solution of the SDE

$$\mathrm{d}X_t \;=\; b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \quad X_0 = \xi,\ t \geq 0, \tag{3.1.2}$$

where W is a standard Brownian motion and the initial value ξ is a random variable independent of W. We restrict to the ergodic case where the Markov process X admits an invariant measure, and we denote by ρ_b and μ_b the invariant density and the associated invariant measure, respectively. Furthermore, we will always consider stationary solutions of (3.1.2), i.e., we assume that $\xi \sim \mu_b$. In this framework, we will also provide precise uniform concentration inequalities for stochastic integrals

$$\frac{1}{t}\int_0^t f(X_s)\mathrm{d}X_s - \mathbb{E}\left[f(X_0)b(X_0)\right], \quad f \in \mathcal{F}, \tag{3.1.3}$$

which turn out to be essential for statistical investigations.

Main results For a diffusion process given as the stationary solution of (3.1.2), Theorem 23 provides an exponential tail inequality for

$$\sup_{f\in\mathcal{F}} \sqrt{t}\left|\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s - \mathbb{E}\left[f(X_0)\right]\right|$$

as well as bounds on its p-th moments, for any $p \geq 1$, under standard entropy conditions on the function class $\mathcal{F}$. Proposition 24 and Theorem 30 constitute analogue results for the supremum of the stochastic integrals (3.1.3). We emphasise at this point that we allow for unbounded functions $f \in \mathcal{F}$ which is even for nonuniform Bernstein-type results absolutely nonstandard. Furthermore, we introduce a localisation procedure which allows to look at processes on the whole real line instead of compacts.

As a statistical application, we investigate nonparametric invariant density estimation in supremum-norm based on a continuous record of observations $(X_s)_{0\leq s\leq t}$ of the solution of (3.1.2) started in the equilibrium. In the continuous framework, the local time – which can be interpreted as the derivative of an empirical distribution function – naturally qualifies as an estimator of this density. Corresponding upper bounds for all p-th moments of the sup-norm loss are given in Corollary 28. We advocate the investigation of the continuous, scalar case because it serves as a fundament and as a relevant benchmark for further investigations of discrete observation schemes and the multivariate case. With this purpose in mind, the density estimator based on local time is not the preferable choice as it does not open immediate access to discrete-time or multivariate estimators. In contrast, the very classical kernel (invariant) density estimator meets all these requirements, and it achieves the same (optimal) sup-norm rates of convergence which we establish in Corollary 26.

3.1.2 Structure and techniques: an overview

Introducing methods at the concrete example of a tail estimate for the local time

We will start in Section 3.2 with an exponential uniform upper tail inequality for the local time of a continuous semimartingale, stated in Theorem 15. The local time of semimartingales was discussed by Meyer (1976), and we adopt his definition: Given a continuous semimartingale X, denote by $(L_t^a(X))_{t\geq 0}$, $a\in\mathbb{R}$, the *local time of X at level a*, i.e., the increasing process which satisfies the following identity,

$$(X_t-a)^- \;=\; (X_0-a)^- - \int_0^t \mathbb{1}\{X_s\leq a\}\mathrm{d}X_s + \frac{1}{2}\,L_t^a(X), \quad t>0,\ a\in\mathbb{R}. \qquad (3.1.4)$$

We have chosen to begin from this special case not only because of the statistical interest in the local time. It is instructive since, in the process of proving Theorem 15, we will already introduce key ideas and methods, including the generic chaining and localisation procedures that we will resort to for the further analysis of general empirical processes. From the representation (3.1.4) it actually becomes clear that analysing $\sup_{a\in\mathbb{R}} L_t^a(X)$ requires looking at

$$\sup_{f\in\mathcal{F}}\left|\int_0^t f(X_s)\mathrm{d}X_s\right|, \quad \text{for } \mathcal{F} := \{\mathbb{1}\{\,\cdot\,\leq a\}\;:\; a\in\mathbb{R}\}\,.$$

This expression accounts for the connection to the investigation of uniform concentration inequalities for empirical processes and stochastic integrals as in (3.1.3). The proof thus serves as a blueprint and a concrete example that prevents from losing track while handling the technicalities coming up in the general empirical process setting. Under suitable moment conditions, we do not have to restrict to diffusion processes, yet. Instead, the results presented in Section 3.2 hold in a general continuous semimartingale framework.

A central ingredient of the proof of Theorem 15 is the decomposition of the local time into a martingale part and a remaining term induced by (3.1.4). Considering more general additive functionals as in (3.1.1), we carry on this idea and prove a uniform concentration inequality for empirical processes (3.1.1) of general continuous semimartingales, assuming the existence of a martingale approximation.

Martingale approximation In the discrete framework, the technique of martingale approximation was initiated by Gordin and Lifsic (1978), while Bhattacharya (1982) proved the continuous-time analogue. Their basic idea consists in deriving the CLT for processes $\mathbb{G}_t(f)$,

$$\mathbb{G}_t(f) \;:=\; \sqrt{t}\left(\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s - \mathbb{E}[f(X_0)]\right),$$

f some square-integrable function, by decomposing the above partial sums into the sum of a martingale with stationary increments and a remainder term. Asymptotic normality then follows from a martingale CLT. For fixing terminology, suppose that $\mathbb{G}_t(f)$, $f\colon \mathbb{R}\to\mathbb{R}$, lives on a fixed filtered probability space $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\geq 0},\mathbb{P})$. One then says that there exists a *martingale approximation to* $\mathbb{G}_t(f)$, $f\colon\mathbb{R}\to\mathbb{R}$, if there exist two processes $(M_t(f))_{t\geq 0}$ and $(R_t(f))_{t\geq 0}$ on $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\geq 0},\mathbb{P})$ such that

$$\mathbb{G}_t(f) \;=\; \frac{1}{\sqrt{t}}\,M_t(f) + \frac{1}{\sqrt{t}}\,R_t(f), \quad t>0, \tag{3.1.5}$$

where $(M_t(f))_{t\geq 0}$ is a martingale wrt $(\mathscr{F}_t)_{t\geq 0}$ fulfilling $M_0(f)=0$ and the remainder term $(R_t(f))_{t\geq 0}$ is negligible in some sense.

Results on uniform concentration for empirical processes of continuous semimartingales Given the availability of a suitable martingale approximation of the additive functional, we show in Section 3.3 how to derive uniform concentration results on $t^{-1}\int_0^t f(X_s)\mathrm{d}s$, $f\in\mathcal{F}$, in the continuous semimartingale setting. Speaking of uniform concentration results, we refer to inequalities of the form

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}}|\mathbb{G}_t(f)|\geq \mathrm{e}\Phi(u)\right) \;\leq\; \exp(-u), \quad \text{for any } u\geq 1, \tag{3.1.6}$$

e denoting Euler's number, which is an immediate consequence of the moment bound $\left(\mathbb{E}\left[\sup_{f\in\mathcal{F}}|\mathbb{G}_t(f)|\right]^p\right)^{\frac{1}{p}} \leq \Phi(p)$ for some function $\Phi\colon(0,\infty)\to(0,\infty)$ and any $p\geq 1$. Note that this is not a concentration inequality for the random variable $\sup_{f\in\mathcal{F}} t^{-1/2}\int_0^t f(X_s)\mathrm{d}s$ as such. It is rather a uniform or worst case statement on the concentration of $t^{-1/2}\int_0^t f(X_s)\mathrm{d}s$. Nonetheless, it additionally implies an upper exponential deviation inequality for the random variable

$$\sup_{f\in\mathcal{F}}\frac{1}{\sqrt{t}}\left|\int_0^t f(X_s)\mathrm{d}s\right|$$

from its mean. These uniform concentration inequalities given in Theorem 16 are the main result in Section 3.3. The tail behaviour incorporated in the nature of the function Φ in (3.1.6) is described in terms of entropy integrals. This formulation is not the most handy but means a higher degree of generality. Of course, the entropy integrals can further be upper bounded under mild entropy conditions on the function class as known from the i.i.d. set-up (see Lemma 36 in Section 3.6). The proof of Theorem 16 relies on a localised generic chaining procedure that can be applied assuming the existence of a martingale approximation of the empirical process $(\mathbb{G}_t(f))_{f\in\mathcal{F}}$. Let us already note that our results on the concentration of empirical processes of the form (3.1.1) in Section 3.3 do *not* require the existence of a local time process. Though the framework of continuous semimartingales is suitable for our goal of considering diffusion processes, the techniques could also be applied to other models, e.g., more general classes of Markov processes. The only prerequisites consist in a maximal inequality of the form (3.2.14) and a martingale approximation with suitable moment bounds as in (3.3.19). We also advocate our approach as a starting point for the derivation of parallel results for multivariate diffusion processes.

Results on uniform concentration for empirical processes and stochastic integrals of scalar ergodic diffusions The findings of Section 3.2 and Section 3.3 are applied to obtain uniform concentration results for $t^{-1}\int_0^t f(X_s)\mathrm{d}s$ and $t^{-1}\int_0^t f(X_s)\mathrm{d}X_s$, $f\in\mathcal{F}$, in the diffusion framework in Section 3.4. For the concrete case of diffusion processes, we show in Section 3.4.2 that a suitable martingale approximation as described above exists. This fact immediately implies the uniform concentration inequalities for empirical processes stated in Theorem 23. The natural approach of analysing the supremum of these objects by exploiting concentration results such as Bernstein-type deviation inequalities for additive diffusion functionals has severe obstacles which are detailed in Remark 19. In particular, this approach forces one to impose additional conditions on the characteristics of the diffusion process in order to prove the required *uniform* concentration results. Remarkably, the alternative strategy via martingale approximation allows to work under minimal assumptions on the class of diffusion processes. As a consequence, we obtain results on the uniform concentration both of additive functionals and of stochastic integrals.

The uniform concentration inequality for the stochastic integrals of a diffusion process is subject of Proposition 24 and makes use of Theorem 15 on the local time. In Proposition 24, we consider the question of exploring the tail behaviour for quantities of the form

$$\mathbb{H}_t(f) \;:=\; \sqrt{t}\left(\frac{1}{t}\int_0^t f(X_s)\mathrm{d}X_s - \mathbb{E}\left[f(X_0)b(X_0)\right]\right), \quad f\in\mathcal{F},$$

X some diffusion process solving (3.1.2) and $\mathcal{F}$ denoting some (possibly infinite-dimensional) class of integrable functions. For adaptive procedures for estimating the characteristics of X, one generally requires both an upper bound on

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}}|\mathbb{H}_t(f)|\right], \qquad \mathcal{F} \text{ some class of translated kernel functions,}$$

and an upper tail bound for the deviation of the supremum. Using generic chaining methods initiated by Talagrand (cf. Talagrand (2014)), both can be derived by obtaining upper bounds for *all* p-th ($p\geq 1$) moments of $(\mathbb{H}_t(f))_{f\in\mathcal{F}}$.

Uniform moment bounds and exponential inequalities for stochastic integrals via generic chaining Starting from the basic decomposition

$$\begin{aligned}\mathbb{H}_t(f) &= \frac{1}{\sqrt{t}}\int_0^t \left(f(X_s)b(X_s) - \mathbb{E}\left[f(X_0)b(X_0)\right]\right)\mathrm{d}s \;+\; \frac{1}{\sqrt{t}}\int_0^t f(X_s)\sigma(X_s)\mathrm{d}W_s \\ &=: (\mathbf{I}) + (\mathbf{II}),\end{aligned} \tag{3.1.7}$$

we recognise the empirical process $(\mathbf{I})$ which can be treated by means of Theorem 23. The next step then consists in finding upper bounds on the p-th moments of $(\mathbf{II})$. Applying the Burkholder–Davis–Gundy (BDG) inequality and the occupation times formula, one obtains

$$\begin{aligned}\mathbb{E}\left[\left|\frac{1}{\sqrt{t}}\int_0^t f(X_s)\sigma(X_s)\mathrm{d}W_s\right|^p\right] &\leq C_p\mathbb{E}\left[\left(\frac{1}{t}\int_0^t f^2(X_s)\sigma^2(X_s)\mathrm{d}s\right)^{p/2}\right] \\ &= C_p\mathbb{E}\left[\left(\frac{1}{t}\int_{\mathbb{R}} f^2(y)L_t^y(X)\mathrm{d}y\right)^{p/2}\right] \\ &\leq C_p t^{-p/2}\left(\int_{\mathbb{R}} f^2(y)\mathrm{d}y\right)^{p/2}\ \mathbb{E}\left[\left(\sup_{a\in\mathbb{R}}|L_t^a(X)|\right)^{p/2}\right].\end{aligned}$$

At first sight, this upper bound may seem to be very rough, but looking into the details of the proof, it becomes clear that one needs to obtain the L^2 norm of f on the right hand side for the generic chaining procedure which accounts for this estimate. Conveniently, we can then apply Theorem 15. It provides both an upper bound on the p-th moments $\mathbb{E}\left[\|L_t^\bullet(X)\|_\infty^p\right]$ and a corresponding tail estimate. Inspection of the proof of Theorem 15 shows that it relies on three substantial ingredients:

(i) The proof exploits the decomposition of the local time process into a martingale part and a remainder term provided by Tanaka's formula. The analysis of the martingale part then relies on generic chaining methods.

(ii) The latter requires the increments of the martingale to exhibit a subexponential tail behaviour wrt to a suitable metric (cf. (3.6.45)). We discover this relation from a *sharp* formulation of the bound

$$\mathbb{E}\left[\left(\int_0^t \mathbb{1}\{a \leq X_s \leq b\}\mathrm{d}\langle M, M\rangle_s\right)^p\right] \;\leq\; c_p(b-a)^p\mathbb{E}\left[\langle M, M\rangle_t^{\frac{p}{2}} + \left(\int_0^t |\mathrm{d}V_s|\right)^p\right]$$

(see, e.g., Lemma 9.5 in Le Gall (2016)), for $M = (M_t)_{t\geq 0}$ and $V = (V_t)_{t\geq 0}$ denoting the martingale part and the finite variation part of the semimartingale X, respectively. Here, 'sharp' refers to the dependence of the constant c_p on the order p of the moments. This can be obtained by means of Proposition 4.2 in Barlow and Yor (1982) (see (3.2.12) below).

(iii) The supremum taken over the entire real line is dealt with by an investigation of the random, compact support of the local time $L_t^\bullet(X)$. In particular, we rely on a maximal inequality for the process X which allows to control the probability that the support of the process exceeds certain levels.

As already announced, the proof of Theorem 15 also serves as a blueprint for the analysis of the supremum of additive functionals in Theorem 16.

There is some evidence of the statistical relevance of diffusion local time. As one first concrete example, let us mention the deep Donsker-type theorems for diffusion processes in van der Vaart and van Zanten (2005) whose proof relies on a limit theorem for the supremum of diffusion local time. Another instance concerns the completely different context of studying nonparametric Bayesian procedures for one-dimensional SDEs: Pokern *et al.* (2013) investigate a Bayesian approach to nonparametric estimation of the periodic drift of a scalar diffusion from continuous observations and derive bounds on the rate at which the posterior contracts around the true drift in L^2-norm. Their theoretical results in particular rely on functional limit theorems for the local time of diffusions on the circle.

3.1.3 Statistical applications

The concept of local time is deeply rooted in probability theory. As indicated above, it however presents a very interesting object from a statistical point of view, too. For another concrete motivation, let us specify again to the important class of ergodic diffusion process solutions of SDEs of the form (3.1.2) with invariant density ρ_b. Given a set of observations of the solution of (3.1.2) with unknown drift $b\colon \mathbb{R}\to\mathbb{R}$, natural statistical questions concern the estimation of b and of the invariant density ρ_b. In fact, in view of the basic relation $b=(\sigma^2\rho_b)'/(2\rho_b)$, both tasks are obviously related. Note that continuous observations can identify the diffusion coefficient σ^2. Therefore, it is considered to be known, and the focus is on estimation of the drift coefficient b.

Invariant density estimation via local time Alternatively to (3.1.4), $(L_t^a(X))_{t\geq 0}$ may be introduced via the following approximation result, holding a.s. for every $a\in\mathbb{R}$ and $t\geq 0$,

$$L_t^a(X) \;=\; \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_0^t \mathbb{1}\{a\leq X_s\leq a+\varepsilon\}\mathrm{d}\langle X\rangle_s.$$

This representation now already suggests the meaningful interpretation of the local time as the derivative of an empirical distribution function. Assuming that a continuous record of observations $(X_s)_{0\leq s\leq t}$ of the solution of (3.1.2) is available, it thus appears natural to use local time for constructing an estimator ρ_t° of ρ_b by letting

$$\rho_t^\circ(a) \;:=\; \frac{L_t^a(X)}{t\sigma^2(a)},\quad a\in\mathbb{R}. \tag{3.1.8}$$

One might tackle the question of quantifying the quality of the estimator ρ_t° wrt the sup-norm risk, e.g., by deriving upper bounds on the p-th ($p\geq 1$) moments

$$\mathbb{E}\left[\left(\sup_{a\in\mathbb{R}}|\rho_t^\circ(a)-\rho_b(a)|\right)^p\right] \;=\; \mathbb{E}\left[\left\|\frac{L_t^\bullet(X)}{t\sigma^2}-\rho_b\right\|_\infty^p\right].$$

Local time thus presents an object of its own statistical interest. The corresponding investigation is subject of Section 3.5.

Kernel invariant density estimation Apart from the treatment of the local time estimator in sup-norm loss, the statistical relevance of Theorem 23 – which deals with general empirical processes of a diffusion – is demonstrated by a detailed study of the question of invariant density estimation via the *kernel density estimator* (again in sup-norm loss) and its relation to the local time density estimator in Section 3.5. One clear advantage of the local time estimator ρ_t° introduced in (3.1.8) is that it allows for direct application of deep probabilistic results on diffusion local time. For example, weak convergence properties can be deduced in this way. At the same time, ρ_t° is merely of theoretical interest since its implementation in practice requires another approximation procedure. One first step towards finding practically more feasible estimators is to replace ρ_t° by the standard kernel estimator

$$\rho_{t,K}(h)(x) \;:=\; \frac{1}{th}\int_0^t K\left(\frac{x-X_u}{h}\right)\mathrm{d}u, \quad x \in \mathbb{R}, \tag{3.1.9}$$

$K\colon \mathbb{R} \to \mathbb{R}$ some smooth kernel function with compact support and $h > 0$ some bandwidth. The kernel density estimator outperforms the local time density estimator in various important aspects. First of all, from an applied perspective, working with the kernel density estimator serves as a universal, familiar approach to density estimation in all common models. For our particular diffusion framework, it is straightforward to extend the procedure to the case of discrete or multivariate observations. From a more theoretical perspective, the additional smoothness of the kernel estimator is desirable for investigations. The kernel density estimator can be viewed as a convolution operator applied to the local time. Interestingly, this smoothing is exactly what is required for proving the assertion on $\|\rho_t^\circ - \rho_b\|_\infty$ in Corollary 28. Thus, our proof – which makes use of the kernel density estimator – is more natural than it might look at first sight. In addition, we show that our results on the moments of the supremum of empirical processes imply precise upper bounds on $\mathbb{E}\left[\|\rho_{t,K}(h) - \rho_b\|_\infty^p\right]$, $p \geq 1$. These upper bounds in particular verify that, in terms of performance in sup-norm risk, the kernel density estimator with the universal bandwidth choice $t^{-1/2}$ is as good as the local time density estimator ρ_t°. Furthermore, we provide an in-depth analysis of the stochastic behaviour of $\|\rho_{t,K}(h) - \rho_t^\circ\|_\infty$ which in particular allows to transfer results for the local time estimator to the class of kernel estimators.

Application to adaptive (drift) estimation Beyond the question of invariant density estimation, another important statistical motivation for deriving the concentration inequalities in this chapter is their application to adaptive estimation of the unknown drift coefficient b in (3.1.2). This is dealt with in Chapter 4. Using the presented results and techniques, we suggest a fully data-driven procedure which allows for rate-optimal estimation of the unknown drift wrt sup-norm risk and, at the same time, yields an asymptotically efficient estimator of the invariant density of the diffusion. The procedure is based on Lepski's method for adaptive estimation. In Chapter 4, we also deepen the analysis of the kernel density estimator started here. We derive a Donsker-type convergence result as it is relevant for the construction of (adaptive) confidence bands. Furthermore, we deal with the question of semi-parametric efficiency of the local time and the kernel density estimator in $\ell^\infty(\mathbb{R})$. These contributions heavily rely on the exponential inequality for the sup-norm difference between the local time and the kernel density estimator provided in Theorem 27. This result allows to transfer probabilistic knowledge on the local time to the more accessible and smoother kernel density estimator.

Apart from the apparent extensions to discrete observations of diffusion processes and multivariate state variables, further applications of the concentration inequalities derived in the current work could be found in the field of Bayesian statistical approaches, e.g., concerning supremum norm contraction rates. Another very interesting application of the proposed martingale approximation approach to concentration inequalities concerns bifurcating Markov chains. Bitseki Penda *et al.* (2017) construct adaptive nonparametric estimators of various quantities associated to bifurcating Markov chains. Crucial ingredient for their proofs are Bernstein-type deviation inequalities which in particular can be applied to well localised but unbounded functions. The corresponding findings are proven under a quite strong ergodicity assumption, and the authors suggest to use transportation-information inequalities for Markov chains for deriving similar results under more general conditions. Since the idea of martingale approximation is applicable in the Markov chain set-up, too, there is a natural starting point for the machinery developed in this paper, providing another alternative approach to (even uniform) deviation inequalities for bifurcating Markov chains.

3.2 Exponential tail inequality for the supremum of the local time of continuous semimartingales

Throughout this section, we work on a complete filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$, and we consider a continuous semimartingale X with canonical decomposition $X = X_0 + M + V$. Here, X_0 is an $\mathscr{F}_0$-measurable random variable, $M = (M_t)_{t\geq 0}$ denotes a continuous martingale with $M_0 = 0$ and $V = (V_t)_{t\geq 0}$ is a finite variation process with $V_0 = 0$. To shorten notation, we will often abbreviate

$$\|Y\|_p \;:=\; \left(\mathbb{E}\left[|Y|^p\right]\right)^{\frac{1}{p}}, \qquad \text{for } Y \in L^p(\mathbb{P}),\ p \geq 1.$$

For proving concentration inequalities for generalised additive functionals of the semimartingale X, we impose very general assumptions on the behaviour of the moments of the total variation of V and the quadratic covariation of M.

Assumption 2. There exist deterministic functions $\phi_1 \colon \mathbb{R}_+ \to \mathbb{R}_+$, $\phi_2 \colon \mathbb{R}_+ \to \mathbb{R}_+$ such that, for any $p \geq 1$,

$$\|X_0\|_p + \|X_t\|_p + \left\| \int_0^t |\mathrm{d}V_s| \right\|_p \;\leq\; p\phi_1(t), \qquad \left(\mathbb{E}\left[\langle M\rangle_t^{p/2}\right]\right)^{\frac{1}{p}} \;\leq\; \phi_2(t), \quad t > 0. \tag{3.2.10}$$

Here, $(\int_0^t |\mathrm{d}V_s|)_{t\geq 0}$ denotes the total variation process of V, and we write $|\mathrm{d}V_s|$ for integration with respect to the total variation measure of V. Furthermore, we assume that

$$\lim_{t\to\infty} \phi_1(t) \;=\; \infty \quad \text{ and } \quad \phi_2(t) \;\leq\; \sqrt{\phi_1(t)}.$$

With regard to our goal of proving tail estimates of the *supremum* of stochastic processes, we are interested in finding upper bounds for all p-th moments of

$$\sup_{a\in\mathbb{R}}|L_t^a(X)| \;=\; \|L_t^\bullet(X)\|_\infty.$$

The derivation of such uniform bounds is rather involved and comprises several steps. While the complete proof has been deferred to Section 3.7, it is instructive to sketch the main ideas now. A natural starting point is given by Tanaka's formula. Using (3.1.4) and then (3.2.10), one obtains a decomposition of the local time process which allows to derive the upper bound

$$\left(\mathbb{E}\left[\|L_t^\bullet(X)\|_\infty^p\right]\right)^{\frac{1}{p}} \;\leq\; 2p\phi_1(t) + 2\left(\mathbb{E}\left[\left(\sup_{a\in\mathbb{Q}} \mathbb{1}\left\{\max_{0\leq s\leq t}|X_s|\geq |a|\right\}|\mathbb{M}_t^a|\right)^p\right]\right)^{\frac{1}{p}}, \quad (3.2.11)$$

where $\mathbb{M}_t^a := \int_0^t \mathbb{1}\{X_s \leq a\}\mathrm{d}M_s$, $a\in\mathbb{R}$. Dealing with the sup-norm, it is crucial for the analysis to take into account the random, compact support of the local time in inequality (3.2.11). The size of the support depends on the extremal behaviour of the semimartingale, i.e., if $a\in\operatorname{supp}(L_t^\bullet(X))$, then necessarily $\max_{0\leq s\leq t}|X_s|\geq|a|$. This will allow to extend local arguments to the whole real line.

Coming back to (3.2.11), the main task now consists in controlling the martingale part appearing in the last summand, and it is classical to use the BDG inequality in this respect. The best constant in the BDG inequality is of order $O(\sqrt{p})$, and this fact plays an important role in our subsequent developments. More precisely, Proposition 4.2 in Barlow and Yor (1982) states that there exists a constant $\overline{\mathrm{c}}\geq 1$ such that, for any $p\geq 2$ and any continuous martingale $(N_t)_{t\geq 0}$ with $N_0=0$, one has

$$\left(\mathbb{E}\left[\left(\sup_{0\leq s\leq t}|N_s|\right)^p\right]\right)^{\frac{1}{p}} \;\leq\; \overline{\mathrm{c}}\sqrt{p}\left(\mathbb{E}\left[\langle N\rangle_t^{p/2}\right]\right)^{\frac{1}{p}}. \quad (3.2.12)$$

Consequently, whenever Assumption 2 holds true, one obtains for any $p\geq 1$

$$\left(\mathbb{E}\left[\left(\sup_{0\leq s\leq t}|M_s|\right)^p\right]\right)^{\frac{1}{p}} \;\leq\; \mathrm{c}\sqrt{p}\phi_2(t), \quad \text{with } \mathrm{c}:=\max\left\{1,\sqrt{2}\overline{\mathrm{c}}\right\}, \quad (3.2.13)$$

due to Hölder's inequality and (3.2.10). The upper bound (3.2.13) in particular allows to explore the tail behaviour of $(\mathbb{M}_t^a)_{a\in\mathbb{R}}$. A chaining procedure then yields an upper bound on the expectation on the rhs of (3.2.11) in terms of entropy integrals. This chaining procedure has to be done locally first since – in terms of the finiteness of covering numbers – the corresponding metric structure is not well behaved on the whole real line. Therefore, compact intervals of fixed length are considered, and it is taken into account that the probability of the support of the local time exceeding certain levels is vanishing (see Figure 3.1). The following maximal inequality for the process $(X_s)_{s\in[0,t]}$ allows to control this probability. Its short proof nicely illustrates the basic idea of how to exploit the moment bounds given in (3.2.10).

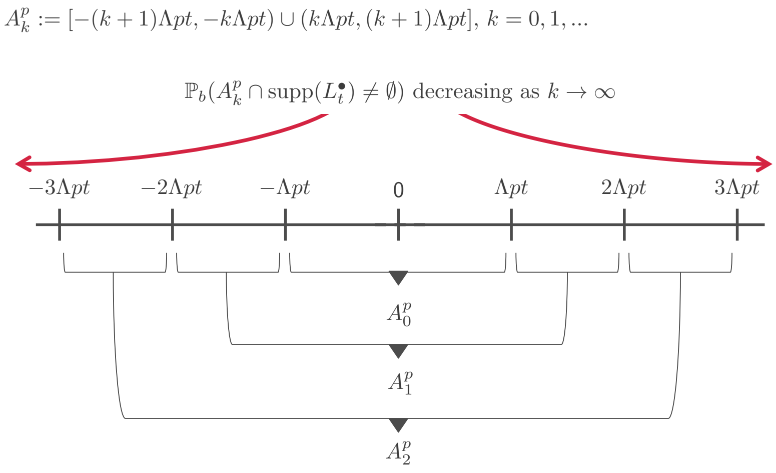

Figure 3.1: Localisation procedure

Lemma 14 (Maximal inequality for X)**.** *Under Assumption 2, it holds, for any $u \geq 1$,*

$$\mathbb{P}\left(\max_{0\leq s\leq t} |X_s| \geq \mathrm{e}\left(u\phi_1(t) + \mathrm{c}\sqrt{u}\phi_2(t)\right)\right) \;\leq\; \mathrm{e}^{-u}. \tag{3.2.14}$$

Proof. Note that

$$\max_{0\leq s\leq t} |X_s| \;\leq\; |X_0| + \int_0^t |\mathrm{d}V_s| + \max_{0\leq s\leq t} |M_s|.$$

Consequently, using (3.2.10) and (3.2.13), for any $p \geq 1$,

$$\left\| \max_{0\leq s\leq t} |X_s| \right\|_p \;\leq\; \phi_1(t)p + \mathrm{c}\sqrt{p}\phi_2(t).$$

Lemma 33 from Section 3.6 then gives (3.2.14). □

In particular, the maximal inequality (3.2.14) provides the final ingredient for verifying the main result of this section. Its complete proof is given in Section 3.7.

Theorem 15. *Consider a continuous semimartingale X with canonical decomposition $X = X_0 + M + V$, and grant Assumption 2. Then, there exists a positive constant κ (not depending on p) such that, for any $p \geq 1$,*

$$\left(\mathbb{E}\left[\|L_t^\bullet(X)\|_\infty^p\right]\right)^{\frac{1}{p}} \;\leq\; \kappa\left(p\phi_1(t) + \sqrt{p}\phi_2(t) + \left(\sqrt{\phi_1(t)} + \sqrt{\phi_2(t)}\right)\log(2p\Lambda(t))\right),$$

where $\Lambda(t) := \mathrm{e}\left(\phi_1(t) + \mathrm{c}\,\phi_2(t)\right)$. Consequently, for any $u \geq 1$,

$$\mathbb{P}\left(\|L_t^\bullet(X)\|_\infty \geq \mathrm{e}\kappa\left(u\phi_1(t) + \sqrt{u}\phi_2(t) + \left(\sqrt{\phi_1(t)} + \sqrt{\phi_2(t)}\right)\log(2u\Lambda(t))\right)\right) \;\leq\; \mathrm{e}^{-u}.$$

3.3 Uniform concentration of empirical processes of continuous semimartingales

In Section 3.2, we focused on analysing the sup-norm of the local time. Rephrasing the problem, we realise why the proof of Theorem 15 is a blueprint for investigating a much more general setting. Letting $\mathcal{F} := \left\{\mathbb{1}_{(-\infty,a]}(\cdot) : a \in \mathbb{R}\right\}$, Tanaka's formula and equation (3.2.11) reveal the core of the investigation: It consists in controlling

$$\sup_{a\in\mathbb{R}} \mathbb{M}_t^a \;=\; \sup_{a\in\mathbb{R}} \int_0^t \mathbb{1}\left\{X_s \le a\right\} \mathrm{d}M_s \;=\; \sup_{f\in\mathcal{F}} \int_0^t f(X_s)\mathrm{d}M_s.$$

Thus, the supremum of the process can be analysed within the framework of empirical processes and related concepts. The purpose of this section is to extend the study from the specific case of local time to additive functionals of the form $\sup_{f\in\mathcal{F}} \int_0^t f(X_s)\mathrm{d}s$ and further to stochastic integrals $\sup_{f\in\mathcal{F}} \int_0^t f(X_s)\mathrm{d}X_s$.

We start by investigating empirical processes of some continuous semimartingale X of the form

$$(\mathbb{G}_t^{b_0}(f))_{f\in\mathcal{F}} \;:=\; \left(\frac{1}{\sqrt{t}}\int_0^t \left(f(X_u)b_0(X_u) - \mathbb{E}[f(X_0)b_0(X_0)]\right)\mathrm{d}u\right)_{f\in\mathcal{F}}, \quad t>0, \tag{3.3.15}$$

indexed by a countable family $\mathcal{F} \subset L^2(\boldsymbol{\lambda})$, $\boldsymbol{\lambda}$ denoting the Lebesgue measure, and for a function $b_0 \colon \mathbb{R} \to \mathbb{R}$ satisfying

$$|b_0(x)| \;\le\; C(1+|x|^\eta), \tag{3.3.16}$$

$\eta \ge 0$, $C \ge 1$ some fixed constants. The main idea for deriving concentration inequalities is to use the technique of martingale approximation which was already introduced in Section 3.1 (cf. (3.1.5)) in a more systematic manner. While Theorem 15 for the local time concerns the supremum taken over the whole real line, we now turn to investigating suprema over general (possibly infinite-dimensional) function classes. For any semi-metric space $(\mathcal{F}, d)$, denote by $N(u, \mathcal{F}, d)$, $u > 0$, the covering number of $\mathcal{F}$ wrt d, i.e., the smallest number of balls of radius u in $(\mathcal{F}, d)$ needed to cover $\mathcal{F}$. Furthermore, we introduce

$$E(\mathcal{F}, d, \alpha) \;:=\; \int_0^\infty \left(\log N(u, \mathcal{F}, d)\right)^{\frac{1}{\alpha}} \mathrm{d}u, \quad \alpha > 0.$$

With regard to the indexing classes of functions $\mathcal{F}$ in (3.3.15), we impose the following basic conditions.

Assumption 3. $\mathcal{F}$ is a countable class of real-valued functions satisfying, for some fixed constants $U, \mathbb{V} > 0$,

$$\sup_{x\in\mathbb{R}} |f(x)| \;\le\; U, \quad \sup_{f\in\mathcal{F}} \|f\|_{L^2(\boldsymbol{\lambda})} \;\le \mathbb{V}.$$

In addition, all $f \in \mathcal{F}$ have compact support with

$$\mathrm{supp}(f) \subset [x_f, x^f], \text{ where } |x^f - x_f| \le \mathcal{S} \text{ and } \mathbb{V} \le \sqrt{\mathcal{S}}, \text{ for some } x_f < x^f,\ \mathcal{S} > 0.$$

Assumption 4. *$\mathcal{F}$ is a countable class of real-valued functions such that there exist constants $\mathrm{e}^2 < \mathbb{A} < \infty$ and $v \geq 2$ such that, for any probability measure $\mathbb{Q}$,*

$$\forall \varepsilon \in (0,1), \quad N\left(\varepsilon, \mathcal{F}, \|\cdot\|_{L^2(\mathbb{Q})}\right) \;\leq\; (\mathbb{A}/\varepsilon)^v. \tag{3.3.17}$$

Throughout the sequel, $\mathbb{C}_{\mathrm{mo}} > 0$ denotes a constant satisfying $\|X_0\|_p = (\mathbb{E}\left[|X_0|^p\right])^{\frac{1}{p}} \leq p\mathbb{C}_{\mathrm{mo}}$, $p \geq 1$. The existence of such a constant follows from Assumption 2. Furthermore, we use the notation $\sup_{f\in\mathcal{F}} |\mathbb{G}_t(f)| =: \|\mathbb{G}_t\|_{\mathcal{F}}$.

Theorem 16. *Let X be a continuous semimartingale as in Assumption 2, and let $b_0\colon \mathbb{R} \to \mathbb{R}$ be a function satisfying (3.3.16) for some constants $\eta \geq 0, C \geq 1$. Suppose that the function class $\mathcal{F}$ satisfies Assumption 3, and define $\mathbb{G}_t^{b_0}(\cdot)$ according to (3.3.15). Assume further that any $f \in \mathcal{F}$ admits a martingale approximation*

$$\mathbb{G}_t^{b_0}(f) \;=\; t^{-1/2}\mathbb{M}_t^f \;+\; t^{-1/2}\mathbb{R}_t^f, \quad t > 0, \tag{3.3.18}$$

for which there exist constants Ψ_1, Ψ_2 and some $\alpha > 0$ such that, for any $f, g \in \mathcal{F}$,

$$\begin{aligned}
\left(\mathbb{E}\left[|\mathbb{M}_t^f|^p\right]\right)^{\frac{1}{p}} &\leq \Psi_1\sqrt{t}p^{\frac{1}{\alpha}}\|f\|_{L^2(\boldsymbol{\lambda})}, \qquad \left(\mathbb{E}\left[\sup_{f\in\mathcal{F}}|\mathbb{R}_t^f|^p\right]\right)^{\frac{1}{p}} \leq \Psi_2 p,\\
\left(\mathbb{E}\left[|\mathbb{M}_t^f - \mathbb{M}_t^g|^p\right]\right)^{\frac{1}{p}} &\leq \Psi_1\sqrt{t}p^{\frac{1}{\alpha}}\|f-g\|_{L^2(\boldsymbol{\lambda})}.
\end{aligned} \tag{3.3.19}$$

For $k \in \mathbb{N}_0$ and fixed $p \geq 1$, define

$$\begin{aligned}
I_k &:= \Big(-2(k+1)p\Lambda(t),\; -2kp\Lambda(t)\Big] \cup \Big(2kp\Lambda(t),\; 2(k+1)p\Lambda(t)\Big] \oplus [-\mathcal{S}, \mathcal{S}],\\
\mathcal{F}_k &:= \{f \in \mathcal{F}\colon\, \mathrm{supp}(f) \subset I_k\},
\end{aligned} \tag{3.3.20}$$

with

$$\Lambda(t) \;:=\; \max\left\{\lambda\mathrm{e}\left(\phi_1(t) + \mathrm{c}\,\phi_2(t)\right), 1\right\} \tag{3.3.21}$$

and $\lambda > 1$ such that $\max\{\mathcal{S}, \mathrm{e}\mathbb{C}_{\mathrm{mo}}\} < p\Lambda(t)$, *for any $p, t \geq 1$. Then, for any $t, p \geq 1$, whenever*

$$\sum_{k=0}^{\infty} E(\mathcal{F}_k, \mathrm{e}\Psi_1\|\cdot\|_{L^2(\boldsymbol{\lambda})}, \alpha)\exp\left(-\frac{k}{2}\right) \;<\; \infty, \tag{3.3.22}$$

it holds

$$\begin{aligned}
\left(\mathbb{E}\left[\|\mathbb{G}_t^{b_0}\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}} \;\leq\; & C_\alpha \sum_{k=0}^{\infty} E(F_k, \mathrm{e}\Psi_1\|\cdot\|_{L^2(\boldsymbol{\lambda})}, \alpha)\exp\left(-\frac{k}{2}\right) + 6\Psi_1(2p)^{\frac{1}{\alpha}}\mathbb{V} + 2\frac{\Psi_2 p}{\sqrt{t}}\\
& + \sqrt{t}CU(1+2\eta\mathbb{C}_{\mathrm{mo}})^{\eta}\exp\left(-\frac{\Lambda(t)}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}\right).
\end{aligned}$$

A few comments on the above result are in order.

Remark 17. (a) It will be shown that there exists a broad class of ergodic diffusion processes admitting a decomposition of the form (3.3.18), with moments satisfying (3.3.19). In most cases, it is not that difficult to bound the moments of the remainder term $\mathbb{R}_t^f$, and usually the corresponding arguments already imply the *uniform* moment bounds required in (3.3.19). The analysis of the martingale part $\mathbb{M}_t^f$ is more challenging. Under the given assumptions, it suffices however to derive *non-uniform* upper bounds on $\|\mathbb{M}_t^f\|_p$. Theorem 16 then allows to translate these bounds into bounds on $\|\mathbf{sup}_{f\in\mathcal{F}}|\mathbb{G}_t^{b_0}(f)|\|_p$.

(b) Assumption (3.3.22) is a very weak one. In fact, we will show that the conditions of Theorem 16 and Assumption 4 on the function class $\mathcal{F}$ imply that (3.3.22) holds true for $\alpha\in\{2/3,1,2\}$ (cf. Lemma 36 in Section 3.6). Whenever $E(\mathcal{F}_k,\mathrm{e}\Psi_1\|\cdot\|_{L^2(\boldsymbol{\lambda})},\alpha)$ can be upper bounded independently of k, say $E(\mathcal{F}_k,\mathrm{e}\Psi_1\|\cdot\|_{L^2(\boldsymbol{\lambda})},\alpha)\leq\mathcal{E}(p,\alpha)$ for all $k\in\mathbb{N}_0$ and some finite constant $\mathcal{E}(p,\alpha)>0$, Theorem 16 yields

$$\left(\mathbb{E}\left[\|\mathbb{G}_t^{b_0}\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}} \leq 3C_\alpha\mathcal{E}(p,\alpha)+6\Psi_1(2p)^{\frac{1}{\alpha}}\mathbb{V}+2\frac{\Psi_2 p}{\sqrt{t}}+\sqrt{t}CU(1+2\eta\mathbb{C}_{\mathrm{mo}})^\eta\mathrm{e}^{-\frac{\Lambda(t)}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}}.$$

Lemma 36 provides such an upper bound $\mathcal{E}(p,\alpha)$ for $\alpha\in\{2/3,1,2\}$. Furthermore, in a lot of interesting instances (e.g., local time or the statistical application in Section 3.5), the function class $\mathcal{F}$ is translation invariant, i.e., for any constant $c\in\mathbb{R}$, $f\in\mathcal{F}$ implies that $f(\cdot+c)\in\mathcal{F}$. In that case, $E(\mathcal{F}_k,\mathrm{e}\Psi_1\|\cdot\|_{L^2(\boldsymbol{\lambda})},\alpha)$ does *not* depend on k, and the finiteness of this quantity entails (3.3.22).

(c) Instead of assuming X to be a continuous semimartingale fulfilling the moment bounds (3.2.10) in Assumption 2, one could also work with other classes of processes satisfying a maximal inequality as in Lemma 14 and allowing for a martingale approximation with moment bounds as in (3.3.19).

Proof of Theorem 16. Fix $p\geq 1$. The definition of $\Lambda(t)$ (cf. (3.3.21)) implies for any $k\in\mathbb{N}$, setting $u=kp$,

$$\mathrm{e}\left(u\phi_1(t)+\mathrm{c}\sqrt{u}\phi_2(t)\right) \leq kp\mathrm{e}\left(\phi_1(t)+\mathrm{c}\,\phi_2(t)\right) \leq kp\Lambda(t),$$

and consequently, according to Lemma 14,

$$\mathbb{P}\left(\max_{0\leq s\leq t}|X_s|>kp\Lambda(t)\right) \leq \exp(-kp).$$

Furthermore, since $\|X_0\|_p \leq p\mathbb{C}_{\mathrm{mo}}$, Lemma 33 yields

$$\mathbb{P}\left(|X_0| \geq \mathrm{e}\mathbb{C}_{\mathrm{mo}}\,u\right) \leq \exp(-u), \quad u\geq 1.$$

Set $A_f := \{\exists\, s\in[0,t]$ such that $X_s\in\mathrm{supp}(f)\}$, and note that, for $f\in\mathcal{F}_k$, $k\in\mathbb{N}$,

$$A_f \subset \left\{\max_{0\leq s\leq t}|X_s|\geq kp\Lambda(t)\right\} =: A_k,$$

since, for any $x \in \operatorname{supp}(f)$, $|x| \geq 2kp\Lambda(t) - \mathcal{S} \geq 2kp\Lambda(t) - kp\Lambda(t) = kp\Lambda(t)$. Let $\mathcal{F}_0^{\mathrm{c}} := \cup_{k=1}^{\infty}\mathcal{F}_k$. Note that, for $f \in \mathcal{F}_0^{\mathrm{c}}$,

$$\begin{aligned}
|\mathbb{E}\left[f(X_0)b_0(X_0)\right]| \leq\ & C\|f\|_\infty \mathbb{E}\left[(1+|X_0|)^\eta \mathbb{1}\left\{|X_0| \geq kp\Lambda(t)\right\}\right] \\
\leq\ & C\|f\|_\infty \left(\mathbb{E}\left[(1+|X_0|)^{2\eta}\right]\right)^{1/2} \left(\mathbb{P}\left(|X_0| \geq kp\Lambda(t)\right)\right)^{1/2} \\
\leq\ & C\|f\|_\infty (1+2\eta\mathbb{C}_{\mathrm{mo}})^\eta \exp\left(-\frac{\Lambda(t)}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}\right).
\end{aligned}$$

Consequently, it holds $\sqrt{t}\,|\mathbb{E}\left[f(X_0)b_0(X_0)\right]| \leq \sqrt{t}CU(1+2\eta\mathbb{C}_{\mathrm{mo}})^\eta \exp\left(-\frac{\Lambda(t)}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}\right)$. We thus obtain the following decomposition:

$$\begin{aligned}
\left(\mathbb{E}\left[\|\mathbb{G}_t^{b_0}\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}} \leq\ & \left(\mathbb{E}\left[\|\mathbb{G}_t^{b_0}\|_{\mathcal{F}_0}^p\right]\right)^{\frac{1}{p}} + \left(\mathbb{E}\left[\|\mathbb{G}_t^{b_0}\ \mathbb{1}(A_f)\|_{\mathcal{F}_0^{\mathrm{c}}}^p\right]\right)^{\frac{1}{p}} + \left(\mathbb{E}\left[\|\mathbb{G}_t^{b_0}\ \mathbb{1}(A_f^{\mathrm{c}})\|_{\mathcal{F}_0^{\mathrm{c}}}^p\right]\right)^{\frac{1}{p}} \\
=\ & \left(\mathbb{E}\left[\|\mathbb{G}_t^{b_0}\|_{\mathcal{F}_0}^p\right]\right)^{\frac{1}{p}} + \left(\mathbb{E}\left[\|\mathbb{G}_t^{b_0}\ \mathbb{1}(A_f)\|_{\mathcal{F}_0^{\mathrm{c}}}^p\right]\right)^{\frac{1}{p}} + \sqrt{t}\|\mathbb{E}\left[f(X_0)b_0(X_0)\right]\|_{\mathcal{F}_0^{\mathrm{c}}} \\
\leq\ & \left(\mathbb{E}\left[\|\mathbb{G}_t^{b_0}\|_{\mathcal{F}_0}^p\right]\right)^{\frac{1}{p}} + \left(\mathbb{E}\left[\|\mathbb{G}_t^{b_0}\ \mathbb{1}(A_f)\|_{\mathcal{F}_0^{\mathrm{c}}}^p\right]\right)^{\frac{1}{p}} \\
& \qquad + \sqrt{t}CU(1+2\eta\mathbb{C}_{\mathrm{mo}})^\eta \mathrm{e}^{-\frac{\Lambda(t)}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}}.
\end{aligned}$$

Regarding the first two terms in the last display, note that

$$\begin{aligned}
\left(\mathbb{E}\left[\|\mathbb{G}_t^{b_0}\|_{\mathcal{F}_0}^p\right]\right)^{\frac{1}{p}} \leq\ & \frac{1}{\sqrt{t}}\left\{\left(\mathbb{E}_b\left[\|\mathbb{M}_t^f\|_{\mathcal{F}_0}^p\right]\right)^{\frac{1}{p}} + \left(\mathbb{E}\left[\|\mathbb{R}_t^f\|_{\mathcal{F}_0}^p\right]\right)^{\frac{1}{p}}\right\}, \\
\left(\mathbb{E}\left[\|\mathbb{G}_t^{b_0}\ \mathbb{1}(A_f)\|_{\mathcal{F}_0^{\mathrm{c}}}^p\right]\right)^{\frac{1}{p}} \leq\ & \frac{1}{\sqrt{t}}\left\{\left(\mathbb{E}\left[\|\mathbb{M}_t^f\ \mathbb{1}(A_f)\|_{\mathcal{F}_0^{\mathrm{c}}}^p\right]\right)^{\frac{1}{p}} + \left(\mathbb{E}\left[\|\mathbb{R}_t^f\|_{\mathcal{F}_0^{\mathrm{c}}}^p\right]\right)^{\frac{1}{p}}\right\}.
\end{aligned}$$

Thus,

$$\left(\mathbb{E}\left[\|\mathbb{G}_t^{b_0}\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}} \leq A + B + \sqrt{t}CU(1+2\eta\mathbb{C}_{\mathrm{mo}})^\eta \exp\left(-\frac{\Lambda(t)}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}\right), \tag{3.3.23}$$

where

$$A := \frac{1}{\sqrt{t}}\left\{\left(\mathbb{E}\left[\|\mathbb{M}_t^f\|_{\mathcal{F}_0}^p\right]\right)^{\frac{1}{p}} + \left(\mathbb{E}\left[\|\mathbb{M}_t^f \mathbb{1}(A_f)\|_{\mathcal{F}_0^{\mathrm{c}}}^p\right]\right)^{\frac{1}{p}}\right\}, \quad B := \frac{2}{\sqrt{t}}\left(\mathbb{E}\left[\|\mathbb{R}_t^f\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}}.$$

Assumption (3.3.19) implies that, for any $f, g \in \mathcal{F}_k$, $\|\mathbb{M}_t^f\|_p \leq \Psi_1\sqrt{t}p^{1/\alpha}\mathbb{V}$, and the following tail estimate,

$$\mathbb{P}\left(|t^{-1/2}(\mathbb{M}_t^f - \mathbb{M}_t^g)| \geq d_2(f,g)u\right) \leq \exp\left(-u^\alpha\right), \quad u \geq 1,$$

where $d_2(f,g) := \mathrm{e}\Psi_1\|f-g\|_{L^2(\boldsymbol{\lambda})}$. Proposition 34 then yields, for any $k \in \mathbb{N}_0$, $q \geq 1$,

$$\begin{aligned}
\frac{1}{\sqrt{t}}\left(\mathbb{E}\left[\|\mathbb{M}_t^f\|_{\mathcal{F}_k}^q\right]\right)^{\frac{1}{q}} \leq\ & C_\alpha \int_0^\infty (\log N(u,\mathcal{F}_k,d_2))^{\frac{1}{\alpha}}\,\mathrm{d}u + \frac{2}{\sqrt{t}}\sup_{f\in\mathcal{F}_k}\|\mathbb{M}_t^f\|_q \\
\leq\ & C_\alpha \int_0^\infty (\log N(u,\mathcal{F}_k,d_2))^{\frac{1}{\alpha}}\,\mathrm{d}u + 2\Psi_1 q^{\frac{1}{\alpha}}\mathbb{V},
\end{aligned} \tag{3.3.24}$$

and, for all $k \in \mathbb{N}$,

$$\begin{aligned}\frac{1}{\sqrt{t}}\left(\mathbb{E}\left[\|\mathbb{M}_t^f \mathbb{1}(A_f)\|_{\mathcal{F}_0^c}^p\right]\right)^{\frac{1}{p}} &\leq \sum_{k=1}^{\infty} \frac{1}{\sqrt{t}}\left(\mathbb{E}\left[\|\mathbb{M}_t^f \mathbb{1}(A_k)\|_{\mathcal{F}_k}^p\right]\right)^{\frac{1}{p}} \\ &\leq \sum_{k=1}^{\infty} \frac{1}{\sqrt{t}}\left(\mathbb{E}\left[\|\mathbb{M}_t^f\|_{\mathcal{F}_k}^{2p}\right]\right)^{\frac{1}{2p}} \mathbb{P}(A_k)^{\frac{1}{2p}} \\ &\leq \sum_{k=1}^{\infty} \frac{1}{\sqrt{t}}\left(\mathbb{E}\left[\|\mathbb{M}_t^f\|_{\mathcal{F}_k}^{2p}\right]\right)^{\frac{1}{2p}} \exp\left(-\frac{k}{2}\right) \\ &\leq \sum_{k=1}^{\infty}\left[C_\alpha \int_0^\infty (\log N(u,\mathcal{F}_k,d_2))^{\frac{1}{\alpha}}\,\mathrm{d}u \mathrm{e}^{-\frac{k}{2}}\right] + 4\Psi_1(2p)^{\frac{1}{\alpha}}\mathbb{V}.\end{aligned} \tag{3.3.25}$$

Finally, the announced moment bound follows from (3.3.23), (3.3.24), (3.3.25) and (3.3.19). □

3.4 Concentration of measure and exponential inequalities for scalar ergodic diffusions

The original motivation for the present study was the question of deriving exponential inequalities for diffusion processes and associated additive functionals as they are constantly used for investigating (adaptive) statistical procedures. The current analysis has a much wider scope, and the results and methods of proof actually apply in a much more general framework. However, for clarity of presentation and in order not to lose the main ideas, we focus in the sequel on a specific class of diffusion processes. The results of this section take up those established in Section 3.2 (for local times) and Section 3.3 (for empirical processes) for the specific diffusion setting. In Section 3.4.3, we even go one step further and establish a concentration result for generalised empirical processes that involve stochastic integrals. We start with introducing our basic class of diffusion processes.

Definition 18. Let $\sigma \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R})$ and assume that, for some constants $\overline{\nu}, \underline{\nu} \in (0,\infty)$, σ satisfies $\underline{\nu} \leq |\sigma(x)| \leq \overline{\nu}$ for all $x \in \mathbb{R}$. For fixed constants $A, \gamma > 0$ and $\mathbb{C} \geq 1$, define the set $\Sigma = \Sigma(\mathbb{C}, A, \gamma, \sigma)$ as

$$\Sigma := \left\{b \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R})\colon |b(x)| \leq \mathbb{C}(1+|x|),\ \forall |x| > A\colon \frac{b(x)}{\sigma^2(x)}\operatorname{sgn}(x) \leq -\gamma\right\}. \tag{3.4.26}$$

Given σ as above and any $b \in \Sigma$, there exists a unique strong solution of the SDE (3.1.2) with ergodic properties and invariant density

$$\rho(x) = \rho_b(x) := \frac{1}{C_{b,\sigma}\sigma^2(x)} \exp\left(\int_0^x \frac{2b(y)}{\sigma^2(y)}\mathrm{d}y\right), \quad x \in \mathbb{R}, \tag{3.4.27}$$

with $C_{b,\sigma} := \int_{\mathbb{R}} \frac{1}{\sigma^2(u)} \exp\left(\int_0^u \frac{2b(y)}{\sigma^2(y)}\mathrm{d}y\right)\mathrm{d}u$ denoting the normalising constant. The invariant measure of the corresponding distribution and its distribution function will be denoted by $\mu = \mu_b$ and $F = F_b$, respectively, and we assume that the process is started in the equilibrium, i.e., $\xi \sim \mu_b$.

Our assumptions on the diffusion characteristics already impose some regularity on the invariant density ρ_b. More precisely, for any $b \in \Sigma(\mathbb{C}, A, \gamma, \sigma)$, $\sigma^2\rho_b$ is continuously differentiable and there exists a constant $\mathcal{L} > 0$ (depending only on $\mathbb{C}, A, \gamma, \underline{\nu}, \overline{\nu}$) such that

$$\sup_{b\in\Sigma(\mathbb{C},A,\gamma,\sigma)} \max\left\{\|\rho_b\|_\infty, \|(\sigma^2\rho_b)'\|_\infty\right\} \ < \ \mathcal{L} \tag{3.4.28}$$

and, for any $\theta > 0$, we have $\sup_{b\in\Sigma(\mathbb{C},A,\gamma,\sigma)} \sup_{x\in\mathbb{R}} \left\{|x|^\theta \rho_b(x)\right\} < \infty$. The analysis of the moments of functionals of the process X relies on upper bounds for the moments of the invariant measure. For any diffusion process X as in Definition 18, it holds

$$\sup_{b\in\Sigma(\mathbb{C},A,\gamma,\sigma)} \|X_0\|_p \ = \ \sup_{b\in\Sigma(\mathbb{C},A,\gamma,\sigma)} \left(\mathbb{E}_b\left[|X_0|^p\right]\right)^{\frac{1}{p}} \ \le \ \mathbb{C}_{\mathrm{mo}}p, \quad p \ge 1, \tag{3.4.29}$$

(cf. Lemma 32 in Section 3.6) for some positive constant $\mathbb{C}_{\mathrm{mo}}$. The above estimates will be used in the sequel without further notice.

Remark 19. A natural approach for analysing the supremum of processes of the form

$$\frac{1}{t}\int_0^t g(X_s)\mathrm{d}s \quad \text{or} \quad \frac{1}{t}\int_0^t g(X_s)\mathrm{d}X_s$$

over entire function classes consists in making use of well-known concentration results for additive diffusion functionals. For any nice diffusion X fulfilling Poincaré's inequality, it is actually known that, for any *bounded* function $g\colon \mathbb{R} \to \mathbb{R}$, one has a Bernstein-type tail estimate of the form

$$\mathbb{P}\left(\frac{1}{t}\int_0^t \left(g(X_s) - \mathbb{E}[g(X_0)]\right)\mathrm{d}s > r\right) \ \le \ \exp\left(-\frac{tr^2}{2(\mathrm{Var}(g) + c_P\|g\|_\infty r)}\right), \tag{3.4.30}$$

for $t, r > 0$ and c_P denoting the Poincaré constant. Given any class $\mathcal{G}$ of bounded functions $g\colon \mathbb{R} \to \mathbb{R}$ fulfilling (3.4.30), the above inequality implies that the process $(\mathbb{G}_t(g))_{g\in\mathcal{G}}$ exhibits a mixed tail behaviour wrt the metrics $d_1(g, g') := \|g - g'\|_\infty$ and $d_2(g, g') := \mathrm{Var}(g - g')$. Chaining procedures as they are used, e.g., for proving Theorem 15 then can be applied to obtain upper bounds of the form

$$\left(\mathbb{E}\left[\left(\sup_{g\in\mathcal{G}}|\mathbb{G}_t(g)|\right)^p\right]\right)^{\frac{1}{p}} \lesssim \frac{1}{\sqrt{t}}\int_0^\infty \log N(\varepsilon, \mathcal{G}, d_1)\mathrm{d}\varepsilon + \int_0^\infty \sqrt{\log N(\varepsilon, \mathcal{G}, d_2)}\mathrm{d}\varepsilon + \sqrt{p} + \frac{p}{\sqrt{t}}. \tag{3.4.31}$$

However, for any bounded $g \in \mathcal{G}$, one can also derive a decomposition of the form (3.1.5) where both the martingale part $M_t(g)$ and the remainder term $R_t(g)$ can be controlled similarly to the local time case.

We do not want to restrict to bounded drift terms $b\colon \mathbb{R} \to \mathbb{R}$. For analysing term (**I**) in (3.1.7), one thus actually requires results for *unbounded* functions $g = fb$. Using the method of transportation-information inequalities, Gao *et al.* (2013) establish Bernstein-type concentration inequalities in the spirit of (3.4.30) for unbounded functions $g\colon \mathbb{R} \to \mathbb{R}$. In principle, one might then deduce upper bounds similarly to (3.4.31). Note however that the results of Gao *et al.* (2013) apply only to a restricted class of diffusion processes. Furthermore, it is far from clear how the corresponding entropy integrals can be controlled, not to say the finiteness of the rhs of (3.4.31) is not at all clear.

In view of the aforementioned obstacles, we return to the alternative approach of proving concentration results via martingale approximation. In the sequel, we will specify the components of the decomposition (3.1.5) and derive upper bounds on the moments of the martingale and the remainder term for a broad class of ergodic diffusion processes.

3.4.1 Moment bounds and tail estimates for diffusion local time

We start with revisiting our result on local time and specifying it for the case of diffusion processes as introduced in Definition 18. Thus, we consider the diffusion local time process $(L_t^a(X))_{a\in\mathbb{R}}$, $t \geq 0$, which is continuous in a and t.

Bounding the moments of $\|L_t^{\bullet}\|_\infty$ by means of Theorem 15 In order to deduce a result by means of Theorem 15, we first argue that Assumption 2 is satisfied for any process X as in Definition 18. Indeed, the finite variation part in this set-up is given by the integrated drift term, i.e., $V_t = \int_0^t b(X_s)\mathrm{d}s$. We thus obtain for the total variation process $\int_0^t |\mathrm{d}V_s| \leq \int_0^t |b(X_s)|\mathrm{d}s\ \forall t \geq 0$. From the moment bounds of the invariant measure (3.4.29) and the at-most-linear-growth condition on $b \in \Sigma(\mathbb{C}, A, \gamma, \sigma)$, one might deduce that

$$\|X_0\|_p + \|X_t\|_p + \left\| \int_0^t |\mathrm{d}V_s| \right\|_p \ \leq\ 2\mathbb{C}_{\mathrm{mo}}p + t\mathbb{C}(1+\mathbb{C}_{\mathrm{mo}}p) \ \leq\ 4pt\mathbb{C}(1+\mathbb{C}_{\mathrm{mo}}).$$

Furthermore, $\left(\mathbb{E}_b[\langle \int_0^{\bullet} \sigma(X_s)\mathrm{d}W_s\rangle_t^{p/2}]\right)^{1/p} \leq \overline{\nu}\sqrt{t}$. Thus, setting $\phi_1(t) := \max(4\mathbb{C}(1+\mathbb{C}_{\mathrm{mo}}), \overline{\nu}^2)t$ and $\phi_2(t) := \overline{\nu}\sqrt{t}$, Assumption 2 is fulfilled. The function $t \mapsto \Lambda(t)$ from Theorem 15 and Theorem 16 takes the form

$$\Lambda(t) \ :=\ \lambda\mathrm{e}\left(\max(4\mathbb{C}(1+\mathbb{C}_{\mathrm{mo}}), \overline{\nu}^2)t + \mathrm{c}\,\overline{\nu}\sqrt{t}\right),$$

with $\lambda > 1$ such that $\max\{\mathcal{S}, \mathrm{e}\mathbb{C}_{\mathrm{mo}}\} < \lambda\mathrm{e}(\max(4\mathbb{C}(1+\mathbb{C}_{\mathrm{mo}}), \overline{\nu}^2) + \mathrm{c}\,\overline{\nu})$. Letting $\Lambda := \lambda\mathrm{e}(\max(4\mathbb{C}(1+\mathbb{C}_{\mathrm{mo}}), \overline{\nu}^2) + \mathrm{c}\,\overline{\nu})$, it holds $\Lambda(t) \leq \Lambda t$, $t \geq 1$, and all the previous proofs also work for Λt instead of $\Lambda(t)$ which we use in the following without further notice. Given these estimates, Corollary 9.10 in Le Gall (2016) now gives, for any $a \in \mathbb{R}$, $p \geq 1$ and $t > 0$,

$$\left(\mathbb{E}_b\left[(L_t^a(X))^p\right]\right)^{\frac{1}{p}} \ \leq\ \widetilde{c}_p\left(pt + \sqrt{t}\right),$$

$\widetilde{c}_p$ some (unspecified) positive constant depending on p. Application of Theorem 15 yields the sup-norm counterpart, namely, the following result for the supremum of diffusion local time.

Corollary 20. *Let X be a diffusion process as in Definition 18. Then, there is a positive constant κ (not depending on p) such that, for any $p, u, t \geq 1$,*

$$\sup_{b\in\Sigma(\mathbb{C},A,\gamma,\sigma)} \left(\mathbb{E}_b\left[\|L_t^\bullet(X)\|_\infty^p\right]\right)^{\frac{1}{p}} \leq \kappa\left(pt+\sqrt{pt}+\sqrt{t}\log t\right),$$
$$\mathbb{P}_b\left(\|L_t^\bullet(X)\|_\infty \geq \mathrm{e}\kappa\left(ut+\sqrt{ut}+\sqrt{t}\log t\right)\right) \leq \exp(-u).$$

3.4.2 Martingale approximation for additive functionals of diffusion processes

We now specify our analysis of empirical processes $(\mathbb{G}_t^{b_0}(f))_{f\in\mathcal{F}}$ as introduced in (3.3.15) to the ergodic diffusion case. Given some function class $\mathcal{F}$, denote $\overline{\mathcal{F}} := \{g-h\colon g,h\in\mathcal{F}\}$.

Proposition 21. *Let X be a diffusion as in Definition 18. Then, for any continuous function b_0 fulfilling $|b_0(x)| \leq C(1+|x|^\eta)$, $C,\eta \geq 0$ some fixed constants, and any class $\mathcal{F}$ of continuous functions $f\colon\mathbb{R}\to\mathbb{R}$ fulfilling Assumption 3, there exists a representation*

$$\mathbb{G}_t^{b_0}(f) \;=\; t^{-1/2}\mathbb{M}_t^f \;+\; t^{-1/2}\mathbb{R}_t^f, \quad t>0, \tag{3.4.32}$$

satisfying, for any $f,g\in\mathcal{F}$, $\mathbb{M}_t^{f-g} = \mathbb{M}_t^f - \mathbb{M}_t^g$. In addition, for any $p\geq 1$ and any $f\in\mathcal{F}\cup\overline{\mathcal{F}}$,

$$\begin{aligned}\left(\mathbb{E}_b\left[|\mathbb{M}_t^f|^p\right]\right)^{\frac{1}{p}} &\leq (2p)^{\eta+1/2}\sqrt{t\mathcal{S}}\|f\|_{L^2(\boldsymbol{\lambda})}\,\overline{\nu}\,\mathrm{c}\,(1+(\mathbb{C}_{\mathrm{mo}}\eta)^\eta)\,\overline{\Lambda}_{\mathrm{prox}},\\ \left(\mathbb{E}_b\left[\left(\sup_{f\in\mathcal{F}}|\mathbb{R}_t^f|\right)^p\right]\right)^{\frac{1}{p}} &\leq p^{\eta+1}\mathcal{S}\;4\max\left\{\mathbb{C}_{\mathrm{mo}}^{\eta+1},1\right\}(\eta+1)^\eta\overline{\Lambda}_{\mathrm{prox}},\end{aligned} \tag{I}$$

with

$$\begin{aligned}\overline{\Lambda}_{\mathrm{prox}}^2 \;:=\;& 16C^2\mathcal{L}C_{b,\sigma}^2\mathrm{e}^{2\underline{\nu}^{-2}\mathbb{C}(2A+A^2)}(1+\sup_{x\in\mathbb{R}}|x|^{2\eta}\rho_b(x)) && (3.4.33)\\ &+4C^2\max\{2^{2\eta},2\}\Big(2K^2\mathcal{L}(1+\sup_{x\in\mathbb{R}}|x|^{2\eta}\rho_b(x))\\ &\qquad+\underline{\nu}^{-2}\sup_{x\geq 0}\exp(-4\gamma x)x^{2\eta}+\underline{\nu}^{-2}\Big), && (3.4.34)\end{aligned}$$

for some constant $K = K(\mathbb{C},A,\gamma,\overline{\nu},\underline{\nu})$. For the particular case $b_0=b$, the representation satisfies, for any $p\geq 1$ and any $f\in\mathcal{F}\cup\overline{\mathcal{F}}$,

$$\begin{aligned}\left(\mathbb{E}_b\left[|\mathbb{M}_t^f|^p\right]\right)^{\frac{1}{p}} &\leq p\sqrt{t}\Gamma_{\mathrm{prox}}\|f\|_{L^2(\boldsymbol{\lambda})}\sqrt{2}\overline{\nu}\,\mathrm{c}(1+\mathcal{S}+\mathbb{C}_{\mathrm{mo}})^{1/2},\\ \left(\mathbb{E}_b\left[\left(\sup_{f\in\mathcal{F}}|\mathbb{R}_t^f|\right)^p\right]\right)^{\frac{1}{p}} &\leq p\Gamma_{\mathrm{prox}},\end{aligned} \tag{II}$$

with

$$\overline{\Gamma}^2_{\text{prox}} := 8\left\{\frac{1}{4}K^2\mathcal{L}^2 + \frac{\underline{\nu}^{-2}\mathbb{C}}{2}\left(1+\sup_{x\geq 0}\exp(-4\gamma x)x\right) + \mathcal{L}^2 C_{b,\sigma}^2 \mathrm{e}^{2\underline{\nu}^{-2}\mathbb{C}(2A+A^2)}\right\},$$

$$\Gamma_{\text{prox}} := 4U\mathbb{C}_{\text{mo}}\Bigg(2K\left(2\overline{\nu}^2\mathcal{L}+\mathbb{C}(1+A)\right) + 2C_{b,\sigma}\mathrm{e}^{\underline{\nu}^{-2}\mathbb{C}(2A+A^2)}\left(A\mathbb{C}(1+A)\mathcal{L}+\frac{\overline{\nu}^2\mathcal{L}}{2}\right)+1\Bigg).$$

Remark 22. The above result should be read carefully. We consider an arbitrary continuous function b_0, not necessarily of compact support, satisfying some polynomial growth condition. Our interest is in bounding the p-th moments of the empirical process $(\mathbb{G}_t^{b_0}(f))_{f\in\mathcal{F}}$, indexed by the functions $\mathcal{F}\ni f\colon\mathbb{R}\to\mathbb{R}$. Neglecting constants, the first approach to analysing the moments of the martingale and of the remainder term shows that, for any $p\geq 1$,

$$\frac{1}{\sqrt{t}}\|\mathbb{M}_t^f\|_p \lesssim p^{\eta+\frac{1}{2}}\sqrt{\mathcal{S}}\|f\|_{L^2(\boldsymbol{\lambda})}, \qquad \left\|\sup_{f\in\mathcal{F}}|\mathbb{R}_t^f|\right\|_p \lesssim p^{\eta+1}\mathcal{S}. \tag{3.4.35}$$

Specifying to the case $b_0=b$, one can exploit the basic relation $(\sigma^2\rho_b)'=2\rho_b b$. One then obtains bounds of the order

$$\frac{1}{\sqrt{t}}\|\mathbb{M}_t^f\|_p \lesssim p\|f\|_{L^2(\boldsymbol{\lambda})}, \qquad \left\|\sup_{f\in\mathcal{F}}|\mathbb{R}_t^f|\right\|_p \lesssim p. \tag{3.4.36}$$

Regarding the exponent of p, (3.4.36) is superior to the bound implied by (3.4.35) for the specific case $\eta=1$ (which corresponds to the standard at-most-linear-growth assumption on the drift term). However, it will be seen below that it might be advantageous to choose the upper bound (3.4.35) with $\eta=1$ for the martingale part. Note that this bound provides the factor $\sqrt{\mathcal{S}}$. In a number of statistical applications (e.g., the procedure that we have in mind), the support of the functions f from the class $\mathcal{F}$ vanishes. Consequently, the contribution of the factor $\sqrt{\mathcal{S}}$ is more beneficial than the improvement in the tail behaviour implied by (3.4.36).

3.4.3 Uniform concentration of empirical processes and stochastic integrals

Consider some diffusion process X as introduced in Definition 18 with invariant measure μ_b, and let us briefly recall our previous outcomes. Proposition 21 gives both a martingale approximation of the empirical process

$$\mathbb{G}_t^b(f) = \frac{1}{\sqrt{t}}\int_0^t \left(f(X_u)b(X_u)-\mathbb{E}_b[f(X_0)b(X_0)]\right)\mathrm{d}u$$

and bounds on the p-th moments of its martingale and remainder term. Theorem 16 allows to translate these bounds into bounds on $\|\mathbb{G}_t^b\|_{\mathcal{F}}^p$, $p\geq 1$, the supremum taken over entire function classes $\mathcal{F}$ and we obtain the following

Theorem 23. *Let X be as in Definition 18. Suppose that $\mathcal{F}$ is a class of continuous functions fulfilling Assumptions 3 and 4, and set* $\Lambda := \lambda \mathrm{e}(\max(4\mathbb{C}(1+\mathbb{C}_{\mathrm{mo}}), \overline{\nu}^2) + \mathrm{c}\,\overline{\nu})$, *where $\lambda > 1$ is chosen such that* $\max\{\mathcal{S}, \mathrm{e}\mathbb{C}_{\mathrm{mo}}\} < \lambda \mathrm{e}(\max(4\mathbb{C}(1+\mathbb{C}_{\mathrm{mo}}), \overline{\nu}^2) + \mathrm{c}\,\overline{\nu})$. *Then, for any $p \geq 1$,*

$$\sup_{b \in \Sigma(\mathbb{C}, A, \gamma, \sigma)} \left(\mathbb{E}_b\left[\|\mathbb{G}_t^1\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}} \;\leq\; \Phi_t(p), \tag{I}$$

for

$$\Phi_t(u) \;:=\; \mathbb{V}\sqrt{\mathcal{S}}\left\{12 C_2 \mathrm{e}\Pi_1 \sqrt{v \log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S} + u\Lambda t}\right)} + 6\Pi_1\sqrt{2u}\right\} + 2\mathcal{S}\frac{\Pi_2 u}{\sqrt{t}} + \sqrt{t} C U \mathrm{e}^{-\frac{\Lambda t}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}},$$

with $\Pi_1 := \sqrt{2}\overline{\nu}\,\mathrm{c}\,\overline{\Lambda}_{\mathrm{prox}}$ *and* $\Pi_2 := 4\max\{\mathbb{C}_{\mathrm{mo}}, 1\} U \overline{\Lambda}_{\mathrm{prox}}$. *Furthermore,*

$$\sup_{b \in \Sigma(\mathbb{C}, A, \gamma, \sigma)} \left(\mathbb{E}_b\left[\|\mathbb{G}_t^b\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}} \;\leq\; \Phi_t^b(p), \tag{II}$$

where

$$\begin{aligned}\Phi_t^b(u) := \mathbb{V}\sqrt{\mathcal{S}}\Bigg\{ 3C_{\frac{2}{3}}\mathrm{e}\Pi_1^b \left(2\left(v \log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S} + u\Lambda t}\right)\right)^{\frac{3}{2}} + 6v^{\frac{3}{2}}\sqrt{\log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S}+u\Lambda t}\right)}\right) \\ + 6\Pi_1^b (2u)^{\frac{3}{2}}\Bigg\} + \frac{2\Pi_2^b u}{\sqrt{t}} + \sqrt{t} C U (1 + 2\mathbb{C}_{\mathrm{mo}})\mathrm{e}^{-\frac{\Lambda t}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}}\end{aligned} \tag{3.4.37}$$

and $\Pi_1^b := 2^{\frac{3}{2}}\overline{\nu}\,\mathrm{c}\,\overline{\Lambda}_{\mathrm{prox}}(1 + \mathbb{C}_{\mathrm{mo}})$, $\Pi_2^b := \Gamma_{\mathrm{prox}}$.

Our interest finally is in formulating exponential inequalities for the process

$$\mathbb{H}_t(f) = \sqrt{t}\left(\frac{1}{t}\int_0^t f(X_s)\mathrm{d}X_s - \int (fb)\mathrm{d}\mu_b\right), \quad f \in \mathcal{F}. \tag{3.4.38}$$

At this point, we can apply several of our previous findings for proving one first uniform moment bound for the general stochastic integral process $(\mathbb{H}_t(f))_{f\in\mathcal{F}}$.

Proposition 24. *Grant the assumptions of Theorem 23. Then, there exists a positive constant $\mathbb{L}$ (depending only on* $\mathrm{c}, \mathbb{C}, \mathbb{C}_{\mathrm{mo}}, \Lambda, U, A, \gamma, v, \mathbb{A}, \underline{\nu}, \overline{\nu}$*) such that, for any $p, t \geq 1$,*

$$\sup_{b \in \Sigma(\mathbb{C}, A, \gamma, \sigma)} \left(\mathbb{E}_b\left[\|\mathbb{H}_t\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}} \;\leq\; \mathbb{L}\left(\mathbb{V}\left(1 + \log\left(\frac{1}{\mathbb{V}}\right) + \log t + p\right) + \frac{p}{\sqrt{t}} + \sqrt{t}\mathrm{e}^{-\frac{\Lambda t}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}}\right).$$

3.5 Statistical applications

This section considers the basic question of density estimation in supremum-norm which, from a general statistical point of view, is of immense theoretical and practical interest. Let us assume that a continuous record of observations $X^t := (X_s)_{0\leq s\leq t}$ of a diffusion process as introduced in Definition 18 is available, and we aim at nonparametric estimation of the associated invariant density ρ_b. Given some smooth kernel function $K\colon \mathbb{R}\to\mathbb{R}$ with compact support, define the standard kernel estimator $\rho_{t,K}(h)$ according to (3.1.9). For our statistical analysis which targets results concerning the risk in sup-norm loss, i.e., the behaviour of the maximal error $\|\rho_{t,K}(h)-\rho_b\|_\infty$, we impose some regularity on b and ρ_b. To be more precise, we look at Hölder classes defined as follows.

Definition 25. Given $\beta,\mathcal{L}>0$, denote by $\mathcal{H}_{\mathbb{R}}(\beta,\mathcal{L})$ the *Hölder class* (on $\mathbb{R}$) as the set of all functions $f\colon\mathbb{R}\to\mathbb{R}$ which are $l:=\lfloor\beta\rfloor$-times differentiable and for which

$$\begin{aligned}\|f^{(k)}\|_\infty &\leq \mathcal{L} \qquad \forall\, k=0,1,...,l,\\ \|f^{(l)}(\cdot+t)-f^{(l)}(\cdot)\|_\infty &\leq \mathcal{L}|t|^{\beta-l} \qquad \forall\, t\in\mathbb{R}.\end{aligned}$$

Set $\Sigma(\beta,\mathcal{L}) := \{b\in\Sigma(\mathfrak{C},A,\gamma,\sigma)\colon\ \rho_b\in\mathcal{H}_{\mathbb{R}}(\beta,\mathcal{L})\}$. Here, $\lfloor\beta\rfloor$ denotes the greatest integer strictly smaller than β.

Considering the class of drift coefficients $\Sigma(\beta,\mathcal{L})$, we use kernel functions satisfying the following assumptions,

- $K:\mathbb{R}\to\mathbb{R}$ is Lipschitz continuous and symmetric,
- $\operatorname{supp}(K)\subset[-1/2,1/2]$,
- K is of order β, i.e., $\int_{\mathbb{R}}K(y)\mathrm{d}y=1$, $\int_{\mathbb{R}}y^jK(y)\mathrm{d}y=0$, $j=1,\dots,\lfloor\beta\rfloor$, $\int_{\mathbb{R}}|y|^\beta|K(y)|\mathrm{d}y<\infty$. (3.5.39)

Corollary 26 (Concentration of the kernel invariant density estimator)**.** *Let X be a diffusion as in Definition 18 with $b\in\Sigma(\beta,\mathcal{L})$, for some $\beta,\mathcal{L}>0$, and let K be a kernel function fulfilling* (3.5.39)*. Given some positive bandwidth h, define the estimator $\rho_{t,K}(h)$ according to* (3.1.9)*. Then, there exist positive constants ν_1,ν_2 (not depending on p) such that, for any $p\geq 1$, $t>0$,*

$$\begin{aligned}\sup_{b\in\Sigma(\beta,\mathcal{L})}\left(\mathbb{E}_b\left[\|\rho_{t,K}(h)-\rho_b\|_\infty^p\right]\right)^{\frac1p} \leq\ & \frac{\nu_1}{\sqrt{t}}\left\{1+\sqrt{\log\left(\frac{1}{\sqrt{h}}\right)}+\sqrt{\log(pt)}+\sqrt{p}\right\}\\ &+\frac{\nu_2 p}{t}+\frac1h\mathrm{e}^{-\frac{\Lambda t}{2\mathrm{e}\mathfrak{C}_{\mathrm{mo}}}}+h^\beta\frac{\mathcal{L}}{\lfloor\beta\rfloor!}\int|u^\beta K(u)|\mathrm{d}u.\end{aligned}$$

Proof. We want to apply Theorem 23 to the class

$$\mathcal{F}:=\left\{K\left(\frac{x-\cdot}{h}\right)\,:\, x\in\mathbb{Q}\right\}.$$

For doing so, note that $\sup_{f\in\mathcal{F}}\|f\|_\infty \leq \|K\|_\infty$, $\sup_{f\in\mathcal{F}}\boldsymbol{\lambda}(\operatorname{supp}(f)) \leq h$ and

$$\left\|K\left(\frac{x-\cdot}{h}\right)\right\|_{L^2(\boldsymbol{\lambda})}^2 \;=\; \int K^2\left(\frac{x-y}{h}\right)\mathrm{d}y \;=\; h\int K^2(z)\mathrm{d}z \;\leq\; h\|K\|_{L^2(\boldsymbol{\lambda})}^2.$$

Setting $\mathcal{S} := h\ \max\{\|K\|_{L^2(\boldsymbol{\lambda})}^2, 1\}$, $\mathbb{V} := \sqrt{h}\|K\|_{L^2(\boldsymbol{\lambda})}$ and taking into account Lemma 35, $\mathcal{F}$ is seen to satisfy Assumptions 3 and 4. Thus, Theorem 23 with $b_0 = 1$ is applicable. In particular, there exist positive constants ν_1 and ν_2 such that, for any $p \geq 1$,

$$\begin{aligned}
&\left(\mathbb{E}_b\left[\left\|\frac{1}{th}\int_0^t K\left(\frac{x-X_u}{h}\right)\mathrm{d}u - \mathbb{E}_b\left[\frac{1}{h}K\left(\frac{x-X_0}{h}\right)\right]\right\|_\infty^p\right]\right)^{\frac{1}{p}}\\
&\quad= \frac{1}{\sqrt{th}}\left(\mathbb{E}_b\left[\left\|\sqrt{t}\left\{\frac{1}{t}\int_0^t K\left(\frac{x-X_u}{h}\right)\mathrm{d}u - \mathbb{E}_b\left[K\left(\frac{x-X_0}{h}\right)\right]\right\}\right\|_\infty^p\right]\right)^{\frac{1}{p}}\\
&\quad\leq \frac{1}{\sqrt{th}}\,h\nu_1\left\{1+\sqrt{\log\left(\frac{1}{\sqrt{h}}\right)}+\sqrt{\log(pt)}+\sqrt{p}\right\}+\frac{\nu_2 p}{t}+\frac{1}{h}\exp\left(-\frac{\Lambda t}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}\right).
\end{aligned}$$

For the bias, we obtain

$$\begin{aligned}
\left|\frac{1}{th}\mathbb{E}_b\left[\int_0^t K\left(\frac{x-X_u}{h}\right)\mathrm{d}u\right]-\rho_b(x)\right| \;&=\; \left|\frac{1}{h}\int K\left(\frac{x-y}{h}\right)(\rho_b(y)-\rho_b(x))\,\mathrm{d}y\right|\\
&\leq\; h^\beta\frac{\mathcal{L}}{\lfloor\beta\rfloor!}\int|u^\beta K(u)|\mathrm{d}u.
\end{aligned}$$

Combining the above estimates, the assertion follows. □

Recall that $L_t^x(X)$ denotes diffusion local time and that $\rho_t^\circ(x) = t^{-1}L_t^x(X)\sigma^{-2}(x)$ is the associated local time estimator of the value of the invariant density $\rho_b(x)$ of X. We now turn to deriving an exponential inequality for the tail probabilities of $\sqrt{t}\|\rho_{t,K}(h)-\rho_t^\circ\|_\infty$ which holds under rather mild assumptions on the diffusion X and the bandwidth h. It can be interpreted as some analogue of Theorem 1 in Giné and Nickl (2009) where the authors investigate the maximum deviation between the classical empirical distribution function (based on i.i.d. observations) and the distribution function obtained from kernel smoothing. The proof of Theorem 27 substantially relies on Proposition 24. Throughout the sequel, we restrict to a constant diffusion coefficient $\sigma^2 \equiv 1$ in order to ease the exposition. Still, our methods are suitable to treat more general diffusion coefficients under Hölder-smoothness conditions on σ^2 that correspond to the conditions on the invariant density.

Theorem 27. *Let X be a diffusion as in Definition 18 with $\sigma^2 \equiv 1$ and $b \in \Sigma(\beta,\mathcal{L})$, for some $\beta,\mathcal{L} > 0$. Consider some kernel function K fulfilling* (3.5.39) *and $h = h_t \in (0,1)$ such that $h_t \geq t^{-1}$. Then, there exist positive constants $\mathcal{V}$, Λ, Λ_0 and $\mathbb{L}$ such that, for all*

$$\begin{aligned}
\lambda \;\geq\; 8\Lambda_0\Bigg[&\sqrt{h}\mathcal{V}\mathrm{e}\mathbb{L}\left\{1+\log\left(\frac{1}{\sqrt{h}\mathcal{V}}\right)+\log t\right\}+\mathrm{e}\mathbb{L}\sqrt{t}\exp\left(-\frac{\Lambda t}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}\right)\\
&+\sqrt{t}h^\beta\frac{\mathcal{L}}{2\lfloor\beta\rfloor!}\int|K(v)v^\beta|\mathrm{d}v\Bigg]
\end{aligned}$$

and any $t > 1$,

$$\sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{P}_b\left(\sqrt{t}\,\|\rho_{t,K}(h) - \rho_t^\circ\|_\infty > \lambda\right) \;\le\; \exp\left(-\frac{\Lambda_1\lambda}{\sqrt{h}}\right). \tag{3.5.40}$$

One first application of Theorem 27 concerns the derivation of an upper bound on the sup-norm risk of the diffusion local time estimator. In fact, it allows to prove the following Corollary which shows that, concerning invariant density estimation, the procedures based on the kernel density and the local time estimator, respectively, are of equal quality in terms of sup-norm rates of convergence.

Corollary 28. *Let X be a diffusion as in Definition 18 with $\sigma^2 \equiv 1$. Then, there is a positive constant ζ such that, for any $p, t \ge 1$,*

$$\sup_{b\in\Sigma(\mathbb{C},A,\gamma,1)} \left(\mathbb{E}_b\left[\left\|\frac{L_t^\bullet(X)}{t} - \rho_b\right\|_\infty^p\right]\right)^{\frac{1}{p}} \;\le\; \zeta\left(\frac{p}{t} + \frac{1+\sqrt{p}+\sqrt{\log(pt)}}{\sqrt{t}} + t\mathrm{e}^{-\frac{\Lambda t}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}}\right). \tag{3.5.41}$$

In addition, for any $u \ge 1$,

$$\mathbb{P}_b\left(\|L_t^\bullet(X) - t\rho_b\|_\infty \ge \mathrm{e}\zeta\left(\sqrt{t}\left(1+\sqrt{\log(ut)}+\sqrt{u}\right) + u + t^2\mathrm{e}^{-\frac{\Lambda t}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}}\right)\right) \;\le\; \mathrm{e}^{-u}. \tag{3.5.42}$$

Remark 29. (a) As already indicated, the results yield the same sup-norm convergence rate for the local time and the kernel density estimator with bandwidth $t^{-1/2}$, i.e.,

$$\sup_{b\in\Sigma(\mathbb{C},A,\gamma,1)} \left(\mathbb{E}_b\left[\|\widetilde{\rho}_t - \rho_b\|_\infty^p\right]\right)^{\frac{1}{p}} = O\left(\left(\frac{\log t}{t}\right)^{1/2}\right), \text{ for } \widetilde{\rho}_t \in \left\{\rho_{t,K}(t^{-1/2}), t^{-1}L_t^\bullet(X)\right\}.$$

(b) The explicit dependence of the minimax upper bounds in Corollary 26 and Corollary 28 on p is crucial for further statistical applications such as adaptive drift estimation. As compared to Corollary 28, we do not have to impose additional smoothness assumptions on the drift coefficient for applying Corollary 26 since $b \in \Sigma(\mathbb{C}, A, \gamma, 1)$ implies that $b \in \Sigma(1, \mathcal{L})$.

(c) Since the local time estimator is unbiased, Corollary 28 can also be interpreted as a result on the centred local time, providing a concentration inequality of the form (3.5.42) which is of its own probabilistic interest.

Once the result for the centred local time stated in (3.5.41) is available, one can derive the following modified version of Proposition 24. In a number of concrete applications, this version can be considered as an improvement, even though we lose the subexponential behaviour. This is our price for obtaining a better upper bound in terms of the size $\mathcal{S}$ of the support of the functions from the function class $\mathcal{F}$. In our statistical application, the support is of size h_t with $h_t \downarrow 0$ as $t \to \infty$. Therefore, gaining another $\sqrt{\mathcal{S}}$ is more beneficial than the subexponential behaviour. Recall the definition of $\mathbb{H}_t$ in (3.4.38).

Theorem 30. *Let X be a diffusion as in Definition 18 with $\sigma^2 \equiv 1$, and grant the assumptions of Theorem 23. Then, for any $p, t \geq 1$, there exist constants $\widetilde{\mathbb{L}}$ and $\widetilde{\mathbb{L}}_0$ such that*

$$\sup_{b\in\Sigma(\mathbb{C},A,\gamma,1)} \left(\mathbb{E}_b\left[\|\mathbb{H}_t\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}} \leq \widetilde{\Psi}_t(p), \tag{3.5.43}$$

where

$$\begin{aligned}\widetilde{\Psi}_t(p) := \widetilde{\mathbb{L}}\Bigg\{ & \mathbb{V}\sqrt{\mathcal{S}}\Big\{\left(\log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S}+p\Lambda t}\right)\right)^{3/2} + \left(\log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S}+p\Lambda t}\right)\right)^{1/2} + p^{3/2}\Big\} + \frac{p}{\sqrt{t}} \\ & + \sqrt{t}\exp\left(-\widetilde{\mathbb{L}}_0 t\right) + \mathbb{V}\left(\log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S}+p\Lambda t}\right)\right)^{1/2} \\ & + \frac{\mathbb{V}}{t^{1/4}}\left(1+\log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S}+p\Lambda t}\right)\right) + \mathbb{V}\left\{\sqrt{p}+\frac{p}{t^{1/4}}\right\}\Bigg\}.\end{aligned}$$

Remark 31. As before, it is straightforward to translate the moment bound (3.5.43) into a corresponding upper tail bound by means of Lemma 33. The effectiveness of the obtained exponential inequalities is reinforced in Chapter 4 where we investigate the question of adaptive drift estimation. In this respect, Theorem 30 on stochastic integrals will be a crucial device.

3.6 Basic auxiliary results

We start with proving two auxiliary results which are frequently used in our analysis.

Lemma 32. *Let X be as in Definition 18. Then, there is a positive constant $\mathbb{C}_{\mathrm{mo}}$, depending only on $\mathbb{C}, A, \gamma, \underline{\nu}, \overline{\nu}$, such that*

$$\sup_{b\in\Sigma(\mathbb{C},A,\gamma,\sigma)} \|X_0\|_p = \sup_{b\in\Sigma(\mathbb{C},A,\gamma,\sigma)} \left(\mathbb{E}_b\left[|X_0|^p\right]\right)^{\frac{1}{p}} \leq \mathbb{C}_{\mathrm{mo}} p, \quad p \geq 1.$$

Proof. Note that

$$\begin{aligned}\sup_{b\in\Sigma(\mathbb{C},A,\gamma,\sigma)} \mathbb{E}_b\left[|X_0|^p\right] &= \sup_{b\in\Sigma(\mathbb{C},A,\gamma,\sigma)} \int |x|^p\rho_b(x)\mathrm{d}x \\ &\leq 2A^{p+1}\mathcal{L} + \int_A^\infty x^p\exp(-2\gamma(x-A))\mathrm{d}x \ \left(\rho_b(A)+\rho_b(-A)\right)\underline{\nu}^{-2}\overline{\nu}^2 \\ &\leq 2A^{p+1}\mathcal{L} + \underline{\nu}^{-2}\overline{\nu}^2(\rho_b(A) \\ &\qquad + \rho_b(-A))\left(2^{p-1}\frac{A^p}{2\gamma} + \frac{2^{p-1}}{(2\gamma)^{p+1}}\int_0^\infty x^p\mathrm{e}^{-x}\mathrm{d}x\right) \\ &= 2A^{p+1}\mathcal{L} + 2\mathcal{L}\underline{\nu}^{-2}\overline{\nu}^2\left(2^{p-1}\frac{A^p}{2\gamma} + \frac{2^{p-1}}{(2\gamma)^{p+1}}\Gamma(p+1)\right).\end{aligned}$$

Due to the formula of Stirling, we have

$$(\Gamma(p+1))^{\frac{1}{p}} \leq \sqrt{2\pi}\mathrm{e}(p+1)^{1+1/p} \leq \sqrt{2\pi}\mathrm{e}(p+1)\widetilde{c} \leq \sqrt{2\pi}2\mathrm{e}\widetilde{c}p$$

for a constant $\widetilde{c}$ such that $\sup_{p\geq 1}(p+1)^{1/p} \leq \widetilde{c}$. This gives the assertion. □

Lemma 33. *Let X be a real-valued random variable satisfying, for any $p \geq 1$, $(\mathbb{E}[|X|^p])^{\frac{1}{p}} \leq f(p)$, for some function $f\colon (0,\infty) \to (0,\infty)$. Then,*

$$\mathbb{P}(|X| \geq \mathrm{e}f(u)) \ \leq \ \exp(-u), \quad u \geq 1. \tag{3.6.44}$$

Proof. Fix $u \geq 1$. Then, for any $p \geq 1$,

$$\mathbb{P}(|X| \geq \mathrm{e}f(u)) \leq \frac{\mathbb{E}[|X|^p]}{\mathrm{e}^p f^p(u)} \ \leq \ \frac{f^p(p)}{\mathrm{e}^p f^p(u)}.$$

Setting $p := u$, we obtain (3.6.44). □

One central ingredient for the proof of our concentration inequalities are generic chaining results which go back to Talagrand (cf. Talagrand (1996) and Talagrand (2014)). We state a version of the results in Dirksen (2015) here which is adjusted to our needs. In particular, we bound the abstract truncated γ-functionals appearing in Dirksen (2015) by entropy integrals.

Proposition 34 (cf. Theorem 3.2 & 3.5 in Dirksen (2015))**.** *Consider a real-valued process $(X_f)_{f\in\mathcal{F}}$, defined on a semi-metric space $(\mathcal{F}, d)$.*

(a) *If there exists some $\alpha \in (0,\infty)$ such that*

$$\mathbb{P}(|X_f - X_g| \geq ud(f,g)) \ \leq \ 2\exp(-u^\alpha) \quad \forall f,g \in \mathcal{F},\ u \geq 1, \tag{3.6.45}$$

then there exists some constant $C_\alpha > 0$ (depending only on α) such that, for any $1 \leq p < \infty$,

$$\left(\mathbb{E}\left[\sup_{f\in\mathcal{F}} |X_f|^p\right]\right)^{\frac{1}{p}} \ \leq \ C_\alpha \int_0^\infty (\log N(u,\mathcal{F},d))^{\frac{1}{\alpha}}\,\mathrm{d}u + 2\sup_{f\in\mathcal{F}} (\mathbb{E}[|X_f|^p])^{\frac{1}{p}}. \tag{3.6.46}$$

(b) *If there exist semi-metrics d_1, d_2 on $\mathcal{F}$ such that*

$$\mathbb{P}\left(|X_f - X_g| \geq ud_1(f,g) + \sqrt{u}d_2(f,g)\right) \ \leq \ 2\mathrm{e}^{-u} \quad \forall f,g \in \mathcal{F},\ u \geq 1,$$

then there exist positive constants $\widetilde{C}_1, \widetilde{C}_2$ such that, for any $1 \leq p < \infty$,

$$\begin{aligned}\left(\mathbb{E}\left[\sup_{f\in\mathcal{F}} |X_f|^p\right]\right)^{\frac{1}{p}} \ \leq \ & \widetilde{C}_1 \int_0^\infty \log N(u,\mathcal{F},d_1)\mathrm{d}u \\ & +\widetilde{C}_2 \int_0^\infty \sqrt{\log N(u,\mathcal{F},d_2)}\mathrm{d}u + 2\sup_{f\in\mathcal{F}} (\mathbb{E}[|X_f|^p])^{\frac{1}{p}}.\end{aligned} \tag{3.6.47}$$

The entropy integrals appearing on the rhs of (3.6.46) and (3.6.47) will be controlled by means of the following lemmata.

Lemma 35. *Given some function of bounded variation $H\colon \mathbb{R}\to\mathbb{R}$ and $h>0$, let*

$$\mathcal{F} \;:=\; \mathcal{F}_h \;=\; \left\{H\left(\frac{x-\cdot}{h}\right) : x\in\mathbb{R}\right\}.$$

Then there exist some constants $A=A(\|H\|_{\mathrm{TV}})<\infty$ and $v\geq 2$, not depending on h, such that, for any probability measure $\mathbb{Q}$ on $\mathbb{R}$ and any $0<\varepsilon<1$,

$$N(\varepsilon,\mathcal{F}_h,\|\cdot\|_{L^2(\mathbb{Q})}) \;\leq\; (A/\varepsilon)^v.$$

The preceding lemma is a consequence of the more general result of Proposition 3.6.12 in Giné and Nickl (2016).

Lemma 36. *Grant the conditions of Theorem 16 and Assumption 4, and define the function classes $\mathcal{F}_k$ according to (3.3.20). Then, for all $k\in\mathbb{N}_0$ and any constant $\Gamma\geq 1$,*

$$\begin{aligned}
\int_0^\infty \log N(u,\mathcal{F}_k,\Gamma\|\cdot\|_{L^2(\boldsymbol{\lambda})})\mathrm{d}u &\leq 2v\mathbb{V}\Gamma\left(1+\log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S}+p\Lambda(t)}\right)\right),\\
\int_0^\infty \sqrt{\log N(u,\mathcal{F}_k,\Gamma\|\cdot\|_{L^2(\boldsymbol{\lambda})})}\mathrm{d}u &\leq 4\mathbb{V}\Gamma\sqrt{v\log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S}+p\Lambda(t)}\right)},\\
\int_0^\infty \left(\log N(u,\mathcal{F}_k,\Gamma\|\cdot\|_{L^2(\boldsymbol{\lambda})})\right)^{3/2}\mathrm{d}u &\leq 2\mathbb{V}\Gamma\left(v\log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S}+p\Lambda(t)}\right)\right)^{3/2}\\
&\quad +6v\mathbb{V}\Gamma\sqrt{v\log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S}+p\Lambda(t)}\right)}.
\end{aligned}$$

Proof. Note that, for $f\in\mathcal{F}_k$,

$$\|f\|_{L^2(\boldsymbol{\lambda})} \;\leq\; \|f\|_{L^2(\nu_k)}\sqrt{4\mathcal{S}+4p\Lambda(t)}, \qquad \text{where } \mathrm{d}\nu_k \;=\; \mathbb{1}\{I_k\}\mathrm{d}\frac{\boldsymbol{\lambda}}{\boldsymbol{\lambda}(I_k)}.$$

Thus, (3.3.17) implies that

$$\begin{aligned}
N\left(u,\mathcal{F}_k,\Gamma\|\cdot\|_{L^2(\boldsymbol{\lambda})}\right) &\leq N\left(u\left(\Gamma\sqrt{4\mathcal{S}+4p\Lambda(t)}\right)^{-1},\mathcal{F}_k,\|\cdot\|_{L^2(\nu_k)}\right)\\
&\leq \left(\frac{\mathbb{A}\Gamma}{u}\sqrt{4\mathcal{S}+4p\Lambda t}\right)^v,
\end{aligned}$$

if $u<2\mathbb{V}\Gamma\leq\Gamma\sqrt{4\mathcal{S}+4p\Lambda(t)}$. Furthermore, since $\sup_{f,g\in\mathcal{F}}\|f-g\|_{L^2(\boldsymbol{\lambda})}\leq 2\mathbb{V}$, it holds that $N(u,\mathcal{F}_k,\Gamma\|\cdot\|_{L^2(\boldsymbol{\lambda})}) \;=\; 1$ for $u\geq 2\mathbb{V}\Gamma$. Thus, for $\alpha=1$, we can upper bound the entropy integral as follows,

$$\begin{aligned}
\int_0^\infty \log N(u,\mathcal{F}_k,\Gamma\|\cdot\|_{L^2(\boldsymbol{\lambda})})\mathrm{d}u &\leq \int_0^{2\mathbb{V}\Gamma} v\log\left(\frac{\mathbb{A}\Gamma}{u}\sqrt{4\mathcal{S}+4p\Lambda(t)}\right)\mathrm{d}u\\
&= v\left[u\log\left(\frac{\mathbb{A}\Gamma}{u}\sqrt{4\mathcal{S}+4p\Lambda(t)}\right)\right]_0^{2\mathbb{V}\Gamma}+2v\mathbb{V}\Gamma\\
&= 2v\mathbb{V}\Gamma\left(1+\log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S}+p\Lambda(t)}\right)\right).
\end{aligned}$$

For $\alpha = 2$, it holds

$$\begin{aligned}\int_0^\infty \sqrt{\log N(u, \mathcal{F}_k, \Gamma\|\cdot\|_{L^2(\boldsymbol{\lambda})})}\mathrm{d}u &\leq \int_0^{2\mathbb{V}\Gamma} \sqrt{v}\sqrt{\log\left(\frac{\mathbb{A}\Gamma}{u}\sqrt{4\mathcal{S}+4p\Lambda(t)}\right)}\mathrm{d}u \\ &\leq \sqrt{v}4\mathbb{V}\Gamma\left(\log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S}+p\Lambda(t)}\right)\right)^{1/2},\end{aligned}$$

where the last estimate is due to the fact that $\int_0^c \sqrt{\log(C/x)}\mathrm{d}x \leq 2c\sqrt{\log(C/c)}$ for $\log(C/c) \geq 2$ (see, e.g., Giné and Nickl (2009), p. 591). This last condition is fulfilled in our situation since $\mathbb{V} \leq \sqrt{\mathcal{S}}$ and $\mathbb{A} > \mathrm{e}^2$. Finally, if $\alpha = 2/3$,

$$\begin{aligned}\int_0^\infty \left(\log N(u, \mathcal{F}_k, \Gamma\|\cdot\|_{L^2(\boldsymbol{\lambda})})\right)^{3/2}\mathrm{d}u &\leq v^{3/2}\int_0^{2\mathbb{V}\Gamma}\left(\log\left(\frac{\mathbb{A}\Gamma}{u}\sqrt{4\mathcal{S}+4p\Lambda(t)}\right)\right)^{3/2}\mathrm{d}u \\ &= v^{3/2}\, u\left(\log\left(\frac{\mathbb{A}\Gamma}{u}\sqrt{4\mathcal{S}+4p\Lambda(t)}\right)\right)^{3/2}\Bigg|_0^{2\mathbb{V}\Gamma} \\ &\quad + v^{3/2}\int_0^{2\mathbb{V}\Gamma}\frac{3}{2}\left(\log\left(\frac{\mathbb{A}}{u}\Gamma\sqrt{4\mathcal{S}+4p\Lambda(t)}\right)\right)^{1/2}\mathrm{d}u \\ &\leq v^{3/2}2\mathbb{V}\Gamma\left(\log\left(\frac{\mathbb{A}}{2\mathbb{V}}\sqrt{4\mathcal{S}+4p\Lambda(t)}\right)\right)^{3/2} \\ &\quad + v^{3/2}\frac{3}{2}4\mathbb{V}\Gamma\left(\log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S}+p\Lambda(t)}\right)\right)^{1/2}.\end{aligned}$$

□

3.7 Proofs for Section 3.2

Proof of Theorem 15. Tanaka's formula (see Proposition 9.2 in Le Gall (2016)) yields the local time representation

$$\begin{aligned}L_t^a(X) &= L_t^a(X)\cdot\mathbb{1}\left\{\max_{0\leq s\leq t}|X_s| \geq |a|\right\} \\ &= 2\left((X_t-a)^- - (X_0-a)^- + \int_0^t \mathbb{1}\{X_s \leq a\}\mathrm{d}X_s\right),\end{aligned}$$

where $x^- := \max\{-x,\ 0\}$. Since semimartingale local time is càdlàg in a, the sup-norm actually refers to a supremum over the rationals $\mathbb{Q}$. In particular, $\|L_t^\bullet(X)\|_\infty$ is measurable. Furthermore, for any $t > 0$ and $p \geq 1$,

$$\begin{aligned}
\left(\mathbb{E}\left[\left(\sup_{a\in\mathbb{Q}}|L_t^a(X)|\right)^p\right]\right)^{\frac{1}{p}} &= \left(\mathbb{E}\left[\left(\sup_{a\in\mathbb{Q}}\left\{|L_t^a(X)|\cdot\mathbb{1}\left\{\max_{0\leq s\leq t}|X_s|\geq|a|\right\}\right\}\right)^p\right]\right)^{\frac{1}{p}} \\
&\leq 2\Bigg(\mathbb{E}\Bigg[\Bigg(\sup_{a\in\mathbb{Q}}\Bigg\{|X_t-X_0|+\int_0^t|\mathrm{d}V_s| \\
&\qquad +\mathbb{1}\left\{\max_{0\leq s\leq t}|X_s|\geq|a|\right\}\left|\int_0^t\mathbb{1}\{X_s\leq a\}\mathrm{d}M_s\right|\Bigg\}\Bigg)^p\Bigg]\Bigg)^{\frac{1}{p}} \qquad (3.7.48)\\
&\leq 2p\phi_1(t)+2\left(\mathbb{E}\left[\sup_{a\in\mathbb{Q}}\mathbb{1}\left\{\max_{0\leq s\leq t}|X_s|\geq|a|\right\}\left|\int_0^t\mathbb{1}\{X_s\leq a\}\mathrm{d}M_s\right|\right\}\right)^p\right]\right)^{\frac{1}{p}},
\end{aligned}$$

where the latter inequality is due to (3.2.10). Recall that $\mathbb{M}_t^a = \int_0^t \mathbb{1}\{X_s \leq a\}\mathrm{d}M_s$, $a \in \mathbb{R}$, and note again that (3.2.12) and (3.2.10) imply that

$$\sup_{a\in\mathbb{Q}}\left(\mathbb{E}\left[|\mathbb{M}_t^a|^p\right]\right)^{\frac{1}{p}} \;\leq\; \sup_{a\in\mathbb{Q}}\sqrt{2}\overline{\mathrm{c}}\sqrt{p}\left(\mathbb{E}\left[\langle\mathbb{M}_t^a\rangle^p\right]\right)^{\frac{1}{2p}} \;\leq\; \mathrm{c}\,\sqrt{p}\phi_2(t), \quad p\geq 1. \qquad (3.7.49)$$

This result provides an upper bound for the expression appearing on the rhs of (3.6.46) in Proposition 34. In order to apply this result, we still have to verify the condition on $\mathbb{M}$, i.e., we have to find a suitable metric structure. For analysing the expression

$$\left|\int_0^t \mathbb{1}\{a < X_s \leq b\}\mathrm{d}M_s\right| \;=\; \left|\mathbb{M}_t^a - \mathbb{M}_t^b\right|, \quad a \leq b,$$

we require an exponential inequality for the tail probability of these increments. We will deduce this inequality by investigating the corresponding moments. The derivation of the upper bounds relies heavily on the following auxiliary result.

Lemma 37 (cf. Lemma 9.5 in Le Gall (2016)). *Consider a continuous semimartingale X satisfying Assumption 2, and write $X = X_0 + M + V$ for its canonical decomposition. Let $p \geq 1$. Then, for every $a, b \in \mathbb{R}$ with $a \leq b$ and every $t \geq 0$, we have*

$$\mathbb{E}\left[\left(\int_0^t \mathbb{1}\{a < X_s \leq b\}\mathrm{d}\langle M\rangle_s\right)^p\right] \;\leq\; 2(16(b-a))^p\left\{\mathrm{c}^p\,p^{p/2}\phi_2^p(t) \;+\; \mathbb{E}\left[\left(\int_0^t|\mathrm{d}V_s|\right)^p\right]\right\}.$$

Now, for any $a \le b \in \mathbb{R}$ and $p \ge 1$, Lemma 37 and (3.2.12) give

$$\begin{aligned}
&\mathbb{E}\left[|\mathbb{M}_t^a - \mathbb{M}_t^b|^p\right] \\
&\quad\le \sqrt{\mathbb{E}\left[\left|\int_0^t \mathbb{1}\{a < X_s \le b\}\mathrm{d}M_s\right|^{2p}\right]} \\
&\quad\le \mathrm{c}^p\, p^{p/2}\sqrt{\mathbb{E}\left[\left(\int_0^t \mathbb{1}\{a < X_s \le b\}\mathrm{d}\langle M\rangle_s\right)^p\right]} \\
&\quad\le \min\Bigg\{2(16(b-a))^{p/2}\left(\mathrm{c}^{3p/2}\, p^{3p/4}\phi_2^{p/2}(t) + \mathrm{c}^p\, p^{p/2}\sqrt{\mathbb{E}\left[\left(\int_0^t |\mathrm{d}V_s|\right)^p\right]}\right), \\
&\qquad\qquad \mathrm{c}^p\, p^{p/2}\phi_2^p(t)\Bigg\} \\
&\quad\le \min\left\{2(16(b-a))^{p/2}\left(\mathrm{c}^{3p/2}\, p^{3p/4}\phi_2^{p/2}(t) + \mathrm{c}^p\, p^p\phi_1^{p/2}(t)\right),\ \mathrm{c}^p\, p^{p/2}\phi_2^p(t)\right\}
\end{aligned}$$

such that

$$\left(\mathbb{E}\left[|\mathbb{M}_t^a - \mathbb{M}_t^b|^p\right]\right)^{1/p} \ \le\ p\min\left\{8\sqrt{|a-b|},\ 1\right\}\left(\mathrm{c}\sqrt{\phi_1(t)} + \mathrm{c}^{3/2}\sqrt{\phi_2(t)}\right).$$

Consequently (cf. Lemma 33), the process $(\mathbb{M}_t^a)_{a\in\mathbb{R}}$ exhibits a subexponential tail behaviour wrt the metric d_1, defined as

$$d_1(a,b) \ :=\ \min\left\{8\sqrt{|a-b|},\ 1\right\}\ \mathrm{e}\left(\mathrm{c}\sqrt{\phi_1(t)} + \mathrm{c}^{3/2}\sqrt{\phi_2(t)}\right), \quad a, b \in \mathbb{R},$$

that is,

$$\mathbb{P}\left(|\mathbb{M}_t^a - \mathbb{M}_t^b| \ge d_1(a,b)u\right) \ \le\ \exp(-u), \quad u \ge 1. \tag{3.7.50}$$

At this point, we would like to apply Proposition 34. Since the entire real line $\mathbb{R}$ cannot be covered with a finite number of d_1-balls, we will use the maximal inequality (3.2.14) in order to apply the chaining procedure *locally* on finite intervals. For setting up the localisation procedure, fix $p_0 \ge 1$, and introduce the intervals

$$\begin{aligned}
A_0^{p_0} &:= [-p_0\Lambda(t),\ p_0\Lambda(t)], \\
A_k^{p_0} &:= \big[-(k+1)p_0\Lambda(t),\ -kp_0\Lambda(t)\big) \cup \big(kp_0\Lambda(t),\ (k+1)p_0\Lambda(t)\big], \quad k \in \mathbb{N},
\end{aligned}$$

with $\Lambda(t) \ \equiv\ \mathrm{e}\,(\phi_1(t) + \mathrm{c}\,\phi_2(t))$.

Lemma 38. *Define* $d_1(t) := \mathrm{e}\left(\mathrm{c}\sqrt{\phi_1(t)} + \mathrm{c}^{3/2}\sqrt{\phi_2(t)}\right)$*. For the* d_1*-entropy integrals of* $A_k^{p_0}$*,* $k \in \mathbb{N}_0$*, the following bound (not depending on* k*) holds true,*

$$\int_0^\infty \log N(u, A_k^{p_0}, d_1)\mathrm{d}u \ \le\ d_1(t)\left(4 + 2\log\left(8\sqrt{2p_0\Lambda(t)}\right)\right).$$

Proof. Fix $k \in \mathbb{N}_0$. Given any $\varepsilon \in (0, d_1(t))$, a decomposition of the sets $A_k^{p_0}$ into intervals of length $(\varepsilon/(8d_1(t)))^2$ gives $N(\varepsilon, A_k, d_1) \ \le\ 2p_0\Lambda(t)\,(\varepsilon/(8d_1(t)))^{-2} + 1$. Moreover, it is clear that $N(\varepsilon, A_k^{p_0}, d_1) = 1$ for any $\varepsilon \geq d_1(t)$. For the entropy integral, we thus obtain the estimate

$$\int_0^\infty \log N(u, A_k^{p_0}, d_1)\mathrm{d}u \;\leq\; 2\int_0^{d_1(t)} \log\left(\sqrt{2p_0\Lambda(t)}\,\frac{8d_1(t)}{u}+1\right)\mathrm{d}u$$
$$\leq\; d_1(t)\left(4+2\log\left(8\sqrt{2p_0\Lambda(t)}\right)\right).$$

□

Taking into account Proposition 34, (3.7.49), (3.7.50) and the previous lemma allow to deduce the local result. For every $k \in \mathbb{N}_0$, $p_0, p \geq 1$, we obtain

$$\left(\mathbb{E}\left[\sup_{a\in A_k^{p_0}}\left|\int_0^t \mathbb{1}\{X_s \leq a\}\mathrm{d}M_s\right|^p\right]\right)^{\frac{1}{p}} \;\leq\; C_1 d_1(t)\left(4+2\log\left(8\sqrt{2p_0\Lambda(t)}\right)\right) + 2\,\mathrm{c}\,\sqrt{p}\phi_2(t).$$

Exploiting the fact that the probability that the support of the local time intersects with the sets $A_k^{p_0}$ vanishes, we can extend this result to the whole real line. Precisely, we use that, for any $k \in \mathbb{N}$, setting $u \equiv kp_0$,

$$\mathrm{e}\left(u\phi_1(t)+\mathrm{c}\,\sqrt{u}\phi_2(t)\right) \;\leq\; kp_0\mathrm{e}\left(\phi_1(t)+\mathrm{c}\,\phi_2(t)\right) \;=\; kp_0\Lambda(t),$$

and consequently, according to Lemma 14,

$$\mathbb{P}\left(\max_{0\leq s\leq t}|X_s| > kp_0\Lambda(t)\right) \;\leq\; \exp\left(-kp_0\right).$$

Moreover, it holds

$$\sum_{k=1}^\infty \exp\left(-\frac{k}{2}\right) = \sum_{k=0}^\infty \exp\left(-\frac{(k+1)}{2}\right) \;\leq\; \int_0^\infty \exp\left(-\frac{x}{2}\right)\mathrm{d}x \;=\; 2\int_0^\infty \mathrm{e}^{-y}\mathrm{d}y \;=\; 2.$$

Coming back to the decomposition (3.7.48), we finish the proof by noting that, for any $p_0 \geq 1$,

$$
\begin{aligned}
&\left(\mathbb{E}\left[\left(\sup_{a\in\mathbb{Q}}\left\{|\mathbb{M}_t^a|\cdot\mathbb{1}\left\{\max_{0\leq s\leq t}|X_s|\geq|a|\right\}\right\}\right)^{p_0}\right]\right)^{\frac{1}{p_0}}\\
&\leq \left(\mathbb{E}\left[\left(\sup_{a\in A_0^{p_0}}|\mathbb{M}_t^a|\right)^{p_0}\right]\right)^{\frac{1}{p_0}}\\
&\quad+\sum_{k=1}^{\infty}\left(\mathbb{E}\left[\left(\sup_{a\in A_k^{p_0}}\left\{|\mathbb{M}_t^a|\cdot\mathbb{1}\left\{\max_{0\leq s\leq t}|X_s|\geq|kp_0\Lambda(t)|\right\}\right\}\right)^{p_0}\right]\right)^{\frac{1}{p_0}}\\
&\leq \left(\mathbb{E}\left[\left(\sup_{a\in A_0^{p_0}}|\mathbb{M}_t^a|\right)^{p_0}\right]\right)^{\frac{1}{p_0}}\\
&\quad+\sum_{k=1}^{\infty}\left[\left(\mathbb{E}\left[\sup_{a\in A_k^{p_0}}|\mathbb{M}_t^a|^{2p_0}\right]\right)^{\frac{1}{2p_0}}\mathbb{P}\left(\max_{0\leq s\leq t}|X_s|\geq|kp_0\Lambda(t)|\right)^{\frac{1}{2p_0}}\right]\\
&\leq C_1 d_1(t)\left(4+2\log\left(8\sqrt{2p_0\Lambda(t)}\right)\right)+2\,\mathrm{c}\,\sqrt{p_0}\phi_2(t)\\
&\quad+\sum_{k=1}^{\infty}\exp\left(-\frac{k}{2}\right)\left[C_1 d_1(t)\left(4+2\log\left(8\sqrt{2p_0\Lambda(t)}\right)\right)+2\,\mathrm{c}\,\sqrt{2p_0}\phi_2(t)\right]\\
&\leq 3C_1 d_1(t)\left(4+2\log\left(8\sqrt{2p_0\Lambda(t)}\right)\right)+10\,\mathrm{c}\,\sqrt{p_0}\phi_2(t).
\end{aligned}
$$

Summing up, we can conclude that, for any $p_0 \geq 1$,

$$
\begin{aligned}
\left(\mathbb{E}\left[\left(\sup_{a\in\mathbb{Q}}|L_t^a(X)|\right)^{p_0}\right]\right)^{\frac{1}{p_0}} &\leq 2p_0\phi_1(t)+3C_1 d_1(t)\left(4+2\log\left(8\sqrt{2p_0\Lambda(t)}\right)\right)\\
&\quad+10\,\mathrm{c}\,\sqrt{p_0}\phi_2(t)\\
&\leq \kappa\left(p_0\phi_1(t)+\left(\sqrt{\phi_1(t)}+\sqrt{\phi_2(t)}\right)\log(2p_0\Lambda(t))+\sqrt{p_0}\phi_2(t)\right),
\end{aligned}
$$

for a positive constant κ depending only on c, C_1. □

3.8 Proofs for Section 3.4

Proof of Proposition 21. Setting for any continuous function $g\colon \mathbb{R}\to\mathbb{R}$

$$
\begin{aligned}
\mathbf{h}^g(u) &:= \frac{2}{\sigma^2(u)\rho_b(u)}\int_{\mathbb{R}} g(y)\rho_b(y)\left(\mathbb{1}\{u>y\}-F_b(u)\right)\mathrm{d}y, && u\in\mathbb{R}, \qquad (3.8.51)\\
G^g(z) &:= \int_0^z \mathbf{h}^g(u)\mathrm{d}u, && z\geq 0,
\end{aligned}
$$

we can apply Itô's formula to $G^{fb_0}(\cdot)$ and X to obtain

$$\begin{aligned}\int_{X_0}^{X_t} \mathbf{h}^{fb_0}(u)\mathrm{d}u &= \int_0^t \mathbf{h}^{fb_0}(X_s)b(X_s)\mathrm{d}s + \int_0^t \mathbf{h}^{fb_0}(X_s)\sigma(X_s)\mathrm{d}W_s + \frac{1}{2}\int_0^t (G^{fb_0})''(X_s)\mathrm{d}s \\ &= \int_0^t \mathbf{h}^{fb_0}(X_s)\sigma(X_s)\mathrm{d}W_s + \int_0^t \Big((fb_0)(X_s) - \mathbb{E}_b\left[(fb_0)(X_0)\right]\Big)\mathrm{d}s.\end{aligned}$$

This gives (3.4.32) for the specifications $\mathbb{M}_t^f := -\int_0^t \mathbf{h}^{fb_0}(X_s)\sigma(X_s)\mathrm{d}W_s$ and $\mathbb{R}_t^f := \int_{X_0}^{X_t} \mathbf{h}^{fb_0}(u)\mathrm{d}u$. The next step consists in bounding the function $\mathbf{h}^{fb_0}(\cdot)$. Note first that the conditions on the class Σ ensure that there exists a constant $K = K(\mathbb{C}, A, \gamma, \overline{\nu}, \underline{\nu})$ such that, for any $b \in \Sigma(\mathbb{C}, A, \gamma, \sigma)$,

$$\sup_{x\geq 0} \frac{1-F_b(x)}{\sigma^2(x)\rho_b(x)} \;\leq\; K \quad \text{ and } \quad \sup_{x\leq 0} \frac{F_b(x)}{\sigma^2(x)\rho_b(x)} \;\leq\; K.$$

For $y \in [0, A]$, we have

$$\begin{aligned}\sigma^2(y)\rho_b(y) \;&=\; C_{b,\sigma}^{-1}\exp\left(2\int_0^y \frac{b(v)}{\sigma^2(v)}\mathrm{d}v\right) \;\geq\; C_{b,\sigma}^{-1}\exp\left(-2\int_0^y \frac{|b(v)|}{\sigma^2(v)}\mathrm{d}v\right) \\ &\geq\; C_{b,\sigma}^{-1}\exp\left(-2\underline{\nu}^{-2}\mathbb{C}\int_0^y (1+v)\mathrm{d}v\right) \;\geq\; C_{b,\sigma}^{-1}\exp\left(-2\underline{\nu}^{-2}\mathbb{C}\int_0^A (1+v)\mathrm{d}v\right) \\ &=\; C_{b,\sigma}^{-1}\mathrm{e}^{-\underline{\nu}^{-2}\mathbb{C}(2A+A^2)}.\end{aligned}$$

Since the same arguments apply to $y \in [-A, 0]$, it holds

$$(\sigma^2\rho_b)^{-1}(y) \;\leq\; C_{b,\sigma}\mathrm{e}^{\underline{\nu}^{-2}\mathbb{C}(2A+A^2)}, \quad y \in [-A, A].$$

We start with analysing the general case. Let $f \in \mathcal{F} \cup \overline{\mathcal{F}}$, and note that $\boldsymbol{\lambda}(\mathrm{supp}(f)) \leq \mathcal{S}$. For any $u \in \mathbb{R}$ and the function $\mathbf{h}^{fb_0}$ defined according to (3.8.51), we have

$$\begin{aligned}|\mathbf{h}^{fb_0}(u)|^2 \;&\leq\; 4\int_{\mathbb{R}} f^2(y)\mathrm{d}y\int_{\mathrm{supp}(f)} C^2(1+|y|^\eta)^2\rho_b^2(y)\frac{(\mathbb{1}\{u>y\} - F_b(u))^2}{\sigma^4(u)\rho_b^2(u)}\mathrm{d}y \\ &=\; 4C^2\|f\|_{L^2(\boldsymbol{\lambda})}^2\left\{\frac{(1-F_b(u))^2}{\sigma^4(u)\rho_b^2(u)}\int_{-\infty}^u \mathbb{1}\{y\in\mathrm{supp}(f)\}(1+|y|^\eta)^2\rho_b^2(y)\mathrm{d}y\right. \\ &\qquad\left. + \frac{F_b^2(u)}{\sigma^4(u)\rho_b^2(u)}\int_u^\infty \mathbb{1}\{y\in\mathrm{supp}(f)\}(1+|y|^\eta)^2\rho_b^2(y)\mathrm{d}y\right\}.\end{aligned}$$

Now, for $u > A$,

$$\begin{aligned}
|\mathbf{h}^{fb_0}(u)|^2 \;&\le\; 4C^2\|f\|^2_{L^2(\boldsymbol{\lambda})}\Bigg\{2K^2\mathcal{L}\mathcal{S}\left(1+\sup_{x\in\mathbb{R}}|x|^{2\eta}\rho_b(x)\right)\\
&\quad+\underline{\nu}^{-2}\int_u^\infty \mathbb{1}\{y\in\operatorname{supp}(f)\}\exp\left(4\int_u^y\frac{b(z)}{\sigma^2(z)}\mathrm{d}z\right)\left(2+2|y|^{2\eta}\right)\mathrm{d}y\Bigg\}\\
&\le\; 4C^2\|f\|^2_{L^2(\boldsymbol{\lambda})}\Bigg\{2K^2\mathcal{L}\mathcal{S}\left(1+\sup_{x\in\mathbb{R}}|x|^{2\eta}\rho_b(x)\right)\\
&\quad+\underline{\nu}^{-2}\int_u^\infty \mathbb{1}\{y\in\operatorname{supp}(f)\}\max\{2^{2\eta},2\}\mathrm{e}^{-4\gamma(y-u)}\left(1+|y-u|^{2\eta}\right)\mathrm{d}y\\
&\quad+\underline{\nu}^{-2}\int_u^\infty \mathbb{1}\{y\in\operatorname{supp}(f)\}\max\{2^{2\eta},2\}\mathrm{e}^{-4\gamma(y-u)}u^{2\eta}\mathrm{d}y\Bigg\}\\
&\le\; 4C^2\|f\|^2_{L^2(\boldsymbol{\lambda})}\Bigg\{2K^2\mathcal{L}\mathcal{S}\left(1+\sup_{x\in\mathbb{R}}|x|^{2\eta}\rho_b(x)\right)\\
&\quad+\underline{\nu}^{-2}\mathcal{S}\max\{2^{2\eta},2\}\sup_{x\ge 0}\left(\exp(-4\gamma x)x^{2\eta}\right)+\underline{\nu}^{-2}\mathcal{S}\max\{2^{2\eta},2\}(1+u^{2\eta})\Bigg\}\\
&\le\; 4C^2\|f\|^2_{L^2(\boldsymbol{\lambda})}\max\{2^{2\eta},2\}\;\mathcal{S}\Bigg\{2K^2\mathcal{L}\left(1+\sup_{x\in\mathbb{R}}|x|^{2\eta}\rho_b(x)\right)\\
&\quad+\underline{\nu}^{-2}\left(\sup_{x\ge 0}\left(\exp(-4\gamma x)x^{2\eta}\right)+1+u^{2\eta}\right)\Bigg\}.
\end{aligned}$$

The case $u < -A$ can be treated analogously. For $-A \le u \le A$, it holds

$$\begin{aligned}
|\mathbf{h}^{fb_0}(u)|^2 \;&\le\; 4C^2\|f\|^2_{L^2(\boldsymbol{\lambda})}\sup_{-A\le x\le A}\frac{4\mathcal{L}\mathcal{S}}{\sigma^4(x)\rho_b^2(x)}\left(1+\sup_{x\in\mathbb{R}}|x|^{2\eta}\rho_b(x)\right)\\
&\le\; 16C^2\|f\|^2_{L^2(\boldsymbol{\lambda})}\mathcal{L}\mathcal{S}C^2_{b,\sigma}\mathrm{e}^{2\underline{\nu}^{-2}\mathbb{C}(2A+A^2)}\left(1+\sup_{x\in\mathbb{R}}|x|^{2\eta}\rho_b(x)\right).
\end{aligned}$$

Thus, for any $u\in\mathbb{R}$, $f\in\mathcal{F}\cup\overline{\mathcal{F}}$ and $\overline{\Lambda}_{\mathrm{prox}}$ defined according to (3.4.33),

$$|\mathbf{h}^{fb_0}(u)|^2 \;\le\; \overline{\Lambda}^2_{\mathrm{prox}}\mathcal{S}\|f\|^2_{L^2(\boldsymbol{\lambda})}\left(1+|u|^{2\eta}\right). \tag{3.8.52}$$

For any $p \ge 2$, it now follows from (3.2.12), (3.8.52) and (3.4.29) that

$$\begin{aligned}
\mathbb{E}_b\left[\left|\frac{1}{\sqrt{t}}\mathbb{M}_t^f\right|^p\right] \;&\le\; \mathrm{c}^p\,p^{p/2}\mathbb{E}_b\left[\left(\frac{1}{t}\int_0^t(\mathbf{h}^{fb_0})^2(X_s)\sigma^2(X_s)\mathrm{d}s\right)^{p/2}\right] \qquad (3.8.53)\\
&\le\; \overline{\nu}^p\,\mathrm{c}^p\,p^{p/2}\left(\overline{\Lambda}^2_{\mathrm{prox}}\mathcal{S}\|f\|^2_{L^2(\boldsymbol{\lambda})}\right)^{p/2}t^{-p/2}\mathbb{E}_b\left[\left(\int_0^t\left(1+|X_s|^{2\eta}\right)\mathrm{d}s\right)^{p/2}\right]\\
&\le\; \|f\|^p_{L^2(\boldsymbol{\lambda})}\overline{\nu}^p\,\mathrm{c}^p\,p^{p/2}\left(\overline{\Lambda}^2_{\mathrm{prox}}\mathcal{S}\right)^{p/2}\left(1+\mathbb{C}^{2\eta}_{\mathrm{mo}}(\eta p)^{2\eta}\right)^{p/2}\\
&\le\; \|f\|^p_{L^2(\boldsymbol{\lambda})}\overline{\nu}^p\,\mathrm{c}^p\left(\overline{\Lambda}^2_{\mathrm{prox}}\mathcal{S}\right)^{p/2}p^{\eta p+p/2}\left(1+(\mathbb{C}_{\mathrm{mo}}\eta)^{2\eta}\right)^{p/2}.
\end{aligned}$$

For $1 \leq p < 2$, one obtains

$$\begin{aligned}
\mathbb{E}_b\left[\left|\frac{1}{\sqrt{t}}\mathbb{M}_t^f\right|^p\right] &\leq (2p)^{p/2}\,\mathrm{c}^p\sqrt{\mathbb{E}_b\left[\left(\frac{1}{t}\int_0^t(\mathbf{h}^{fb_0})^2(X_s)\sigma^2(X_s)\mathrm{d}s\right)^p\right]}\\
&\leq (2p)^{p/2}\overline{\nu}^p\,\mathrm{c}^p\left(\overline{\Lambda}_{\mathrm{prox}}^2\mathcal{S}\|f\|_{L^2(\boldsymbol{\lambda})}^2\right)^{p/2}t^{-p/2}\sqrt{\mathbb{E}_b\left[\left(\int_0^t(1+|X_s|^{2\eta})\mathrm{d}s\right)^p\right]}\\
&\leq (2p)^{p/2}\|f\|_{L^2(\boldsymbol{\lambda})}^p\overline{\nu}^p\,\mathrm{c}^p\left(\overline{\Lambda}_{\mathrm{prox}}^2\mathcal{S}\right)^{p/2}\left(1+\mathbb{C}_{\mathrm{mo}}^{2\eta}(\eta 2p)^{2\eta}\right)^{p/2}\\
&\leq (2p)^{p/2+\eta p}|f\|_{L^2(\boldsymbol{\lambda})}^p\overline{\nu}^p\,\mathrm{c}^p\left(\overline{\Lambda}_{\mathrm{prox}}^2\mathcal{S}\right)^{p/2}\left(1+(\mathbb{C}_{\mathrm{mo}}\eta)^{2\eta}\right)^{p/2}.
\end{aligned}$$

For bounding the remainder term, we start by noting that (3.8.52) implies the upper bound $\sup_{f\in\mathcal{F}}|\mathbf{h}^{fb_0}(u)| \leq \overline{\Lambda}_{\mathrm{prox}}\mathcal{S}(1+|u|^\eta)$. Consequently, for any $p \geq 1$,

$$\begin{aligned}
\left\|\sup_{f\in\mathcal{F}}\left|\int_{X_0}^{X_t}\mathbf{h}^{fb_0}(u)\mathrm{d}u\right|\right\|_p &\leq \mathcal{S}\overline{\Lambda}_{\mathrm{prox}}\left(2\|X_0\|_p + \frac{2}{\eta+1}\|X_0^{\eta+1}\|_p\right)\\
&\leq 2\mathcal{S}\overline{\Lambda}_{\mathrm{prox}}\left(\mathbb{C}_{\mathrm{mo}}p + \frac{1}{\eta+1}\mathbb{C}_{\mathrm{mo}}^{\eta+1}(\eta+1)^{\eta+1}p^{\eta+1}\right)\\
&\leq 4\mathcal{S}\overline{\Lambda}_{\mathrm{prox}}\max\left\{\mathbb{C}_{\mathrm{mo}}^{\eta+1},\ 1\right\}(\eta+1)^\eta p^{\eta+1}.
\end{aligned}$$

We now turn to the particular case $b_0 = b$. For this case, one could use the above results with $\eta = 1$. However, one obtains better estimates by exploiting the relation between ρ_b and b. We start with considering the martingale part. Let $f, g \in \mathcal{F}$, and let $x_f, x_g \in \mathbb{R}$ such that $\mathrm{supp}(f-g) \subset [x_f, x_f+\mathcal{S}] \cup [x_g, x_g+\mathcal{S}]$. Then, for any $u \in \mathbb{R}$,

$$\begin{aligned}
&|\mathbf{h}^{(f-g)b}(u)|^2\\
&\quad\leq 4\int_{\mathbb{R}}(f-g)^2(y)\mathrm{d}y\ \int_{\mathbb{R}}\mathbb{1}\{y\in\mathrm{supp}(f-g)\}b^2(y)\rho_b^2(y)\frac{(\mathbb{1}\{u>y\}-F_b(u))^2}{\sigma^4(u)\rho_b^2(u)}\mathrm{d}y\\
&\quad\leq 4\|f-g\|_{L^2(\boldsymbol{\lambda})}^2\Bigg\{\int_{\mathbb{R}}\mathbb{1}\{y\in[x_f,x_f+\mathcal{S}]\}b^2(y)\rho_b^2(y)\frac{(\mathbb{1}\{u>y\}-F_b(u))^2}{\sigma^4(u)\rho_b^2(u)}\mathrm{d}y\\
&\qquad\qquad+\int_{\mathbb{R}}\mathbb{1}\{y\in[x_g,x_g+\mathcal{S}]\}b^2(y)\rho_b^2(y)\frac{(\mathbb{1}\{u>y\}-F_b(u))^2}{\sigma^4(u)\rho_b^2(u)}\mathrm{d}y\Bigg\}.
\end{aligned}$$

For $u > A$, and any $x \in \mathbb{R}$, exploiting the relation $(\sigma^2\rho_b)' = 2b\rho_b$, it holds

$$\begin{aligned}
&\int_{\mathbb{R}} \mathbb{1}\{y \in [x, x+\mathcal{S}]\} b^2(y)\rho_b^2(y) \frac{(\mathbb{1}\{u > y\} - F_b(u))^2}{\sigma^4(u)\rho_b^2(u)} \mathrm{d}y \\
&\quad\leq \frac{(1-F_b(u))^2}{\sigma^4(u)\rho_b^2(u)} \int_{-\infty}^{u} \mathbb{1}\{y \in [x, x+\mathcal{S}]\} |b^2(y)\rho_b^2(y)| \mathrm{d}y \\
&\qquad + \frac{1}{4\sigma^4(u)\rho_b^2(u)} \int_u^{\infty} \mathbb{1}\{x \leq y \leq x+\mathcal{S}\} |b(y)| \left| \left(\sigma^4\rho_b^2\right)'(y) \right| \sigma^{-2}(y) \mathrm{d}y \\
&\quad\leq \sup_{z \geq 0} \frac{(1-F_b(z))^2}{\sigma^4(z)\rho_b^2(z)} \frac{1}{4} \mathcal{L}^2 \mathcal{S} + \mathbb{1}\{u \leq x\} \frac{\underline{\nu}^{-2}\mathbb{C}(\sigma^2\rho_b)^2(x)}{4(\sigma^2\rho_b)^2(u)} (1+x+\mathcal{S}) \\
&\qquad + \mathbb{1}\{u > x\} \frac{\underline{\nu}^{-2}\mathbb{C}(\sigma^2\rho_b)^2(u)}{4(\sigma^2\rho_b)^2(u)} (1+u+\mathcal{S}) \\
&\quad\leq \frac{K^2\mathcal{L}^2\mathcal{S}}{4} + \mathbb{1}\{u \leq x\} \frac{\underline{\nu}^{-2}\mathbb{C}}{4} \left(\mathrm{e}^{-4\gamma(x-u)}(1+x-u+\mathcal{S}) + \mathrm{e}^{-4\gamma(x-u)} u \right) \\
&\qquad + \frac{\underline{\nu}^{-2}\mathbb{C}}{4}(1+u+\mathcal{S}) \\
&\quad\leq \frac{K^2\mathcal{L}^2\mathcal{S}}{4} + \frac{\underline{\nu}^{-2}\mathbb{C}}{2} \left(1+u+\mathcal{S} + \sup_{z \geq 0} (\exp(-4\gamma z) z) \right).
\end{aligned}$$

For $u < -A$,

$$\begin{aligned}
&\int_{\mathbb{R}} \mathbb{1}\{y \in (x, x+\mathcal{S})\} b^2(y)\rho_b^2(y) \frac{(\mathbb{1}\{u > y\} - F_b(u))^2}{\sigma^4(u)\rho_b^2(u)} \mathrm{d}y \\
&\quad\leq \frac{1}{4\sigma^4(u)\rho_b^2(u)} \int_{-\infty}^{u} \mathbb{1}\{y \in (x, x+\mathcal{S})\} |b(y)| \left| \left(\sigma^4\rho_b^2\right)'(y) \right| \sigma^{-2}(y) \mathrm{d}y \\
&\qquad + \frac{F_b^2(u)}{\sigma^4(u)\rho_b^2(u)} \int_u^{\infty} \mathbb{1}\{x \leq y \leq x+\mathcal{S}\} b^2(y)\rho_b^2(y) \mathrm{d}y \\
&\quad\leq \mathbb{1}\{u \geq x+\mathcal{S}\} \frac{\underline{\nu}^{-2}\mathbb{C}(\sigma^2\rho_b)^2(x+\mathcal{S})}{4\rho_b^2(u)} (1+|x|) \\
&\qquad + \mathbb{1}\{u < x+\mathcal{S}\} \frac{\underline{\nu}^{-2}\mathbb{C}(\sigma^2\rho_b)^2(u)}{4(\sigma^2\rho_b)^2(u)} (1+|u|+\mathcal{S}) + \sup_{z \leq 0} \frac{F_b^2(z)}{\sigma^4(z)\rho_b^2(z)} \frac{1}{4} \mathcal{L}^2 \mathcal{S} \\
&\quad\leq \mathbb{1}\{u \geq x+\mathcal{S}\} \left(\frac{\underline{\nu}^{-2}\mathbb{C}}{4} \mathrm{e}^{-4\gamma(u-(x+\mathcal{S})}(1+u-(x+\mathcal{S})) + \frac{\underline{\nu}^{-2}\mathbb{C}}{4} \mathrm{e}^{-4\gamma(x-u)}(|u|+\mathcal{S}) \right) \\
&\qquad + \frac{\underline{\nu}^{-2}\mathbb{C}}{4}(1+|u|+\mathcal{S}) + \frac{K^2\mathcal{L}^2\mathcal{S}}{4} \\
&\quad\leq \frac{K^2\mathcal{L}^2\mathcal{S}}{4} + \frac{\underline{\nu}^{-2}\mathbb{C}}{2} \left(1+|u|+\mathcal{S} + \sup_{z \geq 0} (\exp(-4\gamma z) z) \right).
\end{aligned}$$

Finally, for $u \in [-A, A]$,

$$\begin{aligned}
|\mathbf{h}^{(f-g)b}(u)|^2 &\leq 4\|f-g\|^2_{L^2(\boldsymbol{\lambda})}\left\{\frac{(1-F_b(u))^2}{\sigma^4(u)\rho_b^2(u)}\int_{-\infty}^u \mathbb{1}\{y \in \operatorname{supp}(f-g)\}b^2(y)\rho_b^2(y)\mathrm{d}y\right.\\
&\quad \left. + \frac{F_b^2(u)}{\sigma^4(u)\rho_b^2(u)}\int_u^\infty \mathbb{1}\{y \in \operatorname{supp}(f-g)\}b^2(y)\rho_b^2(y)\mathrm{d}y\right\}\\
&\leq 4\|f-g\|^2_{L^2(\boldsymbol{\lambda})}\sup_{-A\leq z\leq A}\frac{\mathcal{L}^2\mathcal{S}}{2\sigma^4(z)\rho_b^2(z)}\\
&\leq 4\|f-g\|^2_{L^2(\boldsymbol{\lambda})}\mathcal{L}^2\mathcal{S}C_{b,\sigma}^2\mathrm{e}^{2\underline{\nu}^{-2}\mathbb{C}(2A+A^2)}.
\end{aligned}$$

Summing up, $|\mathbf{h}^{(f-g)b}(u)|^2 \leq \overline{\Gamma}^2_{\mathrm{prox}}\|f-g\|^2_{L^2(\boldsymbol{\lambda})}(1+|u|+\mathcal{S})$. The same arguments give, for any $f \in \mathcal{F}$,

$$|\mathbf{h}^{fb}(u)|^2 \leq \overline{\Gamma}^2_{\mathrm{prox}}\|f\|^2_{L^2(\boldsymbol{\lambda})}(1+|u|+\mathcal{S}).$$

Similarly to (3.8.53), we can conclude for all $f \in \mathcal{F}\cup\overline{\mathcal{F}}$, $p \geq 2$,

$$\begin{aligned}
\mathbb{E}_b\left[\left|\frac{1}{\sqrt{t}}\mathbb{M}_t^f\right|^p\right] &\leq \mathrm{c}^p\, p^{p/2}t^{-p/2}\mathbb{E}_b\left[\left(\int_0^t(\mathbf{h}^{fb})^2(X_s)\sigma^2(X_s)\mathrm{d}s\right)^{p/2}\right]\\
&\leq \overline{\nu}^p\,\mathrm{c}^p\, p^{p/2}t^{-p/2}\mathbb{E}_b\left[\left(\int_0^t\overline{\Gamma}^2_{\mathrm{prox}}\|f\|^2_{L^2(\boldsymbol{\lambda})}(1+|X_s|+\mathcal{S})\mathrm{d}s\right)^{p/2}\right]\\
&\leq \overline{\Gamma}^p_{\mathrm{prox}}\|f\|^p_{L^2(\boldsymbol{\lambda})}\overline{\nu}^p\,\mathrm{c}^p\, p^{p/2}t^{-p/2}\mathbb{E}_b\left[\left(\int_0^t(1+|X_s|+\mathcal{S})\mathrm{d}s\right)^{p/2}\right]\\
&\leq \overline{\Gamma}^p_{\mathrm{prox}}\|f\|^p_{L^2(\boldsymbol{\lambda})}\overline{\nu}^p\,\mathrm{c}^p\, p^p(1+\mathcal{S}+\mathbb{C}_{\mathrm{mo}})^{p/2}.
\end{aligned}$$

For $1 \leq p < 2$, it holds

$$\begin{aligned}
\mathbb{E}_b\left[\left|\frac{1}{\sqrt{t}}\mathbb{M}_t^f\right|^p\right] &\leq (2p)^{p/2}\,\mathrm{c}^p\,t^{-p/2}\sqrt{\mathbb{E}_b\left[\left(\int_0^t(\mathbf{h}^{fb})^2(X_s)\sigma^2(X_s)\mathrm{d}s\right)^p\right]}\\
&\leq (2p)^{p/2}\overline{\nu}^p\,\mathrm{c}^p\,t^{-p/2}\sqrt{\mathbb{E}_b\left[\left(\int_0^t\overline{\Gamma}^2_{\mathrm{prox}}\|f\|^2_{L^2(\boldsymbol{\lambda})}(1+|X_s|+\mathcal{S})\mathrm{d}s\right)^p\right]}\\
&\leq 2^{p/2}\overline{\Gamma}^p_{\mathrm{prox}}\|f\|^p_{L^2(\boldsymbol{\lambda})}\overline{\nu}^p\,\mathrm{c}^p\,p^p(1+\mathcal{S}+\mathbb{C}_{\mathrm{mo}})^{p/2}.
\end{aligned}$$

Hence, we have shown for any $p \geq 1$

$$\left(\mathbb{E}_b\left[\left|\frac{1}{\sqrt{t}}\mathbb{M}_t^f\right|^p\right]\right)^{\frac{1}{p}} \leq \sqrt{2}\overline{\Gamma}_{\mathrm{prox}}\|f\|_{L^2(\boldsymbol{\lambda})}\overline{\nu}\,\mathrm{c}(1+\mathbb{C}_{\mathrm{mo}}+\mathcal{S})^{1/2}p.$$

For bounding the remainder term, let $f \in \mathcal{F}$. We start by decomposing $|\mathbf{h}^{fb}| \leq 2A_1+2A_2$, with

$$\begin{aligned}
A_1(u) &:= \frac{1-F_b(u)}{\sigma^2(u)\rho_b(u)}\left|\int_{-\infty}^u(fb)(y)\rho_b(y)\mathrm{d}y\right|, \quad \text{and}\\
A_2(u) &:= \frac{F_b(u)}{\sigma^2(u)\rho_b(u)}\left|\int_u^\infty(fb)(y)\rho_b(y)\mathrm{d}y\right|.
\end{aligned}$$

For $u \geq 0$,

$$
\begin{aligned}
A_1(u) &\leq K\Big(\int_{-\infty}^{-A} |f(y)|\frac{(\sigma^2\rho_b)'(y)}{2}\mathrm{d}y + \int_{-A}^{A} |f(y)b(y)|\rho_b(y)\mathrm{d}y \\
&\qquad -\mathbb{1}\{u > A\}\int_A^u \frac{(\sigma^2\rho_b)'(y)}{2}|f(y)|\mathrm{d}y\Big) \\
&\leq KU\left(\frac{\overline{\nu}^2\mathcal{L}}{2} + \mathbb{C}(1+A)\mu_b([-A,A]) + \overline{\nu}^2\mathcal{L}\right) \;\leq\; KU\left(2\overline{\nu}^2\mathcal{L} + \mathbb{C}(1+A)\right), \\
A_2(u) &\leq U\left(\mathbb{1}\{u \leq A\}C_{b,\sigma}\mathrm{e}^{\underline{\nu}^{-2}\mathbb{C}(2A+A^2)}\left(\int_u^A \mathbb{C}(1+A)\mathcal{L}\mathrm{d}y - \int_A^\infty \frac{(\sigma^2\rho_b)'(y)}{2}\mathrm{d}y\right)\right) \\
&\qquad -U\mathbb{1}\{u > A\}\frac{1}{\sigma^2(u)\rho_b(u)}\int_u^\infty \frac{(\sigma^2\rho_b)'(y)}{2}\mathrm{d}y \\
&\leq U\left(C_{b,\sigma}\mathrm{e}^{\underline{\nu}^{-2}\mathbb{C}(2A+A^2)}\left(A\mathbb{C}(1+A)\mathcal{L} + \frac{\overline{\nu}^2\mathcal{L}}{2}\right)\right) + \frac{1}{2}U.
\end{aligned}
$$

For $u \leq 0$,

$$
\begin{aligned}
A_1(u) &\leq U\left(\mathbb{1}\{u \geq -A\}C_{b,\sigma}\mathrm{e}^{\underline{\nu}^{-2}\mathbb{C}(2A+A^2)}\left(\int_{-A}^u \mathbb{C}(1+A)\mathcal{L}\mathrm{d}y + \int_{-\infty}^{-A} \frac{(\sigma^2\rho_b)'(y)}{2}\mathrm{d}y\right)\right) \\
&\qquad + U\mathbb{1}\{u < -A\}\frac{1}{\sigma^2(u)\rho_b(u)}\int_{-\infty}^u \frac{(\sigma^2\rho_b)'(y)}{2}\mathrm{d}y \\
&\leq U\left(C_{b,\sigma}\mathrm{e}^{\underline{\nu}^{-2}\mathbb{C}(2A+A^2)}\left(A\mathbb{C}(1+A)\mathcal{L} + \frac{\overline{\nu}^2\mathcal{L}}{2}\right)\right) + \frac{1}{2}U, \\
A_2(u) &\leq KU\left(\mathbb{1}\{u \leq -A\}\int_u^{-A} \frac{(\sigma^2\rho_b)'(y)}{2}\mathrm{d}y + \int_{-A}^{A} |b(y)|\rho_b(y)\mathrm{d}y\right) \\
&\qquad -KU\int_A^\infty \frac{(\sigma^2\rho_b)'(y)}{2}\mathrm{d}y \\
&\leq KU\left(\overline{\nu}^2\mathcal{L} + \mathbb{C}(1+|A|)\mu_b([-A,A]) + \frac{\overline{\nu}^2\mathcal{L}}{2}\right) \;\leq\; KU\left(2\overline{\nu}^2\mathcal{L} + \mathbb{C}(1+A)\right).
\end{aligned}
$$

We have thus shown that

$$
\begin{aligned}
\sup_{f\in\mathcal{F}}\left\|\mathbf{h}^{fb}\right\|_\infty &\leq 4KU\left(2\overline{\nu}^2\mathcal{L} + \mathbb{C}(1+A)\right) \\
&\qquad + 4U\left(C_{b,\sigma}\mathrm{e}^{\underline{\nu}^{-2}\mathbb{C}(2A+A^2)}\left(A\mathbb{C}(1+A)\mathcal{L} + \frac{\overline{\nu}^2\mathcal{L}}{2}\right)\right) + 2U,
\end{aligned}
$$

and, finally,

$$
\begin{aligned}
\left(\mathbb{E}_b\left[\left\|\mathbb{R}_t^f\right\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}} &= \left(\mathbb{E}_b\left[\left\|\int_{X_0}^{X_t}\mathbf{h}^{fb}(u)\mathrm{d}u\right\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}} \;\leq\; \sup_{f\in\mathcal{F}}\left\|\mathbf{h}^{fb}\right\|_\infty\left(\mathbb{E}_b\left[|X_t - X_0|^p\right]\right)^{\frac{1}{p}} \\
&\leq 2\mathbb{C}_{\mathrm{mo}}\,p\sup_{f\in\mathcal{F}}\left\|\mathbf{h}^{fb}\right\|_\infty \;\leq\; \Gamma_{\mathrm{prox}}p.
\end{aligned}
$$

□

Proof of Theorem 23. Under the given assumptions, Proposition 21 implies the decomposition

$$\mathbb{G}_t^{b_0}(f) \;=\; t^{-1/2}\mathbb{M}_t^f + t^{-1/2}\mathbb{R}_t^f, \quad t>0.$$

For $b_0 \equiv 1$, we further obtain, for any $p \geq 1$, $f \in \mathcal{F} \cup \overline{\mathcal{F}}$,

$$\left(\mathbb{E}_b\left[|\mathbb{M}_t^f|^p\right]\right)^{\frac{1}{p}} \;\leq\; \Pi_1\sqrt{pt\mathcal{S}}\|f\|_{L^2(\boldsymbol{\lambda})}, \quad \left(\mathbb{E}_b\left[\|\mathbb{R}_t^{fb}\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}} \;\leq\; \Pi_2 p\mathcal{S}.$$

This corresponds to the case $\alpha = 2$ in Theorem 16 which then yields (**I**). For $b_0 = b$, the upper bounds on the p-th moments, $p \geq 1$, of the martingale and remainder term are specified as

$$\left(\mathbb{E}_b\left[|\mathbb{M}_t^f|^p\right]\right)^{\frac{1}{p}} \;\leq\; \Pi_1^b p^{\frac{3}{2}}\sqrt{t\mathcal{S}}\|f\|_{L^2(\boldsymbol{\lambda})}, \; f \in \mathcal{F}\cup\overline{\mathcal{F}}, \qquad \left(\mathbb{E}_b\left[\|\mathbb{R}_t^{fb}\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}} \;\leq\; \Pi_2^b p, \; f \in \mathcal{F}.$$

Here, we combined the upper bound for the moments of the martingale part for the general case (letting $\eta = 1$) in Proposition 21 with the upper bound on the moments of the remainder term for the specific drift part (equation (**II**) of the proposition). The upper bounds correspond to the case $\alpha = \frac{2}{3}$ in Theorem 16. The assertion follows together with Lemma 36. □

Proof of Proposition 24. The proof substantially relies on Proposition 21 and Theorem 16.

Martingale approximation. Our first step is the martingale approximation of the non-martingale part of

$$\mathbb{H}_t(f) \;=\; \sqrt{t}\left(\frac{1}{t}\int_0^t f(X_s)b(X_s)\mathrm{d}s - \int (fb)d\mu_b + \frac{1}{t}\int_0^t f(X_s)\sigma(X_s)dW_s\right).$$

Proposition 21 gives the representation $\sqrt{t}\mathbb{H}_t(f) = \mathbb{M}_t^f + \mathbb{R}_t^f + \int_0^t f(X_s)\sigma(X_s)\mathrm{d}W_s$, and equation (**II**) yields, for any $p \geq 1$,

$$\begin{aligned}\left(\mathbb{E}_b\left[|\mathbb{M}_t^f|^p\right]\right)^{\frac{1}{p}} &\leq \Phi_1\sqrt{tp}\|f\|_{L^2(\boldsymbol{\lambda})}, &&\text{for } \Phi_1 := \sqrt{2}\overline{\Gamma}_{\mathrm{prox}}\overline{\nu}\,\mathrm{c}\sqrt{1+\mathbb{C}_{\mathrm{mo}}+\mathcal{S}},\\ \left(\mathbb{E}_b\left[\left(\sup_{f\in\mathcal{F}}|\mathbb{R}_t^{fb}|\right)^p\right]\right)^{\frac{1}{p}} &\leq \Phi_2 p, &&\text{for } \Phi_2 := \Gamma_{\mathrm{prox}}.\end{aligned}$$

Application of Theorem 16. Plugging the above estimates of the moments of $\mathbb{M}_t^f$ and $\mathbb{R}_t^f$ into the moment bound of Theorem 16, one gets together with Lemma 36, for any $p \geq 1$,

$$\begin{aligned}\frac{1}{\sqrt{t}}\left(\mathbb{E}_b\left[\left\|\mathbb{M}_t^f + \mathbb{R}_t^f\right\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}} \;&\leq\; C_1\sum_{k=0}^{\infty} E(F_k, \mathrm{e}\Phi_1\; L^2, 1)\exp\left(-\frac{k}{2}\right) + 12\Phi_1 p\mathbb{V} + 2\frac{\Phi_2 p}{\sqrt{t}}\\ &\quad+\sqrt{t}CU(1+2\mathbb{C}_{\mathrm{mo}})\exp\left(-\frac{\Lambda t}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}\right) \;\leq\; \widetilde{\Phi}_t^b(p),\end{aligned}$$

with

$$\widetilde{\Phi}_t^b(p) \ := \ 6C_1 v\mathbb{V}\mathrm{e}\Phi_1\left(1+\log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S}+p\Lambda t}\right)\right)+12\Phi_1 p\mathbb{V}+2\frac{\Phi_2 p}{\sqrt{t}} \\ +\sqrt{t}CU(1+2\mathbb{C}_{\mathrm{mo}})\exp\left(-\frac{\Lambda t}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}\right).$$

Bounding the p-th moments of the original stochastic integral term. Corollary 20 yields a constant $\widetilde{\kappa}$ such that, for any $p\geq 1$ and $t\geq 1$,

$$\max\left\{\left(\mathbb{E}_b\left[\sup_{a\in\mathbb{Q}}\left|\frac{1}{t}L_t^a(X)\right|^{p/2}\right]\right)^{\frac{1}{p}},\left(\mathbb{E}_b\left[\sup_{a\in\mathbb{Q}}\left|\frac{1}{t}L_t^a(X)\right|^{p}\right]\right)^{\frac{1}{2p}}\right\} \\ \leq \widetilde{\kappa}\left(\sqrt{p}+t^{-1/4}\sqrt{\log t}+p^{1/4}t^{-1/4}\right).$$

Consequently, as in the proof of Theorem 30, for any $p\geq 1$,

$$\left(\mathbb{E}_b\left[\left|\frac{1}{\sqrt{t}}\int_0^t f(X_s)\sigma(X_s)\mathrm{d}W_s\right|^p\right]\right)^{\frac{1}{p}} \ \leq \ \overline{\nu}\,\mathrm{c}\sqrt{p}\|f\|_{L^2(\boldsymbol{\lambda})}\widetilde{\kappa}\left(\sqrt{p}+t^{-1/4}\sqrt{\log t}+p^{1/4}t^{-1/4}\right).$$

In particular, this last estimate implies that the p-th moments are uniformly bounded over $\mathcal{F}$ and that the process $\left(t^{-1/2}\int_0^t f(X_s)\sigma(X_s)\mathrm{d}W_s\right)_{f\in\mathcal{F}}$ exhibits a subexponential tail behaviour. Precisely, we have

$$\mathbb{P}_b\left(\left|t^{-1/2}\int_0^t(f-g)(X_s)\sigma(X_s)\mathrm{d}W_s\right| \ \geq \ d(f,g)u\right) \ \leq \ \exp(-u),$$

for $d(f,g):=\max\{\mathrm{e}\Lambda_3,1\}\,\|f-g\|_{L^2(\boldsymbol{\lambda})}$, $\Lambda_3:=\max\{4\widetilde{\kappa}\overline{\nu}\,\mathrm{c}+2\widetilde{\kappa}\overline{\nu}\,\mathrm{c}\sup_{t\geq 1}t^{-1/4}\sqrt{\log t},1\}$. Following the scheme of the proof of Theorem 16, we apply the chaining procedure from Proposition 34 locally and obtain, for all $p,q\geq 1$, $k\in\mathbb{N}_0$,

$$\frac{1}{\sqrt{t}}\left\|\sup_{f\in\mathcal{F}_k}\int_0^t f(X_s)\sigma(X_s)\mathrm{d}W_s\right\|_q \ \leq \ C_1\int_0^\infty\log N(u,\mathcal{F}_k,d)\mathrm{d}u \\ +2\sup_{f\in\mathcal{F}_k}\frac{1}{\sqrt{t}}\left\|\int_0^t f(X_s)\sigma(X_s)\mathrm{d}W_s\right\|_q \\ \leq \ C_1\int_0^\infty\log N(u,\mathcal{F}_k,d)\mathrm{d}u+\Lambda_3\mathbb{V}q.$$

From the local result, we deduce, for any $p \geq 1$,

$$\begin{aligned}
&\frac{1}{\sqrt{t}}\left(\mathbb{E}_b\left[\left\|\int_0^t f(X_s)\sigma(X_s)\mathrm{d}W_s\right\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}} \\
&\quad\leq \frac{1}{\sqrt{t}}\left(\mathbb{E}_b\left[\left\|\int_0^t f(X_s)\sigma(X_s)\mathrm{d}W_s\right\|_{\mathcal{F}_0}^p\right]\right)^{\frac{1}{p}} \\
&\quad\quad+\sum_{k=1}^{\infty}\frac{1}{\sqrt{t}}\left(\mathbb{E}_b\left[\left\|\int_0^t f(X_s)\sigma(X_s)\mathrm{d}W_s\right\|_{\mathcal{F}_k}^{2p}\right]\right)^{\frac{1}{2p}}(\mathbb{P}_b(A_k))^{\frac{1}{2p}} \\
&\quad\leq C_1\sum_{k=0}^{\infty}E(\mathcal{F}_k,d,1)\exp\left(-\frac{k}{2}\right)+6\Lambda_3\mathbb{V}p.
\end{aligned}$$

The entropy integrals can be bounded independently of $k \in \mathbb{N}_0$ by means of Lemma 36, allowing us to conclude finally that, for any $p \geq 1$,

$$\left(\mathbb{E}_b\left[\|\mathbb{H}_t\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}} \quad\leq\quad \widetilde{\Phi}_t^b(p)+\widetilde{\Psi}_t^b(p),$$

with

$$\widetilde{\Psi}_t^b(p) \;:=\; 6C_1 v\mathbb{V}\max\{\mathrm{e}\Lambda_3,1\}\left(1+\log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S}+p\Lambda t}\right)\right)+6\Lambda_3\mathbb{V}p.$$

□

3.9 Proofs for Section 3.5

Proof of Theorem 27. Note first that, for each $x \in \mathbb{R}$, $\rho_{t,K}(h)(x) - \rho_t^\circ(x)$ is a random variable which is right-continuous in x. Thus, $\|\rho_{t,K}(h)-\rho_t^\circ\|_\infty = \sup_{x\in\mathbb{Q}}|\rho_{t,K}(h)(x)-\rho_t^\circ(x)|$ is also measurable as a supremum over a countable set. Introduce

$$\Psi_1(x,y) \;:=\; (x-y)\cdot\mathbb{1}_{(-\infty,x]}(y),\quad \Psi_2(x,X^t) \;:=\; \int_0^t \mathbb{1}_{(-\infty,x]}(X_u)\mathrm{d}X_u,\quad x,y\in\mathbb{R},\ t>0,$$

and abbreviate $K_h(\cdot) := h^{-1}K(\cdot/h)$. Using the occupation times formula and Tanaka's formula for diffusion local time, we obtain

$$\begin{aligned}
\rho_t^\circ(x) &= t^{-1}L_t^x(X) \;=\; 2t^{-1}\left((X_t-x)^- - (X_0-x)^- + \int_0^t \mathbb{1}_{(-\infty,x]}(X_s)\mathrm{d}X_s\right) \\
&= 2t^{-1}\left(\Psi_1(x,X_t)-\Psi_1(x,X_0)+\Psi_2(x,X^t)\right),
\end{aligned}$$

and, since $\int K_h(x-y)\mathrm{d}y = 1$,

$$\begin{aligned}\rho_{t,K}(h)(x) - \rho_t^\circ(x) &= t^{-1}\int_{\mathbb{R}} K_h(x-y)L_t^y(X)\mathrm{d}y - t^{-1}L_t^x(X)\\ &= 2t^{-1}\int_{\mathbb{R}} K_h(x-y)\left(\Psi_1(y,X_t)-\Psi_1(x,X_t)\right)\mathrm{d}y\\ &\qquad +2t^{-1}\int_{\mathbb{R}} K_h(x-y)\left(\Psi_1(x,X_0)-\Psi_1(y,X_0)\right)\mathrm{d}y\\ &\qquad +2t^{-1}\int_{\mathbb{R}} K_h(x-y)\left(\Psi_2(y,X^t)-\Psi_2(x,X^t)\right)\mathrm{d}y\\ &=: A_{1,x}(t,h)+A_{2,x}(t,h)+B_x(t,h).\end{aligned}$$

We start by rewriting

$$\begin{aligned}tA_{1,x}(t,h) &= 2\int_{\mathbb{R}} K_h(x-y)\left\{(y-X_t)\mathbb{1}_{(-\infty,y]}(X_t)-(x-X_t)\mathbb{1}_{(-\infty,x]}(X_t)\right\}\mathrm{d}y\\ &= 2\int_{\mathbb{R}} K(z)\left\{(x-X_t)\mathbb{1}_{(-\infty,x-zh]}(X_t)-(x-X_t)\mathbb{1}_{(-\infty,x]}(X_t)\right\}\mathrm{d}z\\ &\qquad -2\int_{\mathbb{R}} K(z)zh\mathbb{1}_{(-\infty,x-zh]}(X_t)\mathrm{d}z.\end{aligned}$$

Note that $|(x-X_t)\left(\mathbb{1}_{(-\infty,x-zh]}(X_t)-\mathbb{1}_{(-\infty,x]}(X_t)\right)| \leq |z|h$ for $h \geq 0$. Thus, for any $x \in \mathbb{Q}$, $A_{1,x}(t,h) \leq 4ht^{-1}\int_{\mathbb{R}}|K(z)z|\mathrm{d}z$. Since $A_{2,x}(t,h)$ can be treated analogously, it follows

$$\sup_{x\in\mathbb{Q}}|A_{1,x}(t,h)+A_{2,x}(t,h)| \leq 8ht^{-1}\int_{\mathbb{R}}|K(z)z|\mathrm{d}z. \tag{3.9.54}$$

It remains to consider $B_x(t,h)$. For any fixed $x \in \mathbb{Q}$, we have

$$\begin{aligned}tB_x(t,h) &= 2\int_{\mathbb{R}} K_h(x-y)\int_0^t\left\{\mathbb{1}_{(-\infty,y]}(X_s)-\mathbb{1}_{(-\infty,x]}(X_s)\right\}\mathrm{d}X_s\mathrm{d}y\\ &= 2\int_{\mathbb{R}} K(z)\int_0^t\left\{\mathbb{1}_{(-\infty,x-zh]}(X_s)-\mathbb{1}_{(-\infty,x]}(X_s)\right\}\mathrm{d}X_s\mathrm{d}z \qquad (3.9.55)\\ &= 2\int_0^t\left\{\int_{\mathbb{R}} K(z)\mathbb{1}_{(-\infty,x-zh]}(X_s)\mathrm{d}z-\mathbb{1}_{(-\infty,x]}(X_s)\right\}\mathrm{d}X_s\\ &= 2\int_0^t\left\{\int_{\mathbb{R}} K_h(y-X_s)\mathbb{1}_{(-\infty,x-y+X_s]}(X_s)\mathrm{d}y-\mathbb{1}_{(-\infty,x]}(X_s)\right\}\mathrm{d}X_s\\ &= 2\int_0^t\left\{K_h*\mathbb{1}_{(-\infty,x]}-\mathbb{1}_{(-\infty,x]}\right\}(X_s)\mathrm{d}X_s. \qquad (3.9.56)\end{aligned}$$

Here we used a Fubini-type theorem for stochastic integrals (cf. Kailath *et al.* (1978)), allowing us to change the order of integration in (3.9.55). We proceed by applying Proposition 24 to the function class

$$\mathcal{F}_{K,h} := \left\{K_h*\mathbb{1}_{(-\infty,x]}-\mathbb{1}_{(-\infty,x]}\colon\ x\in\mathbb{Q}\right\}. \tag{3.9.57}$$

Following the lines of the proof of Theorem 1 in Giné and Nickl (2009), note first that, for any $x \in \mathbb{R}$, $h > 0$,

$$K_h * \mathbb{1}_{(-\infty,x]}(\cdot) - \mathbb{1}_{(-\infty,x]}(\cdot) \;=\; H\left(\frac{x-\cdot}{h}\right),$$

for $H(u) := \int_{-\infty}^{u} K(z)\mathrm{d}z - \mathbb{1}_{[0,\infty)}(u)$, $u \in \mathbb{R}$. Since H is of bounded variation, Lemma 35 ensures that the entropy condition from Assumption 4 holds true. Moreover,

$$\sup_{x\in\mathbb{Q}} \left\| K_h * \mathbb{1}_{(-\infty,x]}(\cdot) - \mathbb{1}_{(-\infty,x]}(\cdot)\right\|_\infty \;\le\; \int_{\mathbb{R}} |K(z)|\mathrm{d}z + 1 \;\le\; 2\|K\|_{L^1(\boldsymbol{\lambda})} \;=:\; \mathbb{K},$$

i.e., $\mathcal{F}_{K,h}$ is uniformly bounded. Let us now investigate the $L^2(\boldsymbol{\lambda})$-norm of $\mathcal{F}_{K,h}$. To this end, fix $x \in \mathbb{Q}$ and note that, for any $z > 0$,

$$\begin{aligned}\int \left(\mathbb{1}_{(-\infty,x]}(y+z) - \mathbb{1}_{(-\infty,x]}(y)\right)^2 \mathrm{d}y \;&=\; \int \left|\mathbb{1}_{(-\infty,x]}(y) - \mathbb{1}_{(-\infty,x]}(y+z)\right| \mathrm{d}y \\ &=\; \int \mathbb{1}_{(x-z,x]}(y)\mathrm{d}y \;=\; z.\end{aligned}$$

A similar argument for $z \le 0$ yields

$$\int \left(\mathbb{1}_{(-\infty,x]}(y+z) - \mathbb{1}_{(-\infty,x]}(y)\right)^2 \mathrm{d}y \;=\; |z|$$

for all $z \in \mathbb{R}$. This bound implies that

$$\sup_{x\in\mathbb{Q}} \|K_h * \mathbb{1}_{(-\infty,x]} - \mathbb{1}_{(-\infty,x]}\|_{L^2(\boldsymbol{\lambda})} \;\le\; \sqrt{h}\int |K(z)\sqrt{z}|\mathrm{d}z$$

since, for any $x \in \mathbb{Q}$, using Minkowski's integral inequality,

$$\begin{aligned}&\left(\int \left(K_h * \mathbb{1}_{(-\infty,x]}(y) - \mathbb{1}_{(-\infty,x]}(y)\right)^2 \mathrm{d}y\right)^{\frac{1}{2}} \\ &\quad= \left(\int \left(\int K_h(z)\mathbb{1}_{(-\infty,x]}(y+z) - \mathbb{1}_{(-\infty,x]}(y)\mathrm{d}z\right)^2 \mathrm{d}y\right)^{\frac{1}{2}} \\ &\quad\le \int |K_h(z)| \left(\int \left(\mathbb{1}_{(-\infty,x]}(y+z) - \mathbb{1}_{(-\infty,x]}(y)\right)^2 \mathrm{d}y\right)^{\frac{1}{2}} \mathrm{d}z \\ &\quad= \int |K_h(z)|\sqrt{|z|}\mathrm{d}z \;=\; \sqrt{h}\int |K(z)|\sqrt{|z|}\mathrm{d}z.\end{aligned}$$

Clearly, $\operatorname{supp}(K_h * \mathbb{1}_{(-\infty,x]} - \mathbb{1}_{(-\infty,x]}) \subset [x-h/2, x+h/2]$ such that

$$\begin{aligned}\sup_{x\in\mathbb{Q}} \|K_h * \mathbb{1}_{(-\infty,x]} - \mathbb{1}_{(-\infty,x]}\|_{L^2(\boldsymbol{\lambda})} \;&\le\; \sqrt{\mathcal{S}}, \\ \sup_{x\in\mathbb{Q}} \lambda(\operatorname{supp}(K_h * \mathbb{1}_{(-\infty,x]} - \mathbb{1}_{(-\infty,x]})) \;&\le\; \mathcal{S} := h\,\max\left\{1,\ \left(\int |K(z)|\sqrt{|z|}\mathrm{d}z\right)^2\right\}.\end{aligned}$$

We have thus shown that $\mathcal{F}_{K,h}$ satisfies Assumption 3.

However, since the functions $\mathbb{1}_{(-\infty,x]}$ are not continuous, Proposition 24 cannot be applied. Inspection of the proof shows that continuity is required in order to use Proposition 21. More precisely, continuity allows to apply Itô's formula which in turn yields the central representation $\mathbb{G}_t = t^{-1/2}(\mathbb{M}_t + \mathbb{R}_t)$. Consequently, Proposition 24 is applicable once we can show that the same representation is valid for the functions $\mathbb{1}_{(-\infty,x]}b$. For deriving this representation, we need to approximate

$$\int_0^t \mathbb{1}\{X_s \le x\} b(X_s)\mathrm{d}s - \mathbb{E}_b\left[\int_0^t \mathbb{1}\{X_s \le x\}\mathrm{d}X_s\right] \;=\; \int_0^t \mathbb{1}\{X_s \le x\} b(X_s)\mathrm{d}s - \frac{t}{2}\rho_b(x).$$

Denote $f_x(\cdot) := \mathbb{1}\{\cdot \le x\}\, b(\cdot)$, $x \in \mathbb{Q}$. We proceed similarly to the proof of Proposition 21 by setting

$$\begin{aligned}
\mathbf{h}^{f_x}(u) &:= \frac{2}{\rho_b(u)} \int f_x(y)\rho_b(y)(\mathbb{1}\{u > y\} - F_b(u))\mathrm{d}y \\
&= \mathbb{1}\{u > x\}\frac{1}{\rho_b(u)}\rho_b(x)(1 - F_b(u)) + \mathbb{1}\{u \le x\}\left(1 - \frac{F_b(u)\rho_b(x)}{\rho_b(u)}\right) \\
&= \frac{1}{\rho_b(u)}\rho_b(x)(\mathbb{1}\{u > x\} - F_b(u)) + \mathbb{1}\{u \le x\}, \\
\mathbf{h}_n(u) &:= \frac{\rho_b(x)}{\rho_b(u)}(\phi_n(u) - F_b(u)) + (1 - \phi_n(u)),
\end{aligned}$$

for $\phi_n(u)$ denoting a smooth approximation of $\mathbb{1}\{u > x\}$, given as

$$\phi_n(u) \;:=\; \frac{n}{\sqrt{2\pi}} \int_{-\infty}^{u} \exp\left(\frac{-(v-x)^2 n^2}{2}\right)\mathrm{d}v$$

(cf. the proof of Proposition 1.11 in Kutoyants (2004)). Note that $\lim_{n\to\infty} \phi_n(u) = \mathbb{1}\{u > x\}$ and, for any continuous function $g\colon \mathbb{R} \to \mathbb{R}$, it holds

$$\lim_{n\to\infty} \int \phi_n'(u) g(u)\mathrm{d}u \;=\; g(x). \tag{3.9.58}$$

Set

$$H_n(y) \;:=\; \int_0^y \mathbf{h}_n(u)\mathrm{d}u \quad \text{and} \quad H(y) \;:=\; \int_0^y \mathbf{h}^{f_x}(u)\mathrm{d}u.$$

Then $H_n'(y) \;=\; \mathbf{h}_n(y)$, and

$$\begin{aligned}
H_n''(y) &= -\frac{\rho_b(x)\rho_b'(y)}{\rho_b^2(y)}(\phi_n(y) - F(y)) + \frac{\rho_b(x)}{\rho_b(y)}(\phi_n'(y) - \rho_b(y)) - \phi_n'(y) \\
&= -\frac{2\rho_b(x)b(y)}{\rho_b(y)}(\phi_n(y) - F(y)) + \frac{\rho_b(x)}{\rho_b(y)}(\phi_n'(y) - \rho_b(y)) - \phi_n'(y) \\
&= -2\mathbf{h}_n(y)b(y) + 2b(y)(1 - \phi_n(y)) + \frac{\rho_b(x)}{\rho_b(y)}(\phi_n'(y) - \rho_b(y)) - \phi_n'(y).
\end{aligned}$$

Itô's formula yields

$$\begin{aligned}
&H_n(X_t) - H_n(X_0)\\
&= \int H_n'(X_s)\mathrm{d}X_s + \frac{1}{2}\int H_n''(X_s)\mathrm{d}s\\
&= \int_0^t \mathbf{h}_n(X_s)b(X_s)\mathrm{d}s + \int_0^t \mathbf{h}_n(X_s)\mathrm{d}W_s - \int_0^t \mathbf{h}_n(X_s)b(X_s)\mathrm{d}s\\
&\qquad + \int_0^t \left\{b(X_s)(1-\phi_n(X_s)) + \frac{\rho_b(x)}{2\rho_b(X_s)}(\phi_n'(X_s) - \rho_b(X_s)) - \frac{\phi_n'(X_s)}{2}\right\}\mathrm{d}s\\
&= \int_0^t \mathbf{h}_n(X_s)\mathrm{d}W_s\\
&\qquad + \int_0^t \left\{b(X_s)(1-\phi_n(X_s)) + \frac{\rho_b(x)}{2\rho_b(X_s)}(\phi_n'(X_s) - \rho_b(X_s)) - \frac{\phi_n'(X_s)}{2}\right\}\mathrm{d}s.
\end{aligned}$$

Continuity of diffusion local time $(L_t^a)_{a\in\mathbb{R}}$ and (3.9.58) imply that

$$\begin{aligned}
\int_0^t \left\{\frac{\rho_b(x)}{2\rho_b(X_s)}\phi_n'(X_s) - \frac{1}{2}\phi_n'(X_s)\right\}\mathrm{d}s &= \int \left(\frac{\rho_b(x)}{2\rho_b(y)}\phi_n'(y) - \frac{1}{2}\phi_n'(y)\right) L_t^y(X)\mathrm{d}y\\
&\overset{n\to\infty}{\longrightarrow} \left(\frac{\rho_b(x)}{2\rho_b(x)} - \frac{1}{2}\right) L_t^x(X) \;=\; 0.
\end{aligned}$$

Using the at-most-linear-growth condition on b, it can be shown that, for fixed $x \in \mathbb{Q}$, there exist constants θ_1,$\theta_2 > 0$ such that, for all $n \in \mathbb{N}$,

$$\theta_2 F_b(u) \;\geq\; \phi_n(u), \quad \forall\, u \leq -\theta_1.$$

Intuitively speaking, this relation reflects the fact that ρ_b has tails at least as heavy as a normal distribution. This implies, for all $n \in \mathbb{N}$,

$$\|\mathbf{h}_n\|_\infty \;\leq\; \frac{2\rho_b(x)}{\inf_{|u|\leq\theta_1}\rho_b(u)} + \sup_{u\geq 0}\frac{1-F_b(u)}{\rho_b(u)}\rho_b(x) + (\theta_2+1)\sup_{u\leq 0}\frac{F_b(u)}{\rho_b(u)} + 3.$$

Thus, taking account of $\lim_{n\to\infty}\phi_n(u) = \mathbb{1}\{u > x\}$ and $\lim_{n\to\infty}\mathbf{h}_n(u) = \mathbf{h}^{f_x}(u)$, we obtain from the dominated convergence theorem and its version for stochastic integrals (see, e.g., Proposition 5.8 in Le Gall (2016)) almost surely

$$\int_{X_0}^{X_t} \mathbf{h}^{f_x}(u)\mathrm{d}u \;=\; \int_0^t \mathbf{h}^{f_x}(X_s)\mathrm{d}W_s + \int_0^t b(X_s)\mathbb{1}\{X_s \leq x\}\mathrm{d}s - \frac{t}{2}\rho_b(x).$$

Thus, the martingale approximation from Proposition 21 and, consequently, Proposition 24 is valid for the class $\mathcal{F}_{K,h}$ introduced in (3.9.57). In particular, there exist positive constants $\mathbb{L}$ and Λ such that

$$\sup_{b\in\Sigma}\mathbb{P}_b\left(\left\|\sqrt{t}\left(\frac{1}{t}\int_0^t f(X_s)\mathrm{d}X_s - \mathbb{E}_b\left[f(X_0)b(X_0)\right]\right)\right\|_{\mathcal{F}_{K,h}} \geq \phi(u)\right) \;\leq\; \mathrm{e}^{-u} \quad \forall\, u \geq 1,$$

where

$$\phi(u) \;=\; \sqrt{h}\mathcal{V}\mathrm{e}\mathbb{L}\left\{1+\log\left(\frac{1}{\sqrt{h}\mathcal{V}}\right)+\log(t)+u\right\}+\mathrm{e}\mathbb{L}\frac{u}{\sqrt{t}}+\mathrm{e}\mathbb{L}\sqrt{t}\exp\left(-\frac{\Lambda t}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}\right),$$

with $\mathcal{V} := \int |K(z)|\sqrt{|z|}\mathrm{d}z$. Furthermore, for any $x \in \mathbb{Q}$,

$$\begin{aligned}
&\left|\mathbb{E}_b\left[\left(K_h * \mathbb{1}_{(-\infty,x]} - \mathbb{1}_{(-\infty,x]}\right)(X_0)b(X_0)\right]\right| \\
&\quad= \frac{1}{2}\left|\int\int K_h(z)\left(\mathbb{1}_{(-\infty,x]}(z+y) - \mathbb{1}_{(-\infty,x]}(y)\right)\rho_b'(y)\mathrm{d}z\mathrm{d}y\right| \\
&\quad= \frac{1}{2}\left|\int\int K(v)\left(\mathbb{1}_{(-\infty,x-vh]}(y) - \mathbb{1}_{(-\infty,x]}(y)\right)\rho_b'(y)\mathrm{d}v\mathrm{d}y\right| \\
&\quad= \frac{1}{2}\left|-\int_0^\infty K(v)\int_{x-vh}^{x}\rho_b'(y)\mathrm{d}y\mathrm{d}v + \int_{-\infty}^{0} K(v)\int_{x}^{x-vh}\rho_b'(y)\mathrm{d}y\mathrm{d}v\right| \\
&\quad= \frac{1}{2}\left|\int K(v)(\rho_b(x-vh)-\rho_b(x))\mathrm{d}v\right|.
\end{aligned}$$

In case $\beta > 1$, we proceed with

$$\begin{aligned}
&\frac{1}{2}\left|\int K(v)(\rho_b(x-vh)-\rho_b(x))\mathrm{d}v\right| \\
&\quad= \frac{1}{2}\left|\int K(v)\sum_{i=1}^{\lfloor\beta\rfloor-1}\frac{\rho_b^{(i)}(x)}{i!}(vh)^i + \frac{\rho_b^{\lfloor\beta\rfloor}(x-\tau_v vh)}{\lfloor\beta\rfloor!}(vh)^{\lfloor\beta\rfloor}\mathrm{d}v\right| \\
&\quad= \frac{1}{2}\left|\int K(v)\frac{\rho_b^{\lfloor\beta\rfloor}(x-\tau_v vh)-\rho_b^{\lfloor\beta\rfloor}(x)}{\lfloor\beta\rfloor!}(vh)^{\lfloor\beta\rfloor}\mathrm{d}v\right| \\
&\quad\le \frac{\mathcal{L}}{2\lfloor\beta\rfloor!}\int |K(v)|\left|(\tau_v vh)^{\beta-\lfloor\beta\rfloor}(vh)^{\lfloor\beta\rfloor}\right|\mathrm{d}v,
\end{aligned}$$

where $\tau_v \in [0,1]$, $v \in \mathbb{R}$. For $\beta \le 1$, ρ_b is Hölder continuous to the exponent β which implies

$$\frac{1}{2}\left|\int K(v)(\rho_b(x-vh)-\rho_b(x))\mathrm{d}v\right| \;\le\; \frac{\mathcal{L}}{2}\int |K(v)|\,|vh|^\beta\,\mathrm{d}v.$$

Thus, for $b \in \Sigma(\beta,\mathcal{L})$ with $\beta > 0$,

$$\left|\mathbb{E}_b\left[\left(K_h * \mathbb{1}_{(-\infty,x]} - \mathbb{1}_{(-\infty,x]}\right)(X_0)\;b(X_0)\right]\right| \;\le\; h^\beta\frac{\mathcal{L}}{2\lfloor\beta\rfloor!}\int |K(v)v^\beta|\mathrm{d}v.$$

In view of (3.9.56) and the above considerations, for any $u \geq 1$,

$$\begin{aligned}
&\sup_{b\in\Sigma}\mathbb{P}_b\left(\sqrt{t}\sup_{x\in\mathbb{Q}}|B_x(t,h)| \ \geq\ 2\left(\phi(u)+\sqrt{t}h^{\beta}\frac{\mathcal{L}}{2\lfloor\beta\rfloor!}\int|K(v)v^{\beta}|\mathrm{d}v\right)\right)\\
&\quad =\ \sup_{b\in\Sigma}\mathbb{P}_b\left(\sup_{f\in\mathcal{F}_{K,h}}\left|\frac{1}{\sqrt{t}}\int_0^t f(X_s)\mathrm{d}X_s\right| \ \geq\ \phi(u)+\sqrt{t}h^{\beta}\frac{\mathcal{L}}{2\lfloor\beta\rfloor!}\int|K(v)v^{\beta}|\mathrm{d}v\right)\\
&\quad \leq\ \sup_{b\in\Sigma}\mathbb{P}_b\Bigg(\sup_{f\in\mathcal{F}_{K,h}}\left|\frac{1}{\sqrt{t}}\int_0^t f(X_s)\mathrm{d}X_s-\sqrt{t}\mathbb{E}_b[f(X_0)b(X_0)]\right|\\
&\qquad\qquad +\sqrt{t}h^{\beta}\frac{\mathcal{L}}{2\lfloor\beta\rfloor!}\int|K(v)v^{\beta}|\mathrm{d}v \ \geq\ \phi(u)+\sqrt{t}h^{\beta}\frac{\mathcal{L}}{2\lfloor\beta\rfloor!}\int|K(v)v^{\beta}|\mathrm{d}v\Bigg)\\
&\quad \leq\ \mathrm{e}^{-u}. \qquad\qquad (3.9.59)
\end{aligned}$$

Set

$$\lambda_0 := \sqrt{h}\mathcal{V}\mathrm{e}\mathbb{L}\left\{1+\log\left(\frac{1}{\sqrt{h}\mathcal{V}}\right)+\log(t)\right\}+\mathrm{e}\mathbb{L}\sqrt{t}\mathrm{e}^{-\frac{\Lambda t}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}}+\sqrt{t}h^{\beta}\frac{\mathcal{L}}{2\lfloor\beta\rfloor!}\int|K(v)v^{\beta}|\mathrm{d}v.$$

Define $\Lambda_1 := (8\mathcal{V}\mathrm{e}\mathbb{L}+8\mathrm{e}\mathbb{L})^{-1}$, and choose $\Lambda_0 \geq 1$ such that, for all $t \geq 1$, $h \in (0,1)$,

$$8ht^{-1/2}\int_{\mathbb{R}}|K(z)z|\mathrm{d}z < 4\Lambda_0\lambda_0$$

and $\mathcal{V}\mathrm{e}\mathbb{L}\Lambda_1\Lambda_0 > 1$. Taking into account (3.9.54), this choice in particular implies that, for any $\lambda \geq 8\Lambda_0\lambda_0$,

$$\begin{aligned}
&\sup_{b\in\Sigma}\ \mathbb{P}_b\left(\sqrt{t}\|\rho_{t,K}(h)-\rho_t^{\circ}\|_\infty > \lambda\right)\\
&\quad \leq\ \sup_{b\in\Sigma}\ \mathbb{P}_b\left(\sqrt{t}\sup_{x\in\mathbb{Q}}|A_{1,x}(t,h)+A_{2,x}(t,h)+B_x(t,h)| > \lambda\right)\\
&\quad \leq\ \sup_{b\in\Sigma}\ \mathbb{P}_b\left(\sqrt{t}\|B_\bullet(t,h)\|_\infty > \lambda-\frac{8h}{\sqrt{t}}\int|K(z)z|\mathrm{d}z\right)\\
&\quad \leq\ \sup_{b\in\Sigma}\ \mathbb{P}_b\left(\sqrt{t}\|B_\bullet(t,h)\|_\infty > \lambda-4\Lambda_0\lambda_0\right)\ \leq\ \sup_{b\in\Sigma}\ \mathbb{P}_b\left(\sqrt{t}\|B_\bullet(t,h)\|_\infty > \lambda/2\right). \qquad (3.9.60)
\end{aligned}$$

Note that, for $u = \Lambda_1\lambda h^{-1/2}$,

$$\begin{aligned}
\phi(u)+\sqrt{t}h^{\beta}\frac{\mathcal{L}}{2\lfloor\beta\rfloor!}\int|K(v)v^{\beta}|\mathrm{d}v \ &\leq\ \lambda_0+\sqrt{h}\mathcal{V}\mathrm{e}\mathbb{L}u+\frac{\mathrm{e}\mathbb{L}u}{\sqrt{t}}\\
&\leq\ \lambda_0\Lambda_0+\frac{u}{8}\Lambda_1^{-1}\sqrt{h}\ \leq\ \lambda_0\Lambda_0+\frac{\lambda}{8}\ \leq\ \frac{\lambda}{4}.
\end{aligned}$$

Summarising, (3.9.59) and (3.9.60) then give the asserted inequality (3.5.40). □

We are now in a position to derive the announced upper bounds on the moments of centered diffusion local time.

Proof of Corollary 28. We point out that the assumption that $b \in \Sigma(\mathbb{C}, A, \gamma, 1)$ already imposes some regularity on the invariant density in the sense of Definition 25. More precisely, if $b \in \Sigma(\mathbb{C}, A, \gamma, 1)$, the invariant density ρ_b is bounded and Lipschitz continuous due to (3.4.28) which in turn means that $b \in \Sigma(1, \mathcal{L})$. Decompose

$$\left(\mathbb{E}_b\left[\left\|\frac{L_t^\bullet(X)}{t} - \rho_b\right\|_\infty^p\right]\right)^{\frac{1}{p}} \leq \left(\mathbb{E}_b\left[\left\|\rho_t^\circ - \rho_{t,K}\left(t^{-1}\right)\right\|_\infty^p\right]\right)^{\frac{1}{p}} + \left(\mathbb{E}_b\left[\left\|\rho_{t,K}\left(t^{-1}\right) - \rho_b\right\|_\infty^p\right]\right)^{\frac{1}{p}}. \tag{3.9.61}$$

Inspection of the proof of Theorem 27 shows that, for any $b \in \Sigma(1, \mathcal{L})$ and $h = h_t \geq t^{-1}$,

$$\begin{aligned}
\left(\mathbb{E}_b\left[\|\rho_{t,K}(h) - \rho_t^\circ\|_\infty^p\right]\right)^{\frac{1}{p}} &\leq \left(\mathbb{E}_b\left[\|A_{1,x}(t,h) + A_{2,x}(t,h) + B_x(t,h)\|_\infty^p\right]\right)^{\frac{1}{p}} \\
&\leq \frac{8h}{t}\int |K(z)z|\mathrm{d}z + \mathrm{e}^{-1}\varphi(p) + \frac{h\mathcal{L}}{2}\int |K(v)v|\mathrm{d}v,
\end{aligned}$$

where

$$\varphi(u) = \mathcal{V}\mathrm{e}\mathbb{L}\sqrt{\frac{h}{t}}\left\{1 + \log\left(\frac{1}{\sqrt{h}\mathcal{V}}\right) + \log(t) + u\right\} + \mathrm{e}\mathbb{L}\frac{u}{t} + \mathrm{e}\mathbb{L}\exp\left(-\frac{\Lambda t}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}\right).$$

The second term on the rhs of (3.9.61) is bounded by means of Corollary 26. Consequently, specifying $h = h_t \sim t^{-1}$, we obtain a constant ζ such that

$$\left(\mathbb{E}_b\left[\left\|\frac{L_t^\bullet(X)}{t} - \rho_b\right\|_\infty^p\right]\right)^{\frac{1}{p}} \leq \zeta\left(\frac{1}{\sqrt{t}}\left\{1 + \sqrt{\log(pt)} + \sqrt{p}\right\} + \frac{p}{t} + t\exp\left(-\frac{\Lambda t}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}\right)\right).$$

□

Proof of Theorem 30. Analogously to the proof of Proposition 24, we start with decomposing $\mathbb{H}_t$ into finite variation and martingale part,

$$\begin{aligned}
\mathbb{H}_t(f) &= \sqrt{t}\left(\frac{1}{t}\int_0^t f(X_s)b(X_s)\mathrm{d}s - \int (fb)\mathrm{d}\mu_b + \frac{1}{t}\int_0^t f(X_s)\mathrm{d}W_s\right) \\
&= \mathbb{G}_t^b(f) + \frac{1}{\sqrt{t}}\int_0^t f(X_s)\mathrm{d}W_s.
\end{aligned}$$

For the finite variation part, part (**II**) of Theorem 23 gives, for any $p \geq 1$,

$$\left(\mathbb{E}\left[\|\mathbb{G}_t^b\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}} \leq \Phi_t^b(p),$$

for Φ_t^b defined as in (3.4.37). It remains to bound the p-th moments of the original stochastic integral term. Given $f \in \mathcal{F} \cup \overline{\mathcal{F}}$ and any $p \geq 2$, it holds

$$\begin{aligned}
\left(\mathbb{E}_b\left[\left|\frac{1}{\sqrt{t}}\int_0^t f(X_s)\mathrm{d}W_s\right|^p\right]\right)^{\frac{1}{p}} &\leq \mathrm{c}\sqrt{p}\left(\mathbb{E}_b\left[\left(\frac{1}{t}\int_0^t f^2(X_s)\mathrm{d}s\right)^{\frac{p}{2}}\right]\right)^{\frac{1}{p}} \\
&\leq \mathrm{c}\sqrt{p}\|f\|_{L^2(\boldsymbol{\lambda})}\left(\mathbb{E}_b\left[\left(\frac{1}{t}\|L_t^\bullet(X)\|_\infty\right)^{\frac{p}{2}}\right]\right)^{\frac{1}{p}}.
\end{aligned}$$

In the same way, we obtain for $1 \leq p < 2$,

$$\begin{aligned}
\left(\mathbb{E}_b\left[\left|\frac{1}{\sqrt{t}}\int_0^t f(X_s)\mathrm{d}W_s\right|^p\right]\right)^{\frac{1}{p}} &\leq \frac{1}{\sqrt{t}}\left(\mathbb{E}_b\left[\left(\int_0^t f(X_s)\mathrm{d}W_s\right)^{2p}\right]\right)^{\frac{1}{2p}}\\
&\leq \bar{\mathrm{c}}\sqrt{\frac{2p}{t}}\left(\mathbb{E}_b\left[\left(\int_0^t f^2(X_s)\mathrm{d}s\right)^p\right]\right)^{\frac{1}{2p}}\\
&= \mathrm{c}\sqrt{\frac{p}{t}}\left(\mathbb{E}_b\left[\left(\int_{\mathbb{R}} f^2(y)L_t^y(X)\mathrm{d}y\right)^p\right]\right)^{\frac{1}{2p}}\\
&\leq \mathrm{c}\sqrt{p}\|f\|_{L^2(\boldsymbol{\lambda})}\left(\mathbb{E}_b\left[\left(\frac{1}{t}\|L_t^{\bullet}(X)\|_\infty\right)^p\right]\right)^{\frac{1}{2p}}.
\end{aligned}$$

It follows from Corollary 28 that there exists positive constants $\bar{\mathbb{L}}_1$, $\widetilde{\mathbb{L}}_1$ such that, for any $p \geq 1$ and $t \geq 1$,

$$\begin{aligned}
&\max\left\{\left(\mathbb{E}_b\left[\left(\frac{1}{t}\sup_{a\in\mathbb{Q}}|L_t^a(X)|\right)^{\frac{p}{2}}\right]\right)^{\frac{1}{p}}, \left(\mathbb{E}_b\left[\left(\frac{1}{t}\sup_{a\in\mathbb{Q}}|L_t^a(X)|\right)^p\right]\right)^{\frac{1}{2p}}\right\}\\
&\qquad\leq \bar{\mathbb{L}}_1\left(1 + t\exp\left(-\frac{\Lambda t}{2\mathrm{e}\mathbb{C}_{\mathrm{mo}}}\right) + \frac{1}{\sqrt{t}}\left\{1 + \sqrt{\log t} + \sqrt{p}\right\} + \frac{p}{t}\right)^{1/2}\\
&\qquad\leq \widetilde{\mathbb{L}}_1\left(1 + \left(\frac{p}{t}\right)^{1/4} + \sqrt{\frac{p}{t}}\right).
\end{aligned}$$

Consequently, for any $p \geq 1$,

$$\begin{aligned}
2\sup_{f\in\mathcal{F}}\left(\mathbb{E}_b\left[\left|\frac{1}{\sqrt{t}}\int_0^t f(X_s)\mathrm{d}W_s\right|^p\right]\right)^{\frac{1}{p}} &\leq \Lambda_3\mathbb{V}\left(\sqrt{p} + \frac{p}{t^{1/4}}\right),\\
\left(\mathbb{E}_b\left[\left|\frac{1}{\sqrt{t}}\int_0^t (f-g)(X_s)\mathrm{d}W_s\right|^p\right]\right)^{\frac{1}{p}} &\leq \Lambda_3\|f-g\|_{L^2(\boldsymbol{\lambda})}\left(\sqrt{p} + \frac{p}{t^{1/4}}\right), \quad f,g\in\mathcal{F},
\end{aligned}$$

with $\Lambda_3 := \max\{4\widetilde{\mathbb{L}}_1\,\mathrm{c}, 1\}$. In view of Lemma 33, this last estimate implies that, for any $u \geq 1$, $f, g \in \mathcal{F}$,

$$\mathbb{P}_b\left(\left|t^{-1/2}\int_0^t (f-g)(X_s)\mathrm{d}W_s\right| \geq d(f,g)\left(\sqrt{u} + t^{-1/4}u\right)\right) \leq \exp(-u),$$

for $d(f,g) := \mathrm{e}\Lambda_3\,\|f-g\|_{L^2(\boldsymbol{\lambda})}$. Analogously to the proof of Proposition 24, we obtain for all $p, q \geq 1$, $k \in \mathbb{N}_0$,

$$\begin{aligned}
\frac{1}{\sqrt{t}}\left\|\sup_{f\in\mathcal{F}_k}\int_0^t f(X_s)\mathrm{d}W_s\right\|_q \leq \frac{\widetilde{C}_1}{t^{1/4}}\int_0^\infty \log N(u,\mathcal{F}_k,d)\mathrm{d}u + \widetilde{C}_2\int_0^\infty \sqrt{\log N(u,\mathcal{F}_k,d)}\mathrm{d}u&\\
+ 2\sup_{f\in\mathcal{F}}\frac{1}{\sqrt{t}}\left\|\int_0^t f(X_s)\mathrm{d}W_s\right\|_q&,
\end{aligned}$$

and from the local result, we infer, for any $p \geq 1$,

$$\frac{1}{\sqrt{t}}\left(\mathbb{E}_b\left[\left\|\int_0^t f(X_s)\mathrm{d}W_s\right\|_{\mathcal{F}}^p\right]\right)^{\frac{1}{p}} \leq \frac{\widetilde{C}_1}{t^{1/4}}\sum_{k=0}^{\infty} E(\mathcal{F}_k, d, 1)\mathrm{e}^{-k/2} + \widetilde{C}_2 \sum_{k=0}^{\infty} E(\mathcal{F}_k, d, 2)\mathrm{e}^{-k/2} + 6\Lambda_3\mathbb{V}\left(\sqrt{p} + \frac{p}{t^{1/4}}\right).$$

The upper bounds for the entropy integrals from Lemma 36 finally imply that, for any $p \geq 1$,

$$(\mathbb{E}_b\|\mathbb{H}_t\|_{\mathcal{F}}^p)^{\frac{1}{p}} \leq \Phi_t^b(p) + \Pi_t^b(p),$$

with

$$\Pi_t^b(p) := 6\mathbb{V}\mathrm{e}\Lambda_3\left\{\frac{\widetilde{C}_1 v}{t^{1/4}}\left(1 + \log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S} + p\Lambda t}\right)\right) + 2\widetilde{C}_2\sqrt{v\log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S} + p\Lambda t}\right)}\right\} + 6\Lambda_3\mathbb{V}\left(\sqrt{p} + \frac{p}{t^{1/4}}\right).$$

□

4 Sup-norm adaptive drift estimation for ergodic diffusions

4.1 Introduction

The field of nonparametric statistics for stochastic processes has become an integral part of statistics. Due to their practical relevance as standard models in many areas of applied science such as genetics, meteorology or financial mathematics to name very few, the statistical analysis of diffusion processes receives special attention. The first contribution of this chapter is an investigation of adaptive sup-norm convergence rates for a nonparametric Nadaraya–Watson-type drift estimator, based on a continuous record of observations $(X_s)_{0\leq s\leq t}$ of a diffusion process on the real line. The suggested data-driven bandwidth choice relies on Lepski's method for adaptive estimation. Characterising upper and lower bounds, we show that the proposed estimation procedure in the asymptotic regime $t \to \infty$ is minimax rate-optimal over nonparametric Hölder classes. Remarkably, we impose only very mild conditions on the drift coefficient, not going far beyond standard assumptions that ensure the existence of ergodic solutions of the underlying SDE over the real line. In particular, we allow for unbounded drift coefficients. Secondly, we prove a Donsker-type theorem for the classical kernel estimator of the invariant density in $\ell^\infty(\mathbb{R})$ and establish its semiparametric efficiency. With regard to the direct relation between drift coefficient and the invariant density, it is clear that the corresponding estimation problems are closely connected. In a last step, we combine both tasks and suggest an adaptive bandwidth choice that simultaneously yields both an asymptotically efficient, asymptotically normal (in $\ell^\infty(\mathbb{R})$) estimator of the invariant density and, at the same time, the corresponding minimax rate-optimal drift estimator.

So far, results analysing the sup-norm risk in the context of diffusion processes are rather scarce, even though quantifying expected maximal errors is of immense relevance, in particular for practical applications. We therefore start in the basic set-up of continuous observations of a scalar ergodic diffusion process. While the idealised framework of continuous observations of the process may be considered as being far from the reality, it is indisputably of substantial theoretical interest because the statistical results incorporate the very nature of the diffusion process, not being influenced by any discretisation errors. Consequently, they serve as relevant benchmarks for further investigations. Moreover, our approach is attractive in the sense that it provides a reasonable starting point for extending the statistical analysis to discrete observation schemes and even multivariate diffusion processes. A second, very concrete motivation for our framework is the idea of bringing together methods from stochastic control and nonparametric statistics. Diffusion processes serve as a prototype model in stochastic optimal control problems which are solved under the long-standing assumption of continuous observations of a process driven by known dynamics. Relaxing this assumption to the framework of continuous observations of a process driven by an *unknown* drift coefficient, imposing merely mild regularity assumptions, raises interesting questions on how to learn the dynamics by means of nonparametric estimation procedures and to control in an optimal way at the same time. With respect to the statistical methods, these applications typically require optimal bounds on sup-norm errors. This chapter provides these tools for a large class of scalar diffusion processes.

Taking a look at the evolution of the area of statistical estimation for diffusions up to the mid 2000's, we refer to Gobet *et al.* (2004) for a very nice summary. The monograph Kutoyants (2004) provides a comprehensive overview on inference for one-dimensional ergodic diffusion processes on the basis of continuous observations considering pointwise and L^2-risk measures. Banon (1978) is commonly mentioned as the first article addressing the question of nonparametric identification of diffusion processes from continuous data. In nonparametric models, asymptotically efficient estimators typically involve the optimal choice of a tuning parameter that depends on the smoothness of the nonparametric class of targets. From a practical perspective, this is not satisfying at all because the smoothness is usually not known. One thus aims at *adaptive* estimation procedures which are based on purely data-driven estimators adapting to the unknown smoothness.

Spokoiny (2000) and Dalalyan (2005) were the first to study adaptive drift estimation in the diffusion model based on continuous observations. Spokoiny (2000) considers pointwise estimation whereas Dalalyan (2005) investigates a weighted $L^2(\mathbb{R})$-norm. Hoffmann (1999) initiated adaptive estimation in a high-frequency setting, proposing a data driven estimator of the diffusion coefficient based on wavelet thresholding which is rate optimal wrt $L^\gamma(D)$-loss, for $\gamma \in [1, \infty)$ and a compact set D. With regard to low-frequency data, we refer to the seminal paper by Gobet *et al.* (2004). Their objective is inference on the drift and diffusion coefficient of diffusion processes with boundary reflections. The quality of the proposed estimators is measured in the $L^2([a, b])$ distance for any $0 < a < b < 1$. Like restricting to estimation on arbitrary but fixed compact sets, looking at processes with boundary reflections constitutes a possibility to circumvent highly technical issues that we face in our investigation of diffusions on the entire real line. Gobet *et al.* postulate that allowing diffusions on the real line would require to introduce a weighting in the risk measure given by the invariant density. This phenomenon will become visible in our results, as well. The same weighting function can be found in Dalalyan (2005). Intuitively, it seems natural that the estimation risk would explode without a weighting since the observations of the continuous process during a finite period of time do not contain information about the behaviour outside the compact set where the path lives in. For a more detailed heuristic account on the choice of the weight function for $L^2(\mathbb{R})$-risk, we refer to Remark 4.1 in Dalalyan (2005). Sharp adaptive estimation of the drift vector for multidimensional diffusion processes from continuous observations for the L^2- and the pointwise risk has been addressed in Strauch (2015) and Strauch (2016), respectively.

As illustrated, the pointwise and L^2-risks are already well-understood in different frameworks. This thesis complements these developments by an investigation of the sup-norm risk in the continuous observation scheme for diffusion processes living on the whole real line. In the low-frequency framework, this strong norm was studied in Söhl and Trabs (2016) who construct both an adaptive estimator of the drift and adaptive confidence bands. They prove a functional central limit theorem for wavelet estimators in a multi-scale space, i.e., considering a weaker norm that still allows to construct adaptive confidence bands for the invariant density and the drift with optimal $\ell^\infty([a,b])$-diameter. Still, there exist a lot of challenging open questions, and in view of the growing field of applications, there is a clear need for developing and adding techniques and tools for the statistical analysis of stochastic processes under sup-norm risk. Ideally, these tools should include the probabilistic features of the processes and, at the same time, allow for an in-depth analysis of issues such as adaptive estimation in a possibly broad class of models.

A common device for the derivation of adaptive estimation procedures in sup-norm loss are uniform Talagrand-type concentration inequalities and moment bounds for empirical processes based on chaining methods. These tools are made available for a broad class of scalar ergodic diffusion processes in Chapter 3. The concentration inequalities derived therein will serve as the central vehicle for our analysis, and we conjecture that they allow for generalisations on discrete observation schemes, multivariate state variables and even more general Markov processes. Therefore, the approach presented in this thesis provides guidance for further statistical investigations of stochastic processes in sup-norm risk.

Besides the frequentist statistical research, the Bayesian approach found a lot of interest, more recently. In particular, addressing the question of estimating the drift and invariant density of a diffusion process in sup-norm risk, this work is closely related to Nickl and Ray (2018) who start from methodologically totally different Bayesian considerations. These allow for a unified analysis of the scalar and multivariate setting up to dimension 4. In contrast to the model considered here, they assume the drift coefficient to be periodic such that the state space of the diffusion process is restricted to a bounded set. A maximum a posteriori estimate for the drift coefficient based on a truncated Gaussian series prior which can be viewed as a penalised least squares estimator is suggested. The obtained convergence rate of the sup-norm risk equals the minimax-optimal rate achieved by our kernel-based estimator up to log-factors. Furthermore, a functional CLT for the drift estimator viewed as a random element from the dual of certain Besov spaces is shown as well as a Donsker-type functional central limit theorem for the implied plug-in estimator of the invariant density paralleling in dimension one our result for the kernel invariant density estimator in Proposition 45. The CLTs are induced by Bayesian Bernstein–von Mises theorems. In the framework of continuous observations, van der Meulen *et al.* (2006) consider the asymptotic behaviour of posterior distributions in a general Brownian semimartingale model which, as a special case, includes the ergodic diffusion model. Pokern *et al.* (2013) investigate a Bayesian approach to nonparametric estimation of the periodic drift of a scalar diffusion from continuous observations and derive bounds on the rate at which the posterior contracts around the true drift in $L^2([0,1))$-norm. Improvements in terms of these convergence rates results and adaptivity are given in van Waaij and van Zanten (2016). Nonparametric Bayes procedures for estimating the drift of one-dimensional ergodic diffusion models from discrete-time low-frequency data are studied in van der Meulen and van Zanten (2013). The authors give conditions for posterior consistency and verify these conditions for concrete priors. Given discrete observations of a scalar reflected diffusion, Nickl and Söhl (2017) derive (and verify) conditions in the low-frequency sampling regime for prior distributions on the diffusion coefficient and the drift function that ensure minimax optimal contraction rates of the posterior distribution over Hölder–Sobolev smoothness classes in $L^2([a,b])$-distance, for any $0 < a < b < 1$.

Basic framework and outline of the chapter

Taking into view the sup-norm risk, the aim of this chapter is to suggest a rate-optimal nonparametric drift estimator, based on continuous observations of an ergodic diffusion process on the real line which is given as the solution of the SDE

$$\mathrm{d}X_t \;=\; b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \quad X_0 = \xi,\ t > 0, \tag{4.1.1}$$

with unknown drift function $b\colon \mathbb{R} \to \mathbb{R}$, dispersion $\sigma\colon \mathbb{R} \to (0,\infty)$ and some standard Brownian motion $W = (W_t)_{t\geq 0}$. The initial value ξ is a random variable independent of W. We restrict to the ergodic case where the Markov process $(X_t)_{t\geq 0}$ admits an invariant measure, and we denote by ρ_b and μ_b the invariant density and the associated invariant measure, respectively. Furthermore, we will always consider stationary solutions of (4.1.1), i.e., we assume $\xi \sim \mu_b$.

In the set-up of continuous observations, there is no interest in estimating the volatility σ^2 since this quantity is identifiable using the quadratic variation of X. We thus focus on recovering the unknown drift. We develop our results in the following classical scalar diffusion model.

Definition 39. Let $\sigma \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R})$ and assume that, for some constants $\underline{\nu}, \overline{\nu} \in (0,\infty)$, σ satisfies $\underline{\nu} \leq |\sigma(x)| \leq \overline{\nu}$, for all $x \in \mathbb{R}$. For fixed constants $A, \gamma > 0$ and $\mathcal{C} \geq 1$, define the set $\Sigma = \Sigma(\mathcal{C}, A, \gamma, \sigma)$ as

$$\begin{aligned}\Sigma := \Big\{ & b \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R})\colon |b(x)| \leq \mathcal{C}(1 + |x|), \\ & \qquad \forall |x| > A\colon \frac{b(x)}{\sigma^2(x)}\,\mathrm{sgn}(x) \leq -\gamma \Big\}.\end{aligned} \tag{4.1.2}$$

Given any $b \in \Sigma$, there exists a unique strong solution of the SDE (4.1.1) with ergodic properties and invariant density

$$\rho(x) = \rho_b(x) \;:=\; \frac{1}{C_{b,\sigma}\sigma^2(x)}\ \exp\left(\int_0^x \frac{2b(y)}{\sigma^2(y)}\mathrm{d}y\right), \quad x \in \mathbb{R}, \tag{4.1.3}$$

with $C_{b,\sigma} := \int_{\mathbb{R}} \frac{1}{\sigma^2(u)}\ \exp\left(\int_0^u \frac{2b(y)}{\sigma^2(y)}\mathrm{d}y\right)\mathrm{d}u$ denoting the normalising constant. Throughout the sequel and for any $b \in \Sigma$, we will denote by $\mathbb{E}_b$ the expected value with respect to the law of X associated with the drift coefficient b. The distribution function corresponding to $\rho = \rho_b$ and the invariant measure of the distribution will be denoted by $F = F_b$ and $\mu = \mu_b$, respectively.

Our statistical analysis relies heavily on uniform concentration inequalities for continuous-time analogues of empirical processes of the form

$$t^{-1}\int_0^t f(X_s)\mathrm{d}s - \mathbb{E}_b[f(X_0)], \quad f \in \mathcal{F},$$

as well as stochastic integrals

$$t^{-1}\int_0^t f(X_s)\mathrm{d}X_s - \mathbb{E}_b[b(X_0)f(X_0)], \quad f \in \mathcal{F},$$

indexed by some infinite-dimensional function class $\mathcal{F}$. These key devices are provided in our work on concentration inequalities for scalar ergodic diffusions in Chapter 3. They are tailor-made for the investigation of sup-norm risk criteria and can be considered as continuous-time substitutes for Talagrand-type concentration inequalities and moment bounds for empirical processes in the classical i.i.d. framework. In Chapter 3, upper bounds on the expected sup-norm error for a kernel density estimator of the invariant density (that we will use in the present work) are derived as a first statistical application of the developed concentration inequalities. In Section 4.2, we will present the announced probabilistic tools and statistical results from Chapter 3 that will be of crucial importance in our subsequent developments. The advantage of the methods proposed in Chapter 3 is that the martingale approximation approach - which is at the heart of the derivations - yields very elementary simple proofs, working under minimal assumptions on the diffusion process.

The estimators Given continuous observations $X^t = (X_s)_{0\leq s\leq t}$ of a diffusion process as described in Definition 39, first basic statistical questions concern the estimation of the invariant density ρ_b and the drift coefficient b and the investigation of the respective convergence properties. Since $b = (\rho_b\sigma^2)'/(2\rho_b)$, the question of drift estimation is obviously closely related to estimation of the invariant density ρ_b and its derivative ρ_b'. For some smooth kernel function $K\colon \mathbb{R} \to \mathbb{R}$ with compact support, introduce the standard kernel invariant density estimator

$$\rho_{t,K}(h)(x) \;:=\; \frac{1}{th}\int_0^t K\left(\frac{x - X_u}{h}\right)\mathrm{d}u, \quad x \in \mathbb{R}. \tag{4.1.4}$$

A natural estimator of the drift coefficient $b \in \Sigma(\mathcal{C}, A, \gamma, 1)$, which relies on the analogy between the drift estimation problem and the model of regression with random design, is given by a Nadaraya–Watson-type estimator of the form

$$b_{t,K}(h)(x) := \frac{\overline{\rho}_{t,K}(h)(x)}{\rho^+_{t,K}(t^{-1/2})(x) + \sqrt{\frac{\log t}{t}}\exp\left(\sqrt{\log t}\right)}, \tag{4.1.5}$$

$$\text{where } \overline{\rho}_{t,K}(h)(x) := \frac{1}{th}\int_0^t K\left(\frac{x - X_s}{h}\right)\mathrm{d}X_s, \tag{4.1.6}$$

and $\rho^+_{t,K}(h)(x) := \max\{0, \rho_{t,K}(h)(x)\}$. We recognize the kernel density estimator in the denominator, and we will see that $\overline{\rho}_{t,K}$ with the proposed (adaptive) bandwidth choice serves as a rate-optimal estimator of $b\rho_b$. The additive term in the denominator prevents it from becoming small too fast in the tails.

Given a record of continuous observations of a scalar diffusion process X with $b \in \Sigma(\mathcal{C}, A, \gamma, 1)$, the local time estimator $\rho_t^\circ(\bullet) := t^{-1}L_t^\bullet(X)$, for $(L_t^a(X), t \geq 0, a \in \mathbb{R})$ denoting the local time process of X, is available. This is a natural density estimator since diffusion local time can be interpreted as the derivative of the empirical measure. In the past, the latter was exhaustively studied for pointwise estimation and in L^2-risk unlike the sup-norm case. In (Kutoyants, 1998, Sec. 7), weak convergence of the local time estimator to a Gaussian process in $\ell^\infty(\mathbb{R})$ is shown. The same is done for more general diffusion processes in van der Vaart and van Zanten (2005). Having provided the required tools from empirical process theory, upper bounds on all moments of the sup-norm error of ρ_t° are proven in Chapter 3. Unfortunately, the local time estimator is viewed as not being very feasible in practical applications. In addition, it does not offer straightforward extensions to the case of discretely observed or multivariate diffusions, in sharp contrast to the classical kernel-based density estimator. We therefore advocate the usage of the kernel density estimator introduced in (4.1.4) which can be viewed as a universal approach in nonparametric statistics, performing an optimal behaviour over a wide range of models. Furthermore, the kernel density estimator naturally appears in the denominator of our Nadaraya–Watson-type drift estimator defined according to (4.1.5).

Asymptotically efficient density estimation In this chapter, we will complement the sup-norm analysis started in Chapter 3 with an investigation of the asymptotic distribution of the kernel density estimator in a functional sense. We will prove a Donsker-type theorem for the kernel density estimator, thereby demonstrating that this estimator for an appropriate choice of bandwidth behaves asymptotically like the local time estimator. We then go one step further and establish optimality of the limiting distribution, optimality seen in the sense of the general convolution theorem 3.11.2 for the estimation of Banach space valued parameters presented in van der Vaart and Wellner (1996). Their theorem states that, for an asymptotically normal sequence of experiments and any regular estimator, the limiting distribution is the convolution of a specific Gaussian process and a noise factor. This Gaussian process is viewed as the optimal limit law, and we refer to it as the *semiparametric lower bound*. We establish this lower bound and verify that it is achieved by the kernel density estimator. The Donsker-type theorem and the verification of semiparametric efficiency of the kernel-based estimator are the main results on density estimation in the present chapter. They are presented in Section 4.3. Donsker-type theorems can be regarded as frequentist versions of functional Bernstein–von Mises theorems to some extent. In particular, our methods and techniques are interesting for both the frequentist and Bayesian community. The optimal limiting distribution in the sense of the convolution theorem is relevant in the context of Bayesian Bernstein–von Mises theorems in the following sense: If this lower bound is attained, Bayesian credible sets are optimal asymptotic frequentist confidence sets as argued in Castillo and Nickl (2014); see also (Nickl and Söhl, 2017, p. 12) who address Bernstein–von Mises theorems in the context of compound Poisson processes. Our approach concerning the question of efficiency is connected to the recent work Nickl and Ray (2018) where a Bernstein–von Mises theorem for multidimensional diffusions and efficiency of the limit distribution is established. We thank Richard Nickl for the private communication that motivated the derivation of the semiparametric lower bound in this work.

Minimax optimal adaptive drift estimation in sup-norm Subject of Section 4.4 is an adaptive scheme for the sup-norm rate-optimal estimation of the drift coefficient. This is a main contribution and initial motivation of this thesis. Our approach for estimating the drift coefficient is based on Lepski's method for adaptive estimation and the exponential inequalities presented in Section 4.2. For proving upper bounds on the expected sup-norm loss, we follow closely the ideas developed in Giné and Nickl (2009) for the estimation of the density and the distribution function in the classical i.i.d. setting. We suggest a purely data-driven bandwidth choice $\widehat{h}_t$ for the estimator $b_{t,K}$ defined in (4.1.5) and derive upper bounds on the convergence rate of the expected sup-norm risk uniformly over Hölder balls in Theorem 52, imposing very mild conditions on the drift coefficient. To establish minimax optimality of the rate, we prove lower bounds presented in Theorem 53.

Simultaneous adaptive density and drift estimation Observing from (4.1.3) that the invariant density is a transformation of the integrated drift coefficient, it is not surprising that we can carry over the aforementioned approach in Giné and Nickl (2009) (which aims at simultaneous estimation of the distribution function and density in the i.i.d. framework) to the problems of invariant density and drift estimation. We suggest a simultaneous bandwidth selection procedure that allows to derive a result in the spirit of their Theorem 2. Adjusting the procedure from Section 4.4 for choosing the bandwidth $\widehat{h}_t$ in a data-driven way, we can find a bandwidth $\widehat{h}_t$ such that $\rho_{t,K}(\widehat{h}_t)$ is an asymptotically efficient estimator in $\ell^\infty(\mathbb{R})$ for the invariant density and, at the same time, $b_{t,K}(\widehat{h}_t)$ estimates the drift coefficient with minimax optimal rate of convergence wrt sup-norm risk. We formulate this result in Theorem 54.

All proofs are deferred to Section 4.6 and Section 4.7.

4.2 Preliminaries

We will investigate the question of adapting to unknown Hölder smoothness. For ease of presentation, we will suppose in the sequel that $\sigma \equiv 1$. The subsequent results however can be extended to the case of a general diffusion coefficient fulfilling standard regularity and boundedness assumptions. We refer to Chapter 3 where central tools for the investigation of the present work are derived for more general σ as in Definition 39. Recall the definition of the class $\Sigma = \Sigma(\mathcal{C}, A, \gamma, \sigma)$ of drift functions in (4.1.2).

Definition 40. Given $\beta, \mathcal{L} > 0$, denote by $\mathcal{H}_{\mathbb{R}}(\beta, \mathcal{L})$ the *Hölder class* (on $\mathbb{R}$) as the set of all functions $f\colon \mathbb{R} \to \mathbb{R}$ which are $l := \lfloor \beta \rfloor$-times differentiable and for which

$$\begin{aligned} \|f^{(k)}\|_\infty \leq\ & \mathcal{L} && \forall\, k = 0,1,...,l, \\ \|f^{(l)}(\cdot + s) - f^{(l)}(\cdot)\|_\infty \leq\ & \mathcal{L}|s|^{\beta - l} && \forall\, s \in \mathbb{R}. \end{aligned}$$

Set

$$\Sigma(\beta,\mathcal{L}) = \Sigma(\beta,\mathcal{L},\mathcal{C},A,\gamma) := \Big\{ b \in \Sigma(\mathcal{C},A,\gamma,1) \colon \rho_b \in \mathcal{H}_{\mathbb{R}}(\beta+1,\mathcal{L}) \Big\}. \tag{4.2.7}$$

Here, $\lfloor \beta \rfloor$ denotes the greatest integer strictly smaller than β.

Considering the class of drift coefficients $\Sigma(\beta,\mathcal{L})$, we use kernel functions satisfying the following assumptions,

- $K : \mathbb{R} \to \mathbb{R}$ is Lipschitz continuous and symmetric;
- $\operatorname{supp}(K) \subseteq [-1/2, 1/2]$;
- for some $\alpha \geq \beta + 1, K$ is of order α, i.e., $\int_{\mathbb{R}} K(y)\mathrm{d}y = 1$, $\int_{\mathbb{R}} y^j K(y)\mathrm{d}y = 0$, $j = 1, \dots, \lfloor \alpha \rfloor$, $\int_{\mathbb{R}} |y|^\alpha |K(y)|\mathrm{d}y < \infty$.

(4.2.8)

The subsequent deep results from Chapter 3 are fundamental for the investigation of the sup-norm risk. They rely on diffusion specific properties, in particular the existence of local time, on the one hand, and classical empirical process methods like the generic chaining device on the other hand. In the classical setting of statistical inference based on i.i.d. observations $X_1, ..., X_n$, the analysis of sup-norm risks typically requires investigating empirical processes of the form $\left(n^{-1}\sum_{i=1}^n f(X_i)\right)_{f\in\mathcal{F}}$, indexed by a possibly infinite-dimensional class $\mathcal{F}$ of functions which, in many cases, are assumed to be uniformly bounded. Analogously, in the current continuous, non-i.i.d. setting, our analysis raises questions about empirical processes of the form

$$\left(\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s\right)_{f\in\mathcal{F}} \qquad \text{and, more generally,} \qquad \left(\frac{1}{t}\int_0^t f(X_s)\mathrm{d}X_s\right)_{f\in\mathcal{F}}.$$

Clearly, the finite variation part of the stochastic integral entails the need to look at unbounded function classes since we do not want to restrict to bounded drift coefficients. Answers are given in Chapter 3 where we provide exponential tail inequalities both for

$$\sup_{f\in\mathcal{F}}\left[\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s \;-\; \int f\mathrm{d}\mu_b\right] \text{ and } \sup_{f\in\mathcal{F}}\left[\frac{1}{t}\int_0^t f(X_s)\mathrm{d}X_s \;-\; \int bf\mathrm{d}\mu_b\right],$$

imposing merely standard entropy conditions on $\mathcal{F}$. As can be seen from the construction of the estimators, we have to exploit these results in order to deal with both empirical diffusion processes induced by the kernel density estimator $\rho_{t,K}(h)$ (see (4.1.4)) and with stochastic integrals like the estimator $\overline{\rho}_{t,K}(h)$ of the derivative of the invariant density (see (4.1.6)). One first crucial auxiliary result for proving the convergence properties of the estimation schemes proposed in Sections 4.4 and 4.5 is stated in the following

Proposition 41 (Concentration of the estimator $\overline{\rho}_{t,K}(h)$ of $\rho_b'/2$)**.** *Given a continuous record of observations $X^t = (X_s)_{0\leq s\leq t}$ of a diffusion X with $b \in \Sigma = \Sigma(\mathcal{C}, \mathcal{A}, \gamma, 1)$ as introduced in Definition 39 and a kernel K satisfying* (4.2.8)*, define the estimator $\overline{\rho}_{t,K}$ according to* (4.1.6)*. Then, there exist constants $\mathbb{L}, \widetilde{\mathbb{L}}_0$ such that, for any $u, p \geq 1$, $h \in (0,1)$, $t \geq 1$,*

$$\sup_{b\in\Sigma} \left(\mathbb{E}_b\left[\|\overline{\rho}_{t,K}(h) - \mathbb{E}_b\left[\overline{\rho}_{t,K}(h)\right]\|_\infty^p\right]\right)^{\frac{1}{p}} \leq \phi_{t,h}(p),$$
$$\sup_{b\in\Sigma}\mathbb{P}_b\left(\sup_{x\in\mathbb{R}}\left|\overline{\rho}_{t,K}(h)(x) - \mathbb{E}_b\left[\overline{\rho}_{t,K}(h)(x)\right]\right| > \mathrm{e}\phi_{t,h}(u)\right) \leq \mathrm{e}^{-u}, \tag{4.2.9}$$

for

$$\begin{aligned}\phi_{t,h}(u) := \mathbb{L}\Bigg\{&\frac{1}{\sqrt{t}}\Big\{\left(\log\left(\frac{ut}{h}\right)\right)^{3/2} + \left(\log\left(\frac{ut}{h}\right)\right)^{1/2} + u^{3/2}\Big\}\\ &+ \frac{u}{th} + \frac{1}{h}\exp\left(-\widetilde{\mathbb{L}}_0 t\right) + \frac{1}{\sqrt{th}}\left(\log\left(\frac{ut}{h}\right)\right)^{1/2}\\ &+ \frac{1}{t^{3/4}\sqrt{h}}\log\left(\frac{ut}{h}\right) + \frac{1}{\sqrt{th}}\left\{\sqrt{u} + \frac{u}{t^{1/4}}\right\}\Bigg\}.\end{aligned} \tag{4.2.10}$$

The proof of the preceding proposition can be found in Section 4.7.1 and relies on uniform concentration results for stochastic integrals from Chapter 3. These results further allow to prove the following result on the sup-norm distance $\|t^{-1}L_t^\bullet(X) - \rho_{t,K}(h)\|_\infty$ between the local time and the kernel density estimator. The exponential inequality for this distance will be the key to transferring the Donsker theorem for the local time to the kernel density estimator. It can also be interpreted as a result on the uniform approximation error of the scaled local time by its smoothed version, noting that $\rho_{t,K}(h)$ can be seen as a convolution of a mollifier and a scaled version of diffusion local time. The next result actually parallels Theorem 1 in Giné and Nickl (2009) which states a subgaussian inequality for the distribution function in the classical i.i.d. set-up. It serves as an important tool for the analysis of the proposed adaptive scheme for simultaneous estimation of the distribution function and the associated density in Giné and Nickl (2009). The subsequent proposition plays an analogue role for the adaptive scheme for simultaneous estimation of the invariant density and the drift coefficient presented in Section 4.5.

Proposition 42 (Theorem 27 in Chapter 3)**.** *Given a diffusion X with $b \in \Sigma(\beta, \mathcal{L})$, for some $\beta, \mathcal{L} > 0$, consider some kernel function K fulfilling* (4.2.8) *and $h = h_t \in (0,1)$ such that $h_t \geq t^{-1}$. Then, there exist positive constants $\mathcal{V}, \xi_1, \Lambda_0, \Lambda_1$ and $\mathbb{L}$ such that, for all $\lambda \geq \lambda_0(h)$, where*

$$\begin{aligned}\lambda_0(h) := 8\Lambda_0\Bigg[&\sqrt{h}\mathcal{V}\mathrm{e}\mathbb{L}\left\{1 + \log\left(\frac{1}{\sqrt{h}\mathcal{V}}\right) + \log t\right\} + \mathrm{e}\mathbb{L}\sqrt{t}\exp(-\xi_1 t)\\ &+ \sqrt{t}h^{\beta+1}\frac{\mathcal{L}}{2\lfloor\beta+1\rfloor!}\int|K(v)v^{\beta+1}|\mathrm{d}v\Bigg],\end{aligned}$$

and any $t > 1$,

$$\sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{P}_b\left(\sqrt{t}\Big\|\rho_{t,K}(h) - \frac{L_t^\bullet(X)}{t}\Big\|_\infty > \lambda\right) \;\le\; \exp\left(-\frac{\Lambda_1\lambda}{\sqrt{h}}\right).$$

The very first step of our approach to sup-norm adaptive drift estimation consists in estimating the invariant density in sup-norm loss. Corresponding upper bounds on the sup-norm risk have been investigated in Chapter 3. We next cite these bounds for the local time estimator and the kernel density estimator. Our estimation procedure does *not* involve the local time density estimator. For the sake of presenting a complete statistical sup-norm analysis of ergodic scalar diffusions based on continuous observations, we still include it here.

Lemma 43 (Moment bound on the supremum of centred diffusion local time, Corollary 28 of Chapter 3)**.** *Let X be as in Definition 39. Then, there are positive constants ζ, ζ_1 such that, for any $p, t \ge 1$,*

$$\sup_{b\in\Sigma(\mathcal{C},A,\gamma,1)} \left(\mathbb{E}_b\left[\Big\|\frac{L_t^\bullet(X)}{t} - \rho_b\Big\|_\infty^p\right]\right)^{\frac{1}{p}}$$
$$\le\; \zeta\Big(\frac{p}{t} + \frac{1}{\sqrt{t}}\big(1+\sqrt{p}+\sqrt{\log t}\big) \;+\; t\mathrm{e}^{-\zeta_1 t}\Big).$$

In Chapter 3, we have also shown the analogue fundamental result for the sup-norm risk of the kernel density estimator. The following upper bounds will be essential for deriving convergence rates of the Nadaraya–Watson-type drift estimator (see (4.1.5)).

Proposition 44 (Concentration of the kernel invariant density estimator, Corollary 26 of Chapter 3)**.** *Let X be a diffusion with $b \in \Sigma(\beta,\mathcal{L})$, for some $\beta, \mathcal{L} > 0$, and let K be a kernel function fulfilling* (4.2.8)*. Given some positive bandwidth h, define the estimator $\rho_{t,K}(h)$ according to* (4.1.4)*. Then, there exist positive constants ν_1, ν_2, ν_3 such that, for any $p, u \ge 1$, $t > 0$,*

$$\sup_{b\in\Sigma(\beta,\mathcal{L})} \left(\mathbb{E}_b\left[\|\rho_{t,K}(h)-\rho_b\|_\infty^p\right]\right)^{\frac{1}{p}} \le\; \psi_{t,h}(p), \tag{4.2.11}$$
$$\sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{P}_b\left(\|\rho_{t,K}(h)-\rho_b\|_\infty \ge \mathrm{e}\psi_{t,h}(u)\right) \le\; \mathrm{e}^{-u},$$

for

$$\psi_{t,h}(u) :=\; \frac{\nu_1}{\sqrt{t}}\left\{1+\sqrt{\log\left(\frac{1}{\sqrt{h}}\right)}+\sqrt{\log(ut)}+\sqrt{u}\right\}$$
$$+\frac{\nu_2 u}{t}+\frac{1}{h}\mathrm{e}^{-\nu_3 t}+\frac{\mathcal{L}h^{\beta+1}}{\lfloor\beta+1\rfloor!}\int|v^{\beta+1}K(v)|\mathrm{d}v. \tag{4.2.12}$$

Specifying to $h = h_t \sim t^{-1/2}$, an immediate consequence of (4.2.11) is the convergence rate $\sqrt{\log t/t}$ for the risk $\sup_{b\in\Sigma(\beta,\mathcal{L})}\mathbb{E}_b\left[\|\rho_{t,K}(h)-\rho_b\|_\infty\right]$ of the kernel density estimator

$\rho_{t,K}(t^{-1/2})$. Note that we obtain the parametric convergence rate for the bandwidth choice $t^{-1/2}$ which in particular does not depend on the (typically unknown) order of smoothness of the drift coefficient. Thus, there is no extra effort needed for adaptive estimation of the invariant density. This phenomenon appears only in the scalar setting.

4.3 Donsker-type theorems and asymptotic efficiency of kernel invariant density estimators

This section is devoted to the study of weak convergence properties of the kernel density estimator $\rho_{t,K}$. Using the exponential inequality for $\|t^{-1}L_t^\bullet(X) - \rho_{t,K}(h)\|_\infty$ (Proposition 42 from Section 4.2), we derive a uniform CLT for the kernel invariant density estimator. In particular, the result holds for the 'universal' bandwidth choice $h \sim t^{-1/2}$. Furthermore, we use the general theory developed in van der Vaart and Wellner (1996) for establishing asymptotic semiparametric efficiency of $\rho_{t,K}(t^{-1/2})$ in $\ell^\infty(\mathbb{R})$.

4.3.1 Donsker-type theorems

The exponential inequality for the sup-norm difference of the kernel and the local time density estimator stated in Proposition 42 allows to transfer an existing Donsker theorem for the local time density estimator presented in van der Vaart and van Zanten (2005).

Proposition 45. *Given a diffusion X with $b \in \Sigma(\beta, \mathcal{L})$, consider some kernel function K fulfilling* (4.2.8). *Define the estimator $\rho_{t,K}(h)$ according to* (4.1.4) *with bandwidth $h = h_t \in [t^{-1}, 1)$ satisfying $\sqrt{t}h^{\beta+1} \to 0$, as $t \to \infty$. Then,*

$$\sqrt{t}\,(\rho_{t,K}(h) - \rho_b) \overset{\mathbb{P}_b}{\Longrightarrow} \mathbb{H}, \qquad \text{as } t \to \infty,$$

in $\ell^\infty(\mathbb{R})$, where $\mathbb{H}$ is a centred, Gaussian random map with covariance structure

$$\mathbb{E}[\mathbb{H}(x)\mathbb{H}(y)] = 4m(\mathbb{R})\rho_b(x)\rho_b(y) \times \int_{\mathbb{R}} (\mathbb{1}\{[x,\infty)\} - F_b)(\mathbb{1}\{[y,\infty)\} - F_b)\mathrm{d}s, \tag{4.3.13}$$

m and s denoting the speed measure and the scale function of X, respectively.

An analogue result is proven in Nickl and Ray (2018) for a plug-in estimator of the invariant density that is induced by a MAP estimate for a periodic drift coefficient b based on a Bayesian approach.

Remark 46. Donsker-type results turn out to be useful far beyond the question of the behaviour of the density estimator wrt the sup-norm as a specific loss function. In particular, they provide immediate access to solutions of statistical problems concerned with functionals of the invariant density ρ_b. Clearly, this includes the estimation of bounded, linear functionals of ρ_b such as integral functionals, to name just one common class. As an instance, Kutoyants and Yoshida (2007) study the estimation of moments $\mu_b(G)$ for known functions G. The target $\mu_b(G)$ is estimated by the empirical moment estimator $t^{-1}\int_0^t G(X_s)\mathrm{d}s$, and it is shown that this estimator is asymptotically efficient in the sense of local asymptotic minimaxity (LAM) for polynomial loss functions. Parallel results can directly be deduced from the Donsker theorem. Defining the linear functional $\Phi_G\colon \ell^\infty(\mathbb{R}) \to \mathbb{R}$, $h \mapsto \int h(x)G(x)\mathrm{d}x$, the target can be written as $\Phi_G(\rho_b)$, and the empirical moment estimator equals the linear functional applied to the local time estimator, that is,

$$\frac{1}{t}\int_0^t G(X_s)\mathrm{d}s \;=\; \Phi_G(t^{-1}L_t^\bullet(X)).$$

Thus, if Φ_G is bounded, it follows from the results of van der Vaart and van Zanten (2005) (see Lemma 56 in Section 4.6) and from Proposition 45, respectively, that

$$\sqrt{t}\left(\Phi_G(t^{-1}L_t^\bullet(X)) - \Phi_G(\rho_b)\right) \text{ as well as } \sqrt{t}\left(\Phi_G(\rho_{t,K}(t^{-1/2})) - \Phi_G(\rho_b)\right)$$

are asymptotically normal with the limiting distribution $\Phi_G(\mathbb{H})$. Optimality of $\Phi_G(\mathbb{H})$ in the sense of the convolution theorem 3.11.2 in van der Vaart and Wellner (1996) will be shown in the next section.

Not only linear, but also nonlinear functions that allow for suitable linearisations can be analysed, once the required CLTs and optimal rates of convergence are given. This is related to the so-called *plug-in property* introduced in Bickel and Ritov (2003). The suggested connection is explained a bit more detailed in Giné and Nickl (2009).

4.3.2 Semiparametric lower bounds for estimation of the invariant density

We now want to analyse semiparametric optimality aspects of the limiting distribution in Proposition 45 as treated in Chapter 3.11 in van der Vaart and Wellner (1996) or Chapter 25 of van der Vaart (1998). To this end, we first look at lower bounds.

Denote by $\mathbb{P}_{t,h}$ the law of a diffusion process $Y^t := (Y_s)_{0\le s\le t}$ with perturbed drift coefficient $b + t^{-1/2}h$, given as a solution of the SDE

$$\mathrm{d}Y_s \;=\; \left(b(Y_s) + \frac{h(Y_s)}{\sqrt{t}}\right)\mathrm{d}s + \mathrm{d}W_s, \quad Y_0 = X_0,$$

and denote by $\rho_{b+t^{-1/2}h}$ the associated invariant density. Set $\mathbb{P}_b := \mathbb{P}_{t,0}$, and define the set of experiments

$$\left\{C[0,t], \mathcal{B}\left(C[0,t]\right), \mathbb{P}_{t,h}\colon h \in G\right\}, \quad t > 0, \tag{4.3.14}$$

with $G = \ell^\infty(\mathbb{R}) \cap \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R})$ viewed as a linear subspace of $L^2(\mu_b)$. By construction and Girsanov's Theorem (cf. (Liptser and Shiryaev, 2001, Theorem 7.18)), the log-likelihood is given as

$$\begin{aligned}\log\left(\frac{\mathrm{d}\mathbb{P}_{t,h}}{\mathrm{d}\mathbb{P}_b}\right)(X^t) &= \frac{1}{\sqrt{t}}\int_0^t h(X_s)\mathrm{d}W_s - \frac{1}{2t}\int_0^t h^2(X_s)\mathrm{d}s \\ &= \Delta_{t,h} - \frac{1}{2}\|h\|^2_{L^2(\mu_b)} + o_{\mathbb{P}_b}(1),\end{aligned}$$

where $\Delta_{t,h} := t^{-1/2}\int_0^t h(X_s)\mathrm{d}W_s$. Here, the last line follows from the law of large numbers for ergodic diffusions, and the CLT immediately gives $\Delta_{t,h} \overset{\mathbb{P}_b}{\Longrightarrow} \mathcal{N}(0, \|h\|^2_{L^2(\mu_b)})$. Thus, (4.3.14) is an asymptotically normal model. Lemma 57 from Section 4.6 now implies that the sequence $\Psi(\mathbb{P}_{t,h}) := \rho_{b+t^{-1/2}h}$, $t > 0$, is *regular* (or *differentiable*). In fact, it holds

$$\sqrt{t}(\Psi(\mathbb{P}_{t,h}) - \Psi(\mathbb{P}_b)) \ \longrightarrow_{t\to\infty}\ A'h \quad \text{in } \ell^\infty(\mathbb{R}), \text{ for any } h \in G, \tag{4.3.15}$$

for the continuous, linear operator

$$A' \colon (G, L^2(\mu_b)) \to (\ell^\infty(\mathbb{R}), \|\cdot\|_\infty)\,,\ h \mapsto 2\rho_b(H - \mu_b(H)),$$

with $H(\cdot) := \int_0^\cdot h(v)\mathrm{d}v$. We want to determine the optimal limiting distribution for estimating the invariant density $\rho_b = \Psi(\mathbb{P}_{t,0})$ in $\ell^\infty(\mathbb{R})$ in the sense of the convolution theorem 3.11.2 in van der Vaart and Wellner (1996). Since the distribution of a Gaussian process $\mathfrak{G}$ in $\ell^\infty(\mathbb{R})$ is determined by the covariance structure $\mathrm{Cov}(\mathfrak{G}(x), \mathfrak{G}(y))$, $x, y \in \mathbb{R}$, we need to find the Riesz-representer for pointwise evaluations $b_x^* \circ A' \colon G \to \mathbb{R}$, where $b_x^* \colon \ell^\infty(\mathbb{R}) \to \mathbb{R}$, $f \mapsto f(x)$, for any $x \in \mathbb{R}$. Stated differently, we need to find the Cramér–Rao lower bound for pointwise estimation of $\rho_b(x)$, $x \in \mathbb{R}$. Speaking about these one-dimensional targets in $\mathbb{R}$ such as point evaluations $\rho_b(x)$ or linear functionals of the invariant density, we refer to *semiparametric Cramér–Rao lower bounds* as the variance of the optimal limiting distribution from the convolution theorem. This last quantity is a lower bound for the variance of any limiting distribution of a regular estimator.

Our first step towards this goal is to look at integral functionals which we will use to approximate the pointwise evaluations. For any continuous, linear functional $b^* \colon \ell^\infty(\mathbb{R}) \to \mathbb{R}$, we can infer from (4.3.15) that

$$\sqrt{t}(b^*(\Psi(\mathbb{P}_{t,h})) - b^*(\Psi(\mathbb{P}_b))) \ \longrightarrow_{t\to\infty}\ b^*(A'h) \quad \text{in } \mathbb{R}, \text{ for all } h \in G.$$

Considering $\Phi_g \colon \ell^\infty(\mathbb{R}) \to \mathbb{R}$, $f \mapsto \int g(x)f(x)\mathrm{d}x$, for a function $g \in C_c^\infty(\mathbb{R})$, and letting $\Phi_g(\mathbb{P}_{t,h}) := \Phi_g(\rho_{b+t^{-1/2}h})$, this becomes

$$\sqrt{t}(\Phi_g(\mathbb{P}_{t,h}) - \Phi_g(\mathbb{P}_b)) \ \longrightarrow_{t\to\infty}\ \int g(x)(A'h)(x)\mathrm{d}x.$$

The limit defines a continuous, linear map $\kappa \colon (G, L^2(\mu_b)) \to \mathbb{R}$ with representation

$$\begin{aligned}\kappa(h) = &\int g(x)(A'h)(x)\mathrm{d}x \ = \ \int 2g(x)(H(x) - \mu_b(H))\rho_b(x)\mathrm{d}x \\ = &\ 2\langle g_c, H_c\rangle_{\mu_b} \ = \ 2\langle \mathrm{L}_b\mathrm{L}_b^{-1}g_c, H_c\rangle_{\mu_b} \ = \ -\langle \partial \mathrm{L}_b^{-1}g_c, h\rangle_{\mu_b}.\end{aligned} \tag{4.3.16}$$

Here and throughout the sequel, L_b denotes the generator of the diffusion process X with drift coefficient b, i.e., $\mathrm{L}_b f = b\partial f + \partial^2 f/2$, for any $f \in C_c^\infty(\mathbb{R})$, and $f_c := f - \mu_b(f)$ denotes the centred version of f, for any function $f \in L^1(\mu_b)$. Note that $g_c \in \mathrm{Rg}(\mathrm{L}_b)$ due to the following lemma whose proof is deferred to Section 4.6.

Lemma 47. *Let $g \in C_c^\infty(\mathbb{R})$, and set*

$$h(z,x) := \frac{\mathbb{1}\{z \geq x\} - F_b(z)}{\rho_b(z)}, \quad z, x \in \mathbb{R}.$$

Then, g_c is contained in the image of the generator L_b, and

$$L_b^{-1}(g_c) \;=\; \mathcal{T}(z) \;:=\; \int_0^z \int 2g(x)\rho_b(x)h(u,x)\mathrm{d}x\mathrm{d}u.$$

In particular,

$$\iint g(x)H(x,y)g(y)\mathrm{d}y\mathrm{d}x \;=\; \|\partial L_b^{-1}(g_c)\|^2_{L^2(\mu_b)},$$

where $H(x,y) := \mathbb{E}[\mathbb{H}(x)\mathbb{H}(y)]$, $x, y \in \mathbb{R}$, for the Gaussian process $\mathbb{H}$ fulfilling (4.3.13).

We conclude by means of Theorem 3.11.2 in van der Vaart and Wellner (1996) that the *Cramér–Rao lower bound for estimation of $\Phi_g(\mathbb{P}_b)$* is given by $\|\partial \mathrm{L}_b^{-1} g_c\|^2_{L^2(\mu_b)}$. Using an approximation procedure, it then can be shown that the *Cramér–Rao lower bound for pointwise estimation of $\rho_b(y)$* is defined via

$$\mathrm{CR}(y) := \|2\rho_b(y)h(\cdot,y)\|^2_{L^2(\mu_b)}, \quad \text{for any } y \in \mathbb{R}. \tag{4.3.17}$$

For details, see Proposition 58 in Section 4.6. The same arguments apply to estimation of linear combinations $u\rho_b(x) + v\rho_b(y)$, $u, v, x, y \in \mathbb{R}$, and the corresponding Cramér–Rao bound reads

$$\|2u\rho_b(x)h(\cdot,x) + 2v\rho_b(y)h(\cdot,y)\|^2_{L^2(\mu_b)}.$$

It follows that the covariance of the optimal Gaussian process in the convolution theorem is given as

$$\mathrm{CR}(x,y) := 4\rho_b(x)\rho_b(y)\int h(z,x)h(z,y)\rho_b(z)\mathrm{d}z, \quad x, y \in \mathbb{R}. \tag{4.3.18}$$

Summing up, we can deduce from Theorem 3.11.2 in van der Vaart and Wellner (1996) how possible limiting distributions of regular estimators of the invariant density can look like, thereby revealing the optimal limiting distribution. Recall that an estimator $\widehat{\rho}_t$ of the invariant density is called regular if it has a weak limit which is stable with respect to small perturbations of the model, more precisely, if

$$\sqrt{t}\,(\widehat{\rho}_t - \Psi(\mathbb{P}_{t,h})) \overset{\mathbb{P}_{t,h}}{\Longrightarrow} \mathscr{L}, \quad \text{as } t \to \infty, \text{ for any } h \in G,$$

for a fixed, tight Borel probability measure $\mathscr{L}$ in $\ell^\infty(\mathbb{R})$.

Proposition 48. *Let $\widehat{\rho}_t$ be a regular estimator of the invariant density ρ_b. Then, there exist a tight centred, Borel-measurable Gaussian process $\mathfrak{G}$ in $\ell^\infty(\mathbb{R})$ with covariance structure*

$$\mathrm{Cov}(\mathfrak{G}(x), \mathfrak{G}(y)) = \mathrm{CR}(x,y), \quad \textit{for all } x, y \in \mathbb{R},$$

and with $\mathrm{CR}(x,y)$ *as in* (4.3.18) *as well as an independent, tight, Borel-measurable map M in $\ell^\infty(\mathbb{R})$ such that the limit distribution $\mathscr{L}$ of $\sqrt{t}(\widehat{\rho}_t - \rho_b)$ satisfies*

$$\mathscr{L} \sim \mathfrak{G} \;+\; M.$$

4.3.3 Semiparametric efficiency of the kernel density estimator

Having characterised the optimal limit distribution in the previous section, it is natural to ask in a next step for an efficient estimator of linear functionals of the invariant density such as pointwise estimation, functionals of the form $\Phi_g(\mathbb{P}_b) := \mu_b(g) = \int g d\mu_b$ or, even more, for estimation of ρ_b in $\ell^\infty(\mathbb{R})$.

Definition 49. An estimator $\widehat{\rho}_t$ of the invariant density is called *asymptotically efficient* in $\ell^\infty(\mathbb{R})$ if the estimator is regular, i.e.,

$$\sqrt{t}\,(\widehat{\rho}_t - \Psi(\mathbb{P}_{t,h})) \overset{\mathbb{P}_{t,h}}{\Longrightarrow} \mathscr{L}, \qquad \text{as } t \to \infty, \text{ for any } h \in G,$$

for a fixed, tight Borel probability measure $\mathscr{L}$ in $\ell^\infty(\mathbb{R})$, and if $\mathscr{L}$ is the law of the centred Gaussian process $\mathfrak{G}$ specified in Proposition 48, i.e., $\widehat{\rho}_t$ achieves the optimal limiting distribution.

Given an asymptotically efficient estimator $\widehat{\rho}_t$ in $\ell^\infty(\mathbb{R})$ and any bounded, linear functional $b^*\colon \ell^\infty(\mathbb{R}) \to \mathbb{R}$, efficiency of the estimator $b^*(\widehat{\rho}_t)$ for estimation of $b^*(\rho_b)$ then immediately follows. Our next result shows that estimation via $\rho_{t,K}(h)$ for the universal bandwidth choice $h \sim t^{-1/2}$ is suitable for the job.

Theorem 50. *The invariant density estimator* $\rho_{t,K}(t^{-1/2})$ *defined according to* (4.1.4) *is an asymptotically efficient estimator in* $\ell^\infty(\mathbb{R})$.

Remark 51. (a) From the proof of Theorem 50, it can be inferred that the local time estimator $t^{-1}L_t^\bullet(X)$ is an asymptotically efficient estimator, as well.

(b) In terms of earlier research on efficient estimation of the density as a function in $\ell^\infty(\mathbb{R})$, we shall mention Kutoyants (1998) and Negri (2001). The works deal with the efficiency of the local time estimator $t^{-1}L_t^\bullet(X)$ in the LAM sense for certain classes of loss functions. Subject of (Kutoyants, 1998, Section 8) are $L^2(\nu)$ risks for some finite measure ν on $\mathbb{R}$ of the form $t\mathbb{E}_b \int |\widetilde{\rho}_t(x) - \rho_b(x)|^2 \nu(\mathrm{d}x)$, for estimators $\widetilde{\rho}_t$ of ρ_b, whereas Negri complements this work for risks of the form $\mathbb{E}_b\left[g\left(\sqrt{t}\|\widetilde{\rho}_t - \rho_b\|_\infty\right)\right]$ for a class of bounded, positive functions g. Of course, the distribution appearing in the lower bound corresponds to the optimal distribution in the sense of Proposition 48. The derivation in Kutoyants (1998) of the lower bound is based on the van Trees inequality as established in Gill and Levit (1995) as an alternative to the classical approach relying on Hájek–Le Cam theory. On the other hand, Negri's method follows Millar (1983) and makes use of the idea of convergence of experiments originally provided by Le Cam. The optimal distribution in the sense of a convolution theorem is not shown neither is any asymptotic efficiency result in $\ell^\infty(\mathbb{R})$ for the kernel density estimator.

4.4 Minimax optimal adaptive drift estimation wrt sup-norm risk

We now turn to the original question of estimating the drift coefficient in a completely data-driven way. The aim of this section is to suggest a scheme for rate-optimal choice of the bandwidth h, based on a continuous record of observations $X^t \equiv (X_s)_{0\leq s\leq t}$ of a diffusion as introduced in Definition 39, optimality considered in terms of the sup-norm risk. Since we stick to the continuous framework, our previous concentration results are directly applicable, allowing, e.g., for the straightforward derivation of upper bounds on the variance of the estimator $\overline{\rho}_{t,K}(h)$ of the order

$$\overline{\sigma}^2(h,t) := \frac{(\log(t/h))^3}{t} + \frac{\log(t/h)}{th}. \tag{4.4.19}$$

Standard arguments provide for any $b \in \Sigma(\beta,\mathcal{L})$ bounds on the associated bias of order $B(h) \lesssim h^\beta$. In case of known smoothness β, one can then easily derive the optimal bandwidth choice h_t^* by balancing the components of the bias-variance decomposition

$$\sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{E}_b\left[\left\|\overline{\rho}_{t,K}(h) - \frac{\rho_b'}{2}\right\|_\infty\right] \leq B(h) + \mathcal{K}\overline{\sigma}(h,t),$$

resulting in $h_t^* \simeq (\log t/t)^{\frac{1}{2\beta+1}}$. In order to remove the (typically unknown) order of smoothness β from the bandwidth choice, we need to find a data-driven substitute for the upper bound on the bias in the balancing process. Heuristically, this is the idea behind the Lepski-type selection procedure suggested in (4.4.22) below.

1. Specify the discrete grid of candidate bandwidths

$$\mathcal{H} \equiv \mathcal{H}_t := \left\{h_k = \eta^{-k}\colon\ k\in\mathbb{N},\ \eta^{-k} > \frac{(\log t)^2}{t}\right\}, \quad \eta > 1 \text{ arbitrary}, \tag{4.4.20}$$

and define $\overline{\eta}_1 := 24\widetilde{C}_2\|K\|_{L^2(\boldsymbol{\lambda})}\,\mathrm{c}\,\mathrm{e}\sqrt{v}$, $\overline{\eta}_2 := 12\,\mathrm{c}\,\|K\|_{L^2(\boldsymbol{\lambda})}$ and

$$\widetilde{M} = \widetilde{M}_t := C\|\rho_{t,K}(t^{-\frac{1}{2}})\|_\infty, \quad \text{for } C = C(K) := 20\mathrm{e}^2\left(4\overline{\eta}_1 + 2\overline{\eta}_2\right)^2. \tag{4.4.21}$$

2. Set

$$\widehat{h}_t := \max\left\{h\in\mathcal{H}\colon \|\overline{\rho}_{t,K}(h) - \overline{\rho}_{t,K}(g)\|_\infty \leq \sqrt{\widetilde{M}}\overline{\sigma}(g,t) \quad \forall g < h,\ g\in\mathcal{H}\right\}. \tag{4.4.22}$$

The constants involved in the definition of $\overline{\eta}_1$ and $\overline{\eta}_2$ are explained in Remark 63 in Section 4.7.1. For the proposed data-driven scheme for bandwidth choice, we obtain the subsequent

Theorem 52. *For $b \in \Sigma(\beta, \mathcal{L})$ as introduced in Definition 40, consider the SDE* (4.1.1). *Given some kernel K satisfying* (4.2.8), *define the estimators $b_{t,K}(\widehat{h}_t)$ and $\overline{\rho}_{t,K}(\widehat{h}_t)$ according to* (4.1.5), (4.1.6) *and* (4.4.22). *Then, for any $0 < \beta + 1 \leq \alpha$,*

$$\sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{E}_b\left[\left\|\overline{\rho}_{t,K}(\widehat{h}_t) - \frac{\rho_b'}{2}\right\|_\infty\right] \lesssim \left(\frac{\log t}{t}\right)^{\frac{\beta}{2\beta+1}},$$

$$\sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{E}_b\left[\left\|(b_{t,K}(\widehat{h}_t) - b)\cdot \rho_b^2\right\|_\infty\right] \lesssim \left(\frac{\log t}{t}\right)^{\frac{\beta}{2\beta+1}}. \tag{4.4.23}$$

The suggested estimators for ρ_b' and b, respectively, are rate optimal as the following lower bounds imply.

Theorem 53. *Let $\beta, \mathcal{L}, \mathcal{C}, A, \gamma \in (0,\infty)$, and assume that $\Sigma(\beta, \mathcal{L}/2, \mathcal{C}/2, A, \gamma) \neq \emptyset$. Then,*

$$\liminf_{t\to\infty} \inf_{\widetilde{\partial\rho}_t} \sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{E}_b\left[\left(\frac{\log t}{t}\right)^{-\frac{\beta}{2\beta+1}} \|\widetilde{\partial\rho}_t - \rho_b'\|_\infty\right] > 0, \tag{4.4.24}$$

$$\liminf_{t\to\infty} \inf_{\widetilde{b}} \sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{E}_b\left[\left(\frac{\log t}{t}\right)^{-\frac{\beta}{2\beta+1}} \left\|(\widetilde{b} - b)\cdot\rho_b^2\right\|_\infty\right] > 0, \tag{4.4.25}$$

where the infimum is taken over all possible estimators $\widetilde{\partial\rho}_t$ of ρ_b' and $\widetilde{b}$ of the drift coefficient b, respectively.

The proof of the preceding theorem is based on classical tools from minimax theory as laid down in Tsybakov (2009). Precisely, it relies on the Kullback version of Theorem 2.7 in Tsybakov (2009), his main theorem on lower bounds. This result slightly reformulated in terms of our problem is stated in Lemma 64 of Section 4.7.

4.5 Simultaneous estimation

The result presented in the previous section is a classical specification of Lepski's procedure which complements the study of one-dimensional drift estimation in the continuous framework. However, as will be shown in the sequel, our techniques allow for results which go beyond classical issues such as minimax optimality. We will adjust the bandwidth selection procedure from Section 4.4 in such a way that the resulting data-driven bandwidth choice yields an asymptotically efficient estimator of the invariant density and, simultaneously, also gives a drift estimator which achieves the best possible convergence rate in sup-norm loss. The approach is an adaptation of the method developed by Giné and Nickl (2009) to the scalar diffusion set-up.

1. Define the set of candidate bandwidths $\mathcal{H} = \mathcal{H}_t$ according to (4.4.20), and introduce $h_{\min} := \min\{h_k \in \mathcal{H} \colon k \in \mathbb{N}\}$. Set

$$\widehat{M} := \widehat{M}_t := C\|\rho_{t,K}(h_{\min})\|_\infty,$$

for the constant C defined in (4.4.21).

2. Set

$$\widehat{h}_t := \max\left\{h \in \mathcal{H}\colon \|\overline{\rho}_{t,K}(h) - \overline{\rho}_{t,K}(g)\|_\infty \leq \sqrt{\widehat{M}}\overline{\sigma}(g,t)\ \forall g < h,\right.$$
$$\left. g \in \mathcal{H},\ \textbf{and}\ \|\rho_{t,K}(h) - \rho_{t,K}(h_{\min})\|_\infty \leq \frac{\sqrt{h}(\log(1/h))^4}{\sqrt{t}\log t}\right\}. \tag{4.5.26}$$

Our goal is to estimate the invariant density ρ_b and the drift coefficient b via $\rho_{t,K}(\widehat{h}_t)$ and

$$\widetilde{b}_{t,K}(\widehat{h}_t)(x) \;:=\; \frac{\overline{\rho}_{t,K}(\widehat{h}_t)(x)}{\rho^+_{t,K}(\widehat{h}_t)(x) + \sqrt{\frac{\log t}{t}}\exp\left(\sqrt{\log t}\right)}, \tag{4.5.27}$$

respectively, by means of the simultaneous, adaptive bandwidth choice $\widehat{h}_t$.

Theorem 54. *Grant the assumptions of Theorem 52. Given some kernel K satisfying* (4.2.8)*, define the estimators $\rho_{t,K}(\widehat{h}_t)$, $\overline{\rho}_{t,K}(\widehat{h}_t)$ and $\widetilde{b}_{t,K}(\widehat{h}_t)$ according to* (4.1.4), (4.1.6), (4.5.27) *and* (4.5.26)*. Then,*

$$\sqrt{t}\left(\rho_{t,K}(\widehat{h}_t) - \rho_b\right) \overset{\mathbb{P}_b}{\Longrightarrow} \mathbb{H}, \qquad \text{as } t \to \infty, \tag{I}$$

in $\ell^\infty(\mathbb{R})$ for $\mathbb{H}$ as in Proposition 45. Furthermore, for any $0 < \beta + 1 \leq \alpha$,

$$\sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{E}_b\left[\left\|\overline{\rho}_{t,K}(\widehat{h}_t) - \frac{\rho_b'}{2}\right\|_\infty\right] \lesssim \left(\frac{\log t}{t}\right)^{\frac{\beta}{2\beta+1}},$$
$$\sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{E}_b\left[\left\|(\widetilde{b}_{t,K}(\widehat{h}_t) - b)\cdot\rho_b^2\right\|_\infty\right] \lesssim \left(\frac{\log t}{t}\right)^{\frac{\beta}{2\beta+1}}. \tag{II}$$

Remark 55. It could be argued that this simultaneous procedure is of limited relevance because we have verified that the simple choice $h_t = t^{-1/2}$ is optimal for the kernel density estimator. In particular, an adaptive bandwidth selection procedure is not required, such that it suffices to apply the data-driven selection procedure stated in Section 4.4 to obtain an optimal bandwidth for the drift estimator. Besides being of theoretical interest, Theorem 54 still creates added value in relevant aspects. The approach that we demonstrate actually could be extended, e.g., to the framework of multivariate diffusion processes. It is known that in higher dimensions, the kernel invariant density estimator is also rate-optimal, but the optimal bandwidth depends on the unknown smoothness such that it has to be chosen adaptively as it is certainly the case for the drift estimator. In this situation, it is very appealing to have one bandwidth selection procedure that works simultaneously both for invariant density and drift estimation. Our result thus is pioneering work in this direction.

4.6 Proofs for Section 4.3

The following result verifies the general conditions for the uniform CLT for diffusion local time given in van der Vaart and van Zanten (2005) for the class of diffusion processes with $b \in \Sigma(\mathcal{C}, A, \gamma, 1)$.

Lemma 56. *For any $b \in \Sigma(\mathcal{C}, A, \gamma, 1)$, it holds*

$$\sqrt{t}\left(\frac{L_t^\bullet(X)}{t} - \rho_b\right) \overset{\mathbb{P}_b}{\Longrightarrow} \mathbb{H}, \qquad \text{as } t \to \infty, \tag{4.6.28}$$

in $\ell^\infty(\mathbb{R})$, where $\mathbb{H}$ is a centred Gaussian random map with covariance structure specified in (4.3.13).

Proof. We verify the conditions of Corollary 2.7 in van der Vaart and van Zanten (2005) by showing that, for any $b \in \Sigma = \Sigma(\mathcal{C}, A, \gamma, 1)$,

(a) $\int F_b^2(x)(1 - F_b(x))^2 \mathrm{d}s(x) < \infty$ and

(b) $\lim_{x\to-\infty} \rho_b^2(x)|s(x)| \log\log|s(x)| = 0$.

With regard to (a), note first that, for $y > A$,

$$\frac{1 - F_b(y)}{\rho_b(y)} = \int_y^\infty \exp\left(2\int_y^v b(z)\mathrm{d}z\right)\mathrm{d}v \le \int_y^\infty \mathrm{e}^{-2\gamma(v-y)}\mathrm{d}v = \frac{1}{2\gamma}$$

and, for $y < -A$,

$$\frac{F_b(y)}{\rho_b(y)} = \int_{-\infty}^y \exp\left(-2\int_v^y b(z)\mathrm{d}z\right)\mathrm{d}v \le \int_{-\infty}^y \mathrm{e}^{2\gamma(v-y)}\mathrm{d}v = \frac{1}{2\gamma}.$$

Exploiting the relation $\rho_b(\mathrm{d}x) = (s'(x)m(\mathbb{R}))^{-1}\mathrm{d}x$ between the invariant measure and the scale function as well as the speed measure, respectively, we obtain $\mathrm{d}s(x) = (\rho_b(x)m(\mathbb{R}))^{-1}\mathrm{d}x$. Consequently, the above bounds imply that

$$\begin{aligned}
&\int F_b^2(x)(1 - F_b(x))^2\mathrm{d}s(x)\\
&\quad\le \frac{1}{m(\mathbb{R})}\int_{\mathbb{R}} \frac{F_b^2(x)(1-F_b(x))^2}{\rho_b(x)}\mathrm{d}x\\
&\quad\le \frac{1}{m(\mathbb{R})}\left[\int_{-\infty}^{-A} \frac{F_b^2(y)}{\rho_b^2(y)}\rho_b(y)\mathrm{d}y + 2A\sup_{x\in[-A,A]}\rho_b^{-1}(x) + \int_A^\infty \frac{(1-F_b(y))^2}{\rho_b^2(y)}\rho_b(y)\mathrm{d}y\right]\\
&\quad< \infty.
\end{aligned}$$

In order to verify (b), recall first that the scale function s of X is given by

$$s(x) = \int_0^x \exp\left(-2\int_0^y b(z)\mathrm{d}z\right)\mathrm{d}y = C_{b,1}\int_0^x \frac{1}{\rho_b(y)}\mathrm{d}y, \quad x \in \mathbb{R}.$$

Since, for any $b \in \Sigma$, $b(x)\operatorname{sgn}(x) \leq -\gamma$ whenever $|x| > A$, we obtain for $x < -A$

$$\begin{aligned}
\rho_b(x)|s(x)| &= C_{b,1}\rho_b(x)\int_{-A}^{0}\frac{1}{\rho_b(y)}\mathrm{d}y \;+\; C_{b,1}\int_x^{-A}\frac{\rho_b(x)}{\rho_b(y)}\mathrm{d}y \\
&\lesssim o(1) + \int_x^{-A}\exp\left(-2\left(\int_x^0 b(z)\mathrm{d}z - \int_y^0 b(z)\mathrm{d}z\right)\right)\mathrm{d}y \\
&\lesssim o(1) + \int_x^{-A}\exp\left(-2\int_x^y b(z)\mathrm{d}z\right)\mathrm{d}y \\
&\lesssim o(1) + \int_x^{-A}\exp\left(-2\gamma(y-x)\right)\mathrm{d}y \\
&\simeq o(1) + \frac{1}{2\gamma}\left(1-\exp\left(2\gamma(A+x)\right)\right) \;\simeq\; o(1) + \frac{1}{2\gamma} \;=\; O(1),
\end{aligned}$$

as $x \to -\infty$. Furthermore, for $x < -A$,

$$\begin{aligned}
\rho_b(x) &= C_{b,1}^{-1}\exp\left(-2\int_x^0 b(y)\mathrm{d}y\right) \;\lesssim\; \exp\left(-2\int_x^{-A} b(y)\mathrm{d}y\right) \\
&\lesssim \exp(2\gamma(A+x)) \;\lesssim\; \mathrm{e}^{2\gamma x},
\end{aligned}$$

and, using the linear growth condition on b,

$$\begin{aligned}
|s(x)| &= \int_x^0 \exp\left(2\int_y^0 b(z)\mathrm{d}z\right)\mathrm{d}y \;\lesssim\; \int_x^0 \exp\left(2\int_y^{-A} b(z)\mathrm{d}z\right)\mathrm{d}y \\
&\lesssim \int_x^0 \exp\left(2\mathcal{C}\int_y^{-A}(1-z)\mathrm{d}z\right)\mathrm{d}y \;\lesssim\; \int_x^0 \exp(\mathcal{C}(y^2-2y))\mathrm{d}y
\end{aligned}$$

such that $|s(x)| = O\left(1+|x|\exp(4\mathcal{C}x^2)\right)$ and $\log\log|s(x)| = O(x^2)$ as $x \to -\infty$. Finally,

$$\rho_b^2(x)\;|s(x)|\;\log\log|s(x)| \;\lesssim \mathrm{e}^{2\gamma x}\;O(1)\;x^2 \;=\; o(1) \quad \text{as } x \to -\infty.$$

Thus, condition (b) of Theorem 2.6 in van der Vaart and van Zanten (2005) is satisfied. Consequently, there exists a tight version of the Gaussian process $\mathbb{H}$, and (4.6.28) holds true. □

Having verified the conditions for the uniform CLT for diffusion local time, this result can be transferred to the kernel density estimator:

Proof of Proposition 45. We apply Proposition 42 to show that

$$\sqrt{t}\left\|\rho_{t,K}(t^{-1/2}) - \frac{L_t^{\bullet}(X)}{t}\right\|_\infty \;=\; o_{\mathbb{P}_b}(1). \tag{4.6.29}$$

There exists a constant $C > 0$ such that $\lambda_t := C\left(t^{-1/4}\left(1+\log t\right) + t^{-\beta/2}\right)$ fulfills the assumption $\lambda_t \geq \lambda_0(h)$ for $h = h_t = t^{-1/2}$. Since $\lambda_t = o(1)$, for any $\epsilon > 0$ and t

sufficiently large,

$$\begin{aligned}&\mathbb{P}_b\left(\sqrt{t}\|\rho_{t,K}(t^{-1/2})-t^{-1}L_t^\bullet(X)\|_\infty>\epsilon\right)\\&\quad\leq\ \mathbb{P}_b\left(\sqrt{t}\|\rho_{t,K}(t^{-1/2})-t^{-1}L_t^\bullet(X)\|_\infty>\lambda_t\right)\\&\quad\leq\ \exp\left(-\Lambda_1\lambda_t t^{1/4}\right)\ =\ \exp\left(-\Lambda_1 C\left((1+\log t)+t^{\frac14-\frac{\beta}{2}}\right)\right)\\&\quad\longrightarrow\ 0,\quad \text{as } t\to\infty.\end{aligned}$$

Consequently, (4.6.29) holds, and Lemma 56 gives the assertion. □

The remainder of this section is devoted to complementing our study of asymptotic efficiency by stating the remaining proofs. We start with verifying differentiability of the operator $G\ni h\mapsto\rho_{b+h}$.

Lemma 57. *For any $h\in G$, set $H(\cdot):=\int_0^\cdot h(v)\mathrm{d}v$. Then, the operator $h\mapsto\rho_{b+h}$ as a function from $(\ell^\infty(\mathbb{R})\cap\mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}),\|\cdot\|_\infty)$ to $(\ell^\infty(\mathbb{R}),\|\cdot\|_\infty)$ is Fréchet-differentiable at $h=0$ in the sense that*

$$\|\rho_{b+h}-\rho_b-2\rho_b(H-\mu_b(H))\|_\infty\ =\ o(\|h\|_\infty).$$

Proof. Let $h\in\ell^\infty(\mathbb{R})\cap\mathrm{Lip}_{\mathrm{loc}}(\mathbb{R})$, and denote by ρ_{b+h} the invariant density corresponding to the diffusion process with drift $b+h$. Note that, for $\|h\|_\infty$ sufficiently small, $b+h\in\Sigma(\widetilde{\mathcal{C}},\widetilde{A},\widetilde{\gamma},1)$ for some positive constants $\widetilde{\mathcal{C}},\widetilde{A},\widetilde{\gamma}$. Set

$$C_g\ :=\ \int_{\mathbb{R}}\exp\left(2\int_0^x g(v)\mathrm{d}v\right)\mathrm{d}x,\quad g\in\{b,b+h\},\quad B(\cdot)\ :=\ \int_0^\cdot b(v)\mathrm{d}v.$$

Then, for any $x\in\mathbb{R}$,

$$\begin{aligned}\rho_{b+h}(x)-\rho_b(x)=&\ C_b^{-1}\mathrm{e}^{2B(x)}\left(\frac{C_b}{C_{b+h}}\,\mathrm{e}^{2H(x)}-1\right)\\=&\ \rho_b(x)\Bigg\{2H(x)+\log\left(\frac{C_b}{C_{b+h}}\right)\\&\qquad+\frac12\left(2H(x)+\log\left(\frac{C_b}{C_{b+h}}\right)\right)^2\mathrm{e}^{\theta_1(x)\left(\log\left(\frac{C_b}{C_{b+h}}\right)+2H(x)\right)}\Bigg\},\end{aligned}$$

for some $\theta_1(x)\in(0,1)$. Moreover,

$$\log\left(\frac{C_b}{C_{b+h}}-1+1\right)\ =\ \frac{C_b}{C_{b+h}}-1+\frac{1/2}{1+\theta_2\left(C_bC_{b+h}^{-1}-1\right)}\left(\frac{C_b}{C_{b+h}}-1\right)^2,$$

for some $\theta_2\in(0,1)$. Next, we will show that

$$\frac{C_b-C_{b+h}}{C_{b+h}}\ =\ -2\int H(v)\rho_b(v)\mathrm{d}v+o(\|h\|_\infty).$$

Note that

$$\begin{aligned}C_b - C_{b+h} &= \int e^{2B(v)}\left(1 - e^{2H(v)}\right) dv \\ &= -\int e^{2B(v)}\left(2H(v) + \frac{1}{2}e^{2\theta_3(v)H(v)}4H^2(v)\right) dv, \quad \theta_3(v) \in (0,1), \\ &= -C_b \int 2H(v)\rho_b(v)dv + o(\|h\|_\infty),\end{aligned}$$

where we have used $|H(v)| \leq |v|\|h\|_\infty$, $v \in \mathbb{R}$, as well as the fact $\int e^{2\int_0^v b(x)+2|h(x)|dx}|v|^2 dv = O(1)$. We conclude

$$\begin{aligned}\frac{C_b - C_{b+h}}{C_{b+h}} &= \frac{-2C_b\mu_b(H) + o(\|h\|_\infty)}{(C_{b+h} - C_b) + C_b} \\ &= -2\mu_b(H) + \frac{2\mu_b(H)(C_{b+h} - C_b)}{(C_{b+h} - C_b) + C_b} + \frac{o(\|h\|_\infty)}{C_{b+h}} \\ &= -2\mu_b(H) + \frac{o(1)(C_{b+h} - C_b)}{o(1) + C_b} + \frac{o(\|h\|_\infty)}{C_{b+h}} \\ &= -2\mu_b(H) + \frac{o(1)O(\|h\|_\infty) + o(\|h\|_\infty)}{o(1) + C_b} + o(\|h\|_\infty) \\ &= -2\mu_b(H) + o(\|h\|_\infty).\end{aligned}$$

Consequently, $\left(\frac{C_b - C_{b+h}}{C_{b+h}}\right)^2 = o(\|h\|_\infty)$, and it follows

$$\log\left(\frac{C_b}{C_{b+h}}\right) = -2\mu_b(H) + o(\|h\|_\infty) + \frac{1}{2}\, O(1)o(\|h\|_\infty) = -2\mu_b(H) + o(\|h\|_\infty).$$

Taking everything into consideration,

$$\begin{aligned}\rho_{b+h}(x) - \rho_b(x) = \rho_b(x)\Big\{&2(H(x) - \mu_b(H)) + o(\|h\|_\infty) + (2(H(x) - \mu_b(H)) + o(\|h\|_\infty))^2 \\ &\times \frac{1}{2}e^{2\theta_1(x)H(x)}e^{\theta_1(x)(-2\mu_b(H)+o(\|h\|_\infty))}\Big\},\end{aligned}$$

and thus

$$\begin{aligned}&\|\rho_{b+h} - \rho_b - 2\rho_b(H - \mu_b(H))\|_\infty \\ &\quad\leq o(\|h\|_\infty) + \left\|x \mapsto \rho_b(x)\left\{16\|h\|_\infty^2 x^2 + o(\|h\|_\infty)\right\} e^{2\|h\|_\infty|x|}O(1)\right\|_\infty = o(\|h\|_\infty),\end{aligned}$$

using that $\sup_{x\in\mathbb{R}} \rho_b(x)e^{2\|h\|_\infty|x|}(1 + x^2) = O(1)$. □

We proceed with verifying the result on the image of the generator L_b and the expression of $\|\partial \mathrm{L}_b^{-1}(g_c)\|_{L^2(\mu_b)}$ in terms of $H(x,y) = \mathbb{E}[\mathbb{H}(x)\mathbb{H}(y)]$. Recall that $\mathbb{H}$ denotes the centred Gaussian process with covariance structure specified by (4.3.13).

Proof of Lemma 47. Rewriting $H(x,y)$ as

$$H(x,y) = 4\rho_b(x)\rho_b(y) \int \left[(\mathbb{1}\{[x,\infty)\}(z) - F_b(z))\ (\mathbb{1}\{[y,\infty)\}(z) - F_b(z)) \right] \rho_b^{-1}(z)\mathrm{d}z \tag{4.6.30}$$

yields

$$\begin{aligned} \iint g(x)H(x,y)g(y)\mathrm{d}y\mathrm{d}x =&\ 4\iint g(x)g(y)\rho_b(x)\rho_b(y)\int h(z,x)h(z,y)\rho_b(z)\mathrm{d}z\mathrm{d}x\mathrm{d}y \\ =&\ \int \left[\int 2g(x)h(z,x)\rho_b(x)dx\right]^2 \rho_b(z)\mathrm{d}z \\ =&\ \Big\| \int 2g(x)\rho_b(x)h(\cdot,x)\mathrm{d}x\Big\|^2_{L^2(\mu_b)} \ =\ \Big\|\frac{\mathrm{d}}{\mathrm{d}z}\ \mathcal{T}(z)\ \Big\|^2_{L^2(\mu_b)}, \end{aligned}$$

where

$$\begin{aligned} \mathcal{T}(z) =&\ \int_0^z \int 2g(x)\rho_b(x)h(u,x)\mathrm{d}x\mathrm{d}u \\ =&\ \int_0^z \int_{-\infty}^u 2g(x)\rho_b(x)\frac{1-F_b(u)}{\rho_b(u)}\mathrm{d}x\mathrm{d}u - \int_0^z\int_u^\infty 2g(x)\rho_b(x)\mathrm{d}x\frac{F_b(u)}{\rho_b(u)}\mathrm{d}u. \end{aligned}$$

Straightforward calculus gives

$$\begin{aligned} \mathcal{T}'(z) =&\ -\frac{F_b(z)}{\rho_b(z)}\int_z^\infty 2g(x)\rho_b(x)\mathrm{d}x + \frac{1-F_b(z)}{\rho_b(z)}\int_{-\infty}^z 2g(x)\rho_b(x)\mathrm{d}x, \\ \mathcal{T}''(z) =&\ 2g(z) - \int 2g(x)\rho_b(x)\mathrm{d}x - 2b(z)\mathcal{T}'(z). \end{aligned}$$

One can show that $\mathcal{T}$ and its derivatives satisfy an at most linear growth condition, and it is possible to approximate $\mathcal{T}$ by a sequence of functions $\mathcal{T}_n$ in C_c^∞ such that $\lim_{n\to\infty}\|\partial^k\mathcal{T}_n - \partial^k\mathcal{T}\|_{L^4(\mu_b)} = 0$, $k=0,1,2$. In particular, the at most linear growth condition on b implies that $\mathcal{T}_n$ converges to $\mathcal{T}$ in $L^2(\mu_b)$ and $\lim_{n\to\infty}\mathrm{L}_b(\mathcal{T}_n) = g - \mu_b(g)$ in $L^2(\mu_b)$. Since L_b is a closed operator in $L^2(\mu_b)$, we can conclude that $\mathcal{T}\in\mathcal{D}(\mathrm{L}_b)$ and $\mathrm{L}_b(\mathcal{T}) = g - \mu_b(g)$. □

In Section 4.3.2, the Cramér–Rao lower bound for estimating $\Phi_g(\mathbb{P}_b) = \int g(x)\rho_b(x)\mathrm{d}x = \int g\mathrm{d}\mu_b$ is identified as $\|\partial\mathrm{L}_b^{-1}g_c\|^2_{L^2(\mu_b)}$. The following result establishes the corresponding result for pointwise estimation of the invariant density.

Proposition 58. *The Cramér–Rao lower bound for pointwise estimation of $\rho_b(y)$, $y\in\mathbb{R}$ fixed, is defined via* (4.3.17).

Proof. Let $v\colon\mathbb{R}\to\mathbb{R}_+$ be a smooth, symmetric function with support $\mathrm{supp}(v)\subseteq[-1,1]$ and $\int v(z)\mathrm{d}z = 1$, and define $g_\epsilon^y := \epsilon^{-1}v\left((\cdot - y)\epsilon^{-1}\right)$, for any $y\in\mathbb{R}$. Denote by

$$\mathrm{CR}(y,\epsilon)\ :=\ \left\|\frac{\mathrm{d}}{\mathrm{d}z}L_b^{-1}(g_\epsilon^y - \mu_b(g_\epsilon^y))(z)\right\|^2_{L^2(\mu_b)}$$

the Cramér–Rao lower bound for estimation of $\int g_\epsilon^y d\mu_b$. Further, note that

$$\lim_{\epsilon\downarrow 0}\int g_\epsilon^y d\mu_b = \rho_b(y) = b_y^*(\Psi(\mathbb{P}_b)), \text{ for any } y\in\mathbb{R},$$

where the pointwise evaluation $b_y^*:\ell^\infty(\mathbb{R})\longrightarrow\mathbb{R}, f\mapsto f(y)$, is an element of the dual of $\ell^\infty(\mathbb{R})$. We are interested in the Cramér–Rao lower bound for pointwise estimation of $\rho_b(y)=b_y^*(\Psi(\mathbb{P}_b))$, $y\in\mathbb{R}$. This bound is given by the squared $L^2(\mu_b)$-norm of the Riesz representer of b_y^*A'. Since $A'h$ is a continuous function, we have, for any $y\in\mathbb{R}$,

$$b_y^*A'(h) \;=\; \lim_{\epsilon\downarrow 0}\int g_\epsilon^y(x)A'h(x)\mathrm{d}x \;=\; \lim_{\epsilon\downarrow 0} -\langle \partial \mathrm{L}_b^{-1}(g_\epsilon^y-\mu_b(g_\epsilon^y)),\ h\rangle_{\mu_b}$$

due to (4.3.16). We proceed with proving that the limit $\lim_{\epsilon\downarrow 0}\partial L_b^{-1}(g_\epsilon^y-\mu_b(g_\epsilon^y))$ exists in $L^2(\mu_b)$. Fix $y\in\mathbb{R}$. For any $z\in\mathbb{R}$, $z\neq y$, we have for $\epsilon>0$ small enough

$$\begin{aligned}\partial \mathrm{L}_b^{-1}(g_\epsilon^y-\mu_b(g_\epsilon^y))(z) \;&=\; 2\cdot\mathbb{1}\{y\le z\}\frac{1-F_b(z)}{\rho_b(z)}\int_{-\infty}^z g_\epsilon^y(x)\rho_b(x)\mathrm{d}x\\ &\quad -2\cdot\mathbb{1}\{y>z\}\frac{F_b(z)}{\rho_b(z)}\int_z^\infty g_\epsilon^y(x)\rho_b(x)\mathrm{d}x,\end{aligned}$$

due to Lemma 47 and since $\operatorname{supp}(g_\epsilon^y)\subseteq[y-\epsilon,y+\epsilon]$. Moreover, as

$$\max\left\{\sup_{z\in\mathbb{R}}\mathbb{1}\{y\le z\}\frac{1-F_b(z)}{\rho_b(z)},\ \sup_{z\in\mathbb{R}}\mathbb{1}\{y>z\}\frac{F_b(z)}{\rho_b(z)}\right\}<\infty,$$

one obtains

$$\lim_{\epsilon\downarrow 0}\partial \mathrm{L}_b^{-1}(g_\epsilon^y-\mu_b(g_\epsilon^y))(z) \;=\; 2\cdot\frac{\mathbb{1}\{z\ge y\}-F_b(z)}{\rho_b(z)}\rho_b(y) \;=\; 2\rho_b(y)h(z,y)$$

a.e. and in $L^2(\mu_b)$. We conclude that $b_y^*A'(h) = -\langle 2\rho_b(y)h(\cdot,y),h\rangle_{\mu_b}$ such that the assertion follows. □

We are now in a position to prove asymptotic efficiency as defined in Definition 49 for the kernel invariant density estimator with the 'universal' bandwidth choice $t^{-1/2}$.

Proof of Theorem 50. Rewriting the covariance $\mathbb{E}[\mathbb{H}(x),\mathbb{H}(y)]=H(x,y)$ as in (4.6.30), one immediately sees that the law of $\mathbb{H}$ corresponds to the optimal distribution of the convolution theorem due to (4.3.18). It remains to prove regularity of the estimator.

Regularity of the estimator $\rho_{t,K}(t^{-1/2})(y)$ Fix $y\in\mathbb{R}$. As we have already seen in the proof of Proposition 45,

$$\sqrt{t}\left(\rho_{t,K}(t^{-1/2})-t^{-1}L_t^\bullet(X)\right)(y) \;=\; o_{\mathbb{P}_b}(1),$$

(see (4.6.29)). We proceed by exploiting the martingale structure obtained from Proposition 1.11 in Kutoyants (2004),

$$\begin{aligned}
&\sqrt{t}\left(t^{-1}L_t^y(X)-\rho_b(y)\right)\\
&= \frac{2\rho_b(y)}{\sqrt{t}}\int_{X_0}^{X_t}\frac{\mathbb{1}\{v>y\}-F_b(v)}{\rho_b(v)}\mathrm{d}v-\frac{2\rho_b(y)}{\sqrt{t}}\int_0^t\frac{\mathbb{1}\{X_u>y\}-F_b(X_u)}{\rho_b(X_u)}\mathrm{d}W_u\\
&= \frac{R(X_t,y)}{\sqrt{t}}-\frac{R(X_0,y)}{\sqrt{t}}-\frac{2\rho_b(y)}{\sqrt{t}}\int_0^t\frac{\mathbb{1}\{X_u>y\}-F_b(X_u)}{\rho_b(X_u)}\mathrm{d}W_u,
\end{aligned}$$

with $R(x,y):=2\rho_b(y)\int_0^x\frac{\mathbb{1}\{v>y\}-F_b(v)}{\rho_b(v)}\mathrm{d}v$. The process $(X_s)_{s\geq 0}$ is stationary under $\mathbb{P}_b$, and therefore

$$\frac{R(X_t,y)-R(X_0,y)}{\sqrt{t}}=o_{\mathbb{P}_b}(1).$$

Let $h\in G$, and fix $a,c\in\mathbb{R}$. Then,

$$\begin{aligned}
&(a,c)\left(\sqrt{t}\left(\rho_{t,K}(t^{-1/2})(y)-\rho_b(y)\right),\log\left(\frac{\mathrm{d}\mathbb{P}_{t,h}}{\mathrm{d}\mathbb{P}_b}\right)\right)^t\\
&= o_{\mathbb{P}_b}(1)-a\cdot 2\frac{\rho_b(y)}{\sqrt{t}}\int_0^t\frac{\mathbb{1}\{X_u>y\}-F_b(X_u)}{\rho_b(X_u)}\mathrm{d}W_u\\
&\qquad + c\cdot\left(\frac{1}{\sqrt{t}}\int_0^t h(X_s)\mathrm{d}W_s-\frac{1}{2}\int h^2(y)\rho_b(y)\mathrm{d}y\right)\\
&= o_{\mathbb{P}_b}(1)+\frac{1}{\sqrt{t}}\int_0^t\left(-2a\rho_b(y)k(X_u,y)+ch(X_u)\right)\mathrm{d}W_u-\frac{1}{2}c\int h^2(y)\rho_b(y)\mathrm{d}y\\
&\overset{\mathbb{P}_b}{\Longrightarrow}\mathcal{N}\left(-\frac{c}{2}\int h^2(y)\rho_b(y)\mathrm{d}y,\delta^2\right),
\end{aligned}$$

with $k(x,y):=\left(\mathbb{1}\{x>y\}-F_b(x)\right)\rho_b^{-1}(x)$ and

$$\begin{aligned}
\delta^2 &= \mathbb{E}_b\left[(ch(X_0)-2a\rho_b(y)k(X_0,y))^2\right]\\
&= \mathbb{E}_b\left[4a^2\rho_b^2(y)k^2(X_0,y)+c^2h^2(X_0)-4ac\rho_b(y)k(X_0,y)h(X_0)\right]\\
&= (a,c)\begin{pmatrix}\mathbb{E}_b\left[4\rho_b^2(y)k^2(X_0,y)\right] & -\mathbb{E}_b\left[2\rho_b(y)k(X_0,y)h(X_0)\right]\\ -\mathbb{E}_b\left[2\rho_b(y)k(X_0,y)h(X_0)\right] & \mathbb{E}_b\left[h^2(X_0)\right]\end{pmatrix}\begin{pmatrix}a\\ c\end{pmatrix}.
\end{aligned}$$

The Cramér–Wold device then implies that

$$\begin{aligned}
&\left(\sqrt{t}\left(\rho_{t,K}(t^{-1/2})(y)-\rho_b(y)\right),\log\left(\frac{\mathrm{d}\mathbb{P}_{t,h}}{\mathrm{d}\mathbb{P}_b}\right)\right)\\
&\qquad\overset{\mathbb{P}_b}{\Longrightarrow}\mathcal{N}\left(\begin{pmatrix}0\\ -\frac{1}{2}\mathbb{E}_b(h^2(X_0))\end{pmatrix},\begin{pmatrix}\mathbb{E}_b(4\rho_b^2(y)k^2(X_0,y)) & -\tau\\ -\tau & \mathbb{E}_b(h^2(X_0))\end{pmatrix}\right),
\end{aligned}$$

where $\tau:=2\mathbb{E}_b\left[\rho_b(y)k(X_0,y)h(X_0)\right]$. In turn, Le Cam's Third Lemma yields

$$\sqrt{t}\left(\rho_{t,K}(t^{-1/2})(y)-\rho_b(y)\right)\overset{\mathbb{P}_{t,h}}{\Longrightarrow}\mathcal{N}\left(-\tau,4\mathbb{E}_b\left[\rho_b^2(y)k^2(X_0,y)\right]\right).$$

Furthermore,

$$\begin{aligned}
&\sqrt{t}\left(\rho_{t,K}(t^{-1/2})(y) - \rho_{b+t^{-1/2}h}(y)\right) \\
&\quad= \sqrt{t}\left(\rho_{t,K}(t^{-1/2})(y) - \rho_b(y)\right) - \sqrt{t}\left(b_y^*(\Psi(\mathbb{P}_{t,h})) - b_y^*(\Psi(\mathbb{P}_b))\right) \\
&\quad= \sqrt{t}\left(\rho_{t,K}(t^{-1/2})(y) - \rho_b(y)\right) + \langle 2\rho_b(y)h(\cdot,y), h\rangle_{\mu_b} + o(1) \\
&\quad= \sqrt{t}\left(\rho_{t,K}(t^{-1/2})(y) - \rho_b(y)\right) + \tau + o(1) \\
&\quad\overset{\mathbb{P}_{t,h}}{\Longrightarrow} \mathcal{N}\left(0,\ 4\rho_b^2(y)\int \left(\mathbb{1}\{z \geq y\} - F_b(z)\right)^2 \rho_b^{-1}(z)\mathrm{d}z\right) \ = \ \mathcal{N}\left(0,\ \mathbb{E}\left[\mathbb{H}^2(y)\right]\right).
\end{aligned}$$

We conclude that $\rho_{t,K}(t^{-1/2})(y)$ is a regular and consequently efficient estimator of $\rho_b(y)$ for any $y \in \mathbb{R}$.

Regularity in $\ell^\infty(\mathbb{R})$ In an analogous way, it can be shown that all finite-dimensional marginals of

$$\sqrt{t}\left(\rho_{t,K}(t^{-1/2}) - \rho_{b+t^{-1/2}h}\right) \tag{4.6.31}$$

weakly converge to those of $\mathbb{H}$ under $\mathbb{P}_{t,h}$, for any $h \in G$. Therefore, the estimator $\rho_{t,K}(t^{-1/2})$ is also regular in $\ell^\infty(\mathbb{R})$ if we can show that the process in (4.6.31) is asymptotically tight. As we have already seen that the limiting distribution is optimal, this then gives efficiency of $\rho_{t,K}(t^{-1/2})$ in $\ell^\infty(\mathbb{R})$. We proceed as in (Kosorok, 2008, Theorem 11.14). Fix $\epsilon > 0$. Since $\mathbb{P}_{t,h}$ and $\mathbb{P}_b$ are contiguous, $\mathrm{d}\mathbb{P}_{t,h}/\mathrm{d}\mathbb{P}_b$ is stochastically bounded wrt to both $\mathbb{P}_b$ and $\mathbb{P}_{t,h}$. Hence, we find a constant M such that

$$\limsup_{t\to\infty} \mathbb{P}_{t,h}\left(\frac{\mathrm{d}\mathbb{P}_{t,h}}{\mathrm{d}\mathbb{P}_b} > M\right) \ \leq \ \frac{\epsilon}{2}.$$

Furthermore, since $\sqrt{t}(\rho_{t,K}(t^{-1/2}) - \rho_b)$ is asymptotically tight wrt $\mathbb{P}_b$, there exists a compact set $K \subseteq \ell^\infty(\mathbb{R})$ such that, for any $\delta > 0$,

$$\limsup_{t\to\infty} \mathbb{P}_b\left(\sqrt{t}\left(\rho_{t,K}(t^{-1/2}) - \rho_b\right) \in \left(\ell^\infty(\mathbb{R}) \setminus K^\delta\right)^*\right) \ \leq \ \frac{\epsilon}{2M},$$

K^δ denoting the δ-enlargement of K. The superscript * here stands for the minimal measurable cover wrt to both $\mathbb{P}_b$ and $\mathbb{P}_{t,h}$. From these choices, we deduce, for any $\delta > 0$,

$$\begin{aligned}
&\limsup_{t\to\infty} \mathbb{P}_{t,h}\left(\sqrt{t}\left(\rho_{t,K}(t^{-1/2}) - \rho_b\right) \in \left(\ell^\infty(\mathbb{R}) \setminus K^\delta\right)^*\right) \\
&\quad= \ \limsup_{t\to\infty} \int \mathbb{1}\left\{\sqrt{t}\left(\rho_{t,K}(t^{-1/2}) - \rho_b\right) \in \left(\ell^\infty(\mathbb{R}) \setminus K^\delta\right)^*\right\} \mathrm{d}\mathbb{P}_{t,h} \\
&\quad= \ \limsup_{t\to\infty} \int \mathbb{1}\left\{\sqrt{t}\left(\rho_{t,K}(t^{-1/2}) - \rho_b\right) \in \left(\ell^\infty(\mathbb{R}) \setminus K^\delta\right)^*\right\} \frac{\mathrm{d}\mathbb{P}_{t,h}}{\mathrm{d}\mathbb{P}_b}\mathrm{d}\mathbb{P}_b \\
&\quad\leq \ \limsup_{t\to\infty} M \int \mathbb{1}\left\{\sqrt{t}\left(\rho_{t,K}(t^{-1/2}) - \rho_b\right) \in \left(\ell^\infty(\mathbb{R}) \setminus K^\delta\right)^*\right\} \mathrm{d}\mathbb{P}_b \\
&\qquad\qquad + \limsup_{t\to\infty} \mathbb{P}_{t,h}\left(\frac{\mathrm{d}\mathbb{P}_{t,h}}{\mathrm{d}\mathbb{P}_b} > M\right) \ \leq \ \epsilon.
\end{aligned}$$

Due to the differentiability property (4.3.15), we conclude that (4.6.31) is asymptotically tight wrt $\mathbb{P}_{t,h}$, as well. We have thus shown that $\rho_{t,K}(t^{-1/2})$ is an efficient estimator in $\ell^\infty(\mathbb{R})$. □

4.7 Proofs for Section 4.4 and Section 4.5

4.7.1 Preliminaries

We start with the proof of Proposition 41 on the concentration of the estimator $\overline{\rho}_{t,K}(h)$ of $\rho_b'/2$.

Proof of Proposition 41. We apply Theorem 30 in Chapter 3 to the class

$$\mathcal{F} := \left\{K\left(\frac{x-\cdot}{h}\right) : x \in \mathbb{Q}\right\}. \tag{4.7.32}$$

For doing so, note that $\sup_{f\in\mathcal{F}}\|f\|_\infty \leq \|K\|_\infty$, and, for $\boldsymbol{\lambda}$ denoting the Lebesgue measure,

$$\left\|K\left(\frac{x-\cdot}{h}\right)\right\|^2_{L^2(\boldsymbol{\lambda})} \;=\; \int K^2\left(\frac{x-y}{h}\right)\mathrm{d}y \;=\; h\int K^2(z)\mathrm{d}z \;\leq\; h\|K\|^2_{L^2(\boldsymbol{\lambda})}$$

and $\sup_{f\in\mathcal{F}}\boldsymbol{\lambda}(\mathrm{supp}(f)) \leq h$. Due to the Lipschitz continuity of K, Lemma 35 in Chapter 3 yields constants $\mathbb{A} > 0$, $v \geq 2$ (only depending on K) such that, for any probability measure $\mathbb{Q}$ on $\mathbb{R}$ and any $0 < \epsilon < 1$, $N(\epsilon, \mathcal{F}, \|\cdot\|_{L^2(\mathbb{Q})}) \leq (\mathbb{A}/\epsilon)^v$. Here and throughout the sequel, given some semi-metric d, $N(u, \mathcal{F}, d)$, $u > 0$, denotes the covering number of $\mathcal{F}$ wrt d, i.e., the smallest number of balls of radius u in $(\mathcal{F}, d)$ needed to cover $\mathcal{F}$. Since the assumption on the covering numbers of $\mathcal{F}$ in Theorem 30 in Chapter 3 is fulfilled, Theorem 30 can be applied to $\mathcal{F}$ with $\mathcal{S} := h\ \max\{\|K\|^2_{L^2(\boldsymbol{\lambda})}, 1\}$ and $\mathbb{V} := \sqrt{h}\|K\|_{L^2(\boldsymbol{\lambda})}$. In particular, there exist positive constants $\widetilde{\mathbb{L}}$, Λ, $\widetilde{\mathbb{L}}_0$ and $\mathbb{L}$ such that

$$\begin{aligned}
&\sup_{b\in\Sigma}\left(\mathbb{E}_b\left[\|\overline{\rho}_{t,K}(h) - \mathbb{E}_b\left[\overline{\rho}_{t,K}(h)\right]\|_\infty^p\right]\right)^{\frac{1}{p}}\\
&\quad\leq\ \widetilde{\mathbb{L}}\Bigg\{\frac{1}{\sqrt{t}}\Bigg\{\left(\log\left(\sqrt{\frac{h+p\Lambda t}{h}}\right)\right)^{3/2} + \left(\log\left(\sqrt{\frac{h+p\Lambda t}{h}}\right)\right)^{1/2} + p^{3/2}\Bigg\}\\
&\qquad\quad + \frac{p}{th} + \frac{1}{h}\exp\left(-\widetilde{\mathbb{L}}_0 t\right) + \frac{1}{\sqrt{th}}\left(\log\left(\sqrt{\frac{h+p\Lambda t}{h}}\right)\right)^{1/2}\\
&\qquad\quad + \frac{1}{t^{3/4}\sqrt{h}}\left(1+\log\left(\sqrt{\frac{h+p\Lambda t}{h}}\right)\right) + \frac{1}{\sqrt{th}}\left\{\sqrt{p} + \frac{p}{t^{1/4}}\right\}\Bigg\}\\
&\quad\leq \phi_{t,h}(p),
\end{aligned}$$

and (4.2.9) immediately follows. □

We continue with stating and proving a number of auxiliary results required for the investigation of the proposed sup-norm adaptive drift estimation procedure. Recall the definition of $\Sigma(\beta,\mathcal{L})$ in (4.2.7).

Lemma 59. *For $b \in \Sigma(\beta,\mathcal{L})$, $\beta,\mathcal{L} > 0$, and for $\overline{\rho}_{t,K}$ defined according to* (4.1.6) *with some kernel function $K\colon \mathbb{R} \to \mathbb{R}$ satisfying* (4.2.8)*, it holds, for any $h \in \mathcal{H}$,*

$$\sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{E}_b\left[\|\overline{\rho}_{t,K}(h) - \mathbb{E}_b\left[\overline{\rho}_{t,K}(h)\right]\|_\infty^2\right] \leq \mathcal{K}^2\overline{\sigma}^2(h,t), \tag{4.7.33}$$

$$\sup_{b\in\Sigma(\beta,\mathcal{L})} \|\mathbb{E}_b\left[\overline{\rho}_{t,K}(h)\right] - \rho_b b\|_\infty \leq B(h), \tag{4.7.34}$$

where $\mathcal{K}$ denotes some positive constant, $\overline{\sigma}^2(\cdot,\cdot)$ is defined according to (4.4.19) *and*

$$B(h) := h^\beta \frac{\mathcal{L}}{2[\beta]!} \int |K(v)||v|^\beta \mathrm{d}v.$$

Proof. Assertion (4.7.33) follows immediately from Proposition 41. For the bias of $\overline{\rho}_{t,K}$, classical Taylor arguments imply that (see Giné and Nickl (2009))

$$\sup_{b\in\Sigma(\beta,\mathcal{L})} \|\mathbb{E}_b\left[\overline{\rho}_{t,K}(h)\right] - \rho_b b\|_\infty = \frac{1}{2}\sup_{x\in\mathbb{R}}\left|\int_{\mathbb{R}} K_h(x-y)(\rho_b'(y) - \rho_b'(x))\mathrm{d}y\right| \leq B(h).$$

□

The next two auxiliary results give conditions which allow to translate upper and lower bounds on the sup-norm risk of estimators of ρ_b' into corresponding bounds on the weighted risk of drift estimators.

Lemma 60 (Weighted upper bounds for drift estimation)**.** *Given $b \in \Sigma(\beta,\mathcal{L})$, consider estimators ρ_t and $\overline{\rho}_t$ of the invariant density ρ_b and $\rho_b'/2$, respectively, fulfilling the following conditions:*

(E1) $\exists C_1 > 0$ *such that, for any $p \geq 1$,*

$$\sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{E}_b\left[\|\rho_t - \rho_b\|_\infty^p\right] \leq C_1^p t^{-\frac{p}{2}}\left(1 + (\log t)^{\frac{p}{2}} + p^{\frac{p}{2}} + p^p t^{-\frac{p}{2}}\right),$$

and $\rho_t(x) \geq 0$, for any $x \in \mathbb{R}$;

(E2) $\exists C_2 > 0$ *such that* $\sup_{b\in\Sigma(\beta,\mathcal{L})}\mathbb{E}_b\left[\|\overline{\rho}_t\|_\infty^2\right] \leq C_2 t^2$*;*

(E3) $\exists C_3 > 0$ *such that* $\sup_{b\in\Sigma(\beta,\mathcal{L})}\mathbb{E}_b\left[\|\overline{\rho}_t - \rho_b'/2\|_\infty\right] \leq C_3(\log t/t)^{\beta/(2\beta+1)}$.

Then, the drift estimator

$$\widehat{b}_t(x) := \frac{\overline{\rho}_t(x)}{\rho_t^*(x)}, \qquad \text{with } \rho_t^*(x) := \rho_t(x) + \sqrt{\frac{\log t}{t}}\exp\left(\sqrt{\log t}\right),$$

satisfies

$$\sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{E}_b\left[\|(\widehat{b}_t - b)\ \rho_b^2\|_\infty\right] = O\left((\log t/t)^{\beta/(2\beta+1)}\right).$$

Proof. For ease of notation, we refrain in the sequel from carrying the $\sup_{b\in\Sigma(\beta,\mathcal{L})}$ along. All arguments hold for the supremum because constants in the upper bounds do not depend on the specific choice of $b \in \Sigma(\beta,\mathcal{L})$. Introduce the set

$$B_t \;:=\; \left\{\sqrt{t}\|\rho_t-\rho_b\|_\infty \le \sqrt{\log t}\exp(\sqrt{\log t})\right\}, \quad t > \mathrm{e}.$$

For any $p \ge 1$, Markov's inequality and condition (E1) imply that

$$\begin{aligned}
\mathbb{P}_b(B_t^{\mathrm{c}}) \le\; & \mathbb{E}_b\left[\|\rho_t-\rho_b\|_\infty^p\right]\cdot\left(\frac{t}{\log t}\right)^{\frac{p}{2}}\exp(-p\sqrt{\log t})\\
\le\; & C_1^p t^{-\frac{p}{2}}\left(1+(\log t)^{\frac{p}{2}}+p^{\frac{p}{2}}+\left(p/\sqrt{t}\right)^p\right)\left(\frac{t}{\log t}\right)^{\frac{p}{2}}\exp\left(-p\sqrt{\log t}\right)\\
\le\; & C_1^p\left((\log t)^{-\frac{p}{2}}+1+\left(\frac{p}{\log t}\right)^{\frac{p}{2}}+\left(\frac{p}{\sqrt{t\log t}}\right)^p\right)\exp\left(-p\sqrt{\log t}\right).
\end{aligned}$$

Specifying $p = 8\sqrt{\log t}$, one obtains that, for some positive constant C,

$$\mathbb{P}_b(B_t^{\mathrm{c}}) \;\le\; C^{8\sqrt{\log t}}\exp(-8\log t).$$

Thus, on the event B_t^{c},

$$\begin{aligned}
&\mathbb{E}_b\left[\|(\widehat{b}_t-b)\;\rho_b^2\|_\infty\,\mathbb{1}_{B_t^{\mathrm{c}}}\right]\\
&\quad\le\; \left(\mathbb{E}_b\left[\|\widehat{b}_t\rho_b^2\|_\infty^2\right]\;\mathbb{P}_b(B_t^{\mathrm{c}})\right)^{1/2}+\mathbb{E}_b\left[\|b\rho_b^2\|_\infty\cdot\mathbb{1}_{B_t^{\mathrm{c}}}\right]\\
&\quad\le\; \sup_{x\in\mathbb{R}}|\rho_b^2(x)|\left(\mathbb{E}_b\left[\|\overline{\rho}_t\|_\infty^2\right]\;\frac{t}{\log t}\exp(-2\sqrt{\log t})\;\mathbb{P}_b(B_t^{\mathrm{c}})\right)^{1/2}+\mathbb{E}_b\left[\|b\rho_b^2\|_\infty\cdot\mathbb{1}_{B_t^{\mathrm{c}}}\right]\\
&\quad\le\; \mathcal{L}^2\left(\mathbb{E}_b\left[\|\overline{\rho}_t\|_\infty^2\right]\;\frac{t}{\log t}\exp(-2\sqrt{\log t})C^{8\sqrt{\log t}}\exp(-8\log t)\right)^{1/2}+\;\frac{1}{2}\mathcal{L}^2\mathbb{P}_b(B_t^{\mathrm{c}})\\
&\quad=\; O\left(\sqrt{C^{8\sqrt{\log t}}\frac{t^3}{\log t}\exp(-8\log t)}+C^{8\sqrt{\log t}}t^{-8}\right)\;=\;O(t^{-1}).
\end{aligned}$$

On the other hand, $\mathbb{E}_b\left[\|(\widehat{b}_t-b)\;\rho_b^2\|_\infty\,\mathbb{1}_{B_t}\right] \le A_1+A_2$, for

$$A_1 \;:=\; \mathbb{E}_b\left[\left\|\left(\widehat{b}_t-\frac{b\rho_b}{\rho_t^*}\right)\;\rho_b^2\right\|_\infty\cdot\mathbb{1}_{B_t}\right], \qquad A_2 \;:=\; \mathbb{E}_b\left[\left\|\left(\frac{b\rho_b}{\rho_t^*}-b\right)\;\rho_b^2\right\|_\infty\cdot\mathbb{1}_{B_t}\right].$$

Since $\rho_b/\rho_t^* \le 1$ on the event B_t, (E1) and (E3) imply that

$$\begin{aligned}
A_1 \le\; & \sup_{x\in\mathbb{R}}|\rho_b(x)|\;\mathbb{E}_b\left[\|\overline{\rho}_t-\rho_b'/2\|_\infty\right]\;=\;O\left(\left(\frac{\log t}{t}\right)^{\beta/(2\beta+1)}\right),\\
A_2 =\; & \mathbb{E}_b\left[\left\|\frac{\rho_b'\rho_b}{2\rho_t^*}(\rho_b-\rho_t^*)\cdot\mathbb{1}_{B_t}\right\|_\infty\right]\;\le\;\frac{\mathcal{L}}{2}\left\{\mathbb{E}_b\left[\|\rho_b-\rho_t\|_\infty\right]+\left(\frac{\log t}{t}\right)^{1/2}\exp\left(\sqrt{\log t}\right)\right\}.
\end{aligned}$$

Summing up,

$$\mathbb{E}_b\left[\|(\widehat{b}_t-b)\ \rho_b^2\|_\infty\right]\ \leq\ O\left(\frac{1}{t}\right)+O\left(\left(\frac{\log t}{t}\right)^{\frac{\beta}{2\beta+1}}\right)+O\left(\sqrt{\frac{\log t}{t}}\right)+O\left(\sqrt{\frac{\log t}{t}}\mathrm{e}^{\sqrt{\log t}}\right),$$

and the assertion follows. □

Lemma 61 (Weighted lower bounds for drift estimation)**.** *Assume that, for some ψ_t fulfilling $\sqrt{\log t/t}=o(\psi_t)$, it holds*

$$\liminf_{t\to\infty}\inf_{\widetilde{\partial\rho_t^2}}\sup_{b\in\Sigma(\beta,\mathcal{L})}\mathbb{E}_b\left[\psi_t^{-1}\|\widetilde{\partial\rho_t^2}-(\rho_b^2)'\|_\infty\right]\ >\ 0. \tag{4.7.35}$$

Then,

$$\liminf_{t\to\infty}\inf_{\widetilde{b}}\sup_{b\in\Sigma(\beta,\mathcal{L})}\mathbb{E}_b\left[\psi_t^{-1}\left\|(\widetilde{b}-b)\rho_b^2\right\|_\infty\right]\ >\ 0.$$

The infimum in the preceding inequalities is taken over all estimators $\widetilde{\partial\rho_t^2}$ and $\widetilde{b}$ of $(\rho_b^2)'$ and b, respectively.

Proof. Given any estimator $\widetilde{b}$ of the drift coefficient b, define

$$\bar{b}(x):=\begin{cases}\widetilde{b}(x), & \text{if } |\widetilde{b}(x)|\leq\mathcal{C}(1+|x|)\\ \operatorname{sgn}(b(x))\mathcal{C}(1+|x|), & \text{otherwise}\end{cases},\qquad x\in\mathbb{R}.$$

For any $b\in\Sigma(\beta,\mathcal{L})$, it holds $|b(x)|\leq\mathcal{C}(1+|x|)$ for all $x\in\mathbb{R}$. Consequently,

$$\mathbb{E}_b\left[\left\|(\bar{b}-b)\ \rho_b^2\right\|_\infty\right]\ \leq\ \mathbb{E}_b\left[\left\|(\widetilde{b}-b)\ \rho_b^2\right\|_\infty\right].$$

It thus suffices to consider the infimum over all estimators $\widetilde{b}$ satisfying

$$|\widetilde{b}(x)|\leq\mathcal{C}(1+|x|),\quad x\in\mathbb{R}. \tag{4.7.36}$$

In view of the decomposition

$$\widetilde{b}\rho_{t,K}(t^{-\frac{1}{2}})\rho_b-\frac{1}{2}\rho_b'\rho_b\ =\ \frac{1}{2}\rho_b\left(2\widetilde{b}\rho_b-\rho_b'\right)+\widetilde{b}\rho_b\left(\rho_{t,K}(t^{-\frac{1}{2}})-\rho_b\right),$$

it holds

$$\mathbb{E}_b\left[\left\|(\widetilde{b}-b)\ \rho_b^2\right\|_\infty\right]\ =\ \mathbb{E}_b\left[\left\|\frac{1}{2}\rho_b\left(2\widetilde{b}\rho_b-\rho_b'\right)\right\|_\infty\right]\ \geq\ (\mathbf{I})-(\mathbf{II}),$$

with

$$(\mathbf{I}):=\frac{1}{4}\mathbb{E}_b\left[\left\|4\widetilde{b}\rho_{t,K}(t^{-\frac{1}{2}})\rho_b-2\rho_b'\rho_b\right\|_\infty\right],\qquad(\mathbf{II}):=\mathbb{E}_b\left[\left\|\widetilde{b}\rho_b\left(\rho_{t,K}(t^{-\frac{1}{2}})-\rho_b\right)\right\|_\infty\right].$$

Due to (4.7.36), $\sup_{b\in\Sigma(\beta,\mathcal{L})}|\widetilde{b}\rho_b|$ is bounded. Moreover, we can infer from Proposition 44 that there exists a positive constant C_1 such that, for all t sufficiently large, $(\mathbf{II})\ \leq$

$C_1\sqrt{\log t/t}$. Consequently, noting that $4\widetilde{b}\rho_{t,K}(t^{-\frac{1}{2}})\rho_b$ can be viewed as an estimator of $2\rho_b'\rho_b = (\rho_b^2)'$,

$$\begin{aligned}
\liminf_{t\to\infty}\inf_{\widetilde{b}}\sup_{b\in\Sigma(\beta,\mathcal{L})}\mathbb{E}_b\left[\psi_t^{-1}\left\|(\widetilde{b}-b)\ \rho_b^2\right\|_\infty\right] &\geq \liminf_{t\to\infty}\inf_{\widetilde{b}}\sup_{b\in\Sigma(\beta,\mathcal{L})}\psi_t^{-1}((\mathbf{I})-(\mathbf{II}))\\
&\geq \liminf_{t\to\infty}\inf_{\widetilde{b}}\sup_{b\in\Sigma(\beta,\mathcal{L})}\left(\psi_t^{-1}(\mathbf{I})-C_1\psi_t^{-1}\sqrt{\frac{\log t}{t}}\right)\\
&> 0.
\end{aligned}$$

□

Proposition 41 presented in Section 4.2 gives one first result on the concentration behaviour of the estimator $\overline{\rho}_{t,K}(h)$ of $\rho_b'/2$. It follows from a straightforward application of the developments in Chapter 3. Note that, for any bandwidth $h = h_t \in (t^{-1}(\log t)^2,\ (\log t)^{-3})$ and any $u = u_t \in [1, \alpha\log t]$, $\alpha > 0$, the function $\phi_{t,h}(u)$ introduced in (4.2.10) fulfills

$$\phi_{t,h}(u) \ = \ Z_1\sqrt{\frac{\log(ut/h)}{th}} + Z_2\sqrt{\frac{u}{th}} + o\left(\sqrt{\frac{\log(ut/h)}{th}}\right).$$

For the construction of an adaptive estimation procedure which yields rate-optimal drift estimators, we need to specify the constants Z_1 and Z_2. The subsequent Lemma provides the corresponding result. Its proof relies on a modification of the proof of Theorem 30 in Chapter 3.

Lemma 62 (Tail bounds with explicit constants)**.** *Grant the assumptions of Proposition 41, and define the estimator $\overline{\rho}_{t,K}$ according to* (4.1.6)*. Then, for any $h = h_t \in (t^{-1}(\log t)^2, (\log t)^{-3})$, $1 \leq u = u_t \leq \alpha\log t$, for some $\alpha > 0$, and t sufficiently large, it holds*

$$\mathbb{P}_b\left(\sup_{x\in\mathbb{R}}|\overline{\rho}_{t,K}(h_t)(x) - \mathbb{E}_b\left[\overline{\rho}_{t,K}(h_t)(x)\right]| > \mathrm{e}\overline{\psi}_{t,h_t}(u_t)\right) \ \leq \ \mathrm{e}^{-u_t}, \tag{4.7.37}$$

with

$$\overline{\psi}_{t,h}(u) := \sqrt{\|\rho_b\|_\infty}\left\{2\overline{\eta}_1\sqrt{\frac{\log(ut/h)}{th}} + \overline{\eta}_2\sqrt{\frac{u}{th}}\right\},$$

for $\overline{\eta}_1 = 24\widetilde{C}_2\|K\|_{L^2(\boldsymbol{\lambda})}\,\mathrm{c}\,\mathrm{e}\sqrt{v}$ and $\overline{\eta}_2 = 12\,\mathrm{c}\,\|K\|_{L^2(\boldsymbol{\lambda})}$.

Proof. Let us first prove that there exist constants $\mathbb{L}_1, \widetilde{\mathbb{L}}_0$ such that, for any $u, p \geq 1$, $h \in (0,1)$, $t \geq 1$,

$$\begin{aligned}
\left(\mathbb{E}_b\left[\|\overline{\rho}_{t,K}(h) - \mathbb{E}_b\left[\overline{\rho}_{t,K}(h)\right]\|_\infty^p\right]\right)^{\frac{1}{p}} &\leq \ \widetilde{\phi}_{t,h}(p),\\
\mathbb{P}_b\left(\sup_{x\in\mathbb{R}}|\overline{\rho}_{t,K}(h)(x) - \mathbb{E}_b\overline{\rho}_{t,K}(h)(x)| > \mathrm{e}\widetilde{\phi}_{t,h}(u)\right) &\leq \ \mathrm{e}^{-u},
\end{aligned} \tag{4.7.38}$$

for

$$\begin{aligned}\widetilde{\phi}_h(u) \;:=\; & \mathbb{L}_1\Bigg\{\frac{1}{\sqrt{t}}\Big\{\left(\log\left(\frac{ut}{h}\right)\right)^{3/2}+\left(\log\left(\frac{ut}{h}\right)\right)^{1/2}+u^{3/2}\Big\}+\frac{u}{th}+\frac{1}{h}\exp\left(-\widetilde{\mathbb{L}}_0 t\right)\\ & +\frac{1}{t^{3/4}\sqrt{h}}\log\left(\frac{ut}{h}\right)+\frac{1}{\sqrt{th}}\left\{\left(\frac{\log t}{t}\right)^{1/4}\sqrt{u}+\frac{u}{t^{1/4}}\right\}\Bigg\}\\ & +\;\overline{\eta}_1\sqrt{\|\rho_b\|_\infty}\frac{1}{\sqrt{th}}\left(\log\left(\frac{ut}{h}\right)\right)^{1/2}+\overline{\eta}_2\sqrt{\|\rho_b\|_\infty}\left(\frac{u}{th}\right)^{1/2}.\end{aligned}$$

This statement parallels assertion (4.2.9) from Proposition 41. It remains however to identify the constants preceding the terms $(th)^{-1/2}\left(\log(ut/h)\right)^{1/2}$ and $(u/th)^{1/2}$ as $\sqrt{\|\rho_b\|_\infty}\overline{\eta}_1$ and $\sqrt{\|\rho_b\|_\infty}\overline{\eta}_2$, respectively. This requires to look into the details of the proof of Theorem 30 in Chapter 3. In our situation, $\mathcal{F}$ is defined as in (4.7.32). First, note that the term $\frac{1}{\sqrt{th}}\left(\log\left(\frac{ut}{h}\right)\right)^{1/2}$ comes from the analysis of the martingale $(th)^{-1}\int_0^t K\left(\frac{y-X_s}{h}\right)\mathrm{d}W_s$. We repeat the arguments from the proof of Theorem 30 in Chapter 3, here, to discover the required constant. The Burkholder–Davis–Gundy inequality as stated in Proposition 4.2 in Barlow and Yor (1982) and the occupation times formula yield, for any $p \geq 2$, $f \in \mathcal{F}$,

$$\begin{aligned}&\left(\mathbb{E}_b\left[\left|\frac{1}{\sqrt{t}}\int_0^t f(X_s)\mathrm{d}W_s\right|^p\right]\right)^{\frac{1}{p}}\\ &\quad\leq\; \mathrm{c}\sqrt{p}\left(\mathbb{E}_b\left[\left(\frac{1}{t}\int_0^t f^2(X_s)\mathrm{d}s\right)^{\frac{p}{2}}\right]\right)^{\frac{1}{p}}\\ &\quad\leq\; \mathrm{c}\sqrt{p}\|f\|_{L^2(\boldsymbol{\lambda})}\left(\mathbb{E}_b\left[\left(\|t^{-1}L_t^\bullet(X)\|_\infty\right)^{\frac{p}{2}}\right]\right)^{\frac{1}{p}}\\ &\quad\leq\; \mathrm{c}\sqrt{p}\|f\|_{L^2(\boldsymbol{\lambda})}\left\{\left(\mathbb{E}_b\left[\left(\|t^{-1}L_t^\bullet(X)-\rho_b\|_\infty\right)^{\frac{p}{2}}\right]\right)^{\frac{1}{p}}+\sqrt{\|\rho_b\|_\infty}\right\},\end{aligned}$$

where $\mathrm{c} := \sqrt{2}\max\{1,\overline{\mathrm{c}}\}$, for $\overline{\mathrm{c}}$ denoting a universal constant from the BDG inequality. Exploiting the result on centred diffusion local time from Lemma 43, one can deduce that there exists another constant $\mathbb{L}_2 > 0$ such that, for $\mathbb{V} = \sqrt{h}\|K\|_{L^2(\boldsymbol{\lambda})}$,

$$\begin{aligned}&2\sup_{f\in\mathcal{F}}\left(\mathbb{E}_b\left[\left|\frac{1}{\sqrt{t}}\int_0^t f(X_s)\mathrm{d}W_s\right|^p\right]\right)^{\frac{1}{p}}\\ &\quad\leq\; 2\,\mathrm{c}\sqrt{p}\mathbb{V}\left(\sqrt{\|\rho_b\|_\infty}+\mathbb{L}_2\left(\left(\frac{\log t}{t}\right)^{1/4}+\left(\frac{p}{t}\right)^{1/4}+\sqrt{\frac{p}{t}}\right)\right)\\ &\quad\leq\; 2\,\mathrm{c}\,\mathbb{V}\left(\sqrt{\|\rho_b\|_\infty}+\mathbb{L}_2\left(\frac{\log t}{t}\right)^{1/4}\right)\sqrt{p}+4t^{-1/4}\,\mathrm{c}\,\mathbb{V}\mathbb{L}_2 p.\end{aligned}$$

Using similar arguments, the latter is verified for $1 \leq p < 2$. Furthermore, for any $f, g \in \mathcal{F}$, it can be shown analogously that

$$\left(\mathbb{E}_b\left[\left|\frac{1}{\sqrt{t}}\int_0^t (f-g)(X_s)\mathrm{d}W_s\right|^p\right]\right)^{\frac{1}{p}} \leq\ 2\,\mathrm{c}\,\|f-g\|_{L^2(\boldsymbol{\lambda})} \times\left\{\sqrt{p}\left(\sqrt{\|\rho_b\|_\infty} + \mathbb{L}_2\left(\frac{\log t}{t}\right)^{1/4}\right) + \frac{2\mathbb{L}_2 p}{t^{1/4}}\right\}.$$

Applying the generic chaining method and the localisation procedure as in Chapter 3, one can then verify that

$$\begin{aligned}
&\frac{1}{\sqrt{t}}\left(\mathbb{E}_b\left[\left\|\int_0^t K\left(\frac{\cdot - X_s}{h}\right)\mathrm{d}W_s\right\|_\infty^p\right]\right)^{\frac{1}{p}}\\
&\quad\leq \frac{\widetilde{C}_1}{t^{1/4}}\sum_{k=0}^\infty \int_0^\infty \log N(u,\mathcal{F}_k, 4\,\mathrm{c}\,\mathbb{L}_2\mathrm{e}\|\cdot\|_{\mathrm{L}^2(\boldsymbol{\lambda})})\mathrm{d}u\ \mathrm{e}^{-\frac{k}{2}}\\
&\qquad + \widetilde{C}_2\sum_{k=0}^\infty\int_0^\infty \sqrt{\log N\left(u,\mathcal{F}_k, 2\,\mathrm{c}\left(\sqrt{\|\rho_b\|_\infty} + \mathbb{L}_2\left(\frac{\log t}{t}\right)^{1/4}\right)\mathrm{e}\|\cdot\|_{\mathrm{L}^2(\boldsymbol{\lambda})}\right)}\mathrm{d}u\ \mathrm{e}^{-\frac{k}{2}}\\
&\qquad + 12\,\mathrm{c}\,\mathbb{V}\left(\sqrt{\|\rho_b\|_\infty} + \mathbb{L}_2\left(\frac{\log t}{t}\right)^{1/4}\right)\sqrt{p} + \frac{24\,\mathrm{c}\,\mathbb{V}\mathbb{L}_2 p}{t^{1/4}}\\
&\quad\leq \frac{6\widetilde{C}_1}{t^{1/4}} v\mathbb{V}\ 4\,\mathrm{c}\,\mathbb{L}_2\mathrm{e}\left(1 + \log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S} + p\Lambda t}\right)\right) + (\mathbf{A}) + (\mathbf{B}) + \frac{24\,\mathrm{c}\,\mathbb{V}\mathbb{L}_2 p}{t^{1/4}},
\end{aligned}$$

with

$$\begin{aligned}
(\mathbf{A}) &:=\ 12\widetilde{C}_2\mathbb{V}\ 2\,\mathrm{c}\left(\sqrt{\|\rho_b\|_\infty} + \mathbb{L}_2\left(\frac{\log t}{t}\right)^{1/4}\right)\mathrm{e}\sqrt{v\log\left(\frac{\mathbb{A}}{\mathbb{V}}\sqrt{\mathcal{S}+p\Lambda t}\right)},\\
(\mathbf{B}) &:=\ 12\,\mathrm{c}\,\mathbb{V}\left(\sqrt{\|\rho_b\|_\infty} + \mathbb{L}_2\left(\frac{\log t}{t}\right)^{1/4}\right)\sqrt{p}
\end{aligned}$$

for $\mathbb{A}$ and $\mathcal{S}$ defined as in the proof of Proposition 41 and some positive constant Λ. For some further positive constants $\mathbb{L}_3, \mathbb{L}_4$, we can then upper bound

$$\begin{aligned}
\frac{(\mathbf{A})}{\sqrt{th}} &\leq\ \mathbb{L}_3\left(\frac{1}{\sqrt{th}} + \frac{(\log(pt/h))^{3/4}}{t^{3/4}\sqrt{h}}\right) + \overline{\eta}_1\sqrt{\|\rho_b\|_\infty\ \frac{\log(pt/h)}{th}},\\
\frac{(\mathbf{B})}{\sqrt{th}} &\leq\ \mathbb{L}_4\frac{1}{\sqrt{th}}\left(\frac{\log t}{t}\right)^{1/4}\sqrt{p} + \overline{\eta}_2\sqrt{\|\rho_b\|_\infty}\sqrt{\frac{p}{th}}.
\end{aligned}$$

Thus, with regard to the upper bound for $\left(\mathbb{E}_b\left[\|\overline{\rho}_{t,K}(h_t) - \mathbb{E}_b\left[\overline{\rho}_{t,K}(h_t)\right]\|_\infty^p\right]\right)^{\frac{1}{p}}$, the constants preceding the terms $(th)^{-1/2}\sqrt{\log(pt/h)}$ and $\sqrt{p/(th)}$ are identified as $\sqrt{\|\rho_b\|_\infty}\overline{\eta}_1$ and $\sqrt{\|\rho_b\|_\infty}\overline{\eta}_2$, respectively.

Furthermore, we obtained the additional expression $(th)^{-1/2}(\log t/t)^{1/4}\sqrt{p}$, and assertion (4.7.38) follows. In order to show (4.7.37), note finally that in case of $h_t \in ((\log t)^2 t^{-1}, (\log t)^{-3})$ and $1 \leq u_t \leq \alpha \log t$, $\alpha > 0$, it holds

$$\widetilde{\phi}_{t,h_t}(u_t) \;=\; o\left(\frac{1}{\sqrt{th_t}}\sqrt{\log\left(\frac{u_t t}{h_t}\right)}\right) + \overline{\eta}_1\sqrt{\|\rho_b\|_\infty}\frac{1}{\sqrt{th_t}}\sqrt{\log\left(\frac{u_t t}{h_t}\right)} + \overline{\eta}_2\sqrt{\|\rho_b\|_\infty \cdot \frac{u_t}{th_t}}.$$

Thus, $\widetilde{\phi}_{t,h_t}(u_t) \leq \overline{\psi}_{t,h_t}(u_t)$ for t sufficiently large. □

Remark 63. We shortly comment on the constants appearing in the definition of $\overline{\eta}_1$ and $\overline{\eta}_2$ in Lemma 62. The constant c is defined as $\mathrm{c} = \sqrt{2}\max\{1, \overline{\mathrm{c}}\}$, for $\overline{\mathrm{c}}$ denoting the universal constant from the Burkholder–Davis–Gundy inequality as stated in Proposition 4.2 in Barlow and Yor (1982). The constant $\widetilde{C}_2$ originates from the application of Proposition 34 in Chapter 3 and can be specified looking into the details of the proof in Dirksen (2015). The constant v is associated with the entropy condition on $\mathcal{F}$ defined as in (4.7.32) (see Lemma 35 in Chapter 3).

4.7.2 Proof of main results

This section contains the proof of the upper and lower bound results on minimax optimal drift estimation wrt sup-norm risk presented in Section 4.4 and Section 4.5.

Proof of Theorem 52. In order to prove (4.4.23), it suffices to investigate the estimator $\overline{\rho}_{t,K}(\widehat{h}_t)$ since conditions (E1) and (E2) from Lemma 60 are satisfied. Indeed, the first condition refers to the rate of convergence of the invariant density estimator, and it is fulfilled by the kernel density estimator $\rho^+_{t,K}$ with bandwidth $t^{-1/2}$ due to Proposition 44. Note that $b \in \Sigma(\beta, \mathcal{L})$ in the current chapter refers to Hölder continuity of ρ_b with parameter $\beta + 1$ in contrast to the notation in Chapter 3. With regard to (E2), Lemma 59 implies that

$$\begin{aligned}\sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{E}_b\left[\|\overline{\rho}_{t,K}(\widehat{h}_t)\|_\infty^2\right] &\leq \sup_{b\in\Sigma(\beta,\mathcal{L})}\Big(4\mathbb{E}_b\left[\|\overline{\rho}_{t,K}(\widehat{h}_t) - \mathbb{E}_b\left[\overline{\rho}_{t,K}(\widehat{h}_t)\right]\|_\infty^2\right] \\ &\qquad + 4\|\mathbb{E}_b\left[\overline{\rho}_{t,K}(\widehat{h}_t)\right] - b\rho_b\|_\infty^2 + 4\|b\rho_b\|_\infty^2\Big) \\ &\lesssim \overline{\sigma}^2(h_{\min}, t) + B(1) + 1 \;=\; O(1) \quad \text{ as } t \to \infty,\end{aligned}$$

since $1 > \widehat{h}_t \geq h_{\min} = \min\{h_k \in \mathcal{H}\colon k \in \mathbb{N}\} > (\log t)^2/t$ and $\|\rho_b b\|_\infty = \|\rho_b'\|_\infty/2 \leq \mathcal{L}$. Hence, (E2) is satisfied. Consequently, (4.4.23) will follow once we have verified condition (E3) from Lemma 60, i.e., by showing that

$$\sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{E}_b\left[\|\overline{\rho}_{t,K}(\widehat{h}_t) - \rho_b'/2\|_\infty\right] \;=\; O\left(\left(\frac{\log t}{t}\right)^{\frac{\beta}{2\beta+1}}\right).$$

For $C = C(K)$ introduced in (4.4.21), let $M := C\|\rho_b\|_\infty$, and define $\overline{h}_\rho := h(\rho_b')$ as

$$\overline{h}_\rho \;:=\; \max\left\{h \in \mathcal{H}\colon B(h) \;\leq\; \frac{\sqrt{0.8M}}{4}\,\overline{\sigma}(h,t)\right\}.$$

For the given choice of $\overline{h}_\rho$, it holds $B(\overline{h}_\rho) \simeq \sqrt{0.8M}\overline{\sigma}(\overline{h}_\rho,t)/4$, and, since $\mathcal{H} \ni \overline{h}_\rho > (\log t)^2/t$,

$$\overline{h}_\rho^{2\beta+1} \simeq (\log t/t) \quad \text{and} \quad \overline{\sigma}(\overline{h}_\rho,t) \simeq (\log t/t)^{\frac{\beta}{2\beta+1}}. \tag{4.7.39}$$

To see this, first note that, for $h_0 := (\log t/t)^{\frac{1}{2\beta+1}}$, there exists some positive constant L such that $B(h_0) \leq L\overline{\sigma}(h_0,t)$ and $\overline{\sigma}^2(h_0,t) \simeq (\log t/t)^{\frac{2\beta}{2\beta+1}}$. In particular, we have $\overline{h}_\rho \gtrsim h_0$ which is clear by definition of $\overline{h}_\rho$ in case that $L \leq \sqrt{0.8M}/4$. Otherwise, this follows from the fact that, for any $0<\lambda<1$,

$$B(\lambda h_0) = \lambda^\beta B(h_0) \leq \lambda^\beta L\overline{\sigma}(h_0,t) \leq \lambda^\beta L\overline{\sigma}(\lambda h_0,t).$$

For the validity of (4.7.39), it remains to show that $\overline{h}_\rho \lesssim h_0$. We prove that $\overline{h}_\rho^{2\beta+1}h_0^{-(2\beta+1)} = O(1)$. By definition, we have

$$\overline{h}_\rho^{2\beta+1} \simeq \frac{\overline{h}_\rho}{t}\left(\log\left(\frac{t}{\overline{h}_\rho}\right)\right)^3 + \frac{1}{t}\log\left(\frac{t}{\overline{h}_\rho}\right) \lesssim \frac{\overline{h}_\rho}{t}(\log t)^3 + \frac{\log t}{t}, \tag{4.7.40}$$

since, for any $h \in \mathcal{H}$, $h \geq (\log t)^2/t$. This implies that $\overline{h}_\rho^{2\beta+1}h_0^{-(2\beta+1)} \lesssim \overline{h}_\rho(\log t)^2+1$. Again exploiting (4.7.40), we deduce that

$$\begin{aligned}
\overline{h}_\rho(\log t)^2 &\lesssim (\log t)^2\left(\frac{\overline{h}_\rho}{t}(\log t)^3 + \frac{\log t}{t}\right)^{\frac{1}{2\beta+1}}\\
&\lesssim (\log t)^2 t^{-\frac{1}{2\beta+1}}\left(\overline{h}_\rho(\log t)^3 + \log t\right)^{\frac{1}{2\beta+1}}\\
&\lesssim (\log t)^2 t^{-\frac{1}{2\beta+1}}\left((\log t)^3 + \log t\right)^{\frac{1}{2\beta+1}} = o(1).
\end{aligned}$$

Thus, $\overline{h}_\rho^{2\beta+1}h_0^{-(2\beta+1)} = O(1)$, and we have shown (4.7.39).

Case 1: We first consider the situation where $\{\widehat{h}_t \geq \overline{h}_\rho\}$. The definition of $\widehat{h}_t$ according to (4.4.22) and the bias and variance estimates in (4.7.34) and (4.7.33), respectively, imply that

$$\begin{aligned}
&\mathbb{E}_b\left[\|\overline{\rho}_{t,K}(\widehat{h}_t) - \rho_b'/2\|_\infty \mathbb{1}_{\{\widehat{h}_t\geq\overline{h}_\rho\}\cap\{\widetilde{M}\leq 1.2M\}}\right]\\
&\leq \mathbb{E}_b\Bigg[\Big(\|\overline{\rho}_{t,K}(\widehat{h}_t) - \overline{\rho}_{t,K}(\overline{h}_\rho)\|_\infty + \|\overline{\rho}_{t,K}(\overline{h}_\rho) - \mathbb{E}_b\left[\overline{\rho}_{t,K}(\overline{h}_\rho)\right]\|_\infty\\
&\qquad + \|\mathbb{E}_b\left[\overline{\rho}_{t,K}(\overline{h}_\rho)\right] - \rho_b'/2\|_\infty\Big)\mathbb{1}_{\{\widehat{h}_t\geq\overline{h}_\rho\}\cap\{\widetilde{M}\leq 1.2M\}}\Bigg]\\
&\leq \sqrt{1.2M}\,\overline{\sigma}(\overline{h}_\rho,t) + \mathcal{K}\overline{\sigma}(\overline{h}_\rho,t) + \frac{\sqrt{0.8M}}{4}\,\overline{\sigma}(\overline{h}_\rho,t) = O(\overline{\sigma}(\overline{h}_\rho,t)).
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\mathbb{E}_b\left[\|\overline{\rho}_{t,K}(\widehat{h}_t)-\rho_b'/2\|_\infty\ \mathbb{1}_{\{\widehat{h}_t\geq\overline{h}_\rho\}\cap\{\widetilde{M}>1.2M\}}\right]\\
&\leq \sum_{h\in\mathcal{H}\colon h\geq\overline{h}_\rho}\mathbb{E}_b\Big[\left(\|\overline{\rho}_{t,K}(h)-\mathbb{E}_b\left[\overline{\rho}_{t,K}(h)\right]\|_\infty+B(h)\right)\ \mathbb{1}_{\{\widehat{h}_t=h\}\cap\{\widetilde{M}>1.2M\}}\Big]\\
&\lesssim \log t\left(\mathcal{K}\overline{\sigma}(\overline{h}_\rho,t)+B(1)\right)\ \sqrt{\mathbb{E}_b\left[\mathbb{1}\{\widetilde{M}>1.2M\}\right]}.
\end{aligned}$$

The function $\psi_{t,t^{-1/2}}$ introduced in (4.2.12) obviously fulfills $\psi_{t,t^{-1/2}}(\log t)=o(1)$. Plugging in the definition of $\widetilde{M}$ (see (4.4.21)), one thus obtains by means of Proposition 44, for t sufficiently large,

$$\begin{aligned}
\mathbb{P}_b\left(|\widetilde{M}-C\|\rho_b\|_\infty|>0.2C\|\rho_b\|_\infty\right) &= \mathbb{P}_b\left(\Big|\|\rho_{t,K}(t^{-1/2})\|_\infty-\|\rho_b\|_\infty\Big|>0.2\|\rho_b\|_\infty\right)\\
&\leq \mathbb{P}_b\left(\|\rho_{t,K}(t^{-1/2})-\rho_b\|_\infty>\mathrm{e}\psi_{t,t^{-1/2}}(\log t)\right)\\
&\leq t^{-1}.
\end{aligned}\tag{4.7.41}$$

Consequently, we have shown that

$$\mathbb{E}_b\left[\|\overline{\rho}_{t,K}(\widehat{h}_t)-\rho_b'/2\|_\infty\ \mathbb{1}_{\{\widehat{h}_t\geq\overline{h}_\rho\}}\right]\ =\ O(\overline{\sigma}(\overline{h}_\rho,t)).$$

Case 2: It remains to consider the case $\{\widehat{h}_t<\overline{h}_\rho\}$. Decomposing again as in the proof of Theorem 2 in Giné and Nickl (2009), we have

$$\begin{aligned}
&\mathbb{E}_b\left[\|\overline{\rho}_{t,K}(\widehat{h}_t)-\rho_b'/2\|_\infty\ \mathbb{1}_{\{\widehat{h}_t<\overline{h}_\rho\}\cap\{\widetilde{M}<0.8M\}}\right]\\
&\leq \sum_{h\in\mathcal{H}\colon h<\overline{h}_\rho}\mathbb{E}_b\Big[\left(\|\overline{\rho}_{t,K}(h)-\mathbb{E}_b\left[\overline{\rho}_{t,K}(h)\right]\|_\infty+B(h)\right)\ \mathbb{1}_{\{\widehat{h}_t=h\}\cap\{\widetilde{M}<0.8M\}}\Big]\\
&\lesssim \log t\left(\mathcal{K}\overline{\sigma}(h_{\min},t)+B(\overline{h}_\rho)\right)\ \sqrt{\mathbb{E}_b\left[\mathbb{1}\{\widetilde{M}<0.8M\}\right]}\ =\ O(\overline{\sigma}(\overline{h}_\rho,t)),
\end{aligned}$$

where we used (4.7.41) for deriving the last inequality. Furthermore,

$$\begin{aligned}
&\mathbb{E}_b\left[\|\overline{\rho}_{t,K}(\widehat{h}_t)-\rho_b'/2\|_\infty\ \mathbb{1}_{\{\widehat{h}_t<\overline{h}_\rho\}\cap\{\widetilde{M}\geq0.8M\}}\right]\\
&\leq \sum_{h\in\mathcal{H}\colon h<\overline{h}_\rho}\mathbb{E}_b\Big[\left(\|\overline{\rho}_{t,K}(h)-\mathbb{E}_b\left[\overline{\rho}_{t,K}(h)\right]\|_\infty+\|\mathbb{E}_b\left[\overline{\rho}_{t,K}(h)\right]-\rho_b'/2\|_\infty\right)\mathbb{1}_{\{\widehat{h}_t=h\}\cap\{\widetilde{M}\geq0.8M\}}\Big]\\
&\leq \sum_{h\in\mathcal{H}\colon h<\overline{h}_\rho}\mathcal{K}\overline{\sigma}(h,t)\cdot\sqrt{\mathbb{P}_b\left(\{\widehat{h}_t=h\}\cap\{0.8M\leq\widetilde{M}\}\right)}+B(\overline{h}_\rho).
\end{aligned}$$

Since the latter summand is of order $O(\overline{\sigma}(\overline{h}_\rho,t))$, we may focus on bounding the first term. Using again the arguments of Giné and Nickl (2009), the proof boils down to verifying that the term

$$(\mathbf{I}):=\sum_{g\in\mathcal{H}\colon g\leq h}\mathbb{P}_b\left(\|\overline{\rho}_{t,K}(h^+)-\overline{\rho}_{t,K}(g)\|_\infty>\sqrt{0.8M}\overline{\sigma}(g,t)\right),$$

with $h^+ := \min\{g \in \mathcal{H} : g > h\}$, satisfies

$$\sum_{h\in\mathcal{H}:\, h<\overline{h}_\rho} \overline{\sigma}(h,t)\sqrt{(\mathbf{I})} \;=\; O(\overline{\sigma}(\overline{h}_\rho,t)).$$

Analogously to Giné and Nickl (2009), we start by noting that, for $g < h^+ \leq \overline{h}_\rho$,

$$\begin{aligned}
&\|\overline{\rho}_{t,K}(h^+) - \overline{\rho}_{t,K}(g)\|_\infty \\
&\quad\leq\; \|\overline{\rho}_{t,K}(h^+) - \mathbb{E}_b\left[\overline{\rho}_{t,K}(h^+)\right]\|_\infty + \|\overline{\rho}_{t,K}(g) - \mathbb{E}_b\left[\overline{\rho}_{t,K}(g)\right]\|_\infty + B(h^+) + B(g)\\
&\quad\leq\; \|\overline{\rho}_{t,K}(h^+) - \mathbb{E}_b\left[\overline{\rho}_{t,K}(h^+)\right]\|_\infty + \|\overline{\rho}_{t,K}(g) - \mathbb{E}_b\left[\overline{\rho}_{t,K}(g)\right]\|_\infty + \frac{1}{2}\sqrt{0.8M}\overline{\sigma}(g,t).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\mathbb{P}_b\left(\|\overline{\rho}_{t,K}(h^+) - \overline{\rho}_{t,K}(g)\|_\infty > \sqrt{0.8M}\overline{\sigma}(g,t)\right)\\
&\quad\leq\; \mathbb{P}_b\left(\|\overline{\rho}_{t,K}(h^+) - \mathbb{E}_b\left[\overline{\rho}_{t,K}(h^+)\right]\|_\infty > \frac{\sqrt{0.8M}}{4}\overline{\sigma}(h^+,t)\right)\\
&\qquad\qquad + \mathbb{P}_b\left(\|\overline{\rho}_{t,K}(g) - \mathbb{E}_b\left[\overline{\rho}_{t,K}(g)\right]\|_\infty > \frac{\sqrt{0.8M}}{4}\overline{\sigma}(g,t)\right).
\end{aligned}$$

We want to apply Lemma 62 for bounding the last two terms, and for doing so, we verify that

$$\mathrm{e}\overline{\psi}_{t,g}(\log(1/g)) \;\leq\; \frac{\sqrt{0.8M}}{4}\overline{\sigma}(g,t).$$

Indeed,

$$\begin{aligned}
\mathrm{e}\overline{\psi}_{t,g}(\log(1/g)) &=\; \mathrm{e}\sqrt{\|\rho_b\|_\infty}\left\{2\overline{\eta}_1\frac{1}{\sqrt{tg}}\left(\log\left(\frac{\log(1/g)t}{g}\right)\right)^{1/2} + \overline{\eta}_2\left(\frac{\log(1/g)}{tg}\right)^{1/2}\right\}\\
&\leq\; \mathrm{e}(4\overline{\eta}_1 + 2\overline{\eta}_2)\sqrt{\|\rho_b\|_\infty}\left(\frac{\log(t/g)}{tg}\right)^{1/2}\\
&\leq\; \mathrm{e}(4\overline{\eta}_1 + 2\overline{\eta}_2)\sqrt{\|\rho_b\|_\infty}\left(\frac{1}{\sqrt{t}}\left(\log\left(\frac{t}{g}\right)\right)^{3/2} + \left(\frac{\log(t/g)}{tg}\right)^{1/2}\right)\\
&=\; \mathrm{e}(4\overline{\eta}_1 + 2\overline{\eta}_2)\sqrt{\|\rho_b\|_\infty}\overline{\sigma}(g,t) \;\leq\; \frac{\sqrt{0.8M}}{4}\overline{\sigma}(g,t),
\end{aligned}$$

for t sufficiently large and for $M = 20\mathrm{e}^2(4\overline{\eta}_1 + 2\overline{\eta}_2)^2\|\rho_b\|_\infty$ since $h_{\min} \leq g \leq \overline{h}_\rho$. Lemma 62 then implies that, for every $g \leq \overline{h}_\rho$, $g \in \mathcal{H}$ and t large enough,

$$\begin{aligned}
&\mathbb{P}_b\left(\|\overline{\rho}_{t,K}(g) - \mathbb{E}_b\left[\overline{\rho}_{t,K}(g)\right]\|_\infty > \frac{\sqrt{0.8M}}{4}\overline{\sigma}(g,t)\right)\\
&\quad\leq\; \mathbb{P}_b\left(\|\overline{\rho}_{t,K}(g) - \mathbb{E}_b\left[\overline{\rho}_{t,K}(g)\right]\|_\infty > \mathrm{e}\overline{\psi}_{t,g}\left(\log\left(\frac{1}{g}\right)\right)\right) \;\leq\; g.
\end{aligned}$$

Consequently,

$$\sum_{g\in\mathcal{H}\colon g\leq h} \mathbb{P}_b\left(\|\overline{\rho}_{t,K}(h^+)-\overline{\rho}_{t,K}(g)\|_\infty > \sqrt{0.8M}\overline{\sigma}(g,t)\right) \leq \sum_{g\in\mathcal{H}\colon g\leq h}(h^+ + g) \lesssim h\log t$$

and

$$\sum_{h\in\mathcal{H}\colon h<\overline{h}_\rho} \overline{\sigma}(h,t)\sqrt{(\mathbf{I})} \;=\; O\left(t^{-\frac{1}{2}}(\log t)^2\right) \;=\; o(\overline{\sigma}(\overline{h}_\rho,t)). \tag{4.7.42}$$

□

We now turn to proving the result on lower bounds for drift estimation wrt sup-norm risk.

Proof of Theorem 53. We start with stating a crucial auxiliary result.

Lemma 64 (Theorem 2.7 in Tsybakov (2009)). *Fix $\mathcal{C},A,\gamma,\beta,\mathcal{L}\in(0,\infty)$, and assume that there exist a finite set $J_t=\{0,\dots,M_t\}$, $M_t\in\mathbb{N}$, and hypotheses $\{b_j : j\in J_t\}\subseteq\Sigma(\beta,\mathcal{L})$ satisfying*

(a) $\|\rho_j'-\rho_k'\|_\infty \geq 2\psi_t>0,$ *for any $j\neq k$, $j,k\in J_t$, or*

(b) $\|(\rho_j^2)'-(\rho_k^2)'\|_\infty \geq 2\psi_t>0,$ *for any $j\neq k$, $j,k\in J_t$,*

together with the condition that, for any $j\in J_t$, $\mathbb{P}_{b_j}=:\mathbb{P}_j\ll\mathbb{P}_0$, and

$$\frac{1}{|J_t|}\sum_{j\in J_t}\mathrm{KL}(\mathbb{P}_j,\mathbb{P}_0)=\frac{1}{|J_t|}\sum_{j\in J_t}\mathbb{E}_j\left[\log\left(\frac{\mathrm{d}\mathbb{P}_j}{\mathrm{d}\mathbb{P}_0}\left(X^t\right)\right)\right] \;\leq\; \alpha\log(|J_t|),$$

for some $0<\alpha<1/8$. Here, $\mathbb{P}_{b_j}$ is the measure of the ergodic diffusion process defined via the SDE (4.1.1) *with drift coefficient b_j and the corresponding invariant density ρ_j, $j\in J_t$. Then, in case of* (a) *and* (b), *respectively, it follows*

$$\inf_{\widetilde{\partial\rho}_t}\sup_{b\in\Sigma(\beta,\mathcal{L})}\mathbb{E}_b\left[\psi_t^{-1}\|\widetilde{\partial\rho}_t-\rho_b'\|_\infty\right]\geq\; c(\alpha)>0,$$

$$\inf_{\widetilde{\partial\rho_t^2}}\sup_{b\in\Sigma(\beta,\mathcal{L})}\mathbb{E}_b\left[\psi_t^{-1}\|\widetilde{\partial\rho^2}_t-(\rho_b^2)'\|_\infty\right]\geq\; c(\alpha)>0,$$

where the constant $c(\alpha)$ depends only on α and the infimum is taken over all estimators $\widetilde{\partial\rho}_t$ of ρ_b' and estimators $\widetilde{\partial\rho_t^2}$ of $(\rho_b^2)'$, respectively.

Proof of the lower bound for estimating ρ_b' The lower bound (4.4.24) follows by a straightforward application of Lemma 64.

`Step 1:` **Construction of the hypotheses.** Fix $\beta,\mathcal{L},\mathcal{C},\gamma,A\in(0,\infty)$, and let

$$b_0\in\Sigma(\beta,\mathcal{L}/2,\mathcal{C}/2,A,\gamma)\subseteq\Sigma(\beta,\mathcal{L},\mathcal{C},A,\gamma).$$

In addition, set $J_t := \{0, \pm 1, \dots, \pm(\lfloor A(2h_t)^{-1}\rfloor - 1)\}$, $x_j := 2h_t j, j \in J_t$, and

$$h_t \ := \ v\left(\frac{\log t}{t}\right)^{\frac{1}{2\beta+1}},$$

$v < 1$ some positive constant which will be specified later. Let $Q\colon \mathbb{R} \to \mathbb{R}$ be a function that satisfies $Q \in C_c^\infty(\mathbb{R})$, $\operatorname{supp}(Q) \subseteq [-\frac{1}{2}, \frac{1}{2}]$, $Q \in \mathcal{H}(\beta+1, \frac{1}{2})$, $|Q'(0)| > 0$ and $\int Q(x)\mathrm{d}x = 0$. The hypotheses will be constructed via additive perturbations of the invariant density $\rho_0 := \rho_{b_0} \in \mathcal{H}(\beta+1, \mathcal{L}/2)$ associated with the drift coefficient b_0. For this purpose, define

$$G_0 :\equiv 0, \qquad G_j(x) := \mathcal{L}h_t^{\beta+1} Q\left(\frac{x - x_j}{h_t}\right), \quad x \in \mathbb{R},\, j \in J_t \setminus \{0\},$$

and the hypotheses $\rho_j := \rho_0 + G_j$, $j \in J_t$. Fix $j \in J_t \setminus \{0\}$, and note that

$$\partial^k G_j(x) \ = \ \mathcal{L}h_t^{\beta+1-k}\partial^k Q\left(\frac{x - x_j}{h_t}\right), \quad x \in \mathbb{R},\, k = 0, ..., \lfloor \beta+1 \rfloor.$$

We immediately deduce that $\|\partial^k G_j\|_\infty \leq \mathcal{L}/2$ since $h_t \leq 1$ for t sufficiently large and $Q \in \mathcal{H}\left(\beta+1, \frac{1}{2}\right)$. Furthermore, for any $x, y \in \mathbb{R}$,

$$\begin{aligned}
\left|\partial^{\lfloor\beta+1\rfloor} G_j(x) - \partial^{\lfloor\beta+1\rfloor} G_j(y)\right| &= \ \mathcal{L}h_t^{\beta+1-\lfloor\beta+1\rfloor}\left|\partial^{\lfloor\beta+1\rfloor} Q\left(\frac{x - x_j}{h_t}\right) - \partial^{\lfloor\beta+1\rfloor} Q\left(\frac{y - x_j}{h_t}\right)\right| \\
&\leq \ \mathcal{L}h_t^{\beta+1-\lfloor\beta+1\rfloor}\frac{1}{2}\left(\frac{|x-y|}{h_t}\right)^{\beta+1-\lfloor\beta+1\rfloor} \\
&= \ \frac{\mathcal{L}}{2}|x-y|^{\beta+1-\lfloor\beta+1\rfloor}.
\end{aligned}$$

Thus, $G_j \in \mathcal{H}(\beta+1, \mathcal{L}/2)$, and, in particular,

$$\rho_j \ = \ \rho_0 + G_j \in \mathcal{H}(\beta+1, \mathcal{L}), \qquad \text{for all } j \in J_t. \tag{4.7.43}$$

Note that, since $\operatorname{supp}(G_j) \subset (-A, A)$ for all $j \in J_t$, we have $G_j(x) = 0$ for all $|x| \geq A$. Thus, for verifying non-negativity, it suffices to show that $\rho_j(x) \geq 0$ for any $x \in (-A, A)$. Furthermore, $\inf_{-A \leq x \leq A} \rho_0(x) > 0$, and, since

$$\|G_j\|_\infty \ \leq \ \mathcal{L}h_t^{\beta+1}\left\|Q\left(\frac{\cdot - x_j}{h_t}\right)\right\|_\infty \ \leq \ \frac{\mathcal{L}}{2}h_t^{\beta+1} \ = \ o(1),$$

it holds $\liminf_{t\to\infty} \min_{j \in J_t} \inf_{-A \leq x \leq A} \rho_j(x) > 0$. For t sufficiently large, we therefore find a constant c_* such that

$$\min_{j \in J_t} \inf_{-A \leq x \leq A} \rho_j(x) \ \geq \ c_* \ > \ 0. \tag{4.7.44}$$

Since, in addition, $\int G_j(x)\mathrm{d}x = 0$, ρ_j is a probability density for each $j \in J_t$. Let

$$b_j \ := \ \frac{\rho_j'}{2\rho_j}, \quad j \in J_t,$$

and note that $b_j \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R})$. Furthermore, b_j can be rewritten as

$$b_j = \frac{1}{2}\frac{\rho_0' + G_j'}{\rho_0 + G_j} = b_0 + \frac{1}{2}\left(\frac{\rho_0' + G_j'}{\rho_0 + G_j} - \frac{\rho_0'}{\rho_0}\right) = b_0 + \frac{1}{2}\left(\frac{G_j'\rho_0 - \rho_0' G_j}{\rho_0(\rho_0 + G_j)}\right). \tag{4.7.45}$$

From this representation, we directly see that $b_0 = b_j$ on $\mathbb{R} \setminus (-A, A)$. Since $\|G_j\|_\infty + \|G_j'\|_\infty = o(1)$ and due to (4.7.44), we have that

$$\sup_{-A<x<A} \frac{1}{2}\left|\frac{G_j'\rho_0 - \rho_0' G_j}{\rho_0(\rho_0 + G_j)}\right| \;\le\; \frac{\mathcal{C}}{2}, \quad \text{for } t \text{ sufficiently large.}$$

In particular, this implies

$$|b_j(x)| \;\le\; \frac{\mathcal{C}}{2}(1+|x|) + \frac{\mathcal{C}}{2} \;\le\; \mathcal{C}(1+|x|), \quad \forall x \in \mathbb{R},\, t \text{ sufficiently large,}$$

and we can deduce that, for any $j \in J_t$, $b_j \in \Sigma(\mathcal{C}, A, \gamma, 1)$. Therefore, each b_j gives rise to an ergodic diffusion process via the SDE $\mathrm{d}X_t = b_j(X_t)\mathrm{d}t + \mathrm{d}W_t$ with invariant density ρ_j. Taking into consideration (4.7.43), we have thus shown that

$$b_j \in \Sigma(\beta, \mathcal{L}, \mathcal{C}, A, \gamma), \quad \forall j \in J_t,$$

for t sufficiently large.

`Step 2:` **Evaluation of the Kullback–Leibler divergence between the hypotheses.** From Girsanov's theorem, it can be deduced that

$$\mathrm{KL}(\mathbb{P}_j, \mathbb{P}_0) \;=\; \mathbb{E}_j\left[\log\left(\frac{\rho_j(X_0)}{\rho_0(X_0)}\right)\right] + \frac{1}{2}\mathbb{E}_j\left[\int_0^t (b_0(X_u) - b_j(X_u))^2\,\mathrm{d}u\right] \;=:\; (\mathbf{I}_j) + (\mathbf{II}_j).$$

We first prove boundedness of $(\mathbf{I}_j)$. Note that $\rho_j/\rho_0 \equiv 1$ on $\mathbb{R} \setminus (-A, A)$ and

$$\max_{j\in J_t} \sup_{-A\le x\le A} \frac{\rho_j(x)}{\rho_0(x)} \;\le\; \frac{\mathcal{L}}{\inf_{-A\le x\le A}\rho_0(x)}, \qquad \min_{j\in J_t} \inf_{-A\le x\le A} \frac{\rho_j(x)}{\rho_0(x)} \;\ge\; \frac{c_*}{\mathcal{L}}.$$

Therefore, ρ_j/ρ_0 is bounded away from zero and infinity, uniformly for all $j \in J_t$ and t sufficiently large. In particular, $\max_{j\in J_t} |(\mathbf{I}_j)| \;=\; O(1)$. We now turn to analysing $(\mathbf{II}_j)$. From (4.7.45), we can deduce, for any $j \in J_t$,

$$b_0 - b_j \;=\; \frac{b_0 G_j}{\rho_0 + G_j} - \frac{1}{2}\frac{G_j'}{\rho_0 + G_j}. \tag{4.7.46}$$

Since $G_j \equiv 0$ on $\mathbb{R} \setminus (-A, A)$,

$$\max_{j\in J_t}\left\|\frac{b_0 G_j}{\rho_0 + G_j}\right\|_\infty \;\le\; \max_{j\in J_t} \sup_{-A\le x\le A} |b_0(x)| \frac{\|G_j\|_\infty}{c_*} \;=\; O\left(h_t^{\beta+1}\right),$$

and, consequently,

$$\max_{j\in J_t}\mathbb{E}_j\left[\int_0^t \left(\frac{b_0(X_u)G_j(X_u)}{\rho_0(X_u) + G_j(X_u)}\right)^2 \mathrm{d}u\right] = O\left(t h_t^{2(\beta+1)}\right) = O\left(v^{2\beta+2}\log t\right).$$

For the second term on the rhs of (4.7.46), we calculate

$$\begin{aligned}
&\max_{j\in J_t}\mathbb{E}_j\left[\int_0^t\left(\frac{(G_j')^2(X_u)}{4(\rho_0(X_u)+G_j(X_u))^2}\right)\mathrm{d}u\right]\\
&= \max_{j\in J_t}\frac{t}{4}\int_{-A}^{A}\frac{\mathcal{L}^2h_t^{2\beta}(Q')^2\left(\frac{x-x_j}{h_t}\right)}{(\rho_0(x)+G_j(x))^2}\left(\rho_0(x)+\mathcal{L}h_t^{\beta+1}Q\left(\frac{x-x_j}{h_t}\right)\right)\mathrm{d}x\\
&\lesssim \max_{j\in J_t}c_*^{-2}\Bigg[t\int_{-A}^{A}h_t^{2\beta}(Q')^2\left(\frac{x-x_j}{h_t}\right)\rho_0(x)\mathrm{d}x\\
&\qquad+\int_{-A}^{A}(Q')^2\left(\frac{x-x_j}{h_t}\right)\left|Q\left(\frac{x-x_j}{h_t}\right)\right|\mathrm{d}x h_t^{3\beta+1}\Bigg]\\
&\lesssim th_t^{2\beta+1}\mathcal{L}\int(Q')^2(x)\mathrm{d}x+2tA\|Q'\|_\infty^2\|Q\|_\infty h_t^{3\beta+1}\ \lesssim\ th_t^{2\beta+1}\ \simeq\ v^{2\beta+1}\log t.
\end{aligned}$$

This implies that $\max_{j\in J_t}(\mathbf{II}_j)\ =\ O(v^{2\beta+1}\log t)$ such that

$$\max_{j\in J_t}\mathrm{KL}(\mathbb{P}_j,\mathbb{P}_0)\ =\ O(v^{2\beta+1}\log t).$$

`Step 3:` **Deducing the lower bound by application of Lemma 64.** For any $j\neq k, j,k\in J_t$, we have

$$\begin{aligned}
\|\rho_j'-\rho_k'\|_\infty = \ &\|G_j'-G_k'\|_\infty=\mathcal{L}h_t^\beta\sup_{x\in\mathbb{R}}\left|Q'\left(\frac{x-x_j}{h_t}\right)-Q'\left(\frac{x-x_k}{h_t}\right)\right|\\
\geq\ &\mathcal{L}h_t^\beta|Q'(0)|=\mathcal{L}|Q'(0)|v^\beta\left(\frac{\log t}{t}\right)^{\frac{\beta}{2\beta+1}}.
\end{aligned}$$

Here we used that $|x_j-x_k|\geq 2h_t$ implies that $x_j\notin\operatorname{supp}\left(Q'\left(\frac{\cdot-x_k}{h_t}\right)\right)$, noting that

$$\operatorname{supp}\left(Q'\left(\frac{\cdot-x_k}{h_t}\right)\right)\subseteq(x_k-h_t,x_k+h_t).$$

Furthermore, the number of hypotheses $|J_t|$ satisfies

$$|J_t|\ =\ 2\left\lfloor\frac{A}{2h_t}\right\rfloor-1\ \simeq\ v^{-1}\left(\frac{t}{\log t}\right)^{\frac{1}{2\beta+1}}.$$

Consequently, there is a positive constant c_1 such that $\log(|J_t|)\ \geq\ c_1\log t$, for all t sufficiently large. From the arguments in Step 2, it is clear that v can be chosen small enough such that, for some positive constant c_2,

$$\frac{1}{|J_t]}\sum_{j\in J_t}\mathrm{KL}(\mathbb{P}_j,\mathbb{P}_0)\ \leq\ c_2v^{2\beta+1}\log t\ \leq\ \frac{1}{10}c_1\log t\ \leq\ \frac{1}{10}\log(|J_t|),$$

for all t sufficiently large. (4.4.24) now follows immediately from Lemma 64.

Proof of the weighted lower bound for drift estimation For proving (4.4.25), we use Lemma 61 and the following

Proposition 65. *Grant the assumptions of Theorem 53. Then,*

$$\liminf_{t\to\infty} \inf_{\widetilde{\partial\rho_t^2}} \sup_{b\in\Sigma(\beta,\mathcal{L})} \mathbb{E}_b\left[\left(\frac{\log t}{t}\right)^{-\frac{\beta}{2\beta+1}} \|\widetilde{\partial\rho_t^2} - (\rho_b^2)'\|_\infty\right] > 0$$

where the infimum is taken over all possible estimators $\widetilde{\partial\rho_t^2}$ of $(\rho_b^2)'$.

Proof. The proof follows exactly the lines of the proof of lower bound for estimating ρ_b', except from constructing the hypotheses in such a way that, for any $j \neq k$,

$$\|(\rho_j^2)' - (\rho_k^2)'\|_\infty \ \geq\ C\left(\frac{\log t}{t}\right)^{\frac{\beta}{2\beta+1}},$$

C some positive constant. This is achieved by choosing the kernel function Q involved in the construction of the function G_j, $j \in J_t$, a little bit differently. Precisely, specify some function $Q\colon \mathbb{R}\to\mathbb{R}$ such that $Q \in C_c^\infty(\mathbb{R})$, $\operatorname{supp}(Q) \subseteq [-\frac{1}{2},\frac{1}{2}]$, $Q \in \mathcal{H}(\beta+1,\frac{1}{2})$, $\int Q(x)\mathrm{d}x = 0$, $Q(0) = 0$ and $|Q'(0)| > 0$. For any $j \neq k, j,k \in J_t$, one then has

$$\begin{aligned}
\|(\rho_j^2)' - (\rho_k^2)'\|_\infty &= 2\|(\rho_0' + G_j')(\rho_0 + G_j) - (\rho_0' + G_k')(\rho_0 + G_k)\|_\infty\\
&= 2\|\rho_0'G_j + G_j'\rho_0 + G_j'G_j - \rho_0'G_k - G_k'\rho_0 - G_k'G_k\|_\infty\\
&\geq 2\left|\rho_0'(x_j)G_j(x_j) + G_j'(x_j)\rho_0(x_j) + G_j'(x_j)G_j(x_j)\right|\\
&\geq 2\mathcal{L}h_t^\beta\left|\rho_0'(x_j)h_tQ(0) + Q'(0)\rho_0(x_j) + \mathcal{L}h_t^{\beta+1}Q(0)Q'(0)\right|\\
&= 2\mathcal{L}h_t^\beta\left|Q'(0)\right|\rho_0(x_j) \ \geq\ 2\mathcal{L}h_t^\beta\left|Q'(0)\right|\inf_{-A\leq x\leq A}\rho_0(x)\\
&= 2\mathcal{L}\left|Q'(0)\right|\inf_{-A\leq x\leq A}\rho_0(x)v^\beta\left(\frac{\log t}{t}\right)^{\frac{\beta}{2\beta+1}}.
\end{aligned}$$

Here we used the fact that $|x_j - x_k| \geq 2h_t$ implies that

$$x_j \notin \operatorname{supp}\left(Q'\left(\frac{\cdot - x_k}{h_t}\right)\right) \cup \operatorname{supp}\left(Q\left(\frac{\cdot - x_k}{h_t}\right)\right)$$

because $\operatorname{supp}\left(Q'\left(\frac{\cdot - x_k}{h_t}\right)\right) \cup \operatorname{supp}\left(Q\left(\frac{\cdot - x_k}{h_t}\right)\right) \subseteq (x_k - h_t, x_k + h_t)$. The assertion then follows as in the previous proof (see Steps 1-3) from part (b) of Lemma 64. □

In particular, Proposition 65 implies that condition (4.7.35) from Lemma 61 is fulfilled for $\psi_t = (\log t/t)^{\frac{\beta}{2\beta+1}}$, and (4.4.25) follows. □

Proof of Theorem 54. The definition of $\widehat{h}_t$ according to (4.5.26) implies that

$$\|\rho_{t,K}(\widehat{h}_t) - \rho_{t,K}(h_{\min})\|_\infty \lesssim \frac{1}{\sqrt{t}\log t}. \tag{4.7.47}$$

Furthermore, $h_{\min}$ satisfies the assumption of Proposition 45 such that assertion (**I**) of the theorem immediately follows. It remains to verify (**II**). First, we remark that it always holds $\widehat{h}_t \geq h_{\min}$ such that $\widehat{h}_t$ is well-defined. For estimation of ρ_b via $\rho_{t,K}(h_{\min})$, Proposition 44 yields

$$\sup_{b\in\Sigma(\beta,\mathcal{L})} \left(\mathbb{E}_b\left[\|\rho_{t,K}(h_{\min}) - \rho_b\|_\infty^p\right]\right)^{\frac{1}{p}} = O\left(t^{-1/2}(1 + \sqrt{\log t} + \sqrt{p} + pt^{-1/2})\right).$$

We can then deduce from (4.7.47) that

$$\sup_{b\in\Sigma(\beta,\mathcal{L})} \left[\mathbb{E}_b\left(\left\|\rho_{t,K}(\widehat{h}_t) - \rho_b\right\|_\infty\right)^p\right]^{\frac{1}{p}} = O\left(t^{-1/2}(1 + \sqrt{\log t} + \sqrt{p} + pt^{-1/2})\right).$$

Thus, $\rho^+_{t,K}(\widehat{h}_t)$ satisfies assumption (E1) from Lemma 60, since clearly $\|\rho^+_{t,K}(\widehat{h}_t) - \rho_b\|_\infty \leq \|\rho_{t,K}(\widehat{h}_t) - \rho_b\|_\infty$ due to the positivity of ρ_b . We may now follow the proof of Theorem 52. In particular, it again suffices to investigate the estimator $\overline{\rho}_{t,K}(\widehat{h}_t)$ in order to prove (**II**) since conditions (E1) and (E2) from Lemma 60 are satisfied. Note that

$$\mathbb{P}_b\left(|\widehat{M} - C\|\rho_b\|_\infty| > 0.2C\|\rho_b\|_\infty\right) = \mathbb{P}_b\left(|\|\rho_{t,K}(h_{\min})\|_\infty - \|\rho_b\|_\infty| > 0.2\|\rho_b\|_\infty\right) \leq \ t^{-1}.$$

follows exactly as in the proof of Theorem 52 since $\psi_{t,h_{\min}}(\log t) = o(1)$. Additional arguments are required only for the investigation of Case 2 ($\widehat{h}_t < \overline{h}_\rho$). As in the proof of Theorem 52, it is shown that

$$\begin{aligned}&\mathbb{E}_b\left[\|\overline{\rho}_{t,K}(\widehat{h}_t) - \rho_b'/2\|_\infty \,\mathbb{1}_{\{\widehat{h}_t<\overline{h}_\rho\}\cap\{\widehat{M}\geq 0.8M\}}\right]\\ &\quad\leq \sum_{h\in\mathcal{H}\colon\, h<\overline{h}_\rho} \mathcal{K}\overline{\sigma}(h,t)\cdot\sqrt{\mathbb{P}_b\left(\{\widehat{h}_t = h\}\cap\{0.8M \leq \widehat{M}\}\right)} + B(\overline{h}_\rho).\end{aligned}$$

We bound the first term by $\sum_{h\in\mathcal{H}\colon\, h<\overline{h}_\rho}\mathcal{K}\overline{\sigma}(h,t)\left(\sqrt{(\mathbf{I})} + \sqrt{(\mathbf{II})}\right)$, with

$$\begin{aligned}(\mathbf{I}) &:= \sum_{g\in\mathcal{H}\colon\, g\leq h} \mathbb{P}_b\left(\|\overline{\rho}_{t,K}(h^+) - \overline{\rho}_{t,K}(g)\|_\infty > \sqrt{0.8M}\overline{\sigma}(g,t)\right),\\ (\mathbf{II}) &:= \mathbb{P}_b\left(\sqrt{t}\|\rho_{t,K}(h^+) - \rho_{t,K}(h_{\min})\|_\infty > \frac{\sqrt{h^+}(\log(1/h^+))^4}{\log t}\right),\end{aligned}$$

where $h^+ := \min\{g\in\mathcal{H} : g > h\}$. (**I**) is dealt with as in Theorem 52 (see (4.7.42)).

With regard to term (**II**), we argue as before by means of Proposition 42: Since $h_{\min} \leq h^+ \leq \overline{h}_\rho$, it holds for any $\beta > 0$ and an arbitrary positive constant $\mathbb{L}$, for some t onwards,

$$\begin{aligned}\sqrt{h^+}\left(1 + \log(1/\sqrt{h^+}) + \log(t)\right) + \sqrt{t}\mathrm{e}^{-\mathbb{L}t} + \sqrt{t}(h^+)^{\beta+1} &= o(\lambda'),\\ \sqrt{h_{\min}}\left(1 + \log(1/\sqrt{h_{\min}}) + \log(t)\right) + \sqrt{t}\mathrm{e}^{-\mathbb{L}t} + \sqrt{t}(h_{\min})^{\beta+1} &= o(\lambda'),\end{aligned}$$

letting $\lambda' := \sqrt{h^+}(\log(1/h^+))^4/\log t$. Thus, for any $t > 1$ sufficiently large,

$$\begin{aligned}
&\mathbb{P}_b\left(\sqrt{t}\|\rho_{t,K}(h^+) - \rho_{t,K}(h_{\min})\|_\infty > \lambda'\right) \\
&\quad\leq\ \mathbb{P}_b\left(\sqrt{t}\|\rho_{t,K}(h^+) - L_t^\bullet(X)t^{-1}\|_\infty > \frac{\lambda'}{2}\right) \\
&\qquad+\ \mathbb{P}_b\left(\sqrt{t}\|\rho_{t,K}(h_{\min}) - L_t^\bullet(X)t^{-1}\|_\infty > \frac{\lambda'}{2}\right) \\
&\quad\leq\ 2\exp\left(-\frac{\Lambda_1(\log(1/h^+))^4}{\log t}\right)\ \leq\ 2\exp\left(-\widetilde{\Lambda}_1(\log t)^2\right),
\end{aligned}$$

for some positive constant $\widetilde{\Lambda}_1$. Consequently,

$$\sum_{h\in\mathcal{H}\colon h<\overline{h}_\rho} \overline{\sigma}(h,t)\sqrt{(\mathbf{II})}\ \lesssim\ \log t\cdot\overline{\sigma}(h_{\min},t)\sqrt{\exp(-\widetilde{\Lambda}_1(\log t)^2)}\ =\ o(\overline{\sigma}(\overline{h}_\rho,t)).$$

We can then proceed as in the proof of Theorem 52 to finish the proof. □

5 Concluding remarks and outlook

The present work answers very classical questions in nonparametric statistics concerning the class of scalar, ergodic diffusions on the whole real line. A central challenge is the investigation of sup-norm risk which requires the development of deep probabilistic tools from empirical process theory extended to our continuous time framework. Furthermore, considering the natural model of diffusions which live on the whole real line entails technically demanding problems. Their solution lies beyond common strategies only working under standard boundedness conditions. In the literature, these challenges are often circumvented by restricting to compact sets or to diffusions with boundary reflections or periodic drift which confines the states of the process to a bounded set.

The objective of this thesis relies essentially on the assumption of the availability of a continuous record of observations. As has been demonstrated, this approach leads to results that reflect the very nature of the process which is due to the incorporation of their probabilistic structure. This structure has not been blurred by discretisation. On the other hand, the availability of continuous observations is apparently not a realistic assumption since data always comes discrete even if the underlying process is assumed to be continuous. This is one reason why we advocate the kernel type estimators suggested in the present work. They allow for a straightforward extension to the case of discrete observation schemes and studying sup-norm adaptive estimation of diffusion characteristics in this context is an immediate follow-up question.

Convergence rates will depend heavily on the structure of the observation record. In this regard, the results of this thesis can be considered as the fundamental benchmark for what is possible at best which is insightful in terms of deciding in which frequency data should be collected. High frequency data refers to the case of discrete observations at equally spaced time points with the time between two observations vanishing as time increases. High frequency observations can be expected to provide enough information to use and recover the continuous paths properties to some extent which is different from the case of low frequency data, i.e. equidistant, observations imposing different structural properties. Therefore, methods and results presented in this theses can serve as a good starting point for the investigation of high frequency data.

Another straightforward follow-up question is certainly the investigation of multivariate state variables. As for discrete observations, the kernel type estimators are easily adapted to this setting. In the diffusion context, different phenomena as compared to the scalar case can be expected. In particular, local time which plays a central role for the one-dimensional investigation does not exist. For the analysis of sup-norm risk in the described situations, uniform concentration inequalities will again be the key device and we expect our approach to concentration inequalities to be successful in this case, as well.

6 Bibliography

Aeckerle-Willems, C. and Strauch, C. (2018a). Concentration of scalar ergodic diffusions and some statistical implications. arXiv preprint. `arXiv:1807.11331v2`.

Aeckerle-Willems, C. and Strauch, C. (2018b). Sup-norm adaptive simultaneous drift estimation for ergodic diffusions. arXiv preprint. `arXiv:1808.10660`.

Bakry, D., Gentil, I., and Ledoux, M. (2014). *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften*. Springer Cham.

Banon, G. (1978). Nonparametric identification for diffusion processes. *SIAM J. Control Optim.*, **16**(3), 380–395.

Barlow, M. T. and Yor, M. (1982). Semimartingale inequalities via the Garsia–Rodemich–Rumsey lemma, and applications to local times. *J. Funct. Anal.*, **49**(2), 198–229.

Bhattacharya, R. N. (1982). On the functional central limit theorem and the law of the iterated logarithm for Markov processes. *Z. Wahrsch. Verw. Gebiete*, **60**(2), 185–201.

Bickel, P. J. and Ritov, Y. (2003). Nonparametric estimators which can be "plugged-in". *Ann. Statist.*, **31**(4), 1033–1053.

Bitseki Penda, S. V., Hoffmann, M., and Olivier, A. (2017). Adaptive estimation for bifurcating Markov chains. *Bernoulli*, **23**(4B), 3598–3637.

Bousquet, O. (2003). Concentration inequalities for sub-additive functions using the entropy method. In *Stochastic inequalities and applications*, pages 213–247. Springer.

Castillo, I. and Nickl, R. (2014). On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.*, **42**(5), 1941–1969.

Cattiaux, P., Chafai, D., and Guillin, A. (2012). Central limit theorems for additive functionals of ergodic markov diffusions processes. *ALEA*, **9**(2), 337–382.

Dalalyan, A. (2005). Sharp adaptive estimation of the drift function for ergodic diffusions. *Ann. Statist.*, **33**(6), 2507–2528.

Dalalyan, A. and Reiß, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. *Probab. Theory Relat. Fields*, **137**(1), 25–47.

Dirksen, S. (2015). Tail bounds via generic chaining. *Electron. J. Probab.*, **20**, no. 53, 29 pp.

Durrett, R. (1996). *Stochastic Calculus: A Practical Introduction*. Probability and Stochastics Series. Taylor & Francis.

Gao, F., Guillin, A., and Wu, L. (2013). Bernstein type's concentration inequalities for symmetric Markov processes. *Teor. Veroyatnost. i Primenen*, **58**(3), 521–549.

Gill, R. D. and Levit, B. Y. (1995). Applications of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli*, pages 59–79.

Giné, E. and Nickl, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probab. Theory Relat. Fields*, **143**(3-4), 569–596.

Giné, E. and Nickl, R. (2016). *Mathematical foundations of infinite-dimensional statistical models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, New York.

Gobet, E., Hoffmann, M., and Reiß, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.*, **32**(5), 2223–2253.

Gordin, M. I. and Lifsic, B. A. (1978). Central limit theorem for stationary Markov processes. *Dokl. Akad. Nauk SSSR*, **239**(4), 766–767.

Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Process. Appl.*, **79**(1), 135–163.

Kailath, T., Segall, A., and Zakai, M. (1978). Fubini-type theorems for stochastic integrals. *Sankhyā Ser. A*, **40**(2), 138–143.

Kallenberg, O. (1997). *Foundations of modern probability*. Springer Verlag New York Berlin Heidelberg.

Karatzas, I. and Shreve, S. (1988). *Brownian motion and stochastic calculus*. Graduate texts in mathematics. Springer-Verlag.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York.

Kutoyants, Y. A. (1998). Efficient density estimation for ergodic diffusion processes. *Statistical Inference for Stochastic Processes*, **1**(2), 131–155.

Kutoyants, Y. A. (2004). *Statistical inference for ergodic diffusion processes*. Springer Series in Statistics. Springer-Verlag London, Ltd., London.

Kutoyants, Y. A. and Yoshida, N. (2007). Moment estimation for ergodic diffusion processes. *Bernoulli*, **13**(4), 933–951.

Le Gall, J.-F. (2016). *Brownian motion, martingales, and stochastic calculus*, volume 274 of *Graduate Texts in Mathematics*. Springer, [Cham], french edition.

Liptser, R. S. and Shiryaev, A. N. (2001). *Statistics of Random Processes*, volume 1 General Theory of *Applications of Mathematics: Stochastic Modelling and Applied Probability*. Springer, Berlin, second edition.

Meyer, P. A. (1976). Un cours sur les intégrales stochastiques. pages 245–400. Lecture Notes in Math., Vol. 511.

Millar, P. W. (1983). The minimax principle in asymptotic statistical theory. In *Ecole d'Eté de Probabilités de Saint-Flour XI—1981*, pages 75–265. Springer.

Negri, I. (2001). On efficient estimation of invariant density for ergodic diffusion processes. *Statistics & probability letters*, **51**(1), 79–85.

Nickl, R. and Ray, K. (2018). Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. *to appear in Ann. Statist.; arXiv:1810.01702*.

Nickl, R. and Söhl, J. (2017). Bernstein–von Mises theorems for statistical inverse problems II: Compound Poisson processes. arXiv preprint. `arXiv:1709.07752`.

Nickl, R. and Söhl, J. (2017). Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. *Ann. Statist.*, **45**(4), 1664–1693.

Oksendal, B. (2003). *Stochastic differential equations: an introduction with applications*. Springer-Verlag Berlin Heidelberg.

Pokern, Y., Stuart, A. M., and van Zanten, J. H. (2013). Posterior consistency via precision operators for Bayesian nonparametric drift estimation in SDEs. *Stochastic Process. Appl.*, **123**(2), 603–628.

Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag Berlin Heidelberg.

Rogers, L. C. G. and Williams, D. (2000). *Diffusions, Markov processes and martingales: Volume 2, Itô calculus*, volume 2. Cambridge university press.

Söhl, J. and Trabs, M. (2016). Adaptive confidence bands for Markov chains and diffusions: estimating the invariant measure and the drift. *ESAIM Probab. Stat.*, **20**, 432–462.

Spokoiny, V. G. (2000). Adaptive drift estimation for nonparametric diffusion model. *Ann. Statist.*, **33**(3), 815–836.

Strauch, C. (2015). Sharp adaptive drift estimation for ergodic diffusions: the multivariate case. *Stochastic Process. Appl.*, **125**(7), 2562–2602.

Strauch, C. (2016). Exact adaptive pointwise drift estimation for multidimensional ergodic diffusions. *Probab. Theory Related Fields*, **164**(1-2), 361–400.

Talagrand, M. (1996). Majorizing measures: the generic chaining. *Ann. Probab.*, **24**(3), 1049–1103.

Talagrand, M. (1996b). New concentration inequalities in product spaces. *Inventiones mathematicae*, **126**(3), 505–563.

Talagrand, M. (2014). *Upper and lower bounds for stochastic processes*, volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete.* Springer, Heidelberg. Modern methods and classical problems.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

van der Meulen, F. and van Zanten, H. (2013). Consistent nonparametric Bayesian inference for discretely observed scalar diffusions. *Bernoulli*, **19**(1), 44–63.

van der Meulen, F. H., van der Vaart, A. W., and van Zanten, J. H. (2006). Convergence rates of posterior distributions for Brownian semimartingale models. *Bernoulli*, **12**(5), 863–888.

van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

van der Vaart, A. W. and van Zanten, H. (2005). Donsker theorems for diffusions: Necessary and sufficient conditions. *Ann. Probab.*, **33**(4), 1422–1451.

van der Vaart, A. W. and Wellner, J. W. (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer, New York.

van Waaij, J. and van Zanten, H. (2016). Gaussian process methods for one-dimensional diffusions: optimal rates and adaptation. *Electron. J. Stat.*, **10**s(1), 628–645.