

Weierstraß-Institut für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint

ISSN 0946 – 8633

Asymptotic statistical equivalence for ergodic diffusions

Arnak Dalalyan
Laboratoire de Probabilités,
Université Paris VI, Jussieu
75252 Paris Cedex 05, France
dalalyan@ccr.jussieu.fr

Markus Reiß
Weierstraß-Institut Berlin
Mohrenstraße 39
10117 Berlin, Germany
mreiss@wias-berlin.de

submitted: April 2, 2004

No. 916
Berlin 2004



1991 *Mathematics Subject Classification.* 62B15, 62C05, 62G20, 62M99.

Key words and phrases. Asymptotic equivalence, statistical experiment, Le Cam distance, ergodic diffusion, mixed Gaussian white noise.

The authors acknowledge the financial support provided through the European Community's Human Potential Programme under contract HPRN-CT-2000-00100, DYNSTOCH.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: <http://www.wias-berlin.de/>

Abstract

For scalar diffusion models with unknown drift function asymptotic equivalence in the sense of Le Cam's deficiency between statistical experiments is considered for long-time asymptotics. A local asymptotic equivalence result is established with an accompanying sequence of simple Gaussian shift experiments. Corresponding globally asymptotically equivalent experiments are obtained as compound experiments. The results are extended in several directions including time discretisation.

1 Introduction

Different statistical models often exhibit comparable features when they are considered under some natural asymptotics. In nonparametric statistics the problems of estimating a signal in Gaussian white noise, a regression function or a density of i.i.d. observations can all be handled by the same techniques, e.g. using kernel smoothers or projection methods, and the asymptotic minimax properties for the estimation risk usually coincide. The long standing experience that under an asymptotic point of view these models are statistically of the same kind has found its proper mathematical justification in 1996, when Brown and Low (1996) and Nussbaum (1996) proved the asymptotic equivalence of these models in the sense of Le Cam's theory of equivalent statistical experiments. In essence this means that any decision function developed for one model can be carried over, at least in an abstract way, to a decision function in the other models with exactly the same asymptotic risk properties. This is an important conceptual gain compared to the situation before where asymptotic results had to be proved each time separately.

In parametric statistics Le Cam's theory has been successfully applied to a huge variety of experiments because in this case it usually reduces to the property of local asymptotic normality (LAN) and its modifications (Le Cam and Yang 2000). The asymptotic equivalence for nonparametric experiments is conceptually more demanding and by now the class of models that are provably asymptotically equivalent to the three core models of signal detection, regression and density estimation is still limited. Grama and Nussbaum (1998) have proved asymptotic equivalence for generalised linear models, which has recently been extended to a wider nonparametric class in Jähnisch and Nussbaum (2003). Brown, Cai, Low, and Zhang (2002) consider specifically nonparametric regression with random design and provide a constructive asymptotic equivalence result. Certain asymptotic equivalence results have already been obtained for diffusion models. For asymptotically vanishing diffusion coefficients an equivalence result has been proved for diffusions observed on a suitable random time interval by Genon-Catalot, Laredo, and Nussbaum (2002), while Milstein and Nussbaum (1998) obtain asymptotic equivalence for such a diffusion model and its Euler discretisation. More closely related to our work is the study of a null-recurrent diffusion model with long-time asymptotics by Delattre and Hoffmann

(2002). The authors prove asymptotic equivalence to Gaussian models, which have the same structure as ours with the exception of an additional mixing random variable, that can be explained in analogy with the parametric LAMN-property. To overcome technical difficulties for proving analogous results for further nonparametric models, the concept of asymptotic equivalence is sometimes reduced to its weak form, see Drees (2001) for an application to lower bounds in extreme value theory.

One class of standard models in mathematical statistics is certainly given by nonlinear autoregressive processes of the form

$$X_{t+1} = f(X_t) + \varepsilon_{t+1}, \quad t = 0, \dots, T, \quad ((\varepsilon_t)_{t \in \mathbb{N}} \sim N(0, 1) \text{ i.i.d.})$$

and the corresponding continuous time diffusion models

$$dX_t = b(X_t) dt + dW_t, \quad t \in [0, T], \quad (W \text{ Brownian motion}) \quad (1.1)$$

with unknown drift functions f and b , respectively. Under ergodicity assumptions and for large T it is well known that the methodology developed for nonparametric regression can be used for inference on the drift function, for an overview see Taniguchi and Kakizawa (2000) for autoregressive processes and Kutoyants (2003), Fan (2003) for diffusions. In this paper we corroborate the folklore that 'autoregression is just regression' by showing strong asymptotic equivalence of the scalar diffusion model (1.1) with a signal detection or Gaussian shift model, which can be interpreted as a regression model with random design. Our result is established for the scalar diffusion model because we need to employ tools from stochastic analysis that are neither available for time series analysis nor for multidimensional diffusion processes.

Let us briefly introduce some basic notation such that we can announce the main results. For some fixed constants $C, A, \gamma > 0$ we consider the nonparametric drift class

$$\Sigma := \left\{ b \in \text{Lip}_{\text{loc}}(\mathbb{R}) \mid |b(x)| \leq C(1 + |x|), \forall |x| > A : b(x) \text{sgn}(x) \leq -\gamma \right\}, \quad (1.2)$$

where $\text{Lip}_{\text{loc}}(\mathbb{R})$ denotes the set of locally Lipschitz continuous functions $b : \mathbb{R} \rightarrow \mathbb{R}$ and $\text{sgn}(x) := x/|x|$. A standard result in the theory of stochastic differential equations asserts that for $b \in \Sigma$ and a Brownian motion W on some filtered probability space $(\Omega, \mathcal{A}, (\mathcal{A}_t)_{t \geq 0}, \mathbf{P})$ there exists to a given initial value a unique strong solution $(X(t), [0, T])$ of equation (1.1), e.g. Karatzas and Shreve (1991). Moreover, the existence of a stationary solution, unique in law, is ensured with invariant marginal density

$$\mu_b(x) = C_b \exp\left(2 \int_0^x b(y) dy\right), \quad x \in \mathbb{R}, \quad (1.3)$$

where $C_b > 0$ is a normalising constant. Considering in a first step drift functions b in a shrinking neighbourhood around $b_0 \in \Sigma$, we obtain the local asymptotic equivalence result for $T \rightarrow \infty$ of the stationary diffusion experiment given by (1.1) with the accompanying Gaussian shift experiments

$$dZ_x = b(x) \sqrt{\mu_{b_0}(x)} dx + T^{-1/2} dB_x, \quad x \in \mathbb{R}, \quad (1.4)$$

where B denotes a Brownian motion on the real line and μ_{b_0} is the invariant density from (1.3). The analogous regression experiment to (1.4) consists of observing the function b on

a design with density μ_{b_0} , which can be considered random or deterministic in the sense that it determines the distance between two design points. The main idea of the proof is to define a coupling of the original diffusion experiment with another diffusion-type experiment corresponding to a deterministic design. The implementation of this idea is heavily based on the local time of the diffusion process.

The local asymptotic equivalence result has already several implications for the statistical theory of diffusion processes. In particular, it can be used to obtain asymptotically sharp lower risk bounds. For instance, the lower bound of Theorem 1 in Dalalyan and Kutoyants (2002) follows immediately. In order to transfer also global results like upper risk bounds to the diffusion case, an equivalence result should be obtained for all parameters $b \in \Sigma$. We have to impose on the drift functions some minimal regularity larger than $1/2$ to obtain such a global result, cf. Brown and Zhang (1998). Furthermore, since the variance of the local time does not decay sufficiently rapidly, the general global equivalence result can only be established for drift functions that behave nicely far away from the origin. To avoid too much technicalities we therefore consider a global class $\Sigma_0 \subset \Sigma$ of drift functions of regularity larger than $1/2$ that coincide with some known function $b_0 \in \Sigma$ outside a compact interval I .

In absence of a variance stabilising transform the globally equivalent experiments will be of compound type. The first accompanying sequence is given by the observations

$$dX_t = b(X_t) dt + dW_t, \quad t \in [0, S], \quad (1.5)$$

$$dZ_x = b(x)\sqrt{\bar{\mu}_S(x)} dx + (T - S)^{-1/2} dB_x, \quad x \in I, \quad (1.6)$$

where W and B are independent Brownian motions, $\bar{\mu}_S$ is a suitable estimator of μ_b based on the observation $(X(t), t \in [0, S])$ and $S = S(T) \in (0, T)$ satisfies $\lim_{T \rightarrow \infty} (T - S(T))/S(T)^p = 0$ for all $p > 1$. The second accompanying sequence is given solely in terms of Gaussian experiments:

$$U_x = \sqrt{\mu_b(x)} + T^{-1/4}(\tilde{B}_x + \xi), \quad x \in I, \quad (1.7)$$

$$dZ_x = b(x)U_x dx + T^{-1/2} dB_x, \quad x \in I, \quad (1.8)$$

where the Brownian motions B and \tilde{B} and the random variable $\xi \sim N(0, 1)$ are independent.

The local and global equivalence results are derived in Section 2 and 3, respectively. In order to retain a clear presentation, a uniform variance bound on the local time and the construction of global estimators have been deferred to the Appendix. In Section 4 we discuss extensions of the theory developed so far. First, we consider the diffusion model with a general, but known diffusion coefficient σ , for which similar results are obtained. Next, we treat the case of discrete observations of the sample path in the diffusion model and the corresponding Euler discretisation. Conditions for asymptotic equivalence and sufficiency are given for sampling distances tending to zero. Finally, we present a Gaussian shift experiment

$$dZ_x = b(x)\sqrt{\mu_*(x)} dx + T^{-1/2} dB_x, \quad x \in \mathbb{R},$$

with a Brownian motion B and some function μ_* , which is globally less informative than the diffusion model (1.1) such that risk upper bounds obtained for this simple model immediately transfer to the diffusion case. In order to convey concisely the main ideas and to save space, the results in this section are stated in a more informal way.

2 Local equivalence

2.1 The general idea

We shall show that for drift functions b in a shrinking neighbourhood of the drift function b_0 the statistical experiment induced by observing the diffusion process

$$dX_t = b(X_t) dt + dW_t, \quad t \in [0, T], \quad (2.1)$$

is for $T \rightarrow \infty$ asymptotically equivalent to the statistical experiment induced by the observation

$$dZ_x = b(x) \sqrt{\mu_0(x)} dx + T^{-1/2} dB_x, \quad x \in \mathbb{R}, \quad (2.2)$$

where μ_0 is close to the density of the invariant measure of the diffusion process in (2.1) uniformly over the neighbourhood and B denotes a Brownian motion on the real axis.

The main idea of the proof is to perturb the diffusion model (2.1) in such a way that in each state $x \in \mathbb{R}$ the local time, that is the amount of time spent by the process in x until time T , is at least $T\mu_0(x)$ and to provide no information on $b(x)$ after the local time has attained the level $T\mu_0(x)$. At those states x , where the local time does not reach the level $T\mu_0(x)$, additional information on $b(x)$ is revealed. The model thus obtained can be considered as a regression model for b with fixed deterministic design of density $T\mu_0$. It is Gaussian and has the same likelihood process as the model in (2.2), which implies statistical equivalence of the associated statistical experiments.

The intuitive explanation why this approach succeeds is that the diffusion model, like an autoregressive time series model, exhibits two sources of randomness. Firstly, the design, that is how often the states are visited by the process, is random. Secondly, the drift b can merely be observed after contamination by white noise dW . As it turns out, the first source of randomness is less severe than the second and we do not lose too much information by assuming that each state x is visited up to time T with a density according to the approximate expectation $T\mu_0(x)$ of the local time. However, it is evident that this procedure can only work for neighbourhoods around b_0 that shrink with increasing T such that the true expectation $T\mu_b$ converges to $T\mu_0$ in a suitable manner.

2.2 Local experiments

2.1 Definition. For a drift $b_0 \in \Sigma$ and any density $\mu_0 \in L^1(\mathbb{R})$ we introduce their local neighbourhood with parameters $\varepsilon, \zeta, \eta > 0$

$$\Sigma_{\varepsilon, \eta, \zeta}(b_0, \mu_0) := \left\{ b \in \Sigma \mid \left(\int (b - b_0)^2(y) \sqrt{\mu_b(y)} dy \right)^{1/2} \leq \varepsilon, \right. \\ \left. \left(\int (b - b_0)^2(y) |\mu_b(y) - \mu_0(y)| dy \right)^{1/2} \leq \eta, \int |\mu_b(y) - \mu_0(y)| dy \leq \zeta \right\}.$$

Here μ_b denotes the invariant density of the diffusion process with drift b .

2.2 Remark. It is natural to consider neighbourhoods around (b_0, μ_{b_0}) , but it is by no means necessary for the calculations to enforce $\mu_0 = \mu_{b_0}$. For the globalisation the more general approach has the advantage of permitting the usage of separate estimators for functions b and μ_b .

We now define precisely the local experiments \mathbb{E}_1 and \mathbb{F}_1 , for which we shall prove asymptotic equivalence. Note that we define the Gaussian shift experiment on the space $\mathbb{R}^{L^2(\mathbb{R})}$ and not on $C(\mathbb{R})$ via the natural interpretation of the differentials as integrators for $L^2(\mathbb{R})$ -functions. Of course, the law is already characterised by the integration of the functions $\mathbb{1}_{[0,y]}$, $y \in \mathbb{R}$, which corresponds to the signal in white noise interpretation on the space $C(\mathbb{R})$ up to the knowledge of the value at zero.

2.3 Definition. *We define the diffusion experiment localised around (b_0, μ_0)*

$$\mathbb{E}_1 := \mathbb{E}_1(b_0, \mu_0, T, \varepsilon, \eta, \zeta) := (C([0, T]), \mathcal{B}_{C([0, T])}, (\mathbf{P}_b^T)_{b \in \Sigma_{\varepsilon, \eta, \zeta}(b_0, \mu_0)}),$$

where \mathbf{P}_b^T denotes the law of the stationary diffusion process with drift b on the canonical space $C([0, T])$.

The Gaussian shift experiment localised around (b_0, μ_0) is given by

$$\mathbb{F}_1 := \mathbb{F}_1(b_0, \mu_0, T, \varepsilon, \eta, \zeta) := (\mathbb{R}^{L^2(\mathbb{R})}, \mathcal{B}_{\mathbb{R}^{L^2(\mathbb{R})}}, (\mathbf{Q}_b^T)_{b \in \Sigma_{\varepsilon, \eta, \zeta}(b_0, \mu_0)}),$$

where \mathbf{Q}_b^T denotes the law of the Gaussian shift experiment

$$\begin{aligned} dZ_x &= b(x)\sqrt{\mu_0(x)} dx + T^{-1/2} dB_x, \quad x \in \mathbb{R}, \\ \left(\text{i.e., } \left(\int f(x)b(x)\sqrt{\mu_0(x)} dx + T^{-1/2} \int f(x) dB_x, f \in L^2(\mathbb{R}) \right) \right) \end{aligned}$$

with a Brownian motion B on the real line.

In order to pursue our procedure of changing the design appropriately, we need to introduce the so-called local time of a diffusion process X . We refer to Revuz and Yor (1999), Chapter VI for the details. We are going to use that the local time $L_t^y(X)$ of the diffusion process X at the point $y \in \mathbb{R}$ up to time $t \geq 0$ can be constructed such that $(L_t^y, y \in \mathbb{R}, t \geq 0)$ is a process which is continuous in t and càdlàg in y (Revuz and Yor 1999, Theorem VI.1.7). Henceforth we shall work with this process, which satisfies

$$L_t^y(X) = \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} \int_0^t \mathbb{1}_{[y, y+\varepsilon)}(X_s) ds, \quad \mathbf{P}_b\text{-a.s.}$$

By assuming the usual conditions of the filtration $(\mathcal{A}_t)_{t \geq 0}$, we can suppose that $L_t^y(X)$ is \mathcal{A}_t -measurable for $(\mathcal{A}_t)_{t \geq 0}$ -adapted processes X . The main property we need is the following extended occupation time formula (Revuz and Yor 1999, Ex. VI.1.15):

$$\int_0^T f(t, X_t, \omega) dt = \int \left(\int_0^T f(t, y, \omega) d_t L_t^y(X) \right) dy, \quad \mathbf{P}_b\text{-a.s.} \quad (2.3)$$

where $f : \mathbb{R}_+ \times \mathbb{R} \times \Omega \rightarrow \mathbb{R}_+$ is any measurable function and $d_t L_t^y(X)$ denotes integration with respect to the increasing integrator $t \mapsto L_t^y(X)$.

We can now introduce the local experiment \mathbb{F}_2 for which we shall show asymptotic equivalence with \mathbb{E}_1 . We briefly recall the conditions guaranteeing the existence of a weak solution of a stochastic differential equation with a functional form of the drift.

2.4 Proposition. *Consider the stochastic differential equation*

$$dX_t = b(X, t) dt + dW_t, \quad t \in [0, T],$$

with a progressively measurable functional $b : C(\mathbb{R}^+) \times \mathbb{R}^+ \rightarrow \mathbb{R}$. Then a weak solution with some prescribed initial distribution μ exists if $|b(f, t)| \leq K(1 + \|f\|_\infty)$ holds with a suitable constant $K > 0$ for all $f \in C(\mathbb{R}^+)$ and $t \in [0, T]$. The law of the solution is obtained by a change of the Wiener measure on $(C([0, T]), \mathcal{B}_{C([0, T])})$ with initial distribution μ using the Girsanov density

$$Z_T(X) = \exp\left(\int_0^T b(X, s) dX_s - \frac{1}{2} \int_0^T b(X, s)^2 ds\right).$$

Proof. This is the generalisation of Proposition 5.3.6 given in Remark 5.3.8 of Karatzas and Shreve (1991). \square

2.5 Remark. *Under suitable ergodicity assumptions, the linear growth condition on the drift can certainly be dropped and a corresponding uniqueness result will probably hold, but we do not want to deviate further into that direction. We just assume linear growth in the definition of Σ and work with the solution defined in terms of the Girsanov density.*

2.6 Definition. *We define the local experiment \mathbb{F}_2 by*

$$\mathbb{F}_2 := \mathbb{F}_2(b_0, \mu_0, T, \varepsilon, \eta, \zeta) := (C([0, T]) \times \mathbb{R}^{L^2(\mathbb{R})}, \mathcal{B}_{C([0, T])} \otimes \mathcal{B}_{\mathbb{R}}^{\otimes L^2(\mathbb{R})}, (\tilde{\mathbf{Q}}_b^T)_{b \in \Sigma_{\varepsilon, \eta, \zeta}(b_0, \mu_0)}),$$

where $\tilde{\mathbf{Q}}_b^T$ is uniquely defined by

$$\tilde{\mathbf{Q}}_b^T(A \times B) := \int_A K_b^T(f, B) \mathbf{R}_b^T(df), \quad A \in \mathcal{B}_{C([0, T])}, B \in \mathcal{B}_{\mathbb{R}}^{\otimes L^2(\mathbb{R})}.$$

Here, \mathbf{R}_b^T denotes the law of the weak solution Y of the stochastic differential equation

$$dY_t = (b(Y_t) \mathbb{1}_{\{L_t^{Y_t}(Y) \leq T\mu_0(Y_t)\}} + b_0(Y_t) \mathbb{1}_{\{L_t^{Y_t}(Y) > T\mu_0(Y_t)\}}) dt + dW_t, \quad t \in [0, T],$$

on the canonical space $C([0, T])$ with initial distribution $Y_0 \sim \mu_0$, given by Proposition 2.4. The probability $K_b^T(f, \bullet)$ is the law of the Gaussian shift experiment

$$dV_x^f = b(x)(T\mu_0(x) - L_T^x(f))_+^{1/2} dt + dB_x, \quad x \in \mathbb{R},$$

where B denotes a two-sided Brownian motion on \mathbb{R} independent of W and Y_0 and $(A)_+ := \max(A, 0)$.

2.7 Remark. *In the preceding definition we have to choose a measurable version of the mapping $(f, x) \mapsto L_T^x(f)$ on a set of functions f with \mathbf{R}_b^T -probability one in order to have the Markov kernel property of K_b^T . This is certainly possible since by the equivalence of \mathbf{R}_b^T with the Wiener measure this property is satisfied when using a càdlàg-version in x of the Brownian motion local time $L_T^x(W)$ as discussed earlier.*

Finally, we need to introduce yet another experiment which is constructed so as to be equivalent to \mathbb{E}_1 , but to be defined on the same space $C([0, T]) \times \mathbb{R}^{L^2(\mathbb{R})}$ like \mathbb{F}_2 .

2.8 Definition. We define the local experiment \mathbb{E}_2 by

$$\mathbb{E}_2 := \mathbb{E}_2(b_0, \mu_0, T, \varepsilon, \eta, \zeta) := (C([0, T]) \times \mathbb{R}^{L^2(\mathbb{R})}, \mathcal{B}_{C([0, T])} \otimes \mathcal{B}_{\mathbb{R}^{\otimes L^2(\mathbb{R})}}, (\tilde{\mathbf{P}}_b^T)_{b \in \Sigma_{\varepsilon, \eta, \zeta}(b_0, \mu_0)}),$$

where $\tilde{\mathbf{P}}_b^T$ is uniquely defined by

$$\tilde{\mathbf{P}}_b^T(A \times B) := \int_A K_{b_0}^T(f, B) \mathbf{P}_b^T(df), \quad A \in \mathcal{B}_{C([0, T])}, B \in \mathcal{B}_{\mathbb{R}^{\otimes L^2(\mathbb{R})}},$$

with the same notation as for \mathbb{F}_2 .

2.3 Likelihood ratio and equivalent experiments

In the sequel we shall often use the likelihood ratio or Radon-Nikodym derivative for the laws of diffusion-type processes on the space $C([0, T])$. The next theorem is an adaptation of Theorem 7.7 in Liptser and Shiryaev (2001) to our purposes, see also Theorem IV.4.23 in Jacod and Shiryaev (2003).

2.9 Theorem. Suppose $(X_t^{(1)}, t \in [0, T])$ and $(X_t^{(2)}, t \in [0, T])$ are scalar diffusion-type processes that satisfy

$$dX_t^{(i)} = \alpha^{(i)}(X^{(i)}, t) dt + dW_t, \quad t \in [0, T], \quad X_0^{(i)} = \xi^{(i)}, \quad i = 1, 2,$$

with progressively measurable functionals $\alpha^{(i)} : C([0, T]) \times \mathbb{R}^+ \rightarrow \mathbb{R}$ and with a standard Wiener process W .

Then these processes have mutually absolutely continuous distributions $\mathbf{P}_{X^{(i)}}$ on the canonical space $(C([0, T]), \mathcal{B}_{C([0, T])})$ if for $i = 1, 2$

$$\mathbf{P}\left(\int_0^T (\alpha_t^{(i)}(X^{(i)})^2 dt < \infty)\right) = 1 \quad \text{and} \quad \mathbf{P}\left(\int_0^T (\alpha_t^{(1)} - \alpha_t^{(2)})^2 (X^{(i)}) dt < \infty\right) = 1,$$

$\xi^{(1)}$ and $\xi^{(2)}$ are independent of W and have mutually absolutely continuous distributions on \mathbb{R} .

In this case the likelihood ratio $\Lambda_T(X^{(1)}, X^{(2)})(X) = \frac{d\mathbf{P}_{X^{(1)}}}{d\mathbf{P}_{X^{(2)}}}(X)$ is given by

$$\frac{d\mathbf{P}_{\xi^{(1)}}}{d\mathbf{P}_{\xi^{(2)}}}(X_0) \exp\left(\int_0^T (\alpha_t^{(1)} - \alpha_t^{(2)})(X) dX_t - \frac{1}{2} \int_0^T (\alpha_t^{(1)}(X)^2 - \alpha_t^{(2)}(X)^2) dt\right),$$

which under $\mathbf{P}_{X^{(2)}}$ is in law equal to

$$\frac{d\mathbf{P}_{\xi^{(1)}}}{d\mathbf{P}_{\xi^{(2)}}}(\xi^{(2)}) \exp\left(\int_0^T (\alpha_t^{(1)} - \alpha_t^{(2)})(X^{(2)}) dW_t - \frac{1}{2} \int_0^T (\alpha_t^{(1)} - \alpha_t^{(2)})^2 (X^{(2)}) dt\right).$$

2.10 Remark. This representation of the likelihood gives another indication why our limiting experiment \mathbb{F}_1 is natural for the diffusion experiment \mathbb{E}_1 : the Fisher information at b_0 in functional directions h and h' is for $T \rightarrow \infty$ of order

$$I_{h, h'}(b_0) = \mathbf{E}_{b_0} \left[\int_0^T h(X_t) h'(X_t) dt \right] + o(T) = T \int h(x) h'(x) \mu_0(x) dx + o(T).$$

From the definition of the Kullback-Leibler divergence (or relative entropy, denoted by KL) the following result is immediate, compare also with the expression for the Hellinger distance (Jacod and Shiryaev 2003, Theorem IV.4.23).

2.11 Corollary. *Under the conditions of Theorem 2.9 the Kullback-Leibler divergence between the laws of $X^{(1)}$ and $X^{(2)}$ is given by*

$$\text{KL}(\mathbf{P}_{X^{(1)}}, \mathbf{P}_{X^{(2)}}) = -\mathbf{E} \left[\log \left(\frac{d\mathbf{P}_{\xi^{(1)}}}{d\mathbf{P}_{\xi^{(2)}}}(\xi^{(2)}) \right) \right] + \frac{1}{2} \mathbf{E} \left[\int_0^T (\alpha_t^{(1)} - \alpha_t^{(2)})^2(X^{(2)}) dt \right].$$

With these tools at hand we obtain the first equivalency results. We only need to know that two general dominated experiments $\mathbb{G}_1 = (\Omega, \mathcal{A}, (\mathbf{P}_\theta)_{\theta \in \Theta})$ and $\mathbb{G}_2 = (\Omega', \mathcal{A}', (\mathbf{Q}_\theta)_{\theta \in \Theta})$ are statistically equivalent iff the laws of the likelihood processes under the dominating measures \mathbf{P}_0 and \mathbf{Q}_0 coincide (Strasser 1985, Cor. 25.9):

$$\mathcal{L} \left(\left(\frac{d\mathbf{P}_\sigma}{d\mathbf{P}_0} \right)_{\sigma \in \Theta} \mid \mathbf{P}_0 \right) = \mathcal{L} \left(\left(\frac{d\mathbf{Q}_\sigma}{d\mathbf{Q}_0} \right)_{\sigma \in \Theta} \mid \mathbf{Q}_0 \right).$$

2.12 Proposition. *The statistical experiments \mathbb{E}_1 and \mathbb{E}_2 are equivalent.*

Proof. By Theorem 2.9 the measures \mathbf{P}_b^T and $\mathbf{P}_{b_0}^T$ are equivalent for all $b, b_0 \in \Sigma$ such that the likelihood process for \mathbb{E}_1 is well defined. Moreover, in experiment \mathbb{E}_2 we use the kernel $K_{b_0}^T$ which is independent of b such that the Radon-Nikodym derivative

$$\frac{d\tilde{\mathbf{P}}_b^T}{d\tilde{\mathbf{P}}_{b_0}^T}(X, V) = \frac{K_{b_0}^T(X, dV) \mathbf{P}_b(dX)}{K_{b_0}^T(X, dV) \mathbf{P}_{b_0}(dX)} = \frac{d\mathbf{P}_b^T}{d\mathbf{P}_{b_0}^T}(X)$$

depends only on the first coordinate. Consequently, the likelihood processes coincide. \square

2.13 Proposition. *The statistical experiments \mathbb{F}_1 and \mathbb{F}_2 are equivalent.*

Proof. Let us determine the likelihood process for \mathbb{F}_2 under the dominating measure $\tilde{\mathbf{Q}}_{b_0}^T$. We first note that

$$\log \left(\frac{d\tilde{\mathbf{Q}}_b^T}{d\tilde{\mathbf{Q}}_{b_0}^T}(Y, V) \right) = \log \left(\frac{K_b^T(Y, dV) \mathbf{R}_b^T(dY)}{K_{b_0}^T(Y, dV) \mathbf{R}_{b_0}^T(dY)} \right) = \log \left(\frac{dK_b^T(Y, \bullet)}{dK_{b_0}^T(Y, \bullet)}(V) \right) + \log \left(\frac{d\mathbf{R}_b^T}{d\mathbf{R}_{b_0}^T}(Y) \right)$$

holds. Both log-likelihood functions consist of a stochastic integral with respect to a Brownian motion and its quadratic variation term under the dominating measure $\tilde{\mathbf{Q}}_{b_0}^T$. Let us calculate up to sets of probability zero the quadratic variation term in the log-likelihood $\log \left(\frac{d\mathbf{R}_b^T}{d\mathbf{R}_{b_0}^T} \right)$ given in Theorem 2.9 using the occupation time formula (2.3):

$$\begin{aligned} & \int_0^T \left(b(Y_t) \mathbb{1}_{\{L_t^{Y_t}(Y) \leq T\mu_0(Y_t)\}} + b_0(Y_t) \mathbb{1}_{\{L_t^{Y_t}(Y) > T\mu_0(Y_t)\}} - b_0(Y_t) \right)^2 dt \\ &= \int_0^T (b - b_0)^2(Y_t) \mathbb{1}_{\{L_t^{Y_t}(Y) \leq T\mu_0(Y_t)\}} dt \\ &= \int \int_0^T (b - b_0)^2(y) \mathbb{1}_{\{L_t^y(Y) \leq T\mu_0(y)\}} d_t L_t^y(Y) dy \\ &= \int (b - b_0)^2(y) \min(T\mu_0(y), L_T^y(Y)) dy. \end{aligned}$$

Similarly, the quadratic variation term in $\log\left(\frac{dK_b^T(Y, \bullet)}{dK_{b_0}^T(Y, \bullet)}\right)$ is given by

$$\int (b - b_0)^2(x) (T\mu_0(x) - L_T^x(Y))_+ dx.$$

Putting the two identities together, we have proved that the quadratic variation term in $\log\left(\frac{d\tilde{Q}_b^T}{d\tilde{Q}_{b_0}^T}\right)$ equals $\int (b - b_0)^2(y) T\mu_0(y) dy$ and is thus deterministic.

The preceding calculations remain valid when b is replaced by $b_0 + \lambda(b - b_0)$ for any $\lambda \in \mathbb{R}$. Hence using $\mathbf{E}_{b_0}\left[\frac{d\tilde{Q}_b}{d\tilde{Q}_{b_0}}\right] = 1$, we conclude that, under \tilde{Q}_{b_0} , the Laplace transform $\mathbf{E}_{b_0}[\exp(\lambda M(b - b_0))]$ of the stochastic integral term

$$M(h) := \int_0^T h(Y_t) \mathbb{1}_{\{L_t^{Y_t}(Y) \leq T\mu_0(Y_t)\}} dW_t + \int h(x) (T\mu_0(x) - L_T^x(Y))_+^{1/2} dB_x, \quad h \in L^2(\mathbb{R}),$$

is equal $\exp\left(\frac{\lambda^2}{2} \int (b - b_0)^2(y) T\mu_0(y) dy\right)$. Therefore the random variable $M(b - b_0)$ is Gaussian with variance $T \int (b - b_0)^2 \mu_0$. The covariation between two such stochastic integrals with b replaced by b_1 and b_2 , respectively, is by the occupation time formula again $T \int (b_1 - b_0)(b_2 - b_0) \mu_0$. Moreover, using the Cramér-Wold device it then follows that the random process $(M(b - b_0), b \in \Sigma_{\varepsilon, \eta}(b_0))$ under $\tilde{Q}_{b_0}^T$ is Gaussian with zero mean.

Since the likelihood process of the Gaussian shift experiment \mathbb{F}_1 under $Q_{b_0}^T$ is given by

$$\left(\exp\left(T \int (b - b_0)(x) \sqrt{\mu_0(x)} dB_x - \frac{T^2}{2} \int (b - b_0)^2(x) \mu_0(x) dx\right), b \in \Sigma_{\varepsilon, \eta}(b_0) \right),$$

the laws of the two likelihood processes coincide and the experiments are equivalent. \square

2.14 Remark. *The main idea in the preceding proof was to show that the likelihood ratios associated to the experiments \mathbb{E}_2 and \mathbb{F}_2 (as random processes indexed by b) have the same law. At the first look it seems that the experiment generated by the Itô process*

$$\begin{aligned} dY_t &= (b(Y_t) \mathbb{1}_{\{t \leq \tau_b(Y)\}} + b_0(Y_t) \mathbb{1}_{\{t > \tau_b(Y)\}}) dt + dW_t, & t \in [0, T], \\ dY_t &= \left(T \int (b - b_0)^2(y) \mu_0(y) dy - \int_0^T (b - b_0)^2(Y_t) dt \right)_+^{1/2} dt + dW_t, & t \in (T, T + 1), \\ \tau_b(Y) &= \inf \left\{ t \leq T : \int_0^t (b - b_0)^2(Y_s) ds > T \int (b - b_0)^2(y) \mu_0(y) dy \right\}, \\ Y_0 &= \xi_0, \end{aligned}$$

satisfies the same property and therefore can be used instead of \mathbb{F}_2 . Unfortunately, this assertion is false: for a fixed value of b the loglikelihood of the process Y is a Gaussian random variable, but if we consider the same loglikelihood as a process indexed by b , it is no longer Gaussian.

2.15 Remark. *For future reference we list further experiments that are equivalent to \mathbb{F}_1 and \mathbb{F}_2 for parameters $b \in \Sigma_{\varepsilon, \eta}(b_0)$:*

$$\begin{aligned} dY_x &= b(x) dx + T^{-1/2} \mu_0(x)^{-1/2} dB_x, & x \in \mathbb{R}, \\ dY_x &= (b(x) - b_0(x)) \sqrt{\mu_0(x)} dx + T^{-1/2} dB_x, & x \in \mathbb{R}, \\ dY_x &= b(F_{\mu_0}^{-1}(x)) dx + T^{-1/2} dB_x, & x \in (0, 1), \end{aligned}$$

where $F_{\mu}(x) = \int_{-\infty}^x \mu(y) dy$ and dB is Gaussian white noise. For the proof it suffices to check that the laws of the likelihood processes coincide.

2.4 Asymptotic equivalence

By Corollary 59.6 in Strasser (1985) the Le Cam distance Δ between experiments defined on the same measurable space can be estimated by a uniform bound on the total variation distance between the corresponding probability measures. An application of this coupling technique allows to prove the main theorem on local asymptotic equivalence.

2.16 Theorem. *If for $T \rightarrow \infty$ the asymptotics $\varepsilon(T) = o(T^{-1/4})$, $\eta(T) = o(T^{-1/2})$ and $\zeta(T) = o(1)$ hold, then the following convergence holds true uniformly over all $b_0 \in \Sigma$:*

$$\lim_{T \rightarrow \infty} \Delta \left(\mathbb{E}_1(b_0, \mu_0, T, \varepsilon(T), \eta(T), \zeta(T)), \mathbb{F}_1(b_0, \mu_0, T, \varepsilon(T), \eta(T), \zeta(T)) \right) = 0.$$

Proof. By Propositions 2.12 and 2.13 it suffices to prove the asymptotic equivalence for the experiments \mathbb{E}_2 and \mathbb{F}_2 . Their families of measures $(\tilde{\mathbf{P}}_b^T)$ and $(\tilde{\mathbf{Q}}_b^T)$ are defined on the same measurable space $(C([0, T]) \times \mathbb{R}^{L^2(\mathbb{R})}, \mathcal{B}_{C([0, T])} \otimes \mathcal{B}_{\mathbb{R}^{\otimes L^2(\mathbb{R})}})$. We infer (with short-hand notation)

$$\Delta(\mathbb{E}_1, \mathbb{F}_1) = \Delta(\mathbb{E}_2, \mathbb{F}_2) \leq \sup_{b \in \Sigma_{\varepsilon, \eta, \zeta}(b_0, \mu_0)} \|\tilde{\mathbf{P}}_b^T - \tilde{\mathbf{Q}}_b^T\|_{TV},$$

$\|\cdot\|_{TV}$ denoting the total variation norm. Since the measures $(\tilde{\mathbf{P}}_b^T)$ and $(\tilde{\mathbf{Q}}_b^T)$ correspond to diffusion-type processes with different initial distributions, we use the representations (Kallenberg 2002, Theorem 18.10)

$$\tilde{\mathbf{P}}_b^T = \int \tilde{\mathbf{P}}_{b,x}^T \mu_b(x) dx \text{ and } \tilde{\mathbf{Q}}_b^T = \int \tilde{\mathbf{Q}}_{b,x}^T \mu_b(x) dx$$

with the corresponding laws for deterministic initial values $x \in \mathbb{R}$ and infer by the triangle inequality

$$\|\tilde{\mathbf{P}}_b^T - \tilde{\mathbf{Q}}_b^T\|_{TV} \leq \|\mu_b - \mu_0\|_{L^1(\mathbb{R})} + \int \|\tilde{\mathbf{P}}_{b,x}^T - \tilde{\mathbf{Q}}_{b,x}^T\|_{TV} \mu_b(x) dx.$$

Because of $\zeta(T) \rightarrow 0$ the first term tends to zero uniformly.

Since the square of the total variation is bounded by two times the Kullback-Leibler divergence (Deuschel and Stroock 1989, Eq. (3.2.25)), it suffices for the second term to prove that $\int \text{KL}(\tilde{\mathbf{Q}}_{b,x}^T, \tilde{\mathbf{P}}_{b,x}^T) \mu_b(x) dx$ tends to zero uniformly. By Corollary 2.11 this expression equals up to the factor 1/2

$$\begin{aligned} & \int \mathbf{E}_{b,x} \left[\int_0^T (b - b_0)^2(Y_t) \mathbb{1}_{\{L_t^{Y_t}(Y) > T\mu_0(Y_t)\}} + \int (b - b_0)^2(y) (T\mu_0(y) - L_T^y(Y))_+ dy \right] \mu_b(x) dx \\ &= \mathbf{E}_b \left[\int (b - b_0)^2(y) ((L_T^y(Y) - T\mu_0(y))_+ + (T\mu_0(y) - L_T^y(Y))_+) dy \right] \\ &= \int (b - b_0)^2(y) \mathbf{E}_b[|L_T^y(Y) - T\mu_0(y)|] dy. \end{aligned}$$

Since we are in the stationary case, a bias-variance decomposition yields in combination with Proposition 5.1 from Appendix

$$\mathbf{E}_b[|L_T^y(Y) - T\mu_0(y)|] \leq T|\mu_b(y) - \mu_0(y)| + (CT\mu_b(y))^{1/2}.$$

Hence, we obtain the uniform convergence result over $\Sigma_{\varepsilon(T),\eta(T),\zeta(T)}(b_0, \mu_0)$

$$\begin{aligned} \mathbf{E}_b \left[\int (b_0 - b)^2(y) |L_T^y(Y) - T\mu_0(y)| dy \right] \\ \leq \int (b_0 - b)^2(y) \left(T|\mu_b(y) - \mu_0(y)| + \sqrt{CT\mu_b(y)} \right) dy \\ \leq T\eta^2(T) + \sqrt{CT}T^{1/2}\varepsilon^2(T) \xrightarrow{T \rightarrow \infty} 0, \end{aligned}$$

which proves the assertion. \square

2.17 Corollary. *The preceding asymptotic equivalence result holds in particular for the local parameter subclass*

$$\tilde{\Sigma}_{\varepsilon,T}(b_0, \mu_0) := \left\{ b \in \Sigma \mid \left(\int (b - b_0)^2(y) \sqrt{\mu_b(y)} dy \right)^{1/2} \leq \varepsilon, \|\mu_b^{1/2} - \mu_0\mu_b^{-1/2}\|_\infty \leq T^{-1/2} \right\},$$

when $\varepsilon = \varepsilon(T) = o(T^{-1/4})$ for $T \rightarrow \infty$.

Proof. Just note that for any $b \in \tilde{\Sigma}_{\varepsilon,T}(b_0, \mu_0)$

$$\int (b - b_0)^2(y) |\mu_b(y) - \mu_0(y)| dy \leq \|\mu_b^{1/2} - \mu_0\mu_b^{-1/2}\|_\infty \int (b - b_0)^2(y) \sqrt{\mu_b(y)} dy \leq \varepsilon^2(T)T^{-1/2}$$

holds and equally $\int |\mu_b - \mu_0| \leq T^{-1/2} \int \mu_b^{1/2} \lesssim T^{-1/2}$ follows uniformly over b by the uniform exponential decay of μ_b . Therefore $\tilde{\Sigma}_{\varepsilon(T),T}(b_0, \mu_0) \subset \Sigma_{\varepsilon(T),\eta(T),\zeta(T)}(b_0, \mu_0)$ follows with $\eta(T) = \varepsilon(T)T^{-1/4} = o(T^{-1/2})$ and $\zeta(T) = O(T^{-1/2}) = o(1)$. \square

2.18 Proposition. *The statistical experiment $\mathbb{F}_1(b_0, \mu_0, T, \varepsilon, \eta, \zeta)$ is for $\eta = o(T^{-1/2})$ and arbitrary $\varepsilon, \zeta > 0$ asymptotically equivalent to the experiment induced by observing*

$$dY_x = (b(x) - b_0(x))\sqrt{\mu_b(x)} dx + T^{-1/2}dB_x, \quad x \in \mathbb{R},$$

where μ_b is the invariant density corresponding to b , dB is Gaussian white noise on $L^2(\mathbb{R})$ and the parameters b belong to the the same neighbourhood $\Sigma_{\varepsilon,\eta,\zeta}(b_0, \mu_0)$.

Proof. Since the two concerned experiments are defined on the same space, the result follows if we show that the Kullback-Leibler divergence between the likelihood ratios tends to zero. This divergence is given by $\frac{T}{2} \int (b(x) - b_0(x))^2 (\sqrt{\mu_b(x)} - \sqrt{\mu_0(x)})^2 dx$. Using the general inequality $(A - B)^2 \leq |A^2 - B^2|$ for $A, B > 0$, the condition on η yields the result. \square

3 Globalisation

3.1 Main result

A common way of globalising a local equivalence result makes use of the variance stabilising transformation (see Grama and Nussbaum (1998) for the exact definition). In our case this amounts to seeking a functional \mathcal{T} whose differential $D\mathcal{T}(b)[h]$ at the point $b = b_0$

is equal to $\sqrt{\mu_{b_0}}h$. Indeed, for such a functional the Kullback-Leibler divergence between the laws of the Gaussian random measures $dZ_x = (b(x) - b_0(x))\sqrt{\mu_{b_0}(x)}dx + T^{-1/2}dB_x$, $x \in \mathbb{R}$ and $d\bar{Z}_x = (\mathcal{T}(b)(x) - \mathcal{T}(b_0)(x))dx + T^{-1/2}dB_x$, $x \in \mathbb{R}$ is equal to

$$\frac{T}{2} \int_{\mathbb{R}} (\mathcal{T}(b)(x) - \mathcal{T}(b_0)(x) - D\mathcal{T}(b_0)[b - b_0](x))^2 dx$$

and, at a heuristic level, tends to zero if the functional \mathcal{T} is sufficiently regular. This yields the asymptotic equivalence of the two Gaussian shift experiments corresponding to Z and \bar{Z} . Furthermore, it permits to infer the asymptotic equivalence of the experiments characterised by the observations $dZ_x = b(x)\sqrt{\mu_{b_0}(x)}dx + T^{-1/2}dB_x$, $x \in \mathbb{R}$ and $dZ_x = \mathcal{T}(b)(x)dx + T^{-1/2}dB_x$, $x \in \mathbb{R}$, the latter being independent of b_0 .

Unfortunately, following Delattre and Hoffmann (2002) we can show that such a transformation does not exist. Indeed, let us consider the simple case when b is unknown only on a compact interval I . Then the differential of the operator $\mathcal{S} : L^2(I) \rightarrow L^2(I)$, $\mathcal{S}(b) = \sqrt{\mu_b}$ at the point $h \in L^2(I)$ is obviously given by

$$D\mathcal{S}(b)[h] = \lim_{\varepsilon \rightarrow 0} \frac{\sqrt{\mu_{b+\varepsilon h}} - \sqrt{\mu_b}}{\varepsilon},$$

where the convergence is understood in the mean square sense. We find

$$D\mathcal{S}(b)[h](x) = \sqrt{\mu_b(x)} \int_I h(y) [F_b(y) - \mathbb{1}_{\{y \geq x\}}] dy, \quad x \in I,$$

where F_b is the distribution function corresponding to the invariant density μ_b . Therefore, the equality $D\mathcal{T}(b)[h_1] = h_1\sqrt{\mu_b} = h_1\mathcal{S}(b)$ would imply that \mathcal{T} is twice continuously differentiable and $D(D\mathcal{T}(b)[h_1])[h_2] = D(D\mathcal{T}(b)[h_2])[h_1]$ for any $h_1, h_2 \in L^2(I)$. This last equality can be rewritten in the form $h_1 D\mathcal{S}(b)[h_2] = h_2 D\mathcal{S}(b)[h_1]$, $h_1, h_2 \in L^2(I)$, which is evidently not true. This contradiction results essentially from the nonlocal character of the mapping \mathcal{S} . This indicates why the global asymptotic equivalence with a Gaussian shift experiment of the form $dZ_x = \mathcal{T}(b)(x)dx + T^{-1/2}dB_x$ might be impossible to establish.

Nevertheless, we give below an equivalence result which is global and involves a mixed Gaussian white noise experiment. The main idea is to replace in the Gaussian shift experiment $dZ_x = (b(x) - b_0(x))\sqrt{\mu_b(x)}dx + T^{-1/2}dB_x$ the invariant density μ_b by a random approximation, which is independent of B and has the advantage of being observable.

3.1 Definition. *The parameter class $\Sigma_0 = \Sigma_0(\beta, L, m, b_0, I)$ consists of drift functions $b \in \Sigma$ satisfying*

$$b(x) = b_0(x), \quad \forall x \notin I; \quad |b(x) - b(y)| \leq L|x - y|^\beta, \quad \forall x, y \in I,$$

where $I = [-D, D] \subset \mathbb{R}$ is a compact interval, b_0 is a fixed known function and $\beta \in (0, 1)$.

3.2 Remark. *Let us briefly explain why we restrict to the case when $b(x)$ is known for $x \in \mathbb{R} \setminus I$. Since the variance under \mathbf{P}_b^T of the local time $L_T^x - L_S^x$ is of order $(T - S)\mu_b(x)$, condition (3.2) requires the existence of an estimator $\bar{b}_{S(T)}$ such that*

$$\lim_{T \rightarrow \infty} \int_{\mathbb{R}} \mathbf{E}_b \left[(b(x) - \bar{b}_{S(T)}(x))^2 \right] \sqrt{(T - S(T))\mu_b(x)} dx = 0 \quad (3.1)$$

uniformly in b . Standard arguments yield that the Mean Squared Error (MSE) of estimating $b(x)$ by a kernel method with bandwidth h is of order $h^{2\beta} + (S(T)h\mu_b(x))^{-1}$. Therefore the optimal choice of a bandwidth is $h = (S(T)\mu_b(x))^{-1/(2\beta+1)}$. Even for this oracle choice of h the MSE is of order $(S(T)\mu_b(x))^{-2\beta/(2\beta+1)}$ and the integral (3.1) is not finite for $\beta > 1/2$. Fundamentally, this obstruction is due to the relatively slow decay of the variance of local time compared to its expectation: $\lim_{|y| \rightarrow \infty} \sqrt{\text{Var}[L_T^y(Y)]} / \mathbf{E}[L_T^y(Y)] = \infty$.

Note that although the functions in Σ are locally Lipschitz continuous, the Hölder restriction of order β is of different nature: it is uniform over $x \in I$ and over $b \in \Sigma_0$.

3.3 Definition. For any $\varphi > 0$ we denote by $\widehat{\mathbf{Q}}_b^{T,\varphi}$ the measure induced by the process (Z, U) on the canonical space $\mathbb{R}^{L^2(I)} \times C(I)$, where

$$\begin{cases} dZ_x = b(x)U_x dx + T^{-1/2}dB_x, & x \in I, \\ U_x = \sqrt{\mu_b(x)} + \varphi(\tilde{B}_x + \xi), & x \in I. \end{cases}$$

with (B, \tilde{B}) being a two dimensional Brownian motion and $\xi = \varphi^{-1}U_0 \sim \mathcal{N}(0, 1)$ a random variable independent of (B, \tilde{B}) . The accompanying experiment is then $\mathbb{H}(\varphi, T) = (\mathbb{R}^{L^2(I)} \times C(I), \mathcal{B}_{\mathbb{R}}^{\otimes L^2(I)} \otimes \mathcal{B}_{C(I)}, \{\widehat{\mathbf{Q}}_b^{T,\varphi}\}_{b \in \Sigma_0})$.

3.4 Definition. The statistical experiment defined by observing a sample path of the stationary diffusion process (1.1), when the parameter set is Σ_0 , is denoted by $\mathbb{E}(T)$.

We can now announce the main theorem of this section, whose proof is deferred to the end of the section.

3.5 Theorem. Let $b_0 \in \Sigma$ and Σ_0 be defined as above. If $\beta > 1/2$, then the statistical experiments $\mathbb{E}(T)$ and $\mathbb{H}(T^{-1/4}, T)$ are asymptotically equivalent as $T \rightarrow \infty$.

3.6 Remark. The inspection of the proof of Theorem 3.5, combined with the fact that the total variation is bounded by two times the square root of the Kullback-Leibler divergence, shows that the Δ -distance between the experiments $\mathbb{E}(T)$ and $\mathbb{H}(\varphi, T)$ tends to zero at the rate $T^{\frac{1}{4} - \frac{\beta}{2\beta+1}} + \varphi T^{\frac{1}{4\beta+2}} + \varphi^{-1}T^{-\frac{\beta}{2\beta+1}}$. Therefore the rate-optimal choice of φ is $\varphi(T) = T^{-1/4}$ and we have $\Delta(\mathbb{E}(T), \mathbb{H}(T^{-1/4}, T)) \leq CT^{(1/2-\beta)/(4\beta+2)}$.

3.2 Definition of experiments

We start by introducing some probability measures that will be repeatedly used in this section. Some have already been defined in the previous section, but for present purposes we need to specify their dependence not only on b , but also on other parameters. In this section, the substitution of the subscript b of any probability measure by 0 indicates that we consider that measure for b identically equal to zero, e.g. $\widehat{\mathbf{Q}}_0^{T,\varphi} = \widehat{\mathbf{Q}}_b^{T,\varphi}|_{b=0}$, but the meaning of μ_0 has not changed.

Let \mathbf{Q}_{b,μ_0}^T denote the law of the Gaussian shift $dZ_x = b(x)\sqrt{\mu_0(x)}dx + T^{-1/2}dB_x$ on the canonical space $\mathbb{R}^{L^2(\mathbb{R})}$. The log-likelihood of this family of measures is defined by

$$l_T^Q(b, \mu_0, Z) = \log \left(\frac{d\mathbf{Q}_{b,\mu_0}^T}{d\mathbf{Q}_{0,\mu_0}^T}(Z) \right) = \sqrt{T} \int_{\mathbb{R}} b(x)\sqrt{\mu_0(x)}dZ_x - \frac{T}{2} \int_{\mathbb{R}} b^2(x)\mu_0(x)dx.$$

Let $\tilde{\mathbf{Q}}_{b,b_0,y_0,\mu_0}^T$ denote the law of the process (Y, V) given by Definition 2.6 with initial condition $Y_0 = y_0$. The log-likelihood of this family of measures is

$$\begin{aligned} \tilde{l}_T^Q(b, \mu_0, Y, V) &= \log \left(\frac{d\tilde{\mathbf{Q}}_{b,b_0,y_0,\mu_0}^T}{d\tilde{\mathbf{Q}}_{0,b_0,y_0,\mu_0}^T} (Y, V) \right) \\ &= \int_0^T b(Y_t) \mathbb{1}_{\{L_t^{Y_t}(Y) \leq T\mu_0(Y_t)\}} dY_t + \int_{\mathbb{R}} b(x) (T\mu_0(t) - L_T^x(Y))_+^{1/2} dV_x - \frac{T}{2} \int_{\mathbb{R}} b^2(x) \mu_0(x) dx. \end{aligned}$$

It is noteworthy that this log-likelihood does not depend on y_0 and b_0 .

Recall that $\mathbf{P}_{b,x}^T$ and $\tilde{\mathbf{P}}_{b,x}^T$ are defined as in Definitions 2.3 and 2.8, except that the initial condition is $X_0 = x$. The log-likelihoods of the families of measures $(\mathbf{P}_{b,x_0}^T)_{b \in \Sigma_0}$ and $(\tilde{\mathbf{P}}_{b,x_0}^T)_{b \in \Sigma_0}$ will be denoted by $l_T^P(b, X)$ and $\tilde{l}_T^P(b, X)$, respectively. Note that although $\tilde{\mathbf{P}}_{b,x_0}$ is a measure on the product space $C([0, T]) \times \mathbb{R}^{L^2(\mathbb{R})}$, the log-likelihood $\tilde{l}_T^P(b, X)$ depends only on the first component.

3.7 Definition. Let $\tilde{\mathbb{E}} = (C([0, T]) \times \mathbb{R}^{L^2(\mathbb{R})}, \mathcal{B}_{C([0, T])} \otimes \mathcal{B}_{\mathbb{R}^{L^2(\mathbb{R})}}, (\tilde{\mathbf{P}}_b^T)_{b \in \Sigma_0})$.

Let us fix S in the interval $(0, T)$ and define the compound experiment $\mathbb{G} = \mathbb{G}(S, T, \Sigma_0)$ as follows: we observe a sample path of the stationary diffusion process X with drift b up to time S , we compute an estimator $\bar{\mu}_S(\bullet) = \bar{\mu}_S(X, \bullet) \in C(\mathbb{R})$ of the invariant density μ_b , and then we observe a realisation of the conditionally Gaussian process $dZ_x = b(x)\sqrt{\bar{\mu}_S(x)} dx + (T-S)^{-1/2} dB_x$, $x \in \mathbb{R}$. In order to avoid subtle questions of measurability, we assume that $\bar{\mu}_S$ takes its values in a countable set $\mathfrak{M} = \{\mu_1, \mu_2, \dots\} \subset C(\mathbb{R})$.

3.8 Definition. The experiment \mathbb{G} is defined rigorously as

$$\mathbb{G}(S, T) := (C([0, S]) \times \mathbb{R}^{L^2(I)}, \mathcal{B}_{C([0, S])} \otimes \mathcal{B}_{\mathbb{R}^{\otimes L^2(I)}}, (\mathbf{R}_b^{S, T})_{b \in \Sigma_0}),$$

where $\mathbf{R}_b^{S, T}$ is the measure characterised by

$$\mathbf{R}_b^{S, T}(A \times B) = \sum_{i=1}^{\infty} \mathbf{Q}_{b, \mu_i}^{T-S}(B) \mathbf{P}_b^S(A \cap \{\bar{\mu}_S = \mu_i\}), \quad A \in \mathcal{B}_{C([0, S])}, B \in \mathcal{B}_{\mathbb{R}^{\otimes L^2(I)}}.$$

3.3 Asymptotic results

Our program in this section is as follows. We split the diffusion path observed up to time T into two parts: a path observed over $[0, S]$ and another over $[S, T]$. We prove that replacing the second path by a conditionally (to the first path) Gaussian observation we get an asymptotically equivalent experiment. Then we substitute this conditionally Gaussian experiment by another one not involving anymore the observed path over $[0, S]$. In the last step we apply this method in the converse direction, that is, making use of estimators based on the Gaussian observations, we replace the diffusion experiment over $[0, S]$ by a conditionally Gaussian one.

One method of carrying out this program consists in reducing the global equivalence problem to a local one via Lemma 9.3 of Nussbaum (1996), or its extension of Lemma 1, Delattre and Hoffmann (2002). However, this requires a local asymptotic equivalence

result between the diffusion starting at a fixed point x and a Gaussian shift, uniformly in x . Achieving the result by this technique seems to be more technical than what we do below.

3.9 Proposition. *Let $S = S(T) \in (0, T)$ be such that for some estimator \bar{b}_S of b based on the observations $(X_t, t \in [0, S])$ and taking values in a countable set $\mathfrak{B} = \{b_1, b_2, \dots\}$ the following condition is satisfied:*

$$\lim_{T \rightarrow \infty} \sup_{b \in \Sigma_0} \mathbf{E}_b \left[\int_{\mathbb{R}} (b(x) - \bar{b}_{S(T)}(x))^2 |L_T^x(X) - L_{S(T)}^x(X) - \bar{\mu}_{S(T)}(x)(T - S(T))| dx \right] = 0. \quad (3.2)$$

Then the experiments $\mathbb{E}(T)$ and $\mathbb{G}(S(T), T)$ are asymptotically equivalent as T tends to infinity.

Proof. To prove this result we introduce an auxiliary compound experiment $\tilde{\mathbb{G}}$. It is generated by the observation of a sample path of a stationary diffusion with drift b up to time $S = S(T)$ and an Itô process similar to the one of Definition 2.6, except that μ_0 and b_0 are replaced by the estimators $\bar{\mu}_S$ and \bar{b}_S respectively.

More precisely, the statistical experiment $\tilde{\mathbb{G}} = \tilde{\mathbb{G}}(T)$ is defined on the space $(C([0, T]) \times \mathbb{R}^{L^2(\mathbb{R})}, \mathcal{B}_{C([0, T])} \otimes \mathcal{B}_{\mathbb{R}^{\otimes L^2(\mathbb{R})}})$ by the family of probability measures

$$\tilde{\mathbf{R}}_b^{S, T}(A \times B) = \sum_{i, j=1}^{\infty} \int_{A \cap \{(\bar{\mu}_S, \bar{b}_S) = (\mu_i, b_j)\}} \tilde{\mathbf{Q}}_{b, b_j, f(S), \mu_i}^{T-S}(B) \mathbf{P}_b^S(df),$$

for any $A \in \mathcal{B}_{C([0, S])}$, $B \in \mathcal{B}_{C([S, T])} \otimes \mathcal{B}_{\mathbb{R}^{\otimes L^2(\mathbb{R})}}$. The above integral is well defined since the mapping $y \mapsto \tilde{\mathbf{Q}}_{b, b_j, y, \mu_i}^{T-S}(B)$ is measurable (in fact, it is continuous).

It is easy to check (see the proof of Proposition 2.13 for similar calculations) that the log-likelihood $\log \frac{d\mathbf{R}_b^{S, T}}{d\mathbf{R}_0^{S, T}}(X, Z)$ of the family of measures $\{\mathbf{R}_b^{S, T}\}_{b \in \Sigma_0}$ is given by

$$l_{T, S}^R(b, Z, X) = l_{T-S}^Q(b, \bar{\mu}_S(X, \bullet), Z) + l_S^P(b, X), \quad (X, Z) \in C([0, S]) \times \mathbb{R}^{L^2(\mathbb{R})}.$$

Likewise, the log-likelihood of the family of measures $\{\tilde{\mathbf{R}}_b^{S, T}\}_{b \in \Sigma_0}$ is given by

$$\tilde{l}_{T, S}^R(b, Y, V, X) = \tilde{l}_{T-S}^Q(b, \bar{\mu}_S(X, \bullet), Y, V) + l_S^P(b, X),$$

where $(Y, V) \in C([S, T]) \times \mathbb{R}^{L^2(\mathbb{R})}$ and $X \in C([0, S])$. We have proved in Proposition 2.13, that $\mathcal{L}((l_{T-S}^Q(b, \mu_i, Z))_{b \in \Sigma_0} | \mathbf{Q}_{0, \mu_i}^{T-S}) = \mathcal{L}((\tilde{l}_{T-S}^Q(b, \mu_i, Y, V))_{b \in \Sigma_0} | \tilde{\mathbf{Q}}_{0, b_j, y, \mu_i}^{T-S})$ for any $j \in \mathbb{N}$ and for any $y \in \mathbb{R}$. This implies

$$\mathcal{L}((l_{T-S}^Q(b, \mu_i, Z))_{b \in \Sigma_0} | \mathbf{Q}_{0, \bar{\mu}_S(f)}^{T-S}) = \mathcal{L}((\tilde{l}_{T-S}^Q(b, \mu_i, Y, V))_{b \in \Sigma_0} | \tilde{\mathbf{Q}}_{0, \bar{b}_S(f), f(S), \bar{\mu}_S(f)}^{T-S})$$

for any $f \in C([0, S])$, and consequently

$$\mathcal{L}((l_{T, S}^R(b, Z, X))_{b \in \Sigma_0} | \mathbf{R}_0^{T, S}) = \mathcal{L}((\tilde{l}_{T, S}^R(b, Y, V, X))_{b \in \Sigma_0} | \tilde{\mathbf{R}}_0^{T, S}).$$

We infer that the experiments \mathbb{G} and $\tilde{\mathbb{G}}$ are equivalent.

In order to show $\Delta(\mathbb{E}, \mathbb{G}) \rightarrow 0$, it suffices to prove $\Delta(\tilde{\mathbb{E}}, \tilde{\mathbb{G}}) \rightarrow 0$ because the experiments \mathbb{E} and $\tilde{\mathbb{E}}$ are also equivalent: their likelihood ratios coincide (cf. Proposition 2.12). The experiments $\tilde{\mathbb{E}}$ and $\tilde{\mathbb{G}}$ are defined on the same probability space and the Kullback-Leibler divergence between the respective laws is (see Corollary 2.11)

$$\begin{aligned} & \frac{1}{2} \mathbf{E}_b \left[\int_S^T (b(X_t) - \bar{b}_S(X_t))^2 \mathbb{1}_{\{L_t^{X_t}(X) - L_S^{X_t}(X) > \bar{\mu}_S(X_t)(T-S)\}} dt \right. \\ & \quad \left. + \int_{\mathbb{R}} (b(x) - \bar{b}_S(x))^2 (\bar{\mu}_S(x)(T-S) - (L_T^x(X) - L_S^x(X)))_+ dx \right] \\ & = \frac{1}{2} \mathbf{E}_b \left[\int_{\mathbb{R}} (b(x) - \bar{b}_S(x))^2 |L_T^x(X) - L_S^x(X) - \bar{\mu}_S(x)(T-S)| dx \right]. \end{aligned}$$

By condition (3.2), this expression tends to zero uniformly in $b \in \Sigma_0$ when $T \rightarrow \infty$. This completes the proof of the proposition. \square

To pursue the globalisation, we replace the Gaussian shift experiment

$$dZ_x = b(x) \sqrt{\bar{\mu}_{S(T)}(x)} dx + (T - S(T))^{-1/2} dB_x \quad (3.3)$$

by a simple experiment not involving the estimators $\bar{b}_{S(T)}$ and $\bar{\mu}_{S(T)}$. Note also that since b is known outside I , the observations $(Z_x, x \notin I)$ of the Gaussian shift experiment are void (they do not contain any information on the unknown parameter $b(x), x \in I$).

3.10 Definition. *The experiment $\mathbb{G}_1 = \mathbb{G}_1(S, T)$ is for $S \in (0, T)$ defined on $C([0, S]) \times \mathbb{R}^{L^2(I)} \times C(I)$ by the family of product measures $\hat{\mathbf{R}}_b^{S, T}(A \times B) = \mathbf{P}_b^S(A) \hat{\mathbf{Q}}_b^{T-S, \varphi(T)}(B)$, for any $A \in \mathcal{B}_{C([0, S])}$ and for any $B \in \mathcal{B}_{\mathbb{R}^{\otimes L^2(I)}} \otimes \mathcal{B}_{C(I)}$. In other words, \mathbb{G}_1 is induced by observing*

$$\begin{cases} dZ_x = b(x) U_x dx + (T - S)^{-1/2} dB_x, & x \in I, \\ U_x = \sqrt{\mu_b(x)} + \varphi(T) (\tilde{B}_x + \xi), & x \in I, \end{cases} \quad (3.4)$$

where B, \tilde{B} are independent Brownian motions and $\xi \sim N(0, 1)$ is independent of (B, \tilde{B}) .

3.11 Proposition. *Assume that the estimators $\bar{\mu}_S$ are continuously differentiable on I and satisfy the conditions of Proposition 3.9. Moreover, let for $S(T) \in (0, T)$, $\varphi(T) > 0$ the assumptions*

$$\begin{aligned} & \lim_{T \rightarrow \infty} \varphi(T)^{-2} \sup_{b \in \Sigma_0} \int_I \mathbf{E}_b \left[((\sqrt{\mu_b(x)})' - (\sqrt{\bar{\mu}_{S(T)}(x)})')^2 \right] dx = 0, \\ & \lim_{T \rightarrow \infty} \varphi(T)^{-2} \sup_{b \in \Sigma_0} \mathbf{E}_b \left[(\sqrt{\mu_b(0)} - \sqrt{\bar{\mu}_{S(T)}(0)})^2 \right] = 0, \\ & \lim_{T \rightarrow \infty} \varphi(T)^2 (T - S(T)) \sup_{b \in \Sigma_0} \int_I \mathbf{E}_b \left[(\bar{b}_{S(T)}(x) - b(x))^2 \right] dx = 0 \end{aligned}$$

be satisfied. Then the statistical experiments $\mathbb{G}(S(T), T)$ and $\mathbb{G}_1(S(T), T)$ are asymptotically equivalent as T tends to infinity.

Proof. In what follows we consecutively replace experiments by (asymptotically) equivalent ones until we reach the experiment \mathbb{G}_1 .

Recall that \mathbb{G} is defined by observing a diffusion path up to time S and a realisation of the Gaussian process Z given by (3.3). If we replace in \mathbb{G} the observations Z by

$$dZ_x = (b(x) - \bar{b}_S(x)) \sqrt{\bar{\mu}_S(x)} dx + (T - S)^{-1/2} dB_x, \quad x \in I, \quad (3.5)$$

we shall obtain an equivalent experiment, since it has exactly the same likelihood ratio. Then we replace these observations by

$$dZ_x = (b(x) - \bar{b}_S(x)) \sqrt{\mu_b(x)} dx + (T - S)^{-1/2} dB_x, \quad x \in I, \quad (3.6)$$

in view of the fact that the Kullback-Leibler divergence between the corresponding measures is up to some multiplicative constant equal to

$$(T - S) \int_I \mathbf{E}_b \left[(b(x) - \bar{b}_S(x))^2 (\sqrt{\mu_b(x)} - \sqrt{\bar{\mu}_S(x)})^2 \right] dx,$$

which tends to zero as $T \rightarrow \infty$ uniformly in $b \in \Sigma_0$, according to the assumption of Proposition 3.9 and the inequality $(\sqrt{\mu_b} - \sqrt{\bar{\mu}_S})^2 \leq |\mu_b - \bar{\mu}_S|$. It is evident that the statistical experiment

$$\begin{cases} dZ_x = (b(x) - \bar{b}_S(x)) \sqrt{\mu_b(x)} dx + (T - S)^{-1/2} dB_x, & x \in I, \\ dU_x = (\sqrt{\bar{\mu}_S(x)})' dx + \varphi d\tilde{B}_x, & x \in I, \\ U_0 = \sqrt{\bar{\mu}_S(0)} + \varphi \xi. \end{cases} \quad (3.7)$$

with independent Brownian motion \tilde{B} and $\xi \sim \mathcal{N}(0, 1)$ is equivalent to (3.6), since the sample paths of the process U do not contain any information on b . The first two assumptions of the proposition yield the equivalence of experiment (3.7) and

$$\begin{cases} dZ_x = (b(x) - \bar{b}_S(x)) \sqrt{\mu_b(x)} dx + (T - S)^{-1/2} dB_x, & x \in I, \\ dU_x = (\sqrt{\mu_b(x)})' dx + \varphi d\tilde{B}_x, & x \in I, \\ U_0 = \sqrt{\mu_b(0)} + \varphi \xi. \end{cases}$$

An equivalent form of this experiment is

$$\begin{cases} dZ_x = (b(x) - \bar{b}_S(x)) \sqrt{\mu_b(x)} dx + (T - S)^{-1/2} dB_x, & x \in I, \\ U_x = \sqrt{\mu_b(x)} + \varphi (\tilde{B}_x + \xi), & x \in I. \end{cases} \quad (3.8)$$

Computing the Kullback-Leibler divergence and using the third assumption of the proposition, one can easily check that experiment (3.8) is asymptotically equivalent to the experiment

$$\begin{cases} dZ_x = (b(x) - \bar{b}_S(x)) U_x dx + (T - S)^{-1/2} dB_x, & x \in I, \\ U_x = \sqrt{\mu_b(x)} + \varphi (\tilde{B}_x + \xi), & x \in I. \end{cases} \quad (3.9)$$

This completes the proof of the proposition, since the laws of likelihood processes of the experiments (3.9) and (3.4) coincide. \square

Having obtained the asymptotic equivalence between \mathbb{E} and \mathbb{G}_1 , we aim to replace the first part of the compound experiment \mathbb{G}_1 , which is the ergodic diffusion, by a conditionally Gaussian experiment. To do so, we assume that $\hat{\mu}_{T,S(T)}$ and $\hat{b}_{T,S(T)}$ are estimators of μ_b and b based on the observations $(Z_x, U_x, x \in I)$ and taking their values in countable subsets of $C(I)$. We denote by $\hat{\mathbf{E}}_b^{T-S,\varphi}$ the expectation with respect to the measure $\hat{\mathbf{Q}}_b^{T-S,\varphi}$ induced by the processes in (3.4).

3.12 Proposition. *Let $\varphi = \varphi(T)$ and $S = S(T)$ be such that the assumptions*

$$\lim_{T \rightarrow \infty} \sup_{b \in \Sigma_0} \widehat{\mathbf{E}}_b^{T-S, \varphi} \left[\int_I (b(x) - \widehat{b}_{T,S}(x))^2 (\sqrt{S\mu_b(x)} + S|\mu_b(x) - \widehat{\mu}_{T,S}(x)|) dx \right] = 0, \quad (3.10)$$

$$\lim_{T \rightarrow \infty} \sup_{b \in \Sigma_0} \widehat{\mathbf{E}}_b^{T-S, \varphi} \left[\int_{\mathbb{R}} |\mu_b(x) - \widehat{\mu}_{T,S}(x)| dx \right] = 0, \quad (3.11)$$

$$\lim_{T \rightarrow \infty} \sup_{b \in \Sigma_0} S \widehat{\mathbf{E}}_b^{T-S, \varphi} \left[\int_I (b(x) - \widehat{b}_{T,S}(x))^2 (\sqrt{\widehat{\mu}_{T,S}(x)} - U_x)^2 dx \right] = 0, \quad (3.12)$$

are fulfilled. Then the statistical experiments $\mathbb{G}_1(S(T), T)$ and $\mathbb{H}(\varphi(T), T)$ are asymptotically equivalent as T tends to infinity.

Proof. Recall that the experiment \mathbb{G}_1 is characterised by the observations $(X, Z^{(1)}, U)$, where X is defined by (1.1) and $(Z^{(1)}, U)$ are as in (3.4) with B replaced by $B^{(1)}$.

Let \mathbb{G}_2 be the statistical experiment defined by the observations $(Y, V, Z^{(1)}, U)$, where $Z^{(1)}$ and U are as above, $Y_0 \sim \widehat{\mu}_{T,S}$ and

$$\begin{cases} dY_t = (b(Y_t) \mathbb{1}_{\{L_t^{Y_t}(Y) \leq S\widehat{\mu}_{T,S}(Y_t)\}} + \widehat{b}_{T,S}(Y_t) \mathbb{1}_{\{L_t^{Y_t}(Y) > S\widehat{\mu}_{T,S}(Y_t)\}}) dt + dW_t, & t \in [0, S], \\ dV_x = b(x) (S\widehat{\mu}_{T,S}(x) - L_S^x(Y))_+^{1/2} dx + dB'_x, & x \in I. \end{cases}$$

In these formulae, we assume that $\widehat{b}_{T,S}(x)$ is equal to $b_0(x) = b(x)$ for any $x \notin I$ and the Brownian motions W, B' are mutually independent and independent of $(B^{(1)}, \tilde{B}, \xi)$.

The total variation distance between the laws describing the experiments \mathbb{G}_1 and \mathbb{G}_2 is controlled by (see the proof of Theorem 2.16)

$$\widehat{\mathbf{E}}_b^{T-S, \varphi} \left[\int_{\mathbb{R}} (b(x) - \widehat{b}_{T,S}(x))^2 \mathbf{E}_b |L_S^x(X) - S\widehat{\mu}_{T,S}(x)| dx \right] + \widehat{\mathbf{E}}_b^{T-S, \varphi} \|\widehat{\mu}_{T,S} - \mu_b\|_{L^1}.$$

From Proposition 5.1 we know $\text{Var}_b(L_S^x) \leq CS\mu_b(x)$. By the Cauchy-Schwarz inequality $\mathbf{E}_b |L_S^x(X) - S\widehat{\mu}_{T,S}(x)| \leq C\sqrt{S\mu_b(x)}$. So assumptions (3.10) and (3.11) yield the asymptotic equivalence of \mathbb{G}_1 and \mathbb{G}_2 .

Repeating the same arguments as those used in Proposition 3.9 for establishing the equivalence between \mathbb{G} and \mathbb{G}_1 , we can prove that the experiment \mathbb{G}_2 is equivalent to

$$\begin{cases} dZ_x^{(2)} = (b(x) - \widehat{b}_{T,S}(x)) \sqrt{\widehat{\mu}_{T,S}(x)} dx + S^{-1/2} dB_x^{(2)}, & x \in I, \\ dZ_x^{(1)} = b(x) U_x dx + (T - S)^{-1/2} dB_x^{(1)}, & x \in I, \\ U_x = \sqrt{\mu_b(x)} + \varphi(T) (\tilde{B}_x + \xi), & x \in I, \end{cases} \quad (3.13)$$

where $B^{(2)}$ is a Brownian motion independent of $(B^{(1)}, \tilde{B}, \xi)$. Once again considering the Kullback-Leibler divergence, one checks that the statistical experiment (3.13) is asymptotically equivalent to

$$\begin{cases} dZ_x^{(2)} = (b(x) - \widehat{b}_{T,S}(x)) U_x dx + S^{-1/2} dB_x^{(2)}, & x \in I, \\ dZ_x^{(1)} = b(x) U_x dx + (T - S)^{-1/2} dB_x^{(1)}, & x \in I, \\ U_x = \sqrt{\mu_b(x)} + \varphi(T) (\tilde{B}_x + \xi), & x \in I, \end{cases}$$

provided that (3.12) is satisfied. This last experiment, in turn, is equivalent to

$$\begin{cases} dZ_x^{(2)} = b(x) U_x dx + S^{-1/2} dB_x^{(2)}, & x \in I, \\ dZ_x^{(1)} = b(x) U_x dx + (T - S)^{-1/2} dB_x^{(1)}, & x \in I, \\ U_x = \sqrt{\mu_b(x)} + \varphi(T) (\tilde{B}_x + \xi), & x \in I, \end{cases} \quad (3.14)$$

since their likelihood functions coincide. The same argument yields the equivalence of (3.14) and $\mathbb{H}(\varphi, T)$ from Definition 3.3, with $B_x = (\sqrt{S} B_x^{(2)} + \sqrt{T - S} B_x^{(1)})/\sqrt{T}$ and $Z_x = (S Z_x^{(2)} + (T - S) Z_x^{(1)})/T$. \square

3.4 Proof of Theorem 3.5

First of all, note that under the conditions imposed on b and b_0 , all the stochastic differential equations introduced in previous sections have a weak solution. To establish the result of the theorem, it suffices to check that for some $S(T) \in [0, T]$ the conditions of Propositions 3.9-3.12 are fulfilled with $\varphi(T) = T^{-1/4}$.

Set $S(T) = T/2$. Since $\mathbf{E}_b(L_T^x - L_{S(T)}^x) = (T - S(T))\mu_b(x)$ and the variance of the local time at x between time instants $S(T)$ and T is bounded by $\mathcal{O}(T - S(T))\mu_b(x)$, the estimators $\bar{b}_{S(T)}$ and $\bar{\mu}_{S(T)}$ proposed in Appendix satisfy (3.2) as soon as $S(T)^{-2\beta/(2\beta+1)}(\sqrt{T - S(T)} + (T - S(T))/\sqrt{S(T)})$ tends to zero. This convergence holds (for $S(T) = T/2$) if and only if $\beta > 1/2$.

To verify the conditions of Proposition 3.11 we use the obvious relation $(\sqrt{\mu_b(x)})' = b(x)\sqrt{\mu_b(x)}$ and the bounds (5.3)-(5.6). We infer that the desired conditions are fulfilled if $\varphi(T)^{-2}T^{-2\beta/(2\beta+1)}$ and $T^{1-2\beta/(2\beta+1)}\varphi(T)^2$ tend to zero. This is obviously the case for $\varphi(T)^2 = T^{-1/2}$ and $\beta > 1/2$. The verification of the conditions of Proposition 3.12 is achieved similarly using Lemma 5.5.

4 Extensions and generalisations

4.1 General diffusion coefficient

Let us consider the model with non-constant diffusion coefficient

$$dX_t = b(X_t)dt + \sigma(X_t) dW_t, \quad t \in [0, T]. \quad (4.1)$$

If we suppose continuous-time observations, $\sigma^2(x)$ is perfectly identifiable at all points x where the local time $L_T^x(X)$ is nonzero. Let us for simplicity assume that the diffusion coefficient σ is known everywhere, satisfies $\sigma_* \leq \sigma(x) \leq \sigma^*$, $x \in \mathbb{R}$, for some constants $\sigma^* \geq \sigma_* > 0$ and that σ is differentiable with a continuous bounded derivative. Then the stochastic differential equation (4.1) has for any initial value a unique strong solution and for all $b \in \Sigma$ a stationary solution with invariant density

$$\mu_{b,\sigma}(x) = C_{b,\sigma}\sigma(x)^{-2} \exp\left(\int_0^x \frac{2b(y)}{\sigma^2(y)} dy\right), \quad x \in \mathbb{R},$$

where $C_{b,\sigma} > 0$ is a normalising constant.

Since σ is known, the experiment of observing

$$\bar{X}_t := H(X_t), \quad t \in [0, T], \quad \text{with } H(x) = \int_0^x \frac{1}{\sigma(v)} dv,$$

is clearly equivalent to that of observing directly $(X_t, t \in [0, T])$, the stationary solution of (4.1). By Itô's formula we infer that \bar{X} is the stationary solution of the equation

$$d\bar{X}_t = \bar{b}(\bar{X}_t) dt + dW_t, \quad t \in [0, T], \quad \text{with } \bar{b}(x) := \left(\frac{b}{\sigma} - \frac{\sigma'}{2}\right)(G^{-1}(x)).$$

By transformation, the invariant density $\mu_{\bar{b}}$ of \bar{X} is given by $(\mu_{b,\sigma} \circ G^{-1})(G^{-1})'$. Applying Theorem 2.16, we conclude that the experiment of observing (4.1) is asymptotically equivalent to the Gaussian white noise experiment

$$dZ_x = \left(\frac{b}{\sigma} - \frac{\sigma'}{2}\right)(G^{-1}(x)) \sqrt{\mu_{b_0,\sigma}(G^{-1}(x))\sigma(G^{-1}(x))} dx + T^{-1/2} dB_x, \quad x \in \mathbb{R},$$

for shrinking neighbourhoods as in Theorem 2.16 in terms of the transformed quantities \bar{b} , \bar{b}_0 and $\mu_{\bar{b}_0}$. By regarding the likelihood process, it follows that observing dZ is equivalent to observing

$$d\tilde{Z}_x = \left(\frac{b}{\sigma} - \frac{\sigma'}{2}\right)(x) \sqrt{\mu_{b_0,\sigma}(x)} dx + T^{-1/2} dB_x, \quad x \in \mathbb{R},$$

which in turn is equivalent to observing

$$dV_x = b(x) \sqrt{\mu_{b_0,\sigma}(x)} dx + T^{-1/2} \sigma(x) dB_x, \quad x \in \mathbb{R}.$$

The asymptotic equivalence with the latter Gaussian white noise experiment can be explained intuitively by adopting the viewpoint of regression. The random design follows the density $\mu_{b_0,\sigma}$, while the observation noise at the point x has level $T^{-1/2}\sigma(x)$. Finally, note that under our rather restrictive assumptions on σ the same globalisation procedure as for constant diffusion coefficients can be applied.

4.2 Time discretisation

We show that the diffusion experiment based on continuous-time observations is asymptotically equivalent to its discrete counterpart, provided that the sampling distance decreases at the appropriate rate. Let $\mathcal{P}(T) = \{0 = t_0 < t_1 < \dots < t_{N(T)} = T\}$ be a grid on the interval $[0, T]$. Set $d_i = t_{i+1} - t_i$ and $|\mathcal{P}(T)| = \max_{i=0,\dots,N(T)-1} d_i$.

We consider the autoregression experiment defined by observing (y_1, \dots, y_N) from

$$y_{i+1} = y_i + d_i b(y_i) + \sqrt{d_i} \xi_i, \quad i = 0, \dots, N-1, \quad y_0 \sim \mu_b, \quad (4.2)$$

where ξ_i are i. i. d. with $\xi_1 \sim N(0, 1)$ and independent of y_0 . We introduce the function

$$b^{\mathcal{P}} : C([0, T]) \times [0, T] \rightarrow \mathbb{R}, \quad b^{\mathcal{P}}(x, t) = \sum_{i=0}^{N-1} b(x(t_i)) \mathbb{1}_{[t_i, t_{i+1}[}(t).$$

For $b \in \Sigma$ this function is progressively measurable and the stochastic differential equation

$$dX_t^{\mathcal{P}} = b^{\mathcal{P}}(X^{\mathcal{P}}, t) dt + dW_t, \quad t \in [0, T], \quad X_0^{\mathcal{P}} \sim \mu_b, \quad (4.3)$$

has a weak solution defined by Proposition 2.4. One checks that the laws of the likelihood processes of the experiments (4.2) and (4.3) coincide, therefore these experiments are equivalent. According to Corollary 2.11, the Δ -distance between the stationary diffusion experiment (1.1) and the experiment (4.3) tends to zero if

$$\mathbf{E}_b \left[\sum_{i=0}^{N(T)-1} \int_{t_i}^{t_{i+1}} (b(X_t) - b(X_{t_i}))^2 dt \right] \quad (4.4)$$

tends to zero uniformly in b . If we assume that b is Hölder continuous with exponent β , then (4.4) is up to a multiplicative constant bounded by

$$\sum_{i=0}^{N(T)-1} \int_{t_i}^{t_{i+1}} \mathbf{E}_b [(X_t - X_{t_i})^{2\beta}] dt \leq \sum_{i=0}^{N(T)-1} \int_{t_i}^{t_{i+1}} 4 \mathbf{E}_b \left[\left(\int_{t_i}^t b(X_s) ds \right)^{2\beta} + (W_t - W_{t_i})^{2\beta} \right] dt.$$

Using the linear growth condition and the boundedness of all moments of X_t under \mathbf{P}_b uniformly over b , one finds that (4.4) tends to zero as soon as $\sum_{i=0}^{N-1} d_i^{\beta+1} \leq T |\mathcal{P}(T)|^\beta \rightarrow 0$. We infer that if the mesh size $|\mathcal{P}(T)|$ satisfies $T |\mathcal{P}(T)|^\beta \rightarrow 0$ as $T \rightarrow \infty$, then the statistical experiments (1.1), (4.3) and (4.2) are asymptotically equivalent. Since $(X_{t_0}^{\mathcal{P}}, \dots, X_{t_N}^{\mathcal{P}})$ is a sufficient statistics in the experiment (4.3), the asymptotic equivalence between (4.3) and (1.1) entails the asymptotic sufficiency of $(X_{t_0}, \dots, X_{t_N})$ in (1.1). As we have already mentioned in Introduction, analogous results for the diffusion experiment with asymptotically vanishing diffusion coefficient are obtained in Milstein and Nussbaum (1998).

Moreover, since we have shown that the total variation distance between \mathbf{P}_b^T and the law of $X^{\mathcal{P}}$ on $(C([0, T]), \mathcal{B}_C([0, T]))$ tends to zero, the same holds true for the restrictions of these measures to the sub- σ -algebra generated by the discrete observations at time instants t_i , $i = 0, \dots, N(T)$. This implies that the experiments defined by observing $(X_{t_0}^{\mathcal{P}}, \dots, X_{t_N}^{\mathcal{P}})$ and $(X_{t_0}, \dots, X_{t_N})$ are asymptotically equivalent. Since observing $(X_{t_0}^{\mathcal{P}}, \dots, X_{t_N}^{\mathcal{P}})$ is equivalent to observing the whole path $X^{\mathcal{P}}$, and (4.3) is asymptotically equivalent to (1.1), we get the asymptotic equivalence of the discretely sampled diffusion experiment and the diffusion experiment based on continuous-time observations, as soon as $T |\mathcal{P}(T)|^\beta \rightarrow 0$.

Note that, although these kinds of discretisation results can be expected, Brown, Wang, and Zhao (2002) have discovered and studied surprising results on the nonequivalence in the parametric setup for GARCH time series and their diffusion limits in the context of stochastic volatility models.

4.3 A less informative experiment

We present an accompanying sequence of simple white noise experiments that is globally less informative for the asymptotics $T \rightarrow \infty$ than our diffusion experiment. Let us first consider the local experiment

$$\mathbb{F}_1^< := \mathbb{F}_1^<(b_0, T, \varepsilon, \eta, \zeta) := (\mathbb{R}^{L^2(\mathbb{R})}, \mathcal{B}_{\mathbb{R}}^{\otimes L^2(\mathbb{R})}, (\mathbf{Q}_b^{T, <})_{b \in \Sigma_{\varepsilon, \eta, \zeta}(b_0, \mu_{b_0})}),$$

where $\mathbf{Q}_b^{T,<}$ denotes the law of the Gaussian shift experiment

$$dZ_x^< = b(x)\sqrt{\mu_*(x)}dx + T^{-1/2}dB_x, \quad x \in \mathbb{R}, \quad (4.5)$$

with a Brownian motion B on the real line and a measurable function $\mu_* : \mathbb{R} \rightarrow [0, \infty)$ satisfying $\mu_*(x) \leq \mu_{b_0}(x)$ for all $x \in \mathbb{R}$, $b \in \Sigma_{\varepsilon, \eta, \zeta}(b_0, \mu_{b_0})$. Hence, \mathbb{F}_1 with the centre (b_0, μ_{b_0}) and $\mathbb{F}_1^<$ are defined on the same measurable space and only differ in the choice of μ_{b_0} and μ_* , respectively.

We claim that the experiment $\mathbb{F}_1^<$ is less informative than \mathbb{F}_1 . In fact, it suffices to construct a Markov kernel $K : \mathbb{R}^{L^2(\mathbb{R})} \times \mathcal{B}_{\mathbb{R}}^{\otimes L^2(\mathbb{R})} \rightarrow [0, 1]$ such that $\mathbf{Q}_b^{T,<} = K \otimes \mathbf{Q}_b^T$ holds for all $b \in \Sigma_{\varepsilon, \eta, \zeta}(b_0, \mu_{b_0})$, see Section 49 in Strasser (1985) or Theorem 2.2 in Le Cam and Yang (2000). Hence, using $\mu_* \leq \mu_{b_0}$ and setting for $\varphi \in \mathbb{R}^{L^2(\mathbb{R})}$

$$G_\varphi(f) := \varphi\left(\left(\frac{\mu_*}{\mu_{b_0}}\right)^{1/2}f\right) + \int T^{-1/2}\left(1 - \frac{\mu_*(x)}{\mu_{b_0}(x)}\right)^{1/2}f(x)d\bar{B}_x$$

with a Brownian motion \bar{B} on \mathbb{R} , we find that G_φ is a random functional on $L^2(\mathbb{R})$ and we define the Markov kernel $K(\varphi, A) := \mathbf{P}_{\bar{B}}(G_\varphi \in A)$, where $\mathbf{P}_{\bar{B}}$ denotes the law of the Brownian motion \bar{B} . Using short-hand notation for the functionals, we obtain

$$\begin{aligned} \int K(\varphi, A)\mathbf{Q}_b^T(d\varphi) &= \mathbf{Q}_b^T \otimes \mathbf{P}_{\bar{B}}\left(\left(\left(\frac{\mu_*}{\mu_{b_0}}\right)^{1/2}dZ^< + T^{-1/2}\left(1 - \frac{\mu_*}{\mu_{b_0}}\right)^{1/2}d\bar{B}\right) \in A\right) \\ &= \mathbf{P}_B \otimes \mathbf{P}_{\bar{B}}\left(\left(b\mu_*^{1/2}dx + T^{-1/2}\left(\left(\frac{\mu_*}{\mu_{b_0}}\right)^{1/2}dB + \left(1 - \frac{\mu_*}{\mu_{b_0}}\right)^{1/2}d\bar{B}\right)\right) \in A\right) \\ &= \mathbf{P}_{\tilde{B}}\left(\left(b\mu_*^{1/2}dx + T^{-1/2}d\tilde{B}\right) \in A\right) \end{aligned}$$

with a new Brownian motion \tilde{B} . Consequently, $\mathbb{F}_1^<$ is a randomisation of \mathbb{F}_1 and thus \mathbb{F}_1 is more informative than $\mathbb{F}_1^<$. Under the asymptotics of Theorem 2.16, we conclude by transitivity that the localised diffusion experiment \mathbb{E}_1 is asymptotically also more informative than $\mathbb{F}_1^<$ uniformly over $b_0 \in \Sigma$, which means in terms of the Le Cam-deficiency δ

$$\lim_{T \rightarrow \infty} \sup_{b_0 \in \Sigma_0} \delta\left(\mathbb{E}_1(b_0, \mu_{b_0}, T, \varepsilon(T), \eta(T), \zeta(T)), \mathbb{F}_1^<(b_0, \mu_{b_0}, T, \varepsilon(T)\eta(T), \zeta(T))\right) = 0.$$

Of course, for the uniformity in the last result we have to assume $\mu_* \leq \mu_{b_0}$ for all $b_0 \in \Sigma_0$, which we shall also do subsequently.

In a second step, we take advantage of the fact that the laws in $\mathbb{F}_1^<$ do not depend anymore on the centre of localisation so that we can apply the usual globalisation procedure, cf. Lemma 9.3 and its proof in Nussbaum (1996). Let us denote by $\mathbb{F}^<$ the global white noise experiment with the law in (4.5) and with the parameter class Σ_0 from Definition 3.1. We claim that $\mathbb{F}^<$ is asymptotically less informative than \mathbb{E} . Following the proof of Proposition 3.9, we use an estimator $\bar{b}_{T/2}$ from the observation of the diffusion experiment until time $T/2$ such that the local neighbourhood is attained asymptotically:

$$\lim_{T \rightarrow \infty} \inf_{b \in \Sigma_0} \mathbf{P}_b^{T/2}\left(\bar{b}_{T/2} \in \Sigma_{\varepsilon(T), \eta(T), \zeta(T)}(b)\right) = 1.$$

Therefore the diffusion experiment $\mathbb{E}(T)$ is asymptotically more informative than observing

$$\begin{cases} dX_t = b(X_t)dt + dW_t, & t \in [0, T/2], \\ dZ_x^< = b(x)\sqrt{\mu_*(x)}dx + (T/2)^{-1/2}dB_x, & x \in \mathbb{R}. \end{cases}$$

As in Section 5.3, we construct estimators for b based on the observation of $Z^<$ and infer like in Proposition 3.12 that the latter experiment is more informative than $\mathbb{F}^<$.

As an application, the result that the global diffusion experiment \mathbb{E} is asymptotically more informative than the white noise experiment $\mathbb{F}^<$ can be used to infer optimal minimax rates for the drift estimation under an $L^p(I)$ -loss function and with a standard nonparametric class of prescribed regularity contained in Σ_0 . Choosing $\mu_\star = \inf_{b \in \Sigma_0} \mu_b$, we observe that on the compact interval I the function μ_\star is strictly positive such that it can be bounded from above and below by a constant factor times some density μ_b on I . Therefore, when using the local equivalence result for the asymptotic lower bound (cf. the methodology in Korostelev and Tsybakov (1993)) and the global deficiency result for the asymptotic upper bound, the risk bounds will merely differ by a constant factor. We suspect that for a pointwise loss function the asymptotic deficiency yields in many cases even the exact asymptotic constants, provided the lower bound is proved locally around the parameter b_0 fulfilling $\mu_{b_0}(x_0) = \mu_\star(x_0)$ for the point x_0 under consideration. Another possibility to determine exact asymptotic constants is to use the compound experiment \mathbb{G} and to follow the methodology developed by Nussbaum (1996) for the Pinsker constant.

5 Appendix

5.1 Variance of local time

5.1 Proposition. *There is a constant $C > 0$ only depending on the class Σ such that*

$$\text{Var}_{\mathbf{P}_b}[L_T^y(Y)] \leq C \mu_b(y).$$

Proof. For simplicity we shall omit the index b everywhere in the proof. Let us introduce the transition density $p(x, y; t) = \mathbf{P}(Y_t \in dy | Y_0 = x)$ and the Markov transition operators $P_t f(x) = \int f(y) p(x, y; t) dy$ for functions f in the space $L^2(\mu)$ of functions square-integrable with respect to the invariant measure. Its generator L has in our case the divergence form representation $Lf = \frac{1}{2} \mu^{-1} (\mu f')'$ for f in the domain $\mathcal{D}(L)$ of L , see Revuz and Yor (1999) for further details. We find

$$\begin{aligned} \text{Var}(L_T^y(Y)) &= \int_0^T \int_0^T p(y, y; |t-s|) \mu(y) dt ds - T^2 \mu(y)^2 \\ &= \int_0^T 2(T-u) (p(y, y; u) \mu(y) - \mu(y)^2) du \\ &\leq 2T \mu(y) \int_0^T (p(y, y; u) - \mu(y)) du. \end{aligned}$$

In order to evaluate the integral we make use of the formula for operator semigroups in Banach spaces (Engel and Nagel 2000, Lemma II.1.3)

$$\int_0^T P_t Lf dt = P_T f - f, \quad \forall f \in \mathcal{D}(L). \quad (5.1)$$

Let us introduce the distribution function $F_\mu(x) := \int_{-\infty}^x \mu(\xi) d\xi$ and

$$G_y(x) := \int_y^x \frac{2\mu(y)}{\mu(u)} \left(\mathbb{1}_{[y, \infty)}(u) - F_\mu(u) \right) du.$$

Then formally $LG_y(x) = \delta_y(x) - \mu(y)$ holds and thus

$$\int_0^T p(x, y; t) - \mu(y) dt = \int_0^T P_t LG_y(x) dt = P_T G_y(x) - G_y(x). \quad (5.2)$$

This formula can be justified by using a smoothing function $h \in C^\infty(\mathbb{R})$ with $h \geq 0$, $\int h = 1$ and compact support and then setting

$$G_{y, \varepsilon}(x) := \int_y^x \int_{-\infty}^u \frac{2\mu(\xi)}{\mu(z)} \left(\varepsilon^{-1} h(\varepsilon^{-1}(y - \xi)) - \int \varepsilon^{-1} h(\varepsilon^{-1}(y - v)) \mu(v) dv \right) d\xi du.$$

Then $G_{y, \varepsilon}$ lies in the domain of L and satisfies

$$LG_{y, \varepsilon}(x) = \varepsilon^{-1} h(\varepsilon^{-1}(y - x)) - \int \varepsilon^{-1} h(\varepsilon^{-1}(y - v)) \mu(v) dv.$$

Applying formula (5.1) to $G_{y, \varepsilon}$ and then letting $\varepsilon \rightarrow 0$ gives the representation (5.2).

By the contraction property of $(P_t)_{t \geq 0}$ we find

$$\lim_{T \rightarrow \infty} \int_0^T p(x, y; t) - \mu(y) dt = \mathbf{E}_\mu[G_y(X_0)] - G_y(x).$$

Now a simple calculation yields

$$\lim_{T \rightarrow \infty} \int_0^T p(y, y; t) - \mu(y) dt = \int \frac{2\mu(y)}{\mu(u)} \left((1 - F_\mu(u))^2 \mathbb{1}_{[y, \infty)}(u) + F_\mu(u)^2 \mathbb{1}_{(-\infty, y]}(u) \right) du.$$

The conditions on the class Σ ensure

$$\sup_{x \leq 0} \frac{F_\mu(x)}{\mu(x)} \leq K \quad \text{and} \quad \sup_{x \geq 0} \frac{1 - F_\mu(x)}{\mu(x)} \leq K$$

with the constant $K = \frac{1}{2\gamma} + \frac{\exp(2CA(1+A))}{2C(1+A)}$. Using this estimate we obtain for $y \geq 0$

$$\begin{aligned} \int_y^\infty \frac{2\mu(y)}{\mu(u)} (1 - F_\mu(u))^2 du &\leq K^2 \int_y^\infty 2\mu(y)\mu(u) du \leq 2K^2\mu(y), \\ \int_{-\infty}^y \frac{2\mu(y)}{\mu(u)} (F_\mu(u))^2 du &\leq \int_{-\infty}^0 2K^2\mu(y)\mu(u) du + \int_0^y \frac{2\mu(y)}{\mu(u)} du \\ &\leq 2K^2\mu(y) + 2 \int_0^y \exp\left(2 \int_u^y b(v) dv\right) du. \end{aligned}$$

According to the conditions in (1.2), in each estimate the right hand side is uniformly bounded over Σ . We apply the symmetric argument for $y \leq 0$. Finally, we observe that $p_b(y, y; t) - \mu_b(y) \geq 0$ follows for all $y \in \mathbb{R}$ and $t > 0$ from the fact that $p_b(x, y; t) - \mu_b(y)$ is the continuous kernel of the nonnegative operator $P_t(\text{Id} - \Pi_0)$ with $\Pi_0 f = \mathbf{E}_\mu[f(X_0)]$ on $L^2(\mu)$. \square

5.2 Estimators in the model of ergodic diffusion

Let $X = \{X_t, t \in [0, S]\}$ be a sample path of a stationary diffusion process satisfying (1.1). Let $K : \mathbb{R} \rightarrow \mathbb{R}$ be a smooth kernel function with support in $[-1, 1]$. We shall use the standard kernel estimator of μ_b defined as follows:

$$\tilde{\mu}_S(x) = \frac{1}{Sh_S} \int_0^S K\left(\frac{x - X_t}{h_S}\right) dt, \quad x \in \mathbb{R},$$

where $h_S = S^{-1/(2\beta+1)}$ is the bandwidth leading to a rate optimal estimator. It is evident that this estimator is differentiable and its derivative equals:

$$\tilde{\mu}'_S(x) = \frac{1}{Sh_S^2} \int_0^S K'\left(\frac{x - X_t}{h_S}\right) dt, \quad x \in \mathbb{R}.$$

5.2 Lemma. *There exist some positive constants m and L_1 such that*

$$m \leq \mu_b(x) \leq m^{-1}, \quad |\mu'_b(x) - \mu'_b(y)| \leq L_1|x - y|^\beta, \quad \forall x, y \in I, \quad \forall b \in \Sigma_0.$$

Proof. Use the uniform boundedness of b on I and the formula $\mu'_b(x) = 2b(x)\mu_b(x)$. \square

5.3 Lemma. *For any real $p > 1$, there is some positive constant C_1 depending only on β, L_1, m, p such that for any $x \in I$ we have*

$$\mathbf{E}_b[(\tilde{\mu}_S(x) - \mu_b(x))^{2p}] \leq C_1 S^{-p}, \quad \mathbf{E}_b[(\tilde{\mu}'_S(x) - \mu'_b(x))^{2p}] \leq C_1 S^{-2p\beta/(2\beta+1)}.$$

Proof. Using standard arguments, the bias in both cases can be bounded as follows:

$$|\mathbf{E}_b[\tilde{\mu}_S(x)] - \mu_b(x)| \leq Ch_S^{\beta+1}, \quad |\mathbf{E}_b[\tilde{\mu}'_S(x)] - \mu'_b(x)| \leq Ch_S^\beta.$$

Let us evaluate the variance term of $\tilde{\mu}'_S(x)$:

$$\mathbf{E}_b[(\tilde{\mu}'_S(x) - \mathbf{E}_b[\tilde{\mu}'_S(x)])^{2p}] = S^{-2p} h_S^{-4p} \mathbf{E}_b\left[\left(\int_0^S \left[K'\left(\frac{x - X_t}{h_S}\right) - \mathbf{E}_b K'\left(\frac{x - X_t}{h_S}\right)\right] dt\right)^{2p}\right].$$

We introduce the function

$$g_x(u) = \frac{2}{\mu_b(u)} \int_{\mathbb{R}} K'\left(\frac{x - y}{h_S}\right) \mu_b(y) (\mathbb{1}_{\{u > y\}} - F_\mu(u)) dy.$$

The Itô formula applied to the function $\int_0^\bullet g_x(u) du$ and to the diffusion X yields

$$\int_{X_0}^{X_S} g_x(u) du = \int_0^S g_x(X_t) dW_t + \int_0^S \left(K'\left(\frac{x - X_t}{h_S}\right) - \mathbf{E}_b \left[K'\left(\frac{x - X_t}{h_S}\right) \right] \right) dt.$$

This equality combined with the Burkholder-Davis-Gundy inequality (Revuz and Yor 1999, Theorem IV.4.1) leads to the estimate

$$\mathbf{E}_b[(\tilde{\mu}'_S(x) - \mathbf{E}_b[\tilde{\mu}'_S(x)])^{2p}] \leq CS^{-2p} h_S^{-4p} \mathbf{E}_b\left[\left(\int_0^{X_S} g_x(u) du\right)^{2p} + \left(\int_0^S g_x^2(X_t) dt\right)^p\right].$$

The integration by parts formula yields

$$\begin{aligned}
|g_x(u)| &= \frac{2h_S}{\mu_b(u)} \left| \int_{-1}^1 K'(v) \mu_b(x - vh_S) (\mathbb{1}_{\{u > x - vh_S\}} - F_\mu(u)) dv \right| \\
&\leq 2h_S \left| K\left(\frac{x-u}{h_S}\right) \right| + \frac{2h_S^2}{\mu_b(u)} \left| \int_{-1}^1 K(v) \mu_b'(x - vh_S) (\mathbb{1}_{\{u > x - vh_S\}} - F_\mu(u)) dv \right| \\
&\leq 2h_S \left| K\left(\frac{x-u}{h_S}\right) \right| + Ch_S^2.
\end{aligned}$$

This term is uniformly in $b \in \Sigma_0$ bounded by Ch_S . Therefore, we get

$$\mathbf{E}_b [(\tilde{\mu}'_S(x) - \mathbf{E}_b[\tilde{\mu}'_S(x)])^{2p}] \leq \frac{C}{(Sh_S)^{2p}} \left(\mathbf{E}_b[X_S^{2p}] + \mathbf{E}_b \left[\left(\int_0^S \left| K\left(\frac{x-X_t}{h_S}\right) \right|^2 dt \right)^p \right] + S^p h_S^{2p} \right).$$

The first expectation in the right hand side is uniformly bounded, while the second one can be estimated as follows:

$$\begin{aligned}
\mathbf{E}_b \left[\left(\int_0^S \left| K\left(\frac{x-X_t}{h_S}\right) \right|^2 dt \right)^p \right] &= \mathbf{E}_b \left[\left(\int_{\mathbb{R}} \left| K\left(\frac{x-y}{h_S}\right) \right|^2 L_S^y(X) dy \right)^p \right] \\
&= h_S^p \mathbf{E}_b \left[\left(\int_{-1}^1 |K(v)|^2 L_S^{x-vh_S}(X) dv \right)^p \right] \\
&\leq 2h_S^p \sup_{|v| \leq 1} \mathbf{E}_b[L_S^{x-vh_S}(X)^p].
\end{aligned}$$

Using the fact that the expectation of the local time $L_S^y(X)$ is equal to $S\mu_b(y)$ and the centred moments of order p are bounded by $CS^{p/2}$ (Kutoyants 2003, (1.35), page 30,), we get the estimate

$$\mathbf{E}_b [(\tilde{\mu}'_S(x) - \mathbf{E}_b[\tilde{\mu}'_S(x)])^{2p}] \leq C((Sh_S)^{-2p} + (Sh_S)^{-p} + S^{-3p/2} h_S^{-p} + S^{-p}).$$

In the right hand side of this inequality, the largest term is obviously $(Sh_S)^{-p}$ and it decreases to zero at the rate $S^{-2p\beta/(2\beta+1)}$. The evaluation of the variance of $\tilde{\mu}_S(x)$ is even simpler and will be omitted. \square

Due to the compactness property derived from Arzela-Ascoli theorem, there exists a finite subset $\mathfrak{M} = \{f_1, \dots, f_N\}$ of $C^1(I)$ such that for any $b \in \Sigma_0$ there exists $f_b \in \mathfrak{M}$ such that $\sup_{x \in I} |f_b(x) - \mu_b(x)| + \sup_{x \in I} |f_b'(x) - \mu_b'(x)| \leq S^{-1/2}$. Since all the functions $\mu_b, \mu_b', \mu_b^{-1}$, $b \in \Sigma_0$ are uniformly bounded on I , the same holds for the elements of \mathfrak{M} . That is $m \leq f(x) \leq 1/m$ and $|f'(x)| \leq m'$ for any $x \in I$ and for any $f \in \mathfrak{M}$. Let $\bar{\mu}_S$ be the estimator of μ_b defined as follows

$$\bar{\mu}_S = \operatorname{argmin}_{f \in \mathfrak{M}} \left(|f(0) - \tilde{\mu}_S(0)| + \|f - \tilde{\mu}_S\|_{L^4} + \|f' - \tilde{\mu}'_S\|_{L^4}^{(2\beta+1)/2\beta} \right),$$

in the case when the minimiser is not unique, we choose the one having the smallest index. Obviously, this estimator takes a finite number of values.

5.4 Lemma. *There exists a constant C depending only on the parameters describing the set Σ_0 , such that the following inequalities hold:*

$$\mathbf{E}_b [(\bar{\mu}_S(0) - \mu_b(0))^2] \leq CS^{-1}, \tag{5.3}$$

$$\int_I \mathbf{E}_b [(\bar{\mu}_S(x) - \mu_b(x))^4] dx \leq CS^{-2}, \tag{5.4}$$

$$\int_I \mathbf{E}_b [(\bar{\mu}'_S(x) - \mu_b'(x))^4] dx \leq CS^{-4\beta/(2\beta+1)}. \tag{5.5}$$

Proof. The proof of these estimates follows from Lemma 5.3 and the definition of $\bar{\mu}_S$ using standard arguments (see, for instance, Nussbaum (1996), page 2424). \square

Using this estimator $\bar{\mu}_S$ and the formula $b(x) = \mu'_b(x)/2\mu_b(x)$, we define the estimator \bar{b}_S of the drift b by $\bar{b}_S(x) = \bar{\mu}'_S(x)/2\bar{\mu}_S(x)$. Since $\bar{\mu}_S \in \mathfrak{M}$, we have

$$\begin{aligned} |\bar{b}_S(x) - b(x)| &\leq \frac{|\mu'_b(x) - \bar{\mu}'_S(x)|}{2\mu_b(x)} + \frac{|\bar{\mu}'_S(x)| |\mu_b(x) - \bar{\mu}_S(x)|}{2\mu_b(x)\bar{\mu}_S(x)} \\ &\leq C(|\mu'_b(x) - \bar{\mu}'_S(x)| + |\mu_b(x) - \bar{\mu}_S(x)|) \end{aligned}$$

for any $x \in I$. An obvious consequence of this inequality and Lemma 5.4 is

$$\sup_{b \in \Sigma_0} \int_I \mathbf{E}_b [(\bar{b}_S(x) - b(x))^4] dx \leq CS^{-4\beta/(2\beta+1)}. \quad (5.6)$$

5.3 Estimators in the conditionally Gaussian model

Let (Z, U) be given by (3.4). The aim in this subsection is to construct estimators of b and μ_b based on the observations $(U_x, Z_x, x \in I)$. Let K be a kernel defined as in the previous subsection, $h = h_{T-S} = (T-S)^{-1/(2\beta+1)}$ and let m be the constant from Lemma 5.2. We define

$$\hat{b}_{T,S}(x) = \frac{1}{h} \begin{cases} \int_I K\left(\frac{x-y}{h} - 1\right) U_y^{-1} dZ_y, & \text{if } x > 0 \text{ and } \inf_{y \in I} U_y > \sqrt{m}/2, \\ \int_I K\left(\frac{x-y}{h} + 1\right) U_y^{-1} dZ_y, & \text{if } x \leq 0 \text{ and } \inf_{y \in I} U_y > \sqrt{m}/2, \\ 0, & \text{if } \inf_{y \in I} U_y \leq \sqrt{m}/2. \end{cases}$$

Of course, this definition is used when $x \in I$; for $x \notin I$ the estimator $\hat{b}_{T,S}(x)$ is defined to be equal to $b_0(x)$. The estimator of μ_b is simply $\hat{\mu}_{T,S} = \mu_{\hat{b}_{T,S}}$.

5.5 Lemma. *For any $p > 0$ there exist some constants $c, C > 0$ such that*

$$\begin{aligned} \sup_{b \in \Sigma_0} \int_I \hat{\mathbf{E}}_b^{T-S,\varphi} [(\hat{b}_{T,S}(x) - b(x))^{2p}] dx &\leq C((T-S)^{-2p\beta/(2\beta+1)} + e^{-c\varphi^{-2}}), \\ \sup_{b \in \Sigma_0} \int_I \hat{\mathbf{E}}_b^{T-S,\varphi} [(\hat{\mu}_{T,S}(x) - \mu_b(x))^{2p}] dx &\leq C((T-S)^{-p} + e^{-c\varphi^{-2}}). \end{aligned}$$

Proof. The first inequality can be obtained by evaluating the L^{2p} -risk of the classical kernel estimator on the event $\{\inf_{x \in I} U_x > \sqrt{m}/2\}$ and by bounding the probability of the event $\{\inf_x U_x \leq \sqrt{m}/2\}$ using the inclusion $\{\inf_x U_x \leq \sqrt{m}/2\} \subset \{\sup_{x \in I} (\xi + \tilde{B}_x) \geq \varphi^{-1}\sqrt{m}/2\}$ and the fact that $\sup_{x \in [0,D]} B_x$ has the same law as $|B_D|$.

To show the second inequality, one checks first, using standard arguments, that the expectation of any power $2p$ of $\sup_{x \in I} |\int_0^x (b - \hat{b}_{T,S})(v) dv|$ is bounded by $C(T-S)^{-p}$. Afterwards, the explicit expression of the mapping $b \mapsto \mu_b$ yields the estimate

$$|\mu_b(x) - \mu_{\hat{b}}(x)| \leq (\mu_b(x) + \mu_{\hat{b}}(x)) \sup_{x \in \mathbb{R}} |1 - e^{2 \int_0^x (b - \hat{b})(v) dv}|.$$

To end up, one uses the inequality $|e^a - 1| \leq 2a, \forall a \in [-1, 1]$ for bounding the expectation on the event $\mathcal{W} = \{\sup_{x \in \mathbb{R}} |\int_0^x (b - \hat{b}_{T,S})(v) dv| < 1\}$ and the Chebyshev inequality for bounding the same expectation on \mathcal{W}^c . \square

5.6 Remark. *So far, the estimators $\widehat{b}_{T,S}$ and $\widehat{\mu}_{T,S}$ do not take values in countable subsets of $C(I)$, but the corresponding modifications in the previous subsection apply here, too.*

References

- BROWN, L. D., T. CAI, M. G. LOW, AND C.-H. ZHANG (2002): “Asymptotic equivalence theory for nonparametric regression with random design,” *Ann. Stat.*, 30(3), 688–707.
- BROWN, L. D., AND M. G. LOW (1996): “Asymptotic equivalence of nonparametric regression and white noise,” *Ann. Stat.*, 24(6), 2384–2398.
- BROWN, L. D., Y. WANG, AND L. ZHAO (2002): “Statistical equivalence at suitable frequencies of GARCH and stochastic volatility models with the corresponding diffusion model,” *Manuscript*, <http://www-stat.wharton.upenn.edu/~lzhao/papers/>.
- BROWN, L. D., AND C.-H. ZHANG (1998): “Asymptotic nonequivalence of nonparametric experiments when the smoothness index is $1/2$,” *Ann. Stat.*, 26(1), 279–287.
- DALALYAN, A., AND Y. KUTOYANTS (2002): “Asymptotically efficient trend coefficient estimation for ergodic diffusion,” *Math. Methods Statist.*, 11(4), 402–427.
- DELATTRE, S., AND M. HOFFMANN (2002): “Asymptotic equivalence for a null recurrent diffusion,” *Bernoulli*, 8(2), 139–174.
- DEUSCHEL, J.-D., AND D. W. STROOCK (1989): *Large deviations. Rev. ed.* Pure and Applied Mathematics, 137. Boston, MA etc.: Academic Press, Inc.
- DREES, H. (2001): “Minimax risk bounds in extreme value theory,” *Ann. Statist.*, 29(1), 266–294.
- ENGEL, K.-J., AND R. NAGEL (2000): *One-parameter semigroups for linear evolution equations.* Graduate Texts in Mathematics. 194. Berlin: Springer.
- FAN, J. (2003): “A selective overview of nonparametric methods in financial econometrics,” *Manuscript*, <http://www.stat.unc.edu/faculty/fan/manuscripts.html>.
- GENON-CATALOT, V., C. LAREDO, AND M. NUSSBAUM (2002): “Asymptotic equivalence of estimating a Poisson intensity and a positive diffusion drift,” *Ann. Stat.*, 30(3), 731–753.
- GRAMA, I., AND M. NUSSBAUM (1998): “Asymptotic equivalence for nonparametric generalized linear models,” *Probab. Theory Relat. Fields*, 111(2), 167–214.
- JACOD, J., AND A. N. SHIRYAEV (2003): *Limit theorems for stochastic processes. 2nd ed.* Grundlehren der Mathematischen Wissenschaften. 288. Berlin: Springer.
- JÄHNISCH, M., AND M. NUSSBAUM (2003): “Asymptotic equivalence for a model of independent non identically distributed observations,” *Statist. Decisions*, 21(3), 197–218.

- KALLENBERG, O. (2002): *Foundations of modern probability. 2nd ed.* Probability and Its Applications. New York, NY: Springer.
- KARATZAS, I., AND S. E. SHREVE (1991): *Brownian motion and stochastic calculus. 2nd ed.* Graduate Texts in Mathematics, 113. New York etc.: Springer-Verlag.
- KOROSTELEV, A., AND A. TSYBAKOV (1993): *Minimax theory of image reconstruction.* Lecture Notes in Statistics (Springer), 82, New York: Springer-Verlag.
- KUTOYANTS, Y. (2003): *Statistical Inference for Ergodic Diffusion Processes.* Springer Series in Statistics, New York.
- LE CAM, L., AND G. YANG (2000): *Asymptotics in statistics. Some basic concepts.* Springer Series in Statistics, New York.
- LIPTSER, R. S., AND A. N. SHIRYAEV (2001): *Statistics of random processes. 1: General theory. 2nd rev. and exp. ed.* Applications of Mathematics. 5. Berlin: Springer.
- MILSTEIN, G., AND M. NUSSBAUM (1998): "Diffusion approximation for nonparametric autoregression," *Probab. Theory Relat. Fields*, 112(4), 535–543.
- NUSSBAUM, M. (1996): "Asymptotic equivalence of density estimation and Gaussian white noise," *Ann. Stat.*, 24(6), 2399–2430.
- REVUZ, D., AND M. YOR (1999): *Continuous martingales and Brownian motion. 3rd ed.* Graduate Texts in Mathematics. 293. Berlin: Springer.
- STRASSER, H. (1985): *Mathematical theory of statistics. Statistical experiments and asymptotic decision theory.* De Gruyter Studies in Mathematics, 7. Berlin - New York: Walter de Gruyter.
- TANIGUCHI, M., AND Y. KAKIZAWA (2000): *Asymptotic theory of statistical inference for time series.* Springer Series in Statistics. New York, NY: Springer.