Bernoulli **11**(3), 2005, 411–444

Exact asymptotics for estimating the marginal density of discretely observed diffusion processes

CRISTINA BUTUCEA[1,2] and MICHAEL H. NEUMANN[3]

[1]*Modal'X, Université Paris X, Bât. G, 200 avenue de la République, 92001 Nanterre Cedex, France*
[2]*Laboratoire de Probabilités et Modèles Aléatoires, Université Paris VI, 175 rue de Chevaleret, 75013 Paris, France. E-mail: butucea@ccr.jussieu.fr*
[3]*Technische Universität Braunschweig, Institut für Mathematische Stochastik, Pockelsstraße 14, D-38106 Braunschweig, Germany. E-mail: mi.neumann@tu-bs.de*

We derive sharp asymptotic minimax bounds (that is, bounds which concern the exact asymptotic constant of the risk) for nonparametric density estimation based on discretely observed diffusion processes. We study two particular problems for which there already exist such results in the case of independent and identically distributed observations, namely, minimax density estimation in Sobolev classes with L_2-loss and in Hölder classes with L_∞-loss. We derive independently lower and upper bounds for the asymptotic minimax risks and show that they coincide with the classic efficiency bounds. We prove that these bounds can be attained by usual kernel density estimators. The lower bounds are obtained by analysing the problem of estimating the marginal density in certain families of processes, $\{\{X_i^f\}, f \in \mathcal{F}_n\}$, which are shrinking neighbourhoods of some central process, $\{X_i^{f_0}\}$, in the sense that the set of densities $\mathcal{F}_n$ forms a shrinking neighborhood centred around f_0.

Keywords: density estimation; dependent data; diffusion processes; discrete sampling; exact asymptotics; minimax risk; nonparametric estimation

1. Introduction

In this paper we study nonparametric estimation of the marginal density based on discrete observations from a real-valued diffusion process. Such density estimators can either be used themselves for further inference or be used as an intermediate step when one intends to estimate the volatility function on the basis of discrete-time observations; see Aït-Sahalia (1996a). There are already many contributions related to our work from two different communities. On the one hand, probabilists working in the field of stochastic processes have studied parametric and nonparametric estimators of model parameters or related quantities for continuous- or discrete-time observations from diffusion processes. On the other hand, researchers from the statistical community have derived minimax results for nonparametric estimators in various settings (regression, density estimation, spectral density estimation), up to the level of 'exact asymptotics' which concerns both the rate of convergence and the asymptotic constant.

Some motivation for our work emerged from Aït-Sahalia (1996a) who used nonparametric diffusion processes in financial mathematics for modelling interest rate processes. Because of the lack of continuous-time observations it was not possible to estimate the model parameters (a drift function described by a finite-dimensional parameter and a volatility function to be estimated nonparametrically) directly; instead, Aït-Sahalia proposed first to estimate this parameter and the stationary density and then to use these results for estimating the volatility function. To estimate the density, he used standard kernel methods on an *ad hoc* basis; however, theory supporting the appropriateness of these methods is still lacking.

To determine whether the use of standard nonparametric estimators should be recommended in the case of observations from diffusion processes, we intend to derive the asymptotic minimax optimality of these methods. In other frameworks (regression, density estimation, spectral density estimation), such results already exist in the statistical literature. Most of these concern optimal rates of convergence in certain smoothness classes (Hölder, Sobolev, Besov), for different loss functions. For a given class of functions, there is usually a wide range of estimators that achieve such a minimax rate of convergence. For example, for kernel estimators only some qualitative characteristics of the kernel function (number of vanishing moments) are important, whereas there is still much freedom for the particular shape of the kernel. To narrow down the set of methods which deserve the term 'optimal', a minimax theory which also focuses on the optimal asymptotic constant is considered. There are several set-ups in nonparametric curve estimation where exact asymptotic minimax results are known: ellipsoidal restrictions on the class of functions in connection with the L_2-loss (a case first studied by Pinsker 1980); Hölder restrictions together with the L_∞-loss (initiated by Korostelev 1993) and with the Bahadur risk; and finally, analytic functions with the L_p and pointwise losses (Ibragimov and Hasminskii 1983; Golubev and Levit 1996), or more general classes of infinitely differentiable functions known as supersmooth, or functions with rapidly decreasing Fourier transforms (Lepski and Levit 1998).

We will develop exact minimax asymptotics for the first two of these classic cases. The derivation of such results usually consists of two parts: first, a particular estimation procedure is proposed, which attains the presumed optimal asymptotic risk bound; then it is proven that this asymptotic bound cannot be improved by any other estimation method. The analogy of the upper bounds to those known in the case of independent and identically distributed (i.i.d.) observations is perhaps not very surprising. Indeed, it is known that certain kernel estimators are asymptotically minimax in the i.i.d. case, and for such estimators there is (for any fixed density f) an equivalence of the (pointwise) asymptotic behaviour between the two cases of independent and weakly dependent observations. Actually, we derive the upper bounds in the more general context of absolutely regular (β-mixing) processes. We devote some effort to showing that the efficiency bounds can be achieved by standard kernel methods which form the most popular class of nonparametric density estimators.

Lower bounds will be obtained by studying asymptotically least favourable parametric subexperiments. For this purpose, we consider statistical experiments based on discrete observations from families of processes $(X_t^f)_{t\geqslant 0}$, where the density f parametrizing this class varies in certain subclasses $\mathcal{F}_n^{(2)}$ or $\mathcal{F}_n^{(\infty)}$ of the Sobolev or Hölder class, respectively. The link to the case of Gaussian shift experiments is achieved by proving local asymptotic normality for appropriate one-dimensional subexperiments. Since the corresponding families of marginal densities are basically given by perturbations supported on subintervals of decreasing (as $n \to \infty$) length, one could already conjecture that some sort of whitening-by-windowing effect is responsible for a behaviour of the likelihood processes that is asymptotically equivalent to that in the case of independent observations.

The efficiency bounds obtained in this paper complement existing results concerning the similar behaviour of nonparametric estimators in the cases of independent and weakly dependent observations with the assertion that one cannot use partial knowledge about the dependence structure to achieve essentially better results than in the case of independent observations. This is particularly remarkable since Saavedra and Cao (1999) have shown in the case of MA(1) processes that standard kernel estimators can be drastically outperformed by certain convolution-type estimators which are $\sqrt{n}$-consistent. Our results give strong justification for the practical application of nonparametric methods that were usually designed in the context of independent data. Furthermore, the coincidence of the first-order minimax asymptotics sets a limitation to eventual improvements by specific modifications that can possibly be devised in order to make use of some partial knowledge about the dependence structure. As a by-product, since the minimax bounds are achieved by kernel estimators with corresponding uniquely defined kernel functions, one can use such results as an objective criterion for finding kernels that are optimal in some reasonable way. This optimality goes beyond the well-known results on kernels that are optimal under additional side conditions. Moreover, in view of the uniqueness of the optimum kernels, we deduce that these estimators are asymptotically admissible in the class of kernel methods.

The paper is structured as follows. Since it may be of interest to researchers from different communities – theoretical statisticians primarily interested in nonparametric curve estimation as well as probabilists working in the field of statistics for stochastic processes – we present in Section 2 an overview of existing results in both areas. In Section 3 we state asymptotic lower bounds to the minimax risks. After describing appropriate subexperiments that are difficult enough for generating the desired bounds, we first state approximations to the likelihood ratios which are the basis for proving local asymptotic normality. Then we formulate theorems with the lower risk bounds. In Section 4 we describe particular kernel estimators which are asymptotically minimax and state upper risk bounds. Section 5 contains proofs of the technical local asymptotic normality results, of the lower and upper risk bounds, and of a Bernstein-type inequality under mixing.

2. Overview of existing results

There are several set-ups in nonparametric curve estimation where exact asymptotic minimax results are known: ellipsoidal restrictions on the class of functions in connection with the L_2-loss (the Pinsker case); Hölder restrictions together with the L_∞-loss

(Korostelev case) and Hölder restrictions with the Bahadur risk, and supersmooth functions with L_p and pointwise losses.

Research in the first case was initiated by the seminal paper of Pinsker (1980), who had already solved all essential problems in the particular context of signal estimation in Gaussian white noise. Later, these results were transferred to spectral density estimation (Efromovich and Pinsker 1982), density estimation (Efromovich and Pinkser 1983), and nonparametric regression with Gaussian errors (Nussbaum 1985) respectively with non-normal error distributions (Golubev and Nussbaum 1990). Other generalizations, for example in Belitser and Levit (1995), consisted of studying the case of more general ellipsoids and the second-order behaviour of the minimax risk. An extensive overview on this topic is given in Nussbaum (1999).

Developments in the second case of L_∞-loss started with the paper by Korostelov (1993), who found exact asymptotic efficiency bounds for nonparametric estimation in Hölder classes with smoothness index $0 < \beta \leqslant 1$. Using completely different arguments (issuing from the theory of optimal recovery), Donoho (1994) extended these results to Hölder classes with smoothness index $\beta > 1$. The same problem for density estimation from i.i.d. observations and for $\beta > 0$ was solved by Korostelev and Nussbaum (1999).

Exact asymptotics in the third case were first developed by Korostelev (1996) in Gaussian regression, and generalized by Korostelev and Leonov (1996) to the non-Gaussian case. Density functions with analytic continuation on a strip around the real axis were estimated in the minimax sharp approach by Ibragimov and Hasminskii (1983) with the L_p-loss. Efficient estimation of such functions was done in the Gaussian white noise model with L_p risk by Guerre and Tsybakov (1998), with pointwise and L_∞ risks by Golubev *et al*. (1996), while density estimation with pointwise risk was given by Golubev and Levit (1996) and a more general density estimation under random censorship by Belitser (1998). For a good review of these results we refer to Ibragimov (2001). Recently, more general classes of infinitely differentiable functions were considered and sharp estimation was given for the pointwise risk in both the minimax and adaptive approaches by Lepski and Levit (1998) in the Gaussian white noise model, and these results were translated to the density model by Artiles (2001).

Estimation of diffusion processes was studied in both a parametric and a nonparametric set-up. The available data are usually supposed to be continuous in time, or discrete with time gap decreasing to 0, or constant. Efficient estimation in this context is based on the local asymptotic normality (LAN) property. In the case of low-frequency data, the LAN property with rate $1/\sqrt{n}$ was established by Roussas (1972). Nevertheless, efficient maximum likelihood estimators are not available since the explicit forms of transition densities are unknown. Recently, the LAN property for general ergodic diffusions with high-frequency data was proven by Gobet (2002) using Malliavin calculus.

Nonparametric estimation of the coefficients is usually done via the marginal density of the process or the spectral decomposition of the infinitesimal generator. Estimation of the marginal density of a process observed on an entire time interval has been studied intensively since the work of Banon (1978) and Banon and Nguyen (1981), who found consistent estimators. First, Castellanna and Leadbetter (1986) found superoptimal estimation rates (parametric in the length of the time interval T) in the case of very irregular sample path processes. Analogous results were obtained by Leblanc (1997) for certain stationary processes and Kutoyants (1997) who gives efficiency bounds in the context of ergodic diffusion processes. A review of available results on superoptimal rates in density or regression estimation that are preserved under certain discretization models is given in Bosq (1998). More recently, Dalalyan and Kutoyants (2002) established efficient minimax estimation for the drift function, in the case of ergodic diffusion processes. For nonparametric estimation of discretely observed diffusions with a decreasing time gap, we refer to Kessler (1997), Hoffmann (1999) and references therein.

Spectral methods in parametric set-ups were developed in Hansen *et al*. (1998) and in parallel by Kessler and Sørensen (1999). An extension of these methods to diffusions on a compact interval with reflecting boundary conditions, observed discretely at low frequency, can be found in Gobet *et al*. (2002).

3. Asymptotic lower bounds to the minimax risks

In this section, we derive asymptotic lower bounds to the minimax risks in estimating the marginal density of a real-valued and stationary diffusion process which is observed at equidistant time points. We basically show that we cannot do better than in the case of i.i.d. observations, even if we have prior knowledge of the dependence mechanism. This case is adequately modelled by a statistical experiment consisting of a family of discretely observed diffusion processes with different marginal densities but one and the same copula function. We devise efficiency bounds in two classic cases for which analogous bounds are already known in the framework of i.i.d. data, that is, we derive exact asymptotic bounds for the L_2 risk in Sobolev classes as well as for the L_∞ risk in Hölder classes.

In the former case, we assume that the marginal density f is in the class

$$\mathcal{F}_2(\beta, L) = \left\{ f \in L_2(-\infty, \infty) : \int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1,\, f(x) \geqslant 0 \ \forall x, \text{ and } \int_{-\infty}^{\infty} \left(f^{(\beta)}(x)\right)^2 \mathrm{d}x \leqslant L \right\},$$

where β is an integer, $\beta > 0$. In the latter case, we assume that f lies in the class

$$\mathcal{F}_\infty(\beta, L, B) = \left\{ f : \int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1,\, 0 \leqslant f(x) \leqslant B \quad \forall x, \right.$$

$$\left. \text{and } |f^{(\lfloor\beta\rfloor)}(x_1) - f^{(\lfloor\beta\rfloor)}(x_2)| \leqslant L|x_1 - x_2|^{\beta - \lfloor\beta\rfloor} \forall x_1, x_2 \right\},$$

where $\lfloor\beta\rfloor$ denotes the greatest integer strictly less than β ($\beta > 0$). We observe that results similar to ours can be derived for Hölder classes without the additional parameter B. Actually, it was shown in Korostelev and Nussbaum (1999) that densities satisfying a Hölder condition for some β and L are uniformly bounded by some $B_* = B_*(\beta, L)$. We decided to state the results for Hölder classes with the additional parameter B since this constant shows up in the minimax bound, which makes the role played by $\|f\|_\infty$ more transparent.

For any given f, we assume that n observations $X_1^f, \ldots, X_n^f$ are available, where the underlying process (in continuous time), $(X_t^f)_{t\geqslant 0}$, is a real-valued and stationary diffusion process obeying the Itô stochastic differential equation

$$\mathrm{d}X_t^f = \mu^f(X_t^f)\,\mathrm{d}t + \sigma^f(X_t^f)\,\mathrm{d}W_t,$$

where $(W_t)_{t\geqslant 0}$ is standard Brownian motion. Let us mention that usually these models are indexed by the drift and the diffusion functions, viewed as parameters. Our choice is to consider the marginal density as an underlying parameter, which is more convenient for our purpose.

Lower bounds for the minimax risk are obtained by considering appropriate families of diffusion processes, $(X_t^f)_{t\geqslant 0}$, indexed by the corresponding marginal density f which varies in appropriate subclasses $\mathcal{F}_n^{(2)}$ and $\mathcal{F}_n^{(\infty)}$ of the Sobolev or Hölder class, respectively. These subclasses of densities are centred around suitable basic functions, $f_0^{(2)}$ and $f_0^{(\infty)}$, respectively, which are the stationary densities of diffusion processes with sufficiently regular drift and diffusion functions. For notational convenience and since no confusion should occur, we drop the indices 2 and ∞ referring to the two cases of L_2 or L_∞ risk.

We begin by fixing a basic process $(X_t^{f_0})_{t\geqslant 0}$. In order to ensure certain smoothness properties of the conditional densities needed below, it is important to choose drift and diffusion functions that are sufficiently regular. For positive constants σ_0, μ_0 and further constants $-\infty < K_1 < K_2 < K_3 < K_4 < \infty$, we define

$$\sigma^{f_0}(x) \equiv \sigma_0 \tag{3.1}$$

and choose a three times continuously differentiable function μ^{f_0} with

$$\mu^{f_0}(x)\begin{cases} = \mu_0, & x \leqslant K_1, \\ \in [0, \mu_0], & x \in (K_1, K_2), \\ = 0, & x \in [K_2, K_3], \\ \in [-\mu_0, 0], & x \in (K_3, K_4), \\ = -\mu_0, & x \geqslant K_4. \end{cases} \tag{3.2}$$

According to Karlin and Taylor (1981, Section 15.5) and Aït-Sahalia (1996b, p. 390), the stationary density f_0 then satisfies

$$f_0(x) = \frac{C_0}{(\sigma^{f_0}(x))^2}\exp\left\{\int_0^x \frac{2\mu^{f_0}(u)}{(\sigma^{f_0}(u))^2}\,\mathrm{d}u\right\}. \tag{3.3}$$

This definition implies in particular that f_0 is monotonously non-decreasing on $(-\infty, K_2]$, constant on $[K_2, K_3]$, and non-increasing on $[K_3, \infty)$. Choosing the above parameters σ_0 and μ_0 accordingly, we can obtain basic functions that will turn out to be suitable for deriving the desired lower bounds in the two cases, L_2 risk combined with Sobolev classes and L_∞ risk with Hölder classes. In the first case, we choose the parameters such that $K_3 - K_2$ is large and $\int_{-\infty}^{\infty}(f_0^{(\beta)})^2$ is small; for more details see the proof of Theorem 3.1. In the second case, the parameters will be chosen such that

$$|f_0^{(\lfloor\beta\rfloor)}(x_1) - f_0^{(\lfloor\beta\rfloor)}(x_2)| \leqslant L|x_1 - x_2|^{\beta-\lfloor\beta\rfloor} \qquad \forall x_1, x_2$$

and, for some B in accordance with this restriction,

$$f_0(x) \leqslant B \ \forall x \qquad \text{and} \qquad f_0(x) = B \ \forall x \in [K_2, K_3].$$

For any fixed choice of the above parameters, the process $(X_t^{f_0})_{t\geqslant 0}$ is absolutely regular (β-mixing) with exponentially decaying coefficients; see, for example, Veretennikov (1984, Section 2). Furthermore, it is well known that $(X_t^{f_0})_{t\geqslant 0}$ is Markovian.

The processes $(X_t^f)_{t\geqslant 0}$, for $f \in \mathcal{F}_n^{(2)}$ or $f \in \mathcal{F}_n^{(\infty)}$, respectively, are obtained by the quantile transform as

$$X_t^f = F^{-1}(F_0(X_t^{f_0})),$$

where F_0 and F are the cumulative distribution functions corresponding to the densities f_0 and f, respectively. It is clear that $(X_t^f)_{t\geqslant 0}$ is also a diffusion process and from Itô's formula we readily see that its drift and diffusion functions have the form

$$\mu^f(x) = q_f'(q_f^{-1}(x))\mu^{f_0}(q_f^{-1}(x)) + \tfrac{1}{2}q_f''(q_f^{-1}(x))(\sigma^{f_0}(q_f^{-1}(x)))^2,$$

$$\sigma^f(x) = q_f'(q_f^{-1}(x))\sigma^{f_0}(q_f^{-1}(x)),$$

with $q_f(x) = F^{-1}(F_0(x))$. The exact form of these functions is, however, not important for what follows since the behaviour of the process $(X_t^f)_{t\geqslant 0}$ is completely described by the particular quantile transform and the copula function which is, of course, equal to that of the basic process $(X_t^{f_0})_{t\geqslant 0}$. For example, mixing properties of $(X_t^f)_{t\geqslant 0}$ are exactly the same as those of $(X_t^{f_0})_{t\geqslant 0}$ since the dependence mechanism underlying both processes is the same.

3.1. Local asymptotic normality

As already mentioned, we derive lower bounds to the minimax risks by considering the problem of estimating the marginal density for certain subfamilies of processes, $\{(X_i^f)_{i=1,\ldots,n}, f \in \mathcal{F}_n\}$, where $\mathcal{F}_n$ denotes either $\mathcal{F}_n^{(2)}$ or $\mathcal{F}_n^{(\infty)}$, which are appropriate parametric subclasses of $\mathcal{F}_2(\beta, L)$ and $\mathcal{F}_\infty(\beta, L, B)$, respectively.

For the case of L_2 risk in Sobolev classes, a sequence of sufficiently hard subexperiments is given by functions of the type

$$f_\theta = f_0^{(2)} + \sum_{j=1}^{s}\sum_{k=1}^{q_n} \theta_{j,k}\phi_{j,k,n}.$$

Here $\theta = (\theta_{j,k})$ parametrizes the class of functions under consideration based on the perturbations

$$\phi_{j,k,n}(x) = n^{-1/2}h_n^{-1/2}\phi_j((x - a_{k,n})/h_n) - r_{j,n}f_0^{(2)}(x),$$

where $r_{j,n} = n^{-1/2}h_n^{1/2}\int \phi_j(x)\,\mathrm{d}x$. The $a_{k,n}$ are chosen such that $a_{1,n} < \ldots < a_{q_n,n}$, s is large enough and the sequences $(q_n)_{n\in\mathbb{N}}$ and $(h_n^{-1})_{n\in\mathbb{N}}$ increase at rate $n^{1/(2\beta+1)}$. An exact description of the functions ϕ_j and $\phi_{j,k,n}$ is given in the proof of Theorem 3.1 below.

For the case of L_∞ risk in Hölder classes, a sequence of asymptotically least favourable subexperiments is given by a class of functions of the type

$$f_\theta = f_0^{(\infty)} + \sum_{j=1}^{M_n} \theta_j \psi_{j,n},$$

where $\theta = (\theta_1, \ldots, \theta_{M_n})^{\mathrm{T}} \in \{-1, 1\}^{M_n}$ and

$$\psi_{j,n}(x) = (\ln(n))^{1/2} n^{-1/2} h_n^{-1/2} \psi((x - b_{j,n})/h_n),$$

for a suitable function ψ, $b_{1,n} < \ldots < b_{M_n,n}$ and $(h_n)_{n\in\mathbb{N}}$ tending to zero at rate $(\ln(n)/n)^{1/(2\beta+1)}$. The exact choice of the function ψ is described in the proof of Theorem 3.2 below.

Both families of functions $(\phi_j((\cdot - a_{k,n})/h_n))$ and $(\psi_{j,n})$ have shrinking support and shrinking uniform bound. Moreover, their integral is 0, so that the resulting functions f_θ are density functions for large enough n.

The key step in deriving lower bounds to the risks consists of studying likelihood ratios and then proving local asymptotic normality. We will actually show that the likelihood ratios behave asymptotically as in the case of i.i.d. observations. The following lemma provides an approximation which underlines this fact. We denote by $p^f(\cdot|\cdot)$ and $p^f(\cdot, \ldots, \cdot)$ conditional and joint densities of the process $(X_i^f)_{i=1,\ldots,n}$, respectively.

Lemma 3.1. *Let $f \in \mathcal{F}_n$ and $f_u = f + u\phi_n$, where ϕ_n is one of the perturbations, from either the $\phi_{j,k,n}$ or the $\psi_{j,n}$, and where u is bounded. Then*

$$\frac{p^{f_u}(x_i|x_{i-1})}{p^f(x_i|x_{i-1})} = \frac{f_u(x_i)}{f(x_i)} R_n(x_i, x_{i-1}),$$

where

$$|R_n(x_i, x_{i-1}) - 1| \leqslant O\Big(n^{-1/2} h_n^{1/2} \exp\big\{\delta(y_i - y_{i-1})^2 + \delta_n |y_i - y_{i-1}|\big\}\Big),$$

and $y_j = F_0^{-1}(F(x_j))$, for $j \in \{i-1, i\}$. $\delta > 0$ is an arbitrarily small constant, $\delta_n = O(n^{-1/2} h_n^{1/2})$.

The following proposition provides an approximation of the logarithmic likelihood ratio of certain one-dimensional subexperiments. This result will be the basis for the LAN property used in the proof of asymptotic lower risk bounds.

Proposition 3.1. *Suppose that the assumptions of Lemma 3.1 are satisfied. Then*

$$\ln \frac{p^{f_u}(x_1, \ldots, x_n)}{p^f(x_1, \ldots, x_n)} = u \sum_{i=1}^n \frac{\phi_n(x_i)}{f(x_i)} - \frac{u^2}{2} n \mathrm{E}_f \left(\frac{\phi_n(X_1^f)}{f(X_1^f)} \right)^2 + R_{f,u}(x_1, \ldots, x_n),$$

say, where, for any constant $u_0 < \infty$,

$$\sup_{f \in \mathcal{F}_n, |u| \leqslant u_0} \Big\{ P_f\Big(|R_{f,u}(X_1^f, \ldots, X_n^f)| > C_\lambda n^{-\epsilon} \Big) \Big\} = O\big(n^{-\lambda}\big)$$

holds for some $\epsilon > 0$ and arbitrary $\lambda < \infty$.

The validity of this approximation does not depend on the particular form of the perturbations $\phi_{j,k,n}$ and $\psi_{j,n}$. Analogous results can be expected whenever the basic process is 'regular enough' and the perturbations have shrinking support.

3.2. Lower risk bounds

On the basis of the LAN property stated in the previous subsection, we are now in a position to prove the desired asymptotic lower risk bounds in the two cases under consideration.

Theorem 3.1. *Suppose that a family of statistical experiments with observations $(X_i^f)_{i=1,\dots,n}$ is given, where these processes are constructed as described at the beginning of Section 3. Then*

$$\liminf_{n\to\infty} \inf_{\tilde{f}} \sup_{f\in\mathcal{F}_2(\beta,L)} \{n^{2\beta/(2\beta+1)}\mathrm{E}_f\|\tilde{f}-f\|^2_{L_2}\} \geqslant \gamma(\beta)L^{1/(2\beta+1)},$$

where $\gamma(\beta) = (2\beta+1)^{1/(2\beta+1)}[\beta/(\pi(\beta+1))]^{2\beta/(2\beta+1)}$ is Pinsker's constant.

The next theorem states the asymptotic risk bound for the L_∞ case. Because of the degenerate behaviour of the supremum deviation we can state the efficiency bounds for a general loss function that may be not explicit.

Theorem 3.2. *Suppose that a family of statistical experiments with observations $(X_i^f)_{i=1,\dots,n}$ is given, where these processes are constructed as described at the beginning of Section 3. Let w be a continuous and monotone non-decreasing function. Then*

$$\liminf_{n\to\infty} \inf_{\tilde{f}} \sup_{f\in\mathcal{F}_\infty(\beta,L,B)} \mathrm{E}_f w((n/\log(n))^{\beta/(2\beta+1)}\|\tilde{f}-f\|_\infty) \geqslant w(\kappa(\beta, L, B)),$$

where

$$\kappa(\beta, L, B) = A_\beta\left(\frac{2(B\wedge B_*)L^{1/\beta}}{2\beta+1}\right)^{\beta/(2\beta+1)},$$

$$A_\beta = \max\{g(0)|\|g\|_{L_2}\leqslant 1, \quad |g^{(\lfloor\beta\rfloor)}(x_1)-g^{(\lfloor\beta\rfloor)}(x_2)|\leqslant |x_1-x_2|^{\beta-\lfloor\beta\rfloor}\ \forall x_1, x_2\}$$

and

$$B_* = B_*(\beta, L) = \max\{g(0)|\,|g^{(\lfloor\beta\rfloor)}(x_1)-g^{(\lfloor\beta\rfloor)}(x_2)|\leqslant L|x_1-x_2|^{\beta-\lfloor\beta\rfloor}\ \forall x_1, x_2\}.$$

The constant $B\wedge B_*$ that appears in Theorem 3.2 requires a few words of explanation. In the problem of estimating a density from i.i.d. data, this constant would be equal to $B_*(\beta, L)$; see Korostelev and Nussbaum (1999). In our context, we have an additional restriction since we define our basic process for technical reasons with constant diffusion coefficient which, in turn, restricts the set of possible densities that fulfil (3.3).

The asymptotic lower risk bounds in Theorems 3.1 and 3.2 are the analogues to the well-known bounds in density estimation from i.i.d. data. This means that there is no gain in knowing the particular dependence structure of the process; we can at best hope to obtain the same efficiency bounds as in the i.i.d. case.

4. Asymptotic upper bounds to the minimax risks

In this section we describe kernel estimators which are optimal from the exact asymptotic minimax point of view. We stress the fact that these methods attain the same upper bounds in a more general context than in Section 3. The only condition on the dependence structure is the following one.

Assumption 4.1. *Let* $\mathcal{F}_j^k = \sigma(X_j, \ldots, X_k)$ *be the* σ*-field generated by* $X_j, \ldots, X_k$*. We assume that the coefficients of absolute regularity* (β*-mixing*),

$$\beta(k) = \max_j \{\mathrm{E}\, \mathrm{ess\,sup}_{V \in \mathcal{F}_{j+k}^n} \{|P(V|\mathcal{F}_1^j) - P(V)|\}\},$$

satisfy

$$\beta(k) \leqslant C \exp(-C_1 k).$$

We assume, moreover, that the conditional densities are uniformly bounded,

$$\sup_{x, y \in R} p^f(y|x) \leqslant C_2,$$

where $C_2 > 0$ *does not depend on* f *in the smoothness class.*

It is well-known that Pinsker's bound can be attained by certain kernel estimators; see Golubev (1987) for nonparametric regression, and Schipper (1996) for density estimation. To achieve the desired asymptotic bound in our context, we may employ exactly the same estimator. Let us denote

$$\hat{f}_{h_n}^{[2]}(x) = \frac{1}{nh_n} \sum_{i=1}^n K_{\beta,2}\left(\frac{X_i - x}{h_n}\right), \tag{4.1}$$

where $K_{\beta,2}$ is the kernel described by its Fourier transform $F(K_{\beta,2})(t) = (1 - |t|^\beta)_+$, (recall that $F(K)(t) = \int_{-\infty}^\infty K(y)\mathrm{e}^{\mathrm{i}yt}\, \mathrm{d}y$) and the bandwidth has the expression

$$h_n = \left(\frac{\beta}{\pi L(\beta+1)(2\beta+1)}\right)^{1/(2\beta+1)} n^{-1/(2\beta+1)}.$$

The kernel function $K_{\beta,2}$ can be obtained by the inverse Fourier transform

$$K_{\beta,2}(x) = -\frac{\beta!}{\pi} \sum_{j=1}^\beta \frac{\sin^{(j)}(x)}{(\beta - j)! x^{j+1}}. \tag{4.2}$$

Note that $K_{\beta,2}(0) = \beta/(\pi(\beta+1))$, which can be obtained either by the previous formula or by direct integration of its Fourier transform, $K_{\beta,2}(0) = \pi^{-1}\int_0^1(1 - t^{\beta})\mathrm{d}t$. The following theorem states the efficiency of the given kernel estimator.

Theorem 4.1. *Suppose that Assumption 4.1 is satisfied. Then*

$$\limsup_{n\to\infty} \sup_{f\in\mathcal{F}_2(\beta,L)} \{n^{2\beta/(2\beta+1)}\mathrm{E}_f\|\hat{f}^{[2]}_{h_n} - f\|^2_{L_2}\} \leqslant \gamma(\beta)L^{1/(2\beta+1)},$$

where $\gamma(\beta)$ was defined in Theorem 3.1 above.

Similarly to the L_2 case, a kernel estimator,

$$\hat{f}^{[\infty]}_{h_n}(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K_{\beta,\infty}\left(\frac{X_i - x}{h_n}\right),$$

will attain the efficiency bound in the L_∞ case. This time the optimal bandwidth is

$$h_n = \left(\frac{2(B\wedge B_*)}{(2\beta+1)L^2}\frac{\log n}{n}\right)^{1/(2\beta+1)}.$$

An appropriate kernel is given by $K_{\beta,\infty}(t) = \phi(t)/\int\phi(x)\mathrm{d}x$, where ϕ is the solution to the optimization problem

$$\phi = \arg\max_g\{g(0)|\|g\|_{L_2} \leqslant 1,\ |g^{(\lfloor\beta\rfloor)}(x_1) - g^{(\lfloor\beta\rfloor)}(x_2)| \leqslant |x_1 - x_2|^{\beta-\lfloor\beta\rfloor}\forall x_1, x_2\}.$$

Korostelev (1993) and Donoho (1994) have shown that, for $0 < \beta \leqslant 1$,

$$A_\beta = \phi(0) = \left(\frac{(2\beta+1)(\beta+1)}{4\beta^2}\right)^{\beta/(2\beta+1)},$$

$$\phi(t) = (\phi(0) - |t|^{\beta})_+,$$

$$K_{\beta,\infty}(t) = \frac{2\beta+1}{2\beta}(\phi(0))^{-(\beta+1)/\beta}\phi(t).$$

For $\beta > 1$, uniqueness of the solution ϕ and compactness of its support are proven in Leonov (1997). The explicit solution is then known only in the case of $\beta = 2$.

In the context of density estimation from i.i.d. data supported on [0, 1], Korostelev and Nussbaum (1999) used a different type of estimator whose risk attains the minimax bound. They used the same kernel estimator $\hat{f}^{[\infty]}_{h_n}$ as we do at certain grid points and interpolated between these points with the aid of appropriate estimators of the derivatives up to the order $\lfloor\beta\rfloor$.

Theorem 4.2. *Suppose that Assumption 4.1 is satisfied. Let w be a continuous and monotone non-decreasing function satisfying $w(x) \leqslant C(1 + |x|^q)$ for all $x > 0$ and some $q < \infty$. Then*

$$\limsup_{n\to\infty} \sup_{f\in\mathcal{F}_\infty(\beta,L,B)} \mathrm{E}_f w((n/\log(n))^{\beta/(2\beta+1)} \|\hat{f}_{h_n}^{[\infty]} - f\|_\infty) \leqslant w(\kappa(\beta, L, B)),$$

where $\kappa(\beta, L, B)$ *was defined in Theorem 3.2 above.*

Since in both cases the lower and upper efficiency bounds coincide, it follows that certain kernel estimators are asymptotically minimax. In each case, there is a unique (up to scaling) kernel function with which the optimum can be attained. Consequently, our minimax approach defines a family of kernel functions which are optimal with respect to some well-defined criterion. In practice, one still has to opt for one of these kernel functions and to choose the bandwith on the basis of the information given by the data. We suggest using cross-validation to make both choices in a reasonable data-driven way.

5. Proofs

5.1. Proofs of the results on local asymptotic normality

Proof of Lemma 3.1. First of all, from the equality

$$\begin{aligned} &P(X_i^f \leqslant x_i | X_{i-1}^f = x_{i-1}) \\ &\quad = P(F^{-1}(F_0(X_i^{f_0})) \leqslant x_i | F^{-1}(F_0(X_{i-1}^{f_0})) = x_{i-1}) \\ &\quad = P(X_i^{f_0} \leqslant F_0^{-1}(F(x_i)) | X_{i-1}^{f_0} = F_0^{-1}(F(x_{i-1}))), \end{aligned}$$

we immediately obtain that

$$\begin{aligned} p^f(x_i|x_{i-1}) &= \frac{\mathrm{d}}{\mathrm{d}x_i} P(X_i^f \leqslant x_i | X_{i-1} = x_{i-1}) \\ &= p^{f_0}(F_0^{-1}(F(x_i)) | F_0^{-1}(F(x_{i-1}))) \frac{f(x_i)}{f_0(F_0^{-1}(F(x_i)))}. \end{aligned}$$

Analogously, with $F_u(x) = \int_{-\infty}^x f_u(y)\,\mathrm{d}y$, we obtain

$$p^{f_u}(x_i|x_{i-1}) = p^{f_0}(F_0^{-1}(F_u(x_i)) | F_0^{-1}(F_u(x_{i-1}))) \frac{f_u(x_i)}{f_0(F_0^{-1}(F_u(x_i)))}.$$

Writing $y_i = F_0^{-1}(F(x_i))$, $y_i^u = F_0^{-1}(F_u(x_i))$, $y_{i-1} = F_0^{-1}(F(x_{i-1}))$ and $y_{i-1}^u = F_0^{-1}(F_u(x_{i-1}))$, we obtain the following explicit form of the residual term:

$$R_n(x_i, x_{i-1}) = \frac{f_0(y_i)}{f_0(y_i^u)} \frac{p^{f_0}(y_i^u|y_{i-1}^u)}{p^{f_0}(y_i|y_{i-1})}$$

$$= \begin{cases} 1, & \text{if } x_i \notin \operatorname{supp}(\phi_n),\ x_{i-1} \notin \operatorname{supp}(\phi_n), \\ \dfrac{p^{f_0}(y_i|y_{i-1}^u)}{p^{f_0}(y_i|y_{i-1})}, & \text{if } x_i \notin \operatorname{supp}(\phi_n),\ x_{i-1} \in \operatorname{supp}(\phi_n), \\ \dfrac{f_0(y_i)}{f_0(y_i^u)} \dfrac{p^{f_0}(y_i^u|y_{i-1}^u)}{p^{f_0}(y_i|y_{i-1})}, & \text{if } x_i \in \operatorname{supp}(\phi_n). \end{cases} \tag{5.1}$$

We have that

$$\gamma_n := |y_i - y_i^u| + |y_{i-1} - y_{i-1}^u| = \begin{cases} O((\ln(n))^{1/2} n^{-1/2} h_n^{1/2}), & \text{if } \phi_n = \phi_{j,k,n} \\ O(n^{-1/2} h_n^{1/2}), & \text{if } \phi_n = \psi_{j,n}. \end{cases}$$

It thus follows from the construction of f_0 that

$$\frac{f_0(y_i)}{f_0(y_i^u)} = 1 + O((\ln(n))^{1/2} n^{-1/2} h_n^{1/2}) \qquad \forall x_i \in \operatorname{supp}(\phi_n). \tag{5.2}$$

Therefore, all we need to derive the desired bound for R_n are appropriate smoothness properties of the conditional densities $p^{f_0}(\cdot|\cdot)$.

From Azencott (1984, p. 478), we obtain that, for any c_1, c_2 with $c_1 > 1/\sigma_0^2 > c_2$, there exist finite positive constants K_1 and K_2 such that

$$K_1 t^{-(\alpha+1)/2} \exp\left\{-c_1 \frac{(x-y)^2}{2t}\right\} \leqslant \left|\frac{\partial^\alpha}{\partial y^\alpha} p_t^{f_0}(x|y)\right| \leqslant K_2 t^{-(\alpha+1)/2} \exp\left\{-c_2 \frac{(x-y)^2}{2t}\right\}, \tag{5.3}$$

$\alpha \in \{0, 1\}$, $t \in (0, 1]$. We also need such a result for the partial derivative with respect to x; however, we could only find the following upper bound with a certain constant c_3, not necessarily close to the desired value $1/\sigma_0^2$:

$$\left|\frac{\partial}{\partial x} p_t^{f_0}(x|y)\right| \leqslant K_3 t^{-(\alpha+1)/2} \exp\left\{-c_3 \frac{(x-y)^2}{2t}\right\}, \qquad t \in (0, 1];$$

see Friedman (1975, Section 6, Theorems 4.5 and 4.7). Indeed, the diffusion coefficient is a constant function, and the drift together and its first derivative are bounded, continuous functions and Hölder continuous of some exponent $0 < \alpha < 1$ (since the drift is supposed to be three times continuously differentiable and non-constant only on compact sets). However, putting these two estimates together, we readily obtain that

$$\begin{aligned} \left|\frac{\partial}{\partial x} p_1^{f_0}(x|y)\right| &\leqslant \int \left|\frac{\partial}{\partial x} p_t^{f_0}(x|z)\right| p_{1-t}^{f_0}(z|y)\,\mathrm{d}z \\ &\leqslant K_2 K_3 t^{-1/2} \int t^{-1/2} \exp\left\{-c_3 \frac{(x-z)^2}{2t}\right\} (1-t)^{-1/2} \exp\left\{-c_2 \frac{(z-y)^2}{2(1-t)}\right\} \mathrm{d}z. \end{aligned}$$

Choosing t sufficiently small, we obtain for any $c_2' > c_2$ that there exists a finite constant K_4 such that

$$\left|\frac{\partial}{\partial x} p_1^{f_0}(x|y)\right| \leqslant K_4 \exp\left\{-c_2' \frac{(x-y)^2}{2}\right\}. \tag{5.4}$$

Equipped with the estimates (5.3) and (5.4) we can now find estimates for the ratios involving conditional densities.

If $x_i \notin \operatorname{supp}(\phi_n)$ and $x_{i-1} \in \operatorname{supp}(\phi_n)$, it follows from (5.3) that

$$\begin{aligned}
\left|\frac{p^{f_0}(y_i|y_{i-1}^u)}{p^{f_0}(y_i|y_{i-1})} - 1\right| &\leqslant \frac{K_2 \int_{y_{i-1}-\gamma_n}^{y_{i-1}+\gamma_n} \exp\{-c_2(y_i - y)^2/2\} \mathrm{d}y}{K_1 \exp\{-c_1(y_i - y_{i-1})^2/2\}} \\
&= \frac{K_2}{K_1} \int_{y_{i-1}-\gamma_n}^{y_{i-1}+\gamma_n} \exp\{-c_2(y_i - y_{i-1} + y_{i-1} - y)^2/2 + c_1(y_i - y_{i-1})^2/2\} \mathrm{d}y \\
&\leqslant \frac{K_2}{K_1} 2\gamma_n \exp\{(c_1 - c_2)(y_i - y_{i-1})^2/2 + c_2\gamma_n |y_i - y_{i-1}|\}. \qquad (5.5)
\end{aligned}$$

If $x_i \in \operatorname{supp}(\phi_n)$ and $x_{i-1} \in \operatorname{supp}(\phi_n)$, we obtain from (5.3) and (5.4) immediately that

$$\left|\frac{p^{f_0}(y_i^u|y_{i-1}^u)}{p^{f_0}(y_i|y_{i-1})} - 1\right| = O\Big((\ln(n))^{1/2} n^{-1/2} h_n^{1/2}\Big). \tag{5.6}$$

Finally, if $x_i \in \operatorname{supp}(\phi_n)$ and $x_{i-1} \notin \operatorname{supp}(\phi_n)$, we obtain from (5.4) by complete analogy to (5.5) that

$$\left|\frac{p^{f_0}(y_i^u|y_{i-1})}{p^{f_0}(y_i|y_{i-1})} - 1\right| \leqslant \frac{K_2}{K_1} 2\gamma_n \exp\{(c_1 - c_2')(y_i - y_{i-1})^2/2 + c_2'\gamma_n |y_i - y_{i-1}|\}. \tag{5.7}$$

The assertion is now obtained from (5.1), (5.2), and (5.5)–(5.7). □

Proof of Proposition 3.1. Define $\eta_{1,n} = f_u(X_1^f)/f(X_1^f) - 1$ and, for $2 \leqslant i \leqslant n$,

$$\eta_{i,n} = \frac{p^{f_u}(X_i^f | X_{i-1}^f)}{p^f(X_i^f | X_{i-1}^f)} - 1.$$

We obtain from a Taylor series expansion of $\ln(1+x)$ that

$$\begin{aligned}
\ln \frac{p^{f_u}(X_1^f, \ldots, X_n^f)}{p^f(X_1^f, \ldots, X_n^f)} &= \sum_{i=1}^n \ln(1 + \eta_{i,n}) \\
&= \sum_{i=1}^n \eta_{i,n} - \frac{1}{2} \sum_{i=1}^n \eta_{i,n}^2 + \sum_{i=1}^n \alpha_{i,n} \eta_{i,n}^3 \\
&= u \sum_{i=1}^n \frac{\phi_n(X_i^f)}{f(X_i^f)} - \frac{u^2}{2} n \mathrm{E}_f \left(\frac{\phi_n(X_1^f)}{f(X_1^f)}\right)^2 + T_1 + T_2 + T_3,
\end{aligned}$$

say, where $|\alpha_{i,n}| < \frac{1}{3}$ and

$$T_1 = \sum_{i=2}^{n} \eta_{i,n} - u \sum_{i=2}^{n} \frac{\phi_n(X_i^f)}{f(X_i^f)},$$

$$T_2 = \frac{u^2}{2} n \mathrm{E}_f \left(\frac{\phi_n(X_1^f)}{f(X_1^f)} \right)^2 - \frac{1}{2} \sum_{i=1}^{n} \eta_{i,n}^2$$

$$T_3 = \sum_{i=1}^{n} \alpha_{i,n} \eta_{i,n}^3.$$

From Lemma 3.1 we see that

$$\eta_{i,n} - \frac{u\phi_n(X_i^f)}{f(X_i^f)} = \frac{p^{f_u}(X_i^f|X_{i-1}^f)}{p^f(X_i^f|X_{i-1}^f)} - \frac{f_u(X_i^f)}{f(X_i^f)}$$

$$= O((\ln(n))^{1/2} n^{-1/2} h_n^{1/2} \exp\{\delta(X_i^f - X_{i-1}^f)^2 + \delta_n |X_i^f - X_{i-1}^f|\}).$$

Furthermore, it is obvious that $\mathrm{E}_f[\eta_{i,n} - u\phi_n(X_i^f)/f(X_i^f)] = 0$. Therefore, we obtain by a Rosenthal inequality for absolutely regular random variables (see Doukhan 1994, Theorem 1.4.1.2) that

$$T_1 = \tilde{O}(n^{-\epsilon}, n^{-\lambda}).$$

By the Cauchy–Schwarz inequality we obtain

$$\mathrm{E}_f T_2 = \frac{u^2}{2} n \mathrm{E}_f \left(\frac{\phi_n(X_1^f)}{f(X_1^f)} \right)^2 - \frac{1}{2} \sum_{i=1}^{n} \mathrm{E}_f \left[\frac{u\phi_n(X_i^f)}{f(X_i^f)} + \left(\eta_{i,n} - \frac{u\phi_n(X_i^f)}{f(X_i^f)} \right) \right]^2$$

$$= O(n^{-\epsilon}).$$

Again, by the Rosenthal inequality, we have that

$$T_2 - \mathrm{E}_f T_2 = \tilde{O}(n^{-\epsilon}, n^{-\lambda}).$$

Since

$$\frac{f_u(x_i|x_{i-1})}{f(x_i|x_{i-1})} - 1 = \frac{u\phi_n(x_i)}{f(x_i)} + \frac{f_u(x_i)}{f(x_i)} [R(x_i, x_{i-1}) - 1]$$

we obtain, once more using the Rosenthal inequality, that

$$T_3 = \tilde{O}(n^{-\epsilon}, n^{-\lambda}),$$

which completes the proof. □

5.2. Proofs of the theorems on the lower bounds

Proof of Theorem 3.1. The basic idea of this proof is similar to that of the proof of the lower minimax bound in Golubev and Nussbaum (1990). We include it since the case of density estimation from dependent data requires some modifications.

Let $\varepsilon > 0$ be arbitrary. We actually show that

$$R_n = \inf_{\tilde{f}} \sup_{f \in \mathcal{F}_n^{(2)}} \{n^{2\beta/(2\beta+1)} \mathrm{E}_f \|\tilde{f} - f\|_{L_2}^2\} \geqslant \gamma(\beta) L^{1/(2\beta+1)} - \varepsilon, \tag{5.8}$$

for $n \geqslant n_\varepsilon$, where

$$\mathcal{F}_2^{(n)} = \{f_\theta, \theta \in \Theta_n\}$$

is an appropriate sequence of asymptotically least favourable parametric subclasses of $\mathcal{F}_2(\beta, L)$. These densities are of the form

$$f_\theta(x) = f_0(x) + \sum_{j=1}^{s} \sum_{k=1}^{q_n} \theta_{j,k} \phi_{j,k,n}(x), \tag{5.9}$$

where $\phi_{j,k,n}$ are perturbations described below.

To find an appropriate basis function f_0, we first choose a second small constant $\varepsilon' > 0$ and an arbitrarily large constant A. We start with any choice of σ_0 and μ_0 in (3.1) and (3.2), where we only assume that μ_0 is a sufficiently regular odd function. According to (3.3), we obtain that the corresponding stationary density is an even function with $f_0(x) = c_0$, for $x \in [-K, K]$, $K = -K_2 = K_3$. If we now replace $\mu^{f_0}(\cdot)$ in (3.2) by

$$\mu^{f_0'} = \begin{cases} h^{-1}\mu^{f_0}(x/h), & \text{if } x \in [-Kh, Kh], \\ h^{-1}\mu^{f_0}(x - Kh + K), & \text{if } x \geqslant Kh, \\ h^{-1}\mu^{f_0}(x + Kh - K), & \text{if } x \leqslant Kh \end{cases}$$

(take, for example, $h = c_0(2A+1)$), we obtain a stationary density f_0' satisfying

$$f_0'(x) = \begin{cases} c_0/h, & \text{if } x \in [-Kh, Kh], \\ h^{-1} f_0(x - Kh + K), & \text{if } x \geqslant Kh, \\ h^{-1} f_0(x + Kh - K), & \text{if } x \leqslant -Kh. \end{cases}$$

Hence, by choosing μ^{f_0} appropriately, we obtain a stationary density $f_0 \in W_2^\beta$ with $f_0(x) = 1/(2A+1)$, for $|x| \leqslant A$, and $\int (f_0^{(\beta)}(x))^2 \, \mathrm{d}x \leqslant \varepsilon'/4$.

The functions ϕ_j are members of the Sobolev space of periodic functions $\mathring{W}_2^\beta = \{\phi \in L_2([0, 1]) : \phi^{(\beta)} \in L_2([0, 1]), \phi^{(k)}(0) = \phi^{(k)}(1) = 0, k = 0, \ldots, \beta - 1\}$ and are solutions of the eigenvalue problem

$$(-1)^\beta \phi^{(2\beta)}(x) = \lambda \phi(x)$$

with $\mathrm{supp}(\phi) \subseteq [0, 1]$ and the boundary conditions $\phi^{(k)}(0) = \phi^{(k)}(1) = 0$, for $k = 0, \ldots, \beta - 1$. We arrange the solutions in such a way that the eigenvalues $(\lambda_j)_{j \in \mathbb{N}}$ are non-decreasing and choose the corresponding eigenfunctions $(\phi_j)_{j\in\mathbb{N}}$ such that they are orthonormal. (They are automatically so if they belong to different eigenvalues; otherwise we can use the Gram–Schmidt othonormalization algorithm.) It is known that the eigenvalues satisfy the asymptotic relation

$$\lambda_j = (\pi j)^{2\beta}(1 + o(1)) \qquad \text{as } j \to \infty;$$

see, for example, Section II.4.9 in Neumark (1960) for details. From integration by parts we obtain that

$$\int_0^1 \phi_j^{(\beta)}(x)\phi_k^{(\beta)}(x)\,\mathrm{d}x = \lambda_j \delta_{j,k} \qquad \forall j,\ k \in \mathbb{N}.$$

We define $q_n = \lfloor 2AKn^{1/(2\beta+1)} \rfloor$, where

$$K^{2\beta+1} = \frac{L - \frac{3}{4}\varepsilon'}{s^{2\beta+1}\int_0^1 b(x)(1 - b(x))\mathrm{d}x}$$

and $b(x) = [1 - (\pi x)^\beta]_+$. With $h_n = 2A/q_n$ and $a_{k,n} = (2(k-1) - q_n)A/q_n$, for $k = 1, \ldots, q_n$, we define the perturbations as

$$\phi_{j,k,n}(x) = n^{-1/2}h_n^{-1/2}\phi_j((x - a_{k,n})/h_n) - r_{j,n} f_0(x), \tag{5.10}$$

where $r_{j,n} = n^{-1/2}h_n^{1/2}\int \phi_j(x)\,\mathrm{d}x$. Note that the density f_θ belongs to the class $\mathcal{F}_2(\beta,\ L)$ if and only if the parameter $\theta = \left(\theta_{j,k}\right)_{j=1,\ldots,s;k=1,\ldots,q_n}$ is contained in

$$\Theta_n = \left\{\theta \in \mathbb{R}^{sq_n} \middle| \left(1 - \sum_{j=1}^{s}\sum_{k=1}^{q_n} r_{j,n}\theta_{j,k}\right)^2 \|f_0^{(\beta)}\|_{L_2}^2 + \sum_{j=1}^{s}\sum_{k=1}^{q_n} \theta_{j,k}^2 \lambda_j \left(\frac{q_n}{2A}\right)^{2\beta} \leqslant L\right\}. \tag{5.11}$$

The left-hand side of (5.8) will be estimated by a certain Bayesian risk which will enable us to calculate a lower efficiency bound explicitly. A sharp asymptotic risk bound will then be obtained by taking a sequence of asymptotically least favourable prior distributions. In view of available results in related settings, it could be anticipated that this can be achieved by sequences of asymptotically normal priors.

Let $\{\mu_c,\ c > 0\}$ be a family of distributions with $\mathrm{supp}(\mu_c) \subseteq [-c,\ c]$, $\int x\mu_c(x)\mathrm{d}x = 0$, $\int x^2\mu_c(\mathrm{d}x) = 1$ and $\mu_c \xrightarrow{P} \mathcal{N}(0,\ 1)$, as $c \to \infty$. Let $\mu_{c,j}$ be the distribution of a random variable $s_j Z_c$, where Z_c has law μ_c, $s_j^2 = a(j/s)/(2nA)$, $j = 1, \ldots, s$, and $a(x) = (\pi x)^{-\beta}[1 - (\pi x)^\beta]_+$. As prior measure for the parameter vector $\theta = (\theta_{1,1}, \ldots, \theta_{1,q_n}, \ldots, \theta_{s,1}, \ldots, \theta_{s,q_n})^{\mathrm{T}}$, we take the product measure $\mu_c^{(n)} = \otimes_{j=1}^s \mu_{c,j}^{\otimes q_n}$, where $\theta_{j,k} \sim \mu_{c,j}$.

Now we obtain that

$$R_n \geqslant \inf_{\tilde{f}}\left\{n^{2\beta/(2\beta+1)}\int_{\Theta_n} \mathrm{E}_{f_\theta}\|\tilde{f} - f_\theta\|_{L_2}^2 \mu_c^{(n)}(\mathrm{d}\theta)\right\} \geqslant R_{n,1} - R_{n,2}, \tag{5.12}$$

say, where

$$R_{n,1} = n^{2\beta/(2\beta+1)} \inf_{\tilde f} \int_{\mathrm{supp}(\mu_c^{(n)})} \mathrm{E}_{f_\theta} \|\tilde f - f_\theta\|_{L_2}^2 \, \mu_c^{(n)}(\mathrm{d}\theta),$$

$$R_{n,2} = n^{2\beta/(2\beta+1)} \sup_{\theta_1,\theta_2 \in \mathrm{supp}(\mu_c^{(n)})} \left\{ \|f_{\theta_1} - f_{\theta_2}\|_{L_2}^2 \right\} \mu_c^{(n)}\left(\Theta_n^c\right).$$

(The second inequality in (5.12) follows from convexity of $\{f_\theta,\, \theta \in \Theta_n\}$ which implies that the Bayes estimator lies in this set.)

Since $(q_n/(2A))^{2\beta+1}/n \leqslant K^{2\beta+1}$, we obtain that

$$\sum_{j=1}^{s} \sum_{k=1}^{q_n} \mathrm{E}_{\mu_{c,j}} \theta_{j,k}^2 \lambda_j \left(\frac{q_n}{2A}\right)^{2\beta} \leqslant s^{2\beta+1} K^{2\beta+1} \frac{1}{s} \sum_{j=1}^{s} a(j/s) \left(\pi \frac{j}{s}\right)^{2\beta}$$

$$\leqslant s^{2\beta+1} K^{2\beta+1} \frac{1}{s} \sum_{j=1}^{s} \left[1 - \left(\pi \frac{j}{s}\right)^{\beta}\right]_+ \left(\pi \frac{j}{s}\right)^{\beta},$$

which tends to $L - 3\varepsilon'/4$ as $s \to \infty$.

Hence, it follows that, for s sufficiently large,

$$\sum_{j=1}^{s} \sum_{k=1}^{q_n} \mathrm{E}_{\mu_{c,j}} \theta_{j,k}^2 \lambda_j \left(\frac{q_n}{2A}\right)^{2\beta} \leqslant L - \varepsilon'/2,$$

which implies, in conjunction with $\sum_{j=1}^{s} \sum_{k=1}^{q_n} r_{j,n} \theta_{j,k} = O_P(n^{-1/2})$, that

$$\mu_c^{(n)}\left(\Theta_n^c\right) \to 0, \qquad \text{as } n \to \infty. \tag{5.13}$$

Since $n^{2\beta/(2\beta+1)} \sup_{\theta_1,\theta_2 \in \mathrm{supp}(\mu_c^{(n)})} \|f_{\theta_1} - f_{\theta_2}\|_{L_2}^2 = O(1)$, we obtain that

$$R_{n,2} \to 0, \qquad \text{as } n \to \infty. \tag{5.14}$$

We now analyse the term $R_{n,1}$. To this end, we consider first the term

$$\tilde R_{n,1} = n^{2\beta/(2\beta+1)} \inf_{\tilde f} \int_{\mathrm{supp}(\mu_c^{(n)})} \mathrm{E}_{f_\theta} \left\| \tilde f - f_0 - \sum_{j=1}^{s} \sum_{k=1}^{q_n} \frac{\theta_{j,k}}{\sqrt{n h_n}} \phi_j \left(\frac{\cdot - a_{k,n}}{h_n}\right) \right\|_{L_2}^2 \mu_c^{(n)}(\mathrm{d}\theta).$$

Using the orthonormality of the perturbations, we obtain that

$$\tilde R_{n,1} = n^{2\beta/(2\beta+1)} \sum_{j=1}^{s} \sum_{k=1}^{q_n} \inf_{\tilde\theta_{j,k}} \int_{\mathrm{supp}(\mu_c^{(n)})} \mathrm{E}_\theta |\tilde\theta_{j,k} - \theta_{j,k}|^2 \mu_c^{(n)}(\mathrm{d}\theta)$$

$$\geqslant n^{2\beta/(2\beta+1)} q_n \sum_{j=1}^{s} \min_{1 \leqslant k \leqslant q_n} \inf_{\theta_{l,m} \in \mathrm{supp}(\mu_{c,j}), (l,m) \neq (j,k)} \inf_{\tilde\theta_{j,k}} \int_{\mathrm{supp}(\mu_c^{(n)})} \mathrm{E}_\theta |\tilde\theta_{j,k} - \theta_{j,k}|^2 \mu_{c,j}\left(\mathrm{d}\theta_{j,k}\right), \tag{5.15}$$

that is, we can reduce our considerations to a separate analysis of certain one-dimensional estimation problems.

To establish the link to Gaussian shift experiments whose analysis finally leads to explicit lower bounds, we will prove local asymptotic normality for the family of one-dimensional subexperiments given by

$$\left\{ f_u^{(j)} = f_0 + \sum_{(l,m):\,(l,m)\neq(j,k_n)} \theta_{l,m}^{(j,n)} \phi_{l,m,n} + u\phi_{j,k_n,n},\ u \in \left[-cs_j,\ cs_j\right] \right\},$$

where k_n and $(\theta_{l,m}^{(j,n)})_{(l,m)\neq(j,k_n)}$ are sequences through which the minimum is attained on the right-hand side of (5.15)

Using Proposition 3.1, we obtain that

$$\begin{aligned}\Lambda_u^{(j)} &= \log \frac{p^{f_u^{(j)}}(X_1, \ldots, X_n)}{p^{f_0^{(j)}}(X_1, \ldots, X_n)} \\ &= u\sum_{i=1}^{n} \frac{\phi_{j,k_n,n}(X_i)}{p^{f_0^{(j)}}(X_i)} - \frac{u^2}{2} n \mathrm{E}_{f_0^{(j)}} \left(\frac{\phi_{j,k_n,n}(X_1)}{p^{f_0^{(j)}}(X_1)} \right)^2 + R_n^{(j)},\end{aligned}$$

where $\sup_{|u|\leqslant u_0}\{P_{f_0^{(j)}}(|R_n^{(j)}| > \epsilon)\} \to 0$, for all u_0, $\epsilon > 0$.

Applying a central limit theorem for a triangular array of strongly mixing random variables (see Politis *et al*. 1997, Theorem A.1), we obtain that

$$\Lambda_{u/\sqrt{n}}^{(j)} \xrightarrow{d} uZ_j - \frac{u^2}{2}\frac{1}{2A+1}, \tag{5.16}$$

where $Z_j \sim \mathcal{N}(0, 1/(2A+1))$.

Now we can proceed in the same way as Golubev and Nussbaum (1990) in the proof of their Theorem A1. Because of the LAN property (5.16), we obtain, for any fixed truncation parameter c,

$$\inf_{\tilde\theta_{j,k_n}} \int_{\mathrm{supp}(\mu_{c,j})} \mathrm{E}_{\theta^{(j)}} |\tilde\theta_{j,k_n} - \theta_{j,k_n}|^2 \mu_{c,j}(\mathrm{d}\theta_{j,k_n}) \geqslant \inf_{\tilde\theta_j} \int_{\mathrm{supp}(\mu_{c,j})} \tilde{\mathrm{E}}_{\theta_j} |\tilde\theta_j - \theta_j|^2 \mu_{c,j}(\mathrm{d}\theta_j) + o_p(n^{-1}), \tag{5.17}$$

where $\theta^{(j)}$ is the parameter vector consisting of $(\theta_{l,m}^{(j,n)})_{(l,m)\neq(j,k_n)}$ and θ_{j,k_n}, and $\tilde{\mathrm{E}}_{\theta_j}$ is the expectation in a Gaussian shift experiment where

$$Y_j = \theta_j + \varepsilon_j \tag{5.18}$$

is observed and $\varepsilon_j \sim \mathcal{N}(0, 1/(n(2A+1)))$. Moreover, it follows from the arguments given in the proof of Theorem A1 of Golubev and Nussbaum (1990) that the right-hand side in (5.17) converges to the Bayesian risk for experiments (5.18) with normal priors $\mu_j \sim \mathcal{N}(0, s_j^2)$ as the truncation parameter c tends to infinity; see also Theorem 3.1 of Neumann and Spokoiny (1995). Therefore, we obtain that

$$\lim_{c\to\infty}\inf_{\tilde\theta_{j,k_n}}\int_{\operatorname{supp}(\mu_{c,j})}\mathrm{E}_{\theta^{(j)}}|\tilde\theta_{j,k_n}-\theta_{j,k_n}|^2\mu_{c,j}\big(\mathrm{d}\theta_{j,k_n}\big)\geqslant\inf_{\tilde\theta_j}\int\tilde{\mathrm{E}}_{\theta_j}|\tilde\theta_j-\theta_j|^2 p_{\mathcal{N}(0,s_j^2)}\big(\theta_j\big)\mathrm{d}\theta_j$$

$$\geqslant\frac{1-\varepsilon'/2}{nA}\,\frac{a(j/s)}{1+a(j/s)},$$

if A is sufficiently large.

Hence,

$$\begin{aligned}\tilde R_{n,1}&\geqslant\frac{1-\varepsilon'/2}{nA}q_n\sum_{j=1}^{s}\frac{a(j/s)}{1+a(j/s)}\\&\geqslant(1-\varepsilon')\frac{q_n s}{nA}\int\Big(1-(\pi x)^{\beta}\Big)_+\mathrm{d}x\\&\geqslant\gamma(\beta)L^{1/(2\beta+1)}-\varepsilon/2,\end{aligned}\tag{5.19}$$

for ε' sufficiently small and s sufficiently large.

Moreover, it is easy to see that

$$n^{2\beta/(2\beta+1)}\mathrm{E}_\theta\left(\sum_{j=1}^{s}\sum_{k=1}^{q_n}\theta_{j,k}r_{j,k,n}\right)^2\|f_0\|_{L_2}^2=O(n^{-1/(2\beta+1)}),\tag{5.20}$$

which implies, in conjunction with (5.19), that

$$R_{n,1}=\tilde R_{n,1}+O\Big(n^{-1/(4\beta+2)}\Big).\tag{5.21}$$

From (5.19) and (5.21) we obtain, for n sufficiently large,

$$R_{n,1}\;\geqslant\;\gamma(\beta)L^{1/(2\beta+1)}-\varepsilon.\tag{5.22}$$

The assertion follows from (5.12), (5.14) and (5.19). □

Proof of Theorem 3.2. Let $\varepsilon>0$ be arbitrary. We will show that

$$\inf_{\tilde f}\sup_{f\in\mathcal{F}_n^{(\infty)}}\mathrm{E}_f w((n/\log(n))^{\beta/(2\beta+1)}\|\tilde f-f\|_\infty)\geqslant(1-\varepsilon)w((1-\varepsilon)\kappa(\beta,L,B))\tag{5.23}$$

holds for all n sufficiently large. Here the class $\mathcal{F}_n^{(\infty)}\subseteq\mathcal{F}_\infty(\beta,L,B)$ is defined as

$$\mathcal{F}_{n,\infty}=\left\{f_\theta=f_0+\sum_{j=1}^{M_n}\theta_j\psi_{j,n},\,(\theta_1,\ldots,\theta_{M_n})\in\{-1,1\}^{M_n}\right\},$$

where the basic density f_0 and the perturbations $\psi_{j,n}$ are defined below.

Let $\delta_1,\delta_2\in(0,1)$ be constants such that $(1-\varepsilon)/[(1-\delta_1)(1-\delta_2)]<1$. Now we can choose a density $f_0\in\mathcal{F}_\infty(\beta,L,B)$ such that $f_0(x)=(1-\delta_1)(B\wedge B_*)$ for all $x\in[0,T]$ and some $T>0$.

According to Leonov (1997), there exists a compactly supported solution ψ_* to the optimization problem

$$\max\{g(0)|\|g\|_{L_2} \leqslant 1, |g^{(\lfloor\beta\rfloor)}(x_1) - g^{(\lfloor\beta\rfloor)}(x_2)| \leqslant |x_1 - x_2|^{\beta-\lfloor\beta\rfloor}\ \forall x_1, x_2\},$$

where, in particular, $\psi_*(0) = A_\beta$. Hence, we can choose a function ψ with compact support $[-D, D]$, satisfying the properties $\int\psi(x)\mathrm{d}x = 0$, $|\psi^{(\lfloor\beta\rfloor)}(x_1) - \psi^{(\lfloor\beta\rfloor)}(x_2)| \leqslant |x_1 - x_2|^{\beta-\lfloor\beta\rfloor}\ \forall x_1, x_2$, $\|\psi\|_{L_2} = 1$, and $\psi(0) = (1-\delta_2)A_\beta$. We now define the perturbations as

$$\psi_{j,n}(x) = \frac{1-\varepsilon}{1-\delta_2} L h_n^\beta \psi((x - b_{j,n})/h_n),$$

where $b_{j,n} = (2j-1)h_n D$, for $j = 1, \ldots, M_n$,

$$h_n = \left(\frac{2(B \wedge B_*)}{(2\beta+1)L^2}\frac{\log n}{n}\right)^{1/(2\beta+1)}$$

and

$$M_n = \left[\frac{T}{2Dh_n}\right].$$

It follows from the construction that $\mathcal{F}_{n,\infty} \subseteq \mathcal{F}_\infty(\beta, L, B)$.

We have that

$$\begin{aligned}
R_n &= \inf_{\tilde{f}} \sup_{f\in\mathcal{F}_\infty(\beta,L,B)} w((1-\varepsilon)\kappa(\beta, L, B))^{-1}\mathrm{E}_f w((n/\log(n))^{\beta/(2\beta+1)}\|\tilde{f} - f\|_\infty) \\
&\geqslant \inf_{\tilde{f}} \sup_{f\in\mathcal{F}_\infty(\beta,L,B)} P_f[(n/\log(n))^{\beta/(2\beta+1)}\|\tilde{f} - f\|_\infty \geqslant (1-\varepsilon)\kappa(\beta, L, B)] \\
&\geqslant \inf_{\tilde{f}} \sup_{f\in\mathcal{F}_{n,\infty}} P_f\left[\max_{j=1,\ldots,M_n}\{|\tilde{f}(b_{j,n}) - f(b_{j,n})|\} \geqslant (1-\varepsilon)Lh_n^\beta\psi(0)/(1-\delta_2)\right].
\end{aligned}$$

The latter inequality holds since $(\log(n)/n)^{\beta/(2\beta+1)}\kappa(\beta, L, B) = Lh_n^\beta\psi_*(0) = Lh_n^\beta\psi(0)/(1-\delta_2)$.

Since estimation of f at the points $y_{1,n}, \ldots, y_{M_n,n}$ is not harder than estimation of f in the supremum norm, we obtain that

$$\begin{aligned}
R_n &\geqslant \inf_{\tilde{\theta}} \sup_{\theta\in\{-1,1\}^{M_n}} P_\theta\left[\max_{j=1,\ldots,M_n}|\tilde{\theta}_j(X_1, \ldots, X_n) - \theta_j| \geqslant 1\right] \\
&\geqslant \inf_{\tilde{\theta}} \frac{1}{2^{M_n}} \sum_{\theta\in\{-1,1\}^{M_n}} P_\theta\left[\max_{j=1,\ldots,M_n}|\tilde{\theta}_j(X_1, \ldots, X_n) - \theta_j| \geqslant 1\right] \\
&= \inf_{\tilde{\theta}} \frac{1}{2^{M_n}} \int_{\mathbb{R}^n} \sum_{\theta\in\{-1,1\}^{M_n}} I\big(|\tilde{\theta}(x) - \theta|_\infty \geqslant 1\big)\, p^{f_\theta}(x)\,\mathrm{d}x.
\end{aligned}$$

(The latter two terms are just the Bayes risk with a corresponding product prior.) Now it is clear that the right-hand side is minimized by a maximum likelihood estimator. Hence,

$$R_n \geqslant \frac{1}{2^{M_n}} \sum_{\theta \in \{-1,1\}^{M_n}} P_\theta \left[\frac{p^{f_{\theta'}}(X_1, \ldots, X_n)}{p^{f_\theta}(X_1, \ldots, X_n)} > 1 \text{ for at least one } \theta' \in \{-1, 1\}^{M_n} \right]$$

$$\geqslant \min_{\theta \in \{-1,1\}^{M_n}} P_\theta \left[\max_{1 \leqslant l \leqslant M_n} \left\{ \log \frac{p^{f_{\theta^{(-j)}}}(X_1, \ldots, X_n)}{p^{f_\theta}(X_1, \ldots, X_n)} \right\} > 0 \right], \tag{5.24}$$

where $\theta^{(-j)} = (\theta_1, \ldots, \theta_{j-1}, -\theta_j, \theta_{j+1}, \ldots, \theta_{M_n})$. Now we obtain the desired result (5.23) if the term on the right-hand side of (5.24) tends to 1 as $n \to \infty$. To show this, we will use Proposition 3.1 which allows us to approximate the logarithmic likelihoods by

$$ALL(j) = -2\theta_j \sum_{i=1}^{n} \frac{\psi_{j,n}(X_i)}{f_\theta(X_i)} - 2n\mathrm{E}_{f_\theta} \left(\frac{\psi_{j,n}(X_1)}{f_\theta(X_1)} \right)^2.$$

According to the whitening-by-windowing principle, it can be shown that each particular term $ALL(j)$ behaves asymptotically as if the X_i were independent. In order to obtain an asymptotic approximation to the distribution of the maximum of these terms, it would be helpful if we could replace $ALL(1), \ldots, ALL(M_n)$ by independent random variables. The classic Poissonization method is, however, not applicable since the dependence between $X_1, \ldots, X_n$ does not allow it. Nevertheless, it is possible to approximate the *unordered* set $\{X_1, \ldots, X_n\}$ by a set of realizations from a Poisson process with intensity function $nf_\theta(x)$. To this end, we successively embed the observations $X_1, \ldots, X_n$ in a Poisson process N on $(0, \infty) \times \mathbb{R}$ with intensity function $p(x, y) \equiv 1$; see Neumann (1998, Section 2.2) for a detailed description. Let $(T_1, Z_1), (T_2, Z_2), \ldots$ be a realization of N, ordered such that

$$T_1/f_\theta(Z_1) \leqslant T_2/f_\theta(Z_2) \leqslant \ldots.$$

It is clear that $Z_1, Z_2, \ldots$ form a sequence of independent random variables with common density f_θ. Let $\nu = \#\{j : T_j \leqslant nf_\theta(Z_j)\}$. According to Proposition 2.1 in Neumann (1998), we can embed the random variables $X_1, \ldots, X_n$ in N such that

$$P\left(\exists j \in \{1, \ldots, M_n\} : \#\{(\{X_1, \ldots, X_n\} \triangle \{Z_1, \ldots, Z_\nu\}) \cap \mathrm{supp}(\psi_{j,n})\} > C_\lambda \sqrt{n} h_n \log(n) \right)$$
$$= O(n^{-\lambda}), \tag{5.25}$$

where $A_1 \triangle A_2$ denotes the symmetric difference of the two sets A_1 and A_2. (The coupling is organized in such a way that it may well happen that the $X_1, \ldots, X_n$ and the $Z_1, Z_2, \ldots$ appear in a different chronological order; (5.25) merely means that the *unordered* sets $\{X_1, \ldots, X_n\}$ and $\{Z_1, \ldots, Z_\nu\}$ are nearly the same, which is sufficient for our purposes.)

We obtain by Proposition 3.1 that

$$P_\theta \left(\left| \log \frac{p^{f_{\theta^{(-j)}}}(X_1, \ldots, X_n)}{p^{f_\theta}(X_1, \ldots, X_n)} - \left[-2\theta_j \sum_{i=1}^{n} \frac{\psi_{j,n}(X_i)}{f_\theta(X_i)} - 2n\mathrm{E}_f \left(\frac{\psi_{j,n}(X_1)}{f_\theta(X_1)} \right)^2 \right] \right| > C_\lambda n^{-\epsilon} \right)$$
$$= O(n^{-\lambda}). \tag{5.26}$$

Since $\|\psi_{j,n}/f_\theta\|_\infty = O(h_n^\beta)$, (5.25) implies that

$$P_{f_\theta}\left(\left|\sum_{i=1}^{n}\frac{\psi_{j,n}(X_i)}{f_\theta(X_i)}-\sum_{i=1}^{\nu}\frac{\psi_{j,n}(Z_i)}{f_\theta(Z_i)}\right| > C_\lambda(\log(n))^{3/2}h_n^{1/2}\right)=O(n^{-\lambda}). \tag{5.27}$$

Furthermore, it is clear that

$$\begin{aligned} n\mathrm{E}_{f_\theta}\left(\frac{\psi_{j,n}(X_1)}{f_\theta(X_1)}\right)^2 &= n\frac{(1-\varepsilon)^2L^2h_n^{2\beta+1}/(1-\delta_2)^2}{[(1-\delta_1)(B\wedge B_*)]^2} \\ &= \frac{(1-\varepsilon)^2}{(1-\delta_1)^2(1-\delta_2)^2}\,\frac{2\log(n)}{2\beta+1}. \end{aligned} \tag{5.28}$$

Let $Z_i^{(j)}$ be the jth member of the sequence $Z_1, Z_2, \ldots$ which falls into $\mathrm{supp}(\psi_{j,n})$, and $\nu_j = \#\{1\leqslant i\leqslant \nu : Z_i\in \mathrm{supp}(\psi_{j,n})\}$. It is clear that $\nu_j\sim \mathrm{Pois}(n\lambda_j)$, where $\lambda_j = \int_{\mathrm{supp}(\psi_{j,n})} f_\theta(x)\,\mathrm{d}x$. Since $P(|\nu_j-[n\lambda_j]| > C_\lambda\sqrt{\log(n)}\sqrt{nh_n}) = O(n^{-\lambda})$ and $Z_1^{(j)}, Z_2^{(j)}, \ldots$ form a sequence of independent random variables, we obtain by Bernstein's inequality (see, for example, Shorack and Wellner 1986 p. 855) that

$$P_\theta\left(\left|\sum_{i=1}^{\nu}\frac{\psi_{j,n}(Z_i)}{f_\theta(Z_i)}-\sum_{i=1}^{[n\lambda_j]}\frac{\psi_{j,n}(Z_i^{(j)})}{f_\theta(Z_i^{(j)})}\right| > C_\lambda(nh_n)^{-1/4}(\log(n))^{5/4}\right)=O(n^{-\lambda}). \tag{5.29}$$

By (5.26)–(5.29) we obtain that

$$\begin{aligned} &P_{f_\theta}\left(\max_{1\leqslant j\leqslant M_n}\left\{\log\frac{p^{f_{\theta(-j)}}(X_1,\ldots,X_n)}{p^{f_\theta}(X_1,\ldots,X_n)}\right\} > 0\right) \\ &\geqslant P_\theta\left(\max_{1\leqslant j\leqslant M_n}\left\{-\theta_j\sum_{i=1}^{[n\lambda_j]}\frac{\psi_{j,n}(Z_i^{(j)})}{f_\theta(Z_i^{(j)})}\right\} > \frac{(1-\varepsilon)^2}{(1-\delta_1)^2(1-\delta_2)^2}\,\frac{2\log(n)}{2\beta+1}+R_n\right) \\ &\quad +\; O(n^{-\lambda}), \end{aligned} \tag{5.30}$$

where $R_n = C_\lambda[n^{-\epsilon}+(\log(n))^{3/2}h_n^{1/2}+(nh_n)^{-1/4}(\log(n))^{5/4}]$. It is clear from the above construction that the terms $\sum_{i=1}^{[n\lambda_j]}\psi_{j,n}(Z_i^{(j)})/f_\theta(Z_i^{(j)})$, $j=1,\ldots,M_n$, are independent. Moreover, we obtain from Theorem 4 in Nagaev (1965) that

$$P_\theta\left(-\theta_j\sum_{i=1}^{[n\lambda_j]}\frac{\psi_{j,n}(Z_i^{(j)})}{f_\theta(Z_i^{(j)})} > z\frac{1-\varepsilon}{(1-\delta_1)(1-\delta_2)}\sqrt{2\log(n^{1/(2\beta+1)})}\right) = (1-\Phi(z))(1+o(1))$$

holds uniformly in $0\leqslant z\leqslant C\sqrt{\log(n)}$ for any $C<\infty$. This implies that the right-hand side of (5.30) converges to 1. Hence, we obtain the assertion. □

5.3. Proofs of the theorems on the upper bounds

Proof of Theorem 4.1. In order to prove the theorem, we use for the kernel estimator (4.1) the decomposition into bias and variance terms,

$$\mathrm{E}_f\|\hat{f}_{h_n}^{[2]}-f\|_{L_2}^2=\|\mathrm{E}_f\hat{f}_{h_n}^{[2]}-f\|_{L_2}^2+\mathrm{E}_f\|\hat{f}_{h_n}^{[2]}-\mathrm{E}_f\hat{f}_{h_n}^{[2]}\|_{L_2}^2. \tag{5.31}$$

The bias is treated analogously to the case of independent data, that is,

$$\begin{aligned}\|\mathrm{E}_f\hat{f}_{h_n}^{[2]}-f\|_{L_2}^2&=\frac{1}{2\pi}\|F\left(\mathrm{E}_f\hat{f}_{h_n}^{[2]}\right)-F(f)\|_{L_2}^2\\&=\frac{1}{2\pi}\|\left(F\left(K_{\beta,2}\right)(h_n\cdot)-1\right)F(f)\|_{L_2}^2\\&\leqslant\frac{1}{2\pi}\int|h_n\omega|^{2\beta}|F(f)(\omega)|^2\mathrm{d}\omega\leqslant Lh_n^{2\beta},\end{aligned} \tag{5.32}$$

where we have used first the Plancherel formula for Fourier transforms and then the expression for the kernel.

In the variance term, there are covariances that do not appear in the independent case. Nevertheless, due to the weak dependence of data, the dominating term is the same as in the independent case. We have

$$\begin{aligned}\mathrm{E}_f\|\hat{f}_{h_n}^{[2]}-\mathrm{E}_f\hat{f}_{h_n}^{[2]}\|_{L_2}^2&=\frac{1}{n}\mathrm{E}_f\int\left(\frac{1}{h_n}K_{\beta,2}\left(\frac{X_1-x}{h_n}\right)-\mathrm{E}_f\frac{1}{h_n}K_{\beta,2}\left(\frac{X_1-x}{h_n}\right)\right)^2\mathrm{d}x\\&\quad+\frac{1}{(nh_n)^2}\sum_{1\leqslant|i-j|\leqslant C_\lambda\log(n)}\int\mathrm{cov}_f\left(K_{\beta,2}\left(\frac{X_i-x}{h_n}\right),\,K_{\beta,2}\left(\frac{X_j-x}{h_n}\right)\right)\mathrm{d}x\\&\quad+\frac{1}{(nh_n)^2}\sum_{|i-j|>C_\lambda\log(n)}\int\mathrm{cov}_f\left(K_{\beta,2}\left(\frac{X_i-x}{h_n}\right),\,K_{\beta,2}\left(\frac{X_j-x}{h_n}\right)\right)\mathrm{d}x\\&=T_1+T_2+T_3,\end{aligned} \tag{5.33}$$

say. We have

$$T_1=\frac{1}{n}\left(\frac{\|K_{\beta,2}\|_{L_2}^2}{h_n}-\frac{1}{2\pi}\|F(K_{\beta,2})(h_n\cdot)F(f)\|_{L_2}^2\right)\leqslant\frac{\|K_{\beta,2}\|_{L_2}^2}{nh_n}. \tag{5.34}$$

Now $\|K_{\beta,2}\|_\infty\leqslant(2\pi)^{-1}\int(1-|\omega|^\beta)_+\,\mathrm{d}\omega\leqslant 1/\pi$ implies in conjunction with (4.2) that $\int|K_{\beta,2}(x)|\,\mathrm{d}x<\infty$. Therefore, we obtain

$$\begin{aligned}T_2&\leqslant\frac{1}{(nh_n)^2}\sum_{1\leqslant|i-j|\leqslant C_\lambda\log(n)}\iiint\left|K_{\beta,2}\left(\frac{x_i-x}{h_n}\right)K_{\beta,2}\left(\frac{x_j-x}{h_n}\right)\right|\\&\qquad\times\left[p^f(x_j|x_i)+p^f(x_j)\right]p^f(x_i)\,\mathrm{d}x_i\,\mathrm{d}x_j\,\mathrm{d}x\\&=O(\log(n)/n).\end{aligned} \tag{5.35}$$

Define

$$H_n(y, z) = \int \left[K_{\beta,2}\left(\frac{y-x}{h_n}\right) - \mathrm{E}_f K_{\beta,2}\left(\frac{X_1 - x}{h_n}\right)\right]\left[K_{\beta,2}\left(\frac{z-x}{h_n}\right) - \mathrm{E}_f K_{\beta,2}\left(\frac{X_1 - x}{h_n}\right)\right] \mathrm{d}x.$$

We can replace X_j by some X'_j which has the same distribution as X_j, is independent of X_i, and satisfies $P(X_j \neq X'_j) \leqslant \beta(|i-j|)$. Since $\mathrm{E}_f H_n(X_i, X'_j) = 0$ and $\sup_{y,z}\{|H_n(y, z)|\} = O(1)$ we obtain

$$T_3 = \frac{1}{(nh_n)^2} \sum_{|i-j| > C_\lambda \log(n)} \mathrm{E}_f[H_n(X_i, X_j) - H_n(X_i, X'_j)]$$

$$= O(n^{-\lambda}), \tag{5.36}$$

provided C_λ is sufficiently large. Hence, we obtain from (5.33)–(5.36) that

$$\mathrm{E}_f \|\hat{f}^{[2]}_{h_n} - \mathrm{E}_f \hat{f}^{[2]}_{h_n}\|^2_{L_2} \leqslant \frac{\|K_{\beta,2}\|^2_{L_2}}{nh_n} + O\left(\frac{\log(n)}{n}\right). \tag{5.37}$$

Since $\|K_{\beta,2}\|^2_{L_2} = \|F(K_{\beta,2})\|^2_{L_2}/(2\pi) = 2\beta^2/(\pi(\beta+1)(2\beta+1))$, we obtain

$$Lh_n^{2\beta} + \frac{\|K_{\beta,2}\|^2_{L_2}}{nh_n} = \left(\frac{\beta n^{-1}}{\pi L(\beta+1)(2\beta+1)}\right)^{2\beta/(2\beta+1)} (L + 2\beta L) = L^{1/(2\beta+1)}\gamma(\beta) n^{-2\beta/(2\beta+1)},$$

which yields the assertion in conjunction with (5.31), (5.32), and (5.37). □

Proof of Theorem 4.2. Let $\varepsilon > 0$ be arbitrary. We will actually prove that, for $r_n = (\log(n)/n)^{\beta/(2\beta+1)}$,

$$\limsup_{n\to\infty} \sup_{f \in \mathcal{F}_\infty(\beta, L, B)} \mathrm{E}_f w(r_n^{-1}\|\hat{f}^{[\infty]}_{h_n} - f\|_\infty) \leqslant w((1+\varepsilon)\kappa(\beta, L, B)) \tag{5.38}$$

holds for $n \geqslant n_0(\varepsilon)$. The fact that we have the additional factor $1+\varepsilon$ at our disposal facilitates our task essentially: we are therefore not forced to undertake some sort of 'exact calculations', and can actually apply rough exponential estimates in conjunction with a suitable chaining technique.

We divide the constant $\kappa(\beta, L, B)$ describing the asymptotic size of $r_n^{-1}\|\hat{f}^{[\infty]}_{h_n} - f\|_\infty$ into a term caused by stochastic fluctuations and a bias term, that is,

$$\kappa(\beta, L, B) = r_n^{-1}\sigma_{(n)}\sqrt{2\log(1/h_n)} + r_n^{-1} Lh_n^\beta B(\beta),$$

where

$$\sigma_{(n)} = \frac{\sqrt{B \wedge B_*}}{\sqrt{nh_n}} \|K_{\beta,\infty}\|_2.$$

Indeed, for all $f \in \mathcal{F}_\infty(\beta, L, B)$,

$$r_n^{-1}\|\mathrm{E}_f \hat{f}^{[\infty]}_{h_n} - f\|_\infty \leqslant Lh_n^\beta B(\beta), \tag{5.39}$$

where

$$B(\beta) = \sup_{g \in \mathcal{F}_\infty(\beta,1,B)} \left| \int K_{\beta,\infty}(u)[g(u) - g(0)]\mathrm{d}u \right| \tag{5.40}$$

and $\|K_{\beta,\infty}\|_2 + B(\beta) = A_\beta$. We have, for some ε' to be chosen below,

$$\begin{aligned}
&\mathrm{E}_f w(r_n^{-1}\|\hat{f}_{h_n}^{[\infty]} - f\|_\infty) \\
&\quad \leqslant w((1+\varepsilon)\kappa(\beta, L, B))P_f(\|\hat{f}_{h_n}^{[\infty]} - \mathrm{E}_f\hat{f}_{h_n}^{[\infty]}\|_\infty \leqslant (1+\varepsilon)\sigma_{(n)}\sqrt{2\log(1/h_n)}) \\
&\qquad + w((1+\varepsilon')\kappa(\beta, L, B)) \\
&\qquad P_f((1+\varepsilon)\sigma_{(n)}\sqrt{2\log(1/h_n)} < \|\hat{f}_{h_n}^{[\infty]} - \mathrm{E}_f\hat{f}_{h_n}^{[\infty]}\|_\infty \leqslant (1+\varepsilon')\sigma_{(n)}\sqrt{2\log(1/h_n)}) \\
&\qquad + w(r_n^{-1}[h_n^{-1}\|K_{\beta,\infty}\|_\infty + Lh_n^\beta B(\beta)]) \\
&\qquad P_f(\|\hat{f}_{h_n}^{[\infty]} - \mathrm{E}_f\hat{f}_{h_n}^{[\infty]}\|_\infty > (1+\varepsilon')\sigma_{(n)}\sqrt{2\log(1/h_n)})
\end{aligned} \tag{5.41}$$

According to this decomposition, we will prove (5.38), and hence the assertion of the theorem, by deriving exponential inequalities for $r_n^{-1}\|\hat{f}_{h_n}^{[\infty]} - \mathrm{E}_f\hat{f}_{h_n}^{[\infty]}\|_\infty$.

Since the support of f is not necessarily restricted to a set of bounded size, we first reduce the problem to the supremum deviation on a certain set with appropriately bounded Lebesgue measure. This boundedness will be used later on; see the derivation of (5.46) below. To this end, we consider overlapping intervals

$$I_k = [kh_n, (k+1)h_n) \oplus \mathrm{supp}(K_{\beta,\infty}(\cdot/h_n))$$

and restrict our primary attention to the set

$$\mathcal{X}_n = \bigcup_{k \in \mathcal{I}_n} [kh_n, (k+1)h_n), \tag{5.42}$$

where

$$\mathcal{I}_n = \{k : P_f(X_i \in I_k) \geqslant n^{-1}\}.$$

We decompose the set of remaining indices $\mathbb{Z}\backslash\mathcal{I}_n$ into disjoint subsets $\mathcal{J}_1, \ldots, \mathcal{J}_{c_n}$ such that

$$n^{-1} \leqslant P_f\left(X_i \in \bigcup_{j \in \mathcal{J}_k} I_j\right) < 2n^{-1}.$$

Using the Bernstein-type inequality given in Lemma 5.1, we obtain

$$P\left(\#\left\{i : X_i \in \bigcup_{j \in \mathcal{J}_k} I_j\right\} > C_\lambda \log(n)\right) = O(n^{-\lambda}),$$

which implies, by $c_n = O(n^\kappa)$, that

$$P\left(\sup_{x \in \mathbb{R}\backslash\mathcal{X}_n} \{|\hat{f}_{h_n}^{[\infty]}(x) - \mathrm{E}_f\hat{f}_{h_n}^{[\infty]}(x)|\} > C_\lambda (nh_n)^{-1}\log(n)\right) = O(n^{-\lambda}). \tag{5.43}$$

To obtain probabilistic bounds for the supremum deviation on $\mathcal{X}_n$, we apply a simple chaining technique based on two grids, a coarse one with grid size $g_{n,1}$ close to h_n, and a fine one with grid size $g_{n,2} = g_{n,1}/K_n$, for some integer K_n; that is, we define

$$x_j = g_0 + j g_{n,1}, \qquad j \in \mathbb{Z},$$

$$x_{j,k} = x_j + k g_{n,2}, \qquad j \in \mathbb{Z},\ k = 1, \ldots, K_n.$$

Before we apply Lemma 5.1 to $|\hat{f}^{[\infty]}_{h_n}(x_j) - \mathrm{E}_f \hat{f}^{[\infty]}_{h_n}(x_j)|$, we derive an upper estimate for

$$\sigma_x^2 = n \max_{i,m} \left\{ \frac{1}{m} \operatorname{var}\left(\frac{1}{nh_n} \left[K_{\beta,\infty}\left(\frac{x - X_i}{h_n} \right) + \ldots + K_{\beta,\infty}\left(\frac{x - X_{i+m-1}}{h_n} \right) \right] \right) \right\},$$

which determines the constant in the exponent of the Bernstein-type inequality given in Lemma 5.1. Analogously to (5.37), we obtain the pointwise estimate

$$\sigma_x^2 = \frac{1}{nh_n^2} \|K_{\beta,\infty}\|_2^2 \cdot \|f\|_\infty + O(\log(n)/n) \tag{5.44}$$

as well as the estimate

$$\int_{-\infty}^{\infty} \sigma_x^2 \, \mathrm{d}x = \frac{1}{nh_n^2} \|K_{\beta,\infty}\|_2^2 + O(\log(n)/n).$$

The latter relation implies that there exists a g_0 such that

$$\sum_{j \in \mathbb{Z}} \sigma_{x_j}^2 \leqslant C \frac{1}{n g_{n,1}}.$$

Furthermore,

$$\#\{j : x_j \in \mathcal{X}_n\} = O(n^\kappa),$$

for some $\kappa < \infty$. We obtain by Lemma 5.1 that, for $n \geqslant n_0$,

$$P\left(\max_{x_j \in \mathcal{X}_n} |\hat{f}^{[\infty]}_{h_n}(x_j) - \mathrm{E}_f \hat{f}^{[\infty]}_{h_n}(x_j)| > u(1 + \varepsilon/3)\sigma_{(n)} \sqrt{2 \log(1/h_n)} \right)$$

$$\leqslant \sum_{j:\, x_j \in \mathcal{X}_n} 4 \exp\{-T_n\} + O(n^{-\lambda}), \tag{5.45}$$

where

$$T_n = \frac{(1-\delta)(1+\varepsilon/3)^2 u^2 \sigma_{(n)}^2 \log(1/h_n)}{\sigma_{x_j}^2 + C_{\lambda,\delta} u (1+\varepsilon/3) \sigma_{(n)} \sqrt{2\log(1/h_n)} \log(n)/(nh_n)}.$$

Since $\#\{j : x_j \in \mathcal{X}_n\} = O(n^\kappa)$ we can use for those j where σ_{x_j} is below a certain threshold the estimate

$$\sum_{j:\, x_j \in \mathcal{X}_n \text{ and } \sigma_{x_j} \leqslant C_3 \sigma_{(n)}^2} \exp\{-T_n\} = O(n^\kappa \exp\{-C_4 \log(n)\}) = O(n^{-\lambda}), \tag{5.46}$$

provided the constant C_3 is sufficiently small. For $\sigma_{x_j}^2 > C_3\sigma_{(n)}^2$, we have that

$$\begin{aligned}\sigma_{x_j}^2 &+ C_{\lambda,\delta}u(1+\varepsilon/3)\sigma_{(n)}\sqrt{2\log(1/h_n)}\log(n)/(nh_n)\\ &\leqslant \sqrt{1+\varepsilon/3}\sigma_{x_j}^2 \leqslant (1+\varepsilon/3)\sigma_{(n)}^2 \leqslant (1-\delta)(1+\varepsilon/3)^2u^2\sigma_{(n)}^2\log(1/h_n).\end{aligned}$$

This implies, in conjunction with the inequality $(1/x)\exp(-c/x) \leqslant (1/x_0)\exp(-c/x_0)$ which holds for $x \leqslant x_0 \leqslant c$, that

$$\begin{aligned}\sum_{j:\, x_j\in\mathcal{X}_n \text{ and } \sigma_{x_j}>C_3\sigma_{(n)}^2} \exp\{-T_n\} &= O\left(\exp\{-(1-\delta)(1+\varepsilon/3)u^2\log(1/h_n)\}\sum_{j:\, x_j\in\mathcal{X}_n}\frac{\sigma_{x_j}^2}{\sigma_{(n)}^2}\right)\\ &= O\left(\frac{h_n}{g_{n,1}}h_n^{-(1-\delta)(1+\varepsilon/3)u^2}\right). \qquad (5.47)\end{aligned}$$

We choose δ such that $(1-\delta)(1+\varepsilon/3) > 1$. Now we obtain from (5.45)–(5.47) that

$$P\left(\max_{x_j\in\mathcal{X}_n}|\hat{f}_{h_n}^{[\infty]}(x_j) - \mathrm{E}_f\hat{f}_{h_n}^{[\infty]}(x_j)| > (1+\varepsilon/3)\sigma_{(n)}\sqrt{2\log(1/h_n)}\right) = o(1) \qquad (5.48)$$

if $K_n = [n^\kappa]$ and $\kappa > 0$ is sufficiently small. Furthermore, we have

$$P\left(\max_{x_j\in\mathcal{X}_n}|\hat{f}_{h_n}^{[\infty]}(x_j) - \mathrm{E}_f\hat{f}_{h_n}^{[\infty]}(x_j)| > \frac{1+\varepsilon'}{1+\varepsilon}(1+\varepsilon/3)\sigma_{(n)}\sqrt{2\log(1/h_n)}\right) = O(n^{-\lambda}) \qquad (5.49)$$

if ε' is sufficiently large.

Since $|x_j - x_{j,k}| \leqslant g_{n,1}$ and $K_{\beta,\infty}$ is Lipschitz of order β, we obtain that

$$\left\|K_{\beta,\infty}\left(\frac{x_j-\cdot}{h_n}\right) - K_{\beta,\infty}\left(\frac{x_{j,k}-\cdot}{h_n}\right)\right\|_\infty = O\big((g_{n,1}/h_n)^{\beta\wedge 1}\big) \qquad (5.50)$$

as well as

$$\left\|K_{\beta,\infty}\left(\frac{x_j-\cdot}{h_n}\right) - K_{\beta,\infty}\left(\frac{x_{j,k}-\cdot}{h_n}\right)\right\|_2^2 = O\big(h_n(g_{n,1}/h_n)^{2(\beta\wedge 1)}\big).$$

The latter relation implies, by complete analogy with the derivation of (5.37), that

$$\begin{aligned}&\frac{1}{nh_n^2}\sum_{i'=1}^n\left|\operatorname{cov}\left(K_{\beta,\infty}\left(\frac{x_j-X_i}{h_n}\right) - K_{\beta,\infty}\left(\frac{x_{j,k}-X_i}{h_n}\right), K_{\beta,\infty}\left(\frac{x_j-X_{i'}}{h_n}\right) - K_{\beta,\infty}\left(\frac{x_{j,k}-X_{i'}}{h_n}\right)\right)\right|\\ &= O\left(\frac{1}{nh_n}(g_{n,1}/h_n)^{2(\beta\wedge 1)} + \frac{1}{\log(n)}(g_{n,1}/h_n)^{2(\beta\wedge 1)} + n^{-\lambda}\right). \qquad (5.51)\end{aligned}$$

From (5.50), (5.51), and the fact that $\#\{(j, k) : x_{j,k} \in \mathcal{X}_n\} = O(n^\gamma)$ for some fixed $\gamma > 0$, we obtain by Lemma 5.1 that

$$P\left(\max_{x_{j,k}\in\mathcal{X}_n}|[\hat{f}^{[\infty]}_{h_n}(x_{j,k}) - \mathrm{E}_f\hat{f}^{[\infty]}_{h_n}(x_{j,k})] - [\hat{f}^{[\infty]}_{h_n}(x_j) - \mathrm{E}_f\hat{f}^{[\infty]}_{h_n}(x_j)]| > \frac{\varepsilon}{3}\sigma_{(n)}\sqrt{2\log(1/h_n)}\right)$$
$$= O(n^{-\lambda}). \tag{5.52}$$

Finally, for $x \in [x_{j,k}, x_{j,k+1}]$, we have the non-stochastic estimate

$$|\hat{f}^{[\infty]}_{h_n}(x) - \hat{f}^{[\infty]}_{h_n}(x_{j,k})| = O\big(h_n^{-1}(g_{n,2}/h_n)^{2(\beta\wedge 1)}\big) \leqslant \frac{\varepsilon}{3}\sigma_{(n)}\sqrt{2\log(1/h_n)} \tag{5.53}$$

if $g_{n,2} = O(n^{-\kappa})$ and $n \geqslant n_0$. Putting all things together, we obtain that

$$P\left(\sup_{x\in\mathbb{R}}\{|\hat{f}^{[\infty]}_{h_n}(x) - \mathrm{E}_f\hat{f}^{[\infty]}_{h_n}(x)|\} > (1+\varepsilon)\sigma_{(n)}\sqrt{2\log(1/h_n)}\right) = o(1) \tag{5.54}$$

and

$$P\left(\sup_{x\in\mathbb{R}}\{|\hat{f}^{[\infty]}_{h_n}(x) - \mathrm{E}_f\hat{f}^{[\infty]}_{h_n}(x)|\} > (1+\varepsilon')\sigma_{(n)}\sqrt{2\log(1/h_n)}\right) = O(n^{-\lambda}), \tag{5.55}$$

provided ε' is sufficiently large. (5.38) now follows from (5.39), (5.41), (5.54) and (5.55). □

5.4. A Bernstein-type inequality

The following lemma is adapted from Doukhan (1994, Theorem 1.4.2.4).

Lemma 5.1. *Let $X_1, \ldots, X_n$ be geometrically β-mixing. Then, for arbitrary $\lambda < \infty$ and $\delta > 0$, there exist $X'_1, \ldots, X'_n$ with*

$$P((X_1, \ldots, X_n) \neq (X'_1, \ldots, X'_n)) = O(n^{-\lambda})$$

such that

$$P\left(\left|\sum_{i=1}^{n} g(X'_i)\right| \geqslant u\right) \leqslant 4\exp\left\{-\frac{(1-\delta)u^2}{2(n\sigma_g^2 + uC_{\lambda,\delta}\log(n)\|g\|_\infty)}\right\}$$

holds for all functions g with $\mathrm{E}g(X_i) = 0$ for all i, where

$$\sigma_g^2 = \max_{i,m}\left\{\frac{1}{m}\mathrm{E}(g(X_i) + \ldots + g(X_{i+m-1}))^2\right\}.$$

Note that there are two differences from Doukhan's formulation. First, we make explicitly clear that there exists some universal substitution of the sample $X_1, \ldots, X_n$ by $X'_1, \ldots, X'_n$ which does not depend on the particular function g. Therefore, the term of order $n^{-\lambda}$ occurs only once even when we apply the Bernstein-type inequality simultaneously to several different g. Second, in Doukhan's formulation, the term which plays the role of our σ_g^2 was bounded from above by $\max_{i,m}\{(1/m)\mathrm{E}(g(X_i) + \ldots + g(X_{i+m}))^2\}$, that is, the variance of $m+1$ successive observations was divided by m. Our slightly different formulation is actually crucial in our context, since we focus on the exact asymptotic constant.

Proof of Lemma 5.1. The proof is of course similar to that in Doukhan (1994) and is only included for completeness.

The basic reason why we obtain (up to the factor $1-\delta$) basically the same constant in the exponent as in the case of independent random variables is that we split $X_1, \ldots, X_n$ into alternating large and small blocks. Let $l_n = [C_{\lambda,1} \log(n)]$ be the length of the large blocks and $s_n = [C_{\lambda,2} \log(n)]$ be the length of the small blocks ($C_{\lambda,1} \geqslant C_{\lambda,2}$). (An appropriate choice of these constants is described below.)

We define sets of indices

$$\mathcal{J}_k^{(l)} = \{(k-1)(l_n+s_n)+1, \ldots, (kl_n+(k-1)s_n) \wedge n\},$$

$$\mathcal{J}_k^{(s)} = \{kl_n+(k-1)s_n+1, \ldots, (k(l_n+s_n)) \wedge n\}.$$

According to Lemma 2 in Doukhan *et al.* (1995), we can replace the large blocks $(X_i, i \in \mathcal{J}_k^{(l)})$ by independent blocks $(X_i', i \in \mathcal{J}_k^{(l)})$ such that

$$(X_i, i \in \mathcal{J}_k^{(l)}) \stackrel{d}{=} (X_i{}', i \in \mathcal{J}_k^{(l)})$$

and

$$P((X_i, i \in \mathcal{J}_k^{(l)}) \neq (X_i', i \in \mathcal{J}_k^{(l)})) \leqslant \beta(s_n+1).$$

Analogously, we can replace the small blocks $(X_i, i \in \mathcal{J}_k^{(s)})$ by independent blocks $(X_i', i \in \mathcal{J}_k^{(s)})$ such that

$$(X_i, i \in \mathcal{J}_k^{(s)}) \stackrel{d}{=} (X_i', i \in \mathcal{J}_k^{(s)})$$

and

$$P((X_i, i \in \mathcal{J}_k^{(s)}) \neq (X_i', i \in \mathcal{J}_k^{(s)})) \leqslant \beta(l_n+1).$$

We choose $C_{\lambda,2}$ such that $\beta([C_{\lambda,2} \log(n)]+1) = O(n^{-\lambda-1})$, which implies that

$$P((X_1, \ldots, X_n) \neq (X_1', \ldots, X_n')) = O(n^{-\lambda}). \tag{5.56}$$

Applying Bernstein's inequality for independent random variables (see, for example, Shorack and Wellner 1986, p. 855) we obtain, with $n^{(l)} = \sum_k \#\mathcal{J}_k^{(l)}$ and $n^{(s)} = \sum_k \#\mathcal{J}_k^{(s)}$, that

$$P\left(\left|\sum_k \sum_{i \in \mathcal{J}_k^{(l)}} g(X_i')\right| \geqslant v\right) \leqslant 2\exp\left\{-\frac{v^2}{2(n^{(l)}\sigma_g^2 + vl_n\|g\|_\infty/3)}\right\} \tag{5.57}$$

and

$$P\left(\left|\sum_k \sum_{i\in\mathcal{J}_k^{(s)}} g(X_i')\right| \geqslant w\right) \leqslant 2\exp\left\{-\frac{w^2}{2(n^{(s)}\sigma_g^2 + ws_n\|g\|_\infty/3)}\right\}. \tag{5.58}$$

We choose $C_{\lambda,1}$ such that

$$1-\frac{\delta}{2}=\frac{(\sqrt{C_{\lambda,1}}+\sqrt{C_{\lambda,2}})^2}{C_{\lambda,1}+C_{\lambda,2}}.$$

This implies that

$$\delta_n = 1-\frac{(\sqrt{n^{(l)}}+\sqrt{n^{(s)}})^2}{n}\leqslant\delta$$

holds for $n\geqslant n_0(\lambda,\delta)$. Therefore, we obtain from (5.57) and (5.58), with $v = u\sqrt{1-\delta_n}\sqrt{n^{(l)}}/\sqrt{n}$ and $w = u\sqrt{1-\delta_n}\sqrt{n^{(s)}}/\sqrt{n}$, that

$$\begin{aligned}
P\left(\left|\sum_{i=1}^n g(X_i')\right| \geqslant u\right) &\leqslant 2\exp\left\{-\frac{u^2(1-\delta_n)n^{(l)}/n}{2(n^{(l)}\sigma_g^2 + ul_n\|g\|_\infty\sqrt{1-\delta_n}(\sqrt{n^{(l)}}/\sqrt{n})/3)}\right\}\\
&\quad + 2\exp\left\{-\frac{u^2(1-\delta_n)n^{(s)}/n}{2(n^{(s)}\sigma_g^2 + us_n\|g\|_\infty\sqrt{1-\delta_n}(\sqrt{n^{(s)}}/\sqrt{n})/3)}\right\}\\
&\leqslant 4\exp\left\{-\frac{u^2(1-\delta)}{2(n\sigma_g^2 + uC_{\lambda,\delta}\|g\|_\infty)}\right\}.
\end{aligned}$$

□

Acknowledgement

We thank Emmanuel Gobet, Michael Nussbaum and Wolfgang Härdle for helpful discussions. Both authors gratefully acknowledge the hospitality and financial support, respectively, of the SFB 373 at Humboldt University, Berlin and of Modal'X at Paris X University, Nanterre.

References

Aït-Sahalia, Y. (1996a) Nonparametric pricing of interest rate derivative securities. *Econometrica*, **64**, 527–560.

Aït-Sahalia, Y. (1996b) Testing continuous-time models of the spot interest rate. *Rev. Financial Stud.*, **9**, 385–426.

Artiles, L.M. (2001) Adaptive minimax estimation in classes of smooth functions. Doctoral thesis, University of Utrecht.

Azencott, R. (1984) Densité des diffusions en temps petit: développements asymptotiques. I. In J. Azéma and M. Yor (eds), *Séminaire de Probabilités XVIII*, Lecture Notes in Math. 1059, pp. 402–498. Berlin: Springer-Verlag.

Banon, G. (1978) Nonparametric identification for diffusion processes. *SIAM J. Control Optim.*, **16**, 380–395.

Banon, G. and Nguyen, H.T. (1981) Recursive estimation in diffusion model. *SIAM J. Control Optim.*, **19**, 676–685.

Belitser, E. (1998) Efficient estimation of analytic density under random censorship. *Bernoulli*, **4**, 519–543.

Belitser, E.N. and Levit, B.Y. (1995) On minimax filtering over ellipsoids. *Math. Methods Statist.*, **4**, 259–273.

Bosq, D. (1998) *Nonparametric Statistics for Stochastic Processes.* New York: Springer-Verlag.

Castellana, J.V. and Leadbetter, M.R. (1986) On smoothed probability density estimation for stationary processes. *Stochastic Process. Appl.*, **21**, 179–193.

Dalalyan, A.S. and Kutoyants, Yu.A. (2002) Asymptotically efficient trend coefficient estimation for ergodic diffusion. *Math. Methods Statist.*, **11**, 402–427.

Donoho, D.L. (1994) Asymptotic minimax risk for sup-norm loss: Solution via optimal recovery. *Probab. Theory Related Fields*, **99**, 145–170.

Doukhan, P. (1994) *Mixing: Properties and Examples*, Lecture Notes in Statist. 85. New York: Springer-Verlag.

Doukhan, P., Massart, P. and Rio, E. (1995) Invariance principles for absolutely regular empirical processes. *Ann. Inst. H. Poincaré Probab. Statist.*, **31**, 393–427.

Efromovich, S.Yu. and Pinsker, M.S. (1982) Estimating a square integrable spectral density from a sequence of observations. *Problems Inform. Transmission*, **17**, 182–196.

Efromovich, S.Yu. and Pinsker, M.S. (1983) Estimation of square-integrable probability density of a random variable. *Problems Inform. Transmission*, **18**, 175–189.

Friedman, A. (1975) *Stochastic Differential Equations and Applications. Vol. 1*. New York: Academic Press.

Gobet, E. (2002) LAN property for ergodic diffusions with discrete observations. *Ann. Inst. H. Poincaré Probab. Statist.*, **38**, 711–737.

Gobet, E., Hoffmann, M. and Reiss, M. (2002) Nonparametric estimation of scalar diffusions based on low frequency data is ill-posed. Preprint 747, LPMA.

Golubev, G.K. (1987) Adaptive asymptotically minimax estimators of smooth signals. *Problems Inform. Transmission*, **23**, 47–55.

Golubev, G.K. and Levit, B.Y. (1996) Asymptotically efficient estimation for analytic distributions. *Math. Methods Statist.*, **5**, 357–368.

Golubev, G.K. and Nussbaum, M. (1990) A risk bound in Sobolev class regression. *Ann. Statist.*, **18**, 758–778.

Golubev, Y.K., Levit, B.Y. and Tsybakov, A.B. (1996) Asymptotically efficient estimation of analytic functions in Gaussian noise. *Bernoulli*, **2**, 167–181.

Guerre, E. and Tsybakov, A.B. (1998) Exact asymptotic minimax constants for the estimation of analytical functions in L_p. *Probab. Theory Related Fields*, **112**, 33–51.

Hansen, L.P., Scheinkman, J.A. and Touzi, N. (1998) Spectral methods for identifying scalar diffusions. *J. Econometrics*, **86**, 1–32.

Hoffmann, M. (1999) L_p estimation of the diffusion coefficient. *Bernoulli*, **5**, 447–481.

Ibragimov, I. (2001) Estimation of analytic functions. In M. de Gunst, C. Klaassen and A. van der Vaart (eds), *State of the Art in Probability and Statistics: Festschrift for Willem R. van Zwet*, IMS

Lecture Notes Monogr. Ser., 36, pp. 359–383. Beachwood, OH: Institute of Mathematical Statistics.

Ibragimov, I.A. and Hasminskii, R.Z. (1983) Estimation of distribution density. *J. Soviet Math.*, **21**, 40–57.

Karlin, S. and Taylor, H.M. (1981) *A Second Course in Stochastic Processes*. Orlando, FL: Academic Press.

Kessler, M. (1997) Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.*, **24**, 211–229.

Kessler, M. and Sørensen, M. (1999) Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, **5**, 299–314.

Korostelev, A.P. (1993) An asymptotically minimax regression estimator in the uniform norm up to exact constant. *Theory Probab. Appl.*, **38**, 737–743.

Korostelev, A.P. (1996) A minimaxity criterion in nonparametric regression based on large-deviations probabilities. *Ann. Statist.*, **24**, 1075–1083.

Korostelev, A.P. and Leonov, S.L. (1996) Minimax Bahadur efficiency for small confidence levels. *Problemy Peredachi Informatsii*, **32**, 3–15 (in Russian) [English translation: *Problems Inform. Transmission*, **32**, 303–313.]

Korostelev, A.P. and Nussbaum, M. (1999) The asymptotic minimax constant for sup-norm loss in nonparametric density estimation. *Bernoulli*, **5**, 1099–1118.

Kutoyants, Yu.A. (1997) Some problems of nonparametric estimation by observations of ergodic diffusion process. *Statist. Probab. Lett.*, **32**, 311–320.

Leblanc, F. (1997) Density estimation for a class of continuous time processes. *Math. Methods Statist.*, **6**, 171–199.

Leonov, S.L. (1997) On the solution of an optimal recovery problem and its applications in nonparametric regression. *Math. Methods Statist.*, **6**, 476–490.

Lepski, O.V. and Levit, B.Y. (1998) Adaptive minimax estimation of infinitely differentiable functions. *Math. Methods Statist.*, **7**, 123–156.

Nagaev, S.V. (1965) Some limit theorems for large deviations. *Teor. Veroyatnost i Primenen.*, **10**, 231–254 (in Russian) [English translation: *Theory Probab. Appl.*, **10**, 214–235.]

Neumann, M.H. (1998) Strong approximation of density estimators from weakly dependent observations by density estimators from independent observations. *Ann. Statist.*, **26**, 2014–2048.

Neumann, M.H. and Spokoiny, V.G. (1995) On the efficiency of wavelet estimators under arbitrary error distributions. *Math. Methods Statist.*, **4**, 137–166.

Neumark, M.A. (1960) *Lineare Differentialoperatoren.* Berlin: Akademie-Verlag.

Nussbaum, M. (1985) Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.*, **13**, 984–997.

Nussbaum, M. (1999) Minimax risk, Pinsker bound. In: S. Kotz (ed.), *Encyclopedia of Statistical Sciences, Update Volume 3*, pp. 451–460. New York: Wiley.

Pinsker, M.S. (1980) Optimal filtration of square-integrable signals in Gaussian noise. *Problemy Peredachi Informatsii*, **16**, 52–68 (in Russian) [English translation: *Problems Inform. Transmission*, **16**, 120–133.]

Politis, D.N., Romano, J.P. and Wolf, M. (1997) Subsampling for heteroskedastic time series. *J. Econometrics*, **81**, 281–317.

Roussas, G.G. (1972) *Contiguity of Probability Measures: Some Applications in Statistics*. Cambridge: Cambridge University Press.

Saavedra, A. and Cao, R. (1999) Rate of convergence of a convolution-type estimator of the marginal density of a MA(1) process. *Stochastic Process. Appl.*, **80**, 129–155.

Schipper, M. (1996) Optimal rates and constants in L_2-minimax estimation of probability density functions. *Math. Methods Statist.*, **5**, 253–274.

Shorack, G.R. and Wellner, J.A. (1986) *Empirical Processes with Applications to Statistics.* New York: Wiley.

Veretennikov, A.Yu. (1984) Bounds for the mixing rate in the theory of stochastic equations. *Theory Probab. Appl.*, **32**, 273–281.

Received June 2003 and revised June 2004