

POSTERIOR CONTRACTION RATE FOR NON-PARAMETRIC BAYESIAN ESTIMATION OF THE DISPERSION COEFFICIENT OF A STOCHASTIC DIFFERENTIAL EQUATION

SHOTA GUGUSHVILI¹ AND PETER SPREIJ²

Abstract. We consider the problem of non-parametric estimation of the deterministic dispersion coefficient of a linear stochastic differential equation based on discrete time observations on its solution. We take a Bayesian approach to the problem and under suitable regularity assumptions derive the posterior contraction rate. This rate turns out to be the optimal posterior contraction rate.

Mathematics Subject Classification. 62G20, 62M05.

Received September 9, 2014. Revised April 30, 2015. Accepted March 14, 2016.

1. INTRODUCTION

Suppose a simple linear stochastic differential equation

$$dX_t = s(t)dW_t, \quad X_0 = x, \quad t \in [0, 1], \quad (1.1)$$

with a deterministic dispersion coefficient s and a deterministic initial condition $X_0 = x$ is given. Here W is a Brownian motion. Without loss of generality, we take $x = 0$. The process X is Gaussian with mean zero and covariance $\rho(u, v) = \int_0^{u \wedge v} (s(t))^2 dt$. By \mathbb{P}_s we will denote the law of the process X corresponding to the dispersion coefficient s in (1.1). The dispersion coefficient s can be interpreted as a signal passing through a noisy channel, where the noise is multiplicative and is modelled by the Brownian motion.

Suppose that corresponding to the true dispersion coefficient $s = s_0$, a sample $X_{t_{i,n}}, i = 1, \dots, n$, from the process X is at our disposal, where $t_{i,n} = i/n, i = 0, \dots, n$. Our goal is non-parametric Bayesian estimation of s_0 . Related references employing the frequentist approach for a similar model are [2, 8, 14]. For a non-parametric Bayesian approach, see [7]. Note that our model shows obvious similarities to a standard non-parametric regression model, or to the white noise model (see *e.g.* [12] or [16] for these models in the non-parametric Bayesian context), but also possesses distinctive features of its own. Two recent works dealing with theoretical properties of non-parametric Bayesian techniques applied in stochastic differential equation models are [17] and [11],

Keywords and phrases. Dispersion coefficient, non-parametric Bayesian estimation, posterior contraction rate, stochastic differential equation.

¹ Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands.
shota.gugushvili@math.leidenuniv.nl

² Korteweg-de Vries Institute for Mathematics, University of Amsterdam, PO Box 94248, 1090 GE Amsterdam, The Netherlands.
spreij@uva.nl

but we note that both the models and observation schemes in those papers are rather different from the ones considered in this work.

Let \mathcal{X} denote some (non-parametric) class of dispersion coefficients s . The likelihood corresponding to the observations $X_{t_{i,n}}$ is given by

$$L_n(s) = \prod_{i=1}^n \left\{ \frac{1}{\sqrt{2\pi \int_{t_{i-1,n}}^{t_{i,n}} s^2(u) du}} \psi \left(\frac{X_{t_{i,n}} - X_{t_{i-1,n}}}{\sqrt{\int_{t_{i-1,n}}^{t_{i,n}} s^2(u) du}} \right) \right\}, \quad (1.2)$$

where $\psi(u) = \exp(-u^2/2)$. For a prior Π on \mathcal{X} , the posterior measure of any measurable set $\mathcal{S} \subset \mathcal{X}$ can be obtained through Bayes' formula,

$$\Pi(\mathcal{S} | X_{t_{0,n}}, \dots, X_{t_{n,n}}) = \frac{\int_{\mathcal{S}} L_n(s) \Pi(ds)}{\int_{\mathcal{X}} L_n(s) \Pi(ds)}.$$

One can then proceed with the computation of other quantities of interest in the Bayesian paradigm, for instance point estimates of s_0 , credible sets and so on.

A desirable property of a Bayes procedure is posterior consistency. In our context posterior consistency means that for every neighbourhood U_{s_0} of s_0 (in a suitable topology)

$$\Pi(U_{s_0}^c | X_{t_{0,n}}, \dots, X_{t_{n,n}}) \xrightarrow{\mathbb{P}_{s_0}} 0$$

as $n \rightarrow \infty$. In other words, when viewed under the true law \mathbb{P}_{s_0} , a consistent Bayesian procedure asymptotically puts posterior mass equal to one on every fixed neighbourhood of the true parameter s_0 . Study of posterior consistency is similar to study of consistency of frequentist estimators, and in fact, if posterior consistency holds, the center of the posterior distribution (in an appropriate sense) will provide a consistent (in the frequentist sense) estimator of the parameter of interest. For an introduction to consistency issues in Bayesian non-parametric statistics, see *e.g.* [6] and [18]. Posterior consistency for the model (1.1) was shown under suitable conditions in [7].

More generally, instead of a fixed neighbourhood U_{s_0} of the true parameter s_0 , one can also take a sequence of neighbourhoods U_{s_0, ε_n} shrinking to s_0 at a rate $\varepsilon_n \rightarrow 0$ (the sequence ε_n determines the size of the neighbourhood) and ask at what rate is ε_n allowed to decay to zero, so that the neighbourhoods U_{s_0, ε_n} still manage to capture most of the posterior mass. A formal way to state this is

$$\Pi(U_{s_0, \varepsilon_n}^c | X_{t_{0,n}}, \dots, X_{t_{n,n}}) \xrightarrow{\mathbb{P}_{s_0}} 0 \quad (1.3)$$

as $n \rightarrow \infty$. The rate ε_n is called the posterior contraction rate, or the posterior convergence rate. Note that ε_n is not uniquely defined: if ε_n is a posterior contraction rate, then so is *e.g.* $2\varepsilon_n$, because $U_{s_0, 2\varepsilon_n}^c \subset U_{s_0, \varepsilon_n}^c$. This, however, is true also for the convergence rate of frequentist estimators, *cf.* a discussion on page 79 in [15]. In general, we are interested in determination of the 'fastest' rate of decay of ε_n , so that (1.3) still holds. Some references on derivation of posterior convergence rates under various statistical setups are [3, 4, 13]. Study of this question parallels the analysis of convergence rates of various estimators in the frequentist literature. In fact, a property like (1.3) also implies that Bayes point estimates have the convergence rate ε_n (in the frequentist sense), *cf.* pages 506–507 in [3]. It is well-known that in finite-dimensional (*i.e.* parametric) statistical problems under suitable regularity assumptions Bayes procedures yield optimal (in the frequentist sense) estimators. The situation is much more subtle in the infinite-dimensional setting: a careless choice of the prior might violate posterior consistency, or the posterior might concentrate around the true parameter value at a suboptimal rate (here by 'suboptimal' we mean the rate slower than the minimax rate for estimation of s_0). Hence the importance of derivation of the posterior contraction rate.

The general structure of the present work is similar to our earlier paper [7] on posterior consistency, but with a crucial difference. To get out main result on the contraction rate we heavily rely on more sophisticated

results from empirical process theory. The outline of the paper is as follows. In Section 2 we formulate a theorem establishing (1.3) under suitable conditions and provide a discussion on it. The proof of the theorem is given in Section 3, while the Appendix contains a number of technical lemmas used in the proof of the theorem.

Throughout the paper we will use the following notation to compare two sequences a_n and b_n of real numbers: $a_n \lesssim b_n$ will mean that there exists a constant $B > 0$ that is independent of n and is such that $a_n \leq Bb_n$; $a_n \gtrsim b_n$ will mean that there exists a constant $A > 0$ that is independent of n and is such that $Aa_n \geq b_n$; $a_n \asymp b_n$ will mean that a_n and b_n are asymptotically of the same order, *i.e.* $-\infty < \liminf_{n \rightarrow \infty} a_n/b_n \leq \limsup_{n \rightarrow \infty} a_n/b_n < \infty$.

2. MAIN THEOREM

We first specify the non-parametric class \mathcal{X} of dispersion coefficients s .

Definition 2.1. Let \mathcal{X} be some collection of dispersion coefficients $s : [0, 1] \rightarrow [\kappa, \mathcal{K}]$, such that $\|s'\|_\infty \leq M$. Here $0 < \kappa < \mathcal{K} < \infty$ and $0 < M < \infty$ are three constants independent of a particular $s \in \mathcal{X}$, while $\|\cdot\|_\infty$ denotes the L_∞ -norm (supremum norm).

Remark 2.2. Since $\mathbb{P}_s = \mathbb{P}_{-s}$, a positivity assumption on $s \in \mathcal{X}$ in Definition 2.1 is a natural identifiability requirement. Furthermore, strict positivity of s allows one to avoid complications when manipulating the likelihood (1.2). Boundedness and differentiability of s also come in handy in the proof of Theorem 2.4 below.

We summarise the assumptions on our statistical model.

Assumption 2.3. Assume that

- (a) the model (1.1) is given with $x = 0$ and $s \in \mathcal{X}$, where \mathcal{X} is as in Definition 2.1;
- (b) $s_0 \in \mathcal{X}$ denotes the true dispersion coefficient;
- (c) a discrete-time sample $\{X_{t_{i,n}}\}$ from the solution X to (1.1) corresponding to s_0 is available, where $t_{i,n} = i/n$, $i = 0, \dots, n$.

For $\varepsilon > 0$ introduce the notation

$$U_{s_0, \varepsilon} = \{s \in \mathcal{X} : \|s - s_0\|_2 < \varepsilon\}, \quad V_{s_0, \varepsilon} = \{s \in \mathcal{X} : \|s - s_0\|_\infty < \varepsilon\}.$$

Here $\|\cdot\|_2$ denotes the L_2 -norm. We will establish (1.3) for the complements of the neighbourhoods U_{s_0, ε_n} of the true parameter s_0 and determine the corresponding posterior contraction rate ε_n . The choice of the L_2 -norm to define neighbourhoods $U_{s_0, \varepsilon}$ appears to be quite natural, because under Assumption 2.3 the distribution of X_1 has as its standard deviation the L_2 -norm of s_0 . Hence the obvious notion of a distance between two dispersion coefficients should be that norm as well.

Theorem 2.4. *Suppose that Assumption 2.3 holds. Let the sequence $\tilde{\varepsilon}_n$ of strictly positive numbers be such that $\tilde{\varepsilon}_n \asymp n^{-1/3} \log n$, and let the prior Π on \mathcal{X} be such that*

$$\Pi(V_{s_0, \tilde{\varepsilon}_n}) \gtrsim e^{-\bar{C}n\tilde{\varepsilon}_n^2} \tag{2.1}$$

for some constant $\bar{C} > 0$ that is independent of n . Then for a large enough constant \widetilde{M} and a sequence $\varepsilon_n = \widetilde{M}\tilde{\varepsilon}_n$,

$$\Pi(U_{s_0, \varepsilon_n}^c | X_{t_{0,n}}, \dots, X_{t_{n,n}}) \xrightarrow{\mathbb{P}_{s_0}} 0$$

holds.

Remark 2.5. Theorem 2.4 states that under the differentiability assumption on the members s of the class \mathcal{X} of dispersion coefficients, the posterior contracts around the true dispersion coefficient s_0 at the rate $n^{-1/3} \log n$. This implies existence of Bayes estimates that converge (in the frequentist sense) to s_0 at the same rate. By Proposition 1 from [8], the rate $n^{-1/3}$ is the minimax convergence rate for estimation of the diffusion coefficient s_0^2 with L_2 -loss function in essentially the same model as ours. In this sense the rate derived in Theorem 2.4 can be thought of as essentially (up to a logarithmic factor) optimal posterior contraction rate. The logarithmic factor is perhaps an artefact of our proof of Theorem 2.4. We add that the use of the empirical process theory techniques involves the choice of many constants that cannot always be easily controlled, and it is at various instances unclear whether our technical estimates and inequalities continue to hold without the logarithmic term. See for instance the proofs of (A.1) and (A.5) below, to mention two examples.

Remark 2.6. An essential condition in Theorem 2.4 is (2.1). A prior Π satisfying condition (2.1) can be constructed, for instance, through a construction similar to the one given in Section 3 of [3], that is based on finite approximating sets (this type of prior was introduced in [5]). Let $\hat{\varepsilon}_n = cn^{-1/3} \log n$ for a constant $c > 0$ to be chosen later on, and let $\{\lambda_n\}$ be a sequence of weights, such that $\lambda_n \geq 0$, $\sum_{n=1}^{\infty} \lambda_n = 1$, and $\log(\lambda_n) \geq -c' \log n$ (some $c' > 0$). Given a function $u \in \mathcal{X}$, an upper bracket u] (note that our notation is non-standard and that we require $u \in \mathcal{X}$) of size $\hat{\varepsilon}_n$ relative to the supremum norm is defined as the set of all those functions $s \in \mathcal{X}$, such that $s(t) \leq u(t)$, $\forall t \in [0, 1]$, and $\|s - u\|_{\infty} < \hat{\varepsilon}_n$ (as a side remark we mention that we could also have used lower brackets, or ordinary brackets). For every $n \in \mathbb{N}$, let Π_n be a uniform distribution on $N_{] , \infty}(\hat{\varepsilon}_n, \mathcal{X})$ upper bracket functions u_j 's, obtained by covering \mathcal{X} with a minimal number of upper brackets of size $\hat{\varepsilon}_n$. We will call $N_{] , \infty}(\hat{\varepsilon}_n, \mathcal{X})$ the $\hat{\varepsilon}_n$ -upper bracketing number of \mathcal{X} relative to the supremum norm (cf. p. 510 in [3]). Next we define the prior Π by

$$\Pi = \sum_{n=1}^{\infty} \lambda_n \Pi_n.$$

Now consider the set $V_{s_0, \tilde{\varepsilon}_n}$ and note that s_0 is contained in one of the $N_{] , \infty}(\hat{\varepsilon}_n, \mathcal{X})$ upper brackets u_j] of size $\hat{\varepsilon}_n$. Then, by construction,

$$\Pi(V_{s_0, \tilde{\varepsilon}_n}) \geq \lambda_n \frac{1}{N_{] , \infty}(\hat{\varepsilon}_n, \mathcal{X})},$$

provided the constant c is chosen small enough. The $\hat{\varepsilon}_n$ -upper bracketing numbers $N_{] , \infty}(\hat{\varepsilon}_n, \mathcal{X})$ are not larger than bracketing numbers $N_{\infty}(\hat{\varepsilon}_n/2, \mathcal{X})$ of \mathcal{X} relative to the L_{∞} -norm (defined implicitly in Def. 2.3 in [1]). By Lemma 2.3 and Problem 2.5 in [1], we obtain for the $\hat{\varepsilon}_n$ -entropy of \mathcal{X} relative to the supremum norm an upper estimate

$$H_{\infty}(\hat{\varepsilon}_n/2, \mathcal{X}) = \log N_{\infty}(\hat{\varepsilon}_n/2, \mathcal{X}) \lesssim \frac{1}{\hat{\varepsilon}_n},$$

so that as a consequence

$$\log \left(\lambda_n \frac{1}{N_{] , \infty}(\hat{\varepsilon}_n, \mathcal{X})} \right) \gtrsim -n\tilde{\varepsilon}_n^2,$$

and (2.1) follows.

Remark 2.7. Theorem 2.4 can be generalised to the case where the members of the class \mathcal{X} of dispersion coefficients are $\beta > 1$ times differentiable with derivatives satisfying suitable boundedness assumptions, and $s_0 \in \mathcal{X}$. The convergence rate that can be obtained in that case is (up to a logarithmic factor) $n^{-\beta/(2\beta+1)}$. The general structure of the proof of the corresponding statement is similar to that of Theorem 2.4, one notable modification being that in the proof of Lemmas A.1 and A.3 one will have to use different entropy estimates for the class \mathcal{X} . A suitable prior can be exhibited through a construction similar to the one given in Remark 2.6 for the case of Theorem 2.4.

Remark 2.8. Equation (1.1) specifies a model for the observations without drift. This model can be expanded by including a drift term as well, and in Bayesian analysis this naturally calls for a prior on it. However, since the drift cannot be consistently estimated under our observation scheme, such an extension would drastically complicate our analysis. Alternatively, one could work in the spirit of [10] and try to establish convergence under misspecification to a ‘wrong’ model, that is in terms of Kullback–Leibler divergence nearest to the true probability that governs the observations. It seems that in this approach it is possible to establish an analogue of Theorem 2.4. We leave this to future research.

3. PROOF OF THEOREM 2.4

Before proceeding any further, we would like to make a general comment on the Proof of Theorem 2.4: in principle, it is conceivable that its statement could be derived from some general result on the posterior contraction rate, see *e.g.* Sections 2 and 3 in [4]. However, in this work we take an alternative approach, that is similar in some respects to the one in [13] and that relies on results from empirical process theory (see *e.g.* [1]). This alternative approach is not necessarily the shortest or simplest, and the choice of a specific path to the derivation of a posterior convergence rate is perhaps a matter of taste.

Throughout this section and the Appendix, $R_n(s) = L_n(s)/L_n(s_0)$ will denote the likelihood ratio corresponding to the observations $X_{t_{i,n}}$. We will use the notation $P_{i,n,s}$ to denote the law of $Y_{i,n} = X_{t_{i,n}} - X_{t_{i-1,n}}$ corresponding to the parameter value s in (1.1), and $P_{i,n,0}$ to denote the law of $Y_{i,n}$ corresponding to the true parameter value s_0 in (1.1). The corresponding densities will be denoted by $p_{i,n,s}$ and $p_{i,n,0}$. We also set

$$z_i = t_{i-1,n}, \quad \mathcal{W}_i = 1 - \frac{Y_{i,n}^2}{\int_{t_{i-1,n}}^{t_{i,n}} s_0^2(u) du}, \quad f_s(z) = \frac{\int_z^{z+1/n} [s_0^2(u) - s^2(u)] du}{\int_z^{z+1/n} s^2(u) du}.$$

The latter notation is reminiscent of the one used in [1]. Note that the \mathcal{W}_i ’s are i.i.d. with zero mean and variance equal to two. Each \mathcal{W}_i is distributed as a random variable $1 - Z_i^2$, for Z_i having a standard normal distribution. As distributions matter in what follows, this justifies omission of a formally required extra index n in $\mathcal{W}_i = \mathcal{W}_{i,n}$. Furthermore, z_i ’s and f_s also formally require an extra index n , but we omit it as no confusion will arise.

Proof of Theorem 2.4. We have

$$\Pi(U_{s_0, \varepsilon_n}^c | X_{t_{0,n}} \dots, X_{t_{n,n}}) = \frac{\int_{U_{s_0, \varepsilon_n}^c} L_n(s) \Pi(ds)}{\int_{\mathcal{X}} L_n(s) \Pi(ds)} = \frac{\int_{U_{s_0, \varepsilon_n}^c} R_n(s) \Pi(ds)}{\int_{\mathcal{X}} R_n(s) \Pi(ds)} = \frac{N_n}{D_n}.$$

We will establish the theorem by separately bounding D_n and N_n and then combining the bounds.

Let $S_n(s) = n^{-1} \log R_n(s)$. Then $D_n = \int_{\mathcal{X}} \exp(n S_n(s)) \Pi(ds)$. We have

$$S_n(s) = \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \mathcal{W}_i f_s(z_i) + \frac{1}{2} \frac{1}{n} \sum_{i=1}^n [\log(1 + f_s(z_i)) - f_s(z_i)].$$

Let n be large enough and assume that $s \in V_{s_0, \varepsilon_n}$. As a consequence of Lemmas A.1 and A.2 from the Appendix and by condition (2.1) on the prior, we get that with probability tending to one as $n \rightarrow \infty$,

$$\frac{1}{D_n} \leq \left(\int_{V_{s_0, \varepsilon_n}} R_n(s) \Pi(ds) \right)^{-1} \lesssim \exp \left(\left(\frac{8\mathcal{K}^2}{\kappa^4} + \bar{C} \right) n \tilde{\varepsilon}_n^2 \right). \quad (3.1)$$

This finishes derivation of a bound for D_n . We now turn to N_n . In Lemma A.3 from the Appendix we show that with probability tending to one as $n \rightarrow \infty$, for some constant $c_1 > 0$ we have $N_n \leq \exp(-c_1 n \varepsilon_n^2)$. Combination

of this bound with (3.1) gives that with probability tending to one as $n \rightarrow \infty$, the inequality

$$\Pi(U_{s_0, \varepsilon_n}^c | X_{t_{0,n}}, \dots, X_{t_{n,n}}) \lesssim \exp \left(-c_1 n \varepsilon_n^2 + \left(\frac{8\mathcal{K}^2}{\kappa^4} + \overline{C} \right) n \tilde{\varepsilon}_n^2 \right)$$

is valid. From this it immediately follows that for $\varepsilon_n = \widetilde{M} \tilde{\varepsilon}_n$ with a large enough constant \widetilde{M} , the left-hand side of the above display converges to zero in probability. This completes the proof of the theorem. \square

APPENDIX

Throughout the Appendix we will use the following notation: for any $\varepsilon > 0$, M_ε will denote the smallest positive integer, such that $2^{M_\varepsilon} \varepsilon^2 \geq 4\mathcal{K}^2$. Note that by definition $2^{M_\varepsilon} \varepsilon^2 \leq 8\mathcal{K}^2$ (because $2^{M_\varepsilon - 1} \varepsilon^2 < 4\mathcal{K}^2$) and that for $\varepsilon \rightarrow 0$ we have $M_\varepsilon \asymp \log_2(1/\varepsilon)$. We set $A_{j,\varepsilon} = \{s \in \mathcal{X} : 2^j \varepsilon^2 \leq \|s - s_0\|_2^2 < 2^{j+1} \varepsilon^2\}$ and $B_{j,\varepsilon} = \{s \in \mathcal{X} : \|s - s_0\|_2^2 < 2^{j+1} \varepsilon^2\}$ for $j = 0, 1, \dots, M_\varepsilon$. We will also let $Z_{i,n,s}(Y_{i,n}) = \log(p_{i,n,s}(Y_{i,n})/p_{i,n,0}(Y_{i,n}))$ denote the log-likelihood ratio corresponding to one ‘observation’ $Y_{i,n}$.

Lemma A.1. *Let the conditions of Theorem 2.4 hold. Then*

$$\sup_{f_s \in \mathcal{F}_{s_0, \tilde{\varepsilon}_n}} \left| \frac{1}{n} \sum_{i=1}^n \mathcal{W}_i f_s(z_i) \right| = O_{\mathbb{P}_{s_0}}(\delta_n),$$

where $\mathcal{F}_{s_0, \tilde{\varepsilon}_n} = \{f_s : \|s - s_0\|_\infty < \tilde{\varepsilon}_n\}$ and δ_n is an arbitrary sequence of positive numbers, such that $\delta_n \asymp \tilde{\varepsilon}_n^2$.

Proof. We will establish the lemma using empirical process theory. In particular, we will employ Corollary 8.8 from [1]. In light of the fact that $\tilde{\varepsilon}_n \asymp n^{-1/3} \log n$, in order to prove the lemma it suffices to show that

$$\sup_{g_s \in \mathcal{G}_{s_0, \tilde{\varepsilon}_n}} \left| \frac{1}{n} \sum_{i=1}^n \mathcal{W}_i g_s(z_i) \right| = O_{\mathbb{P}_{s_0}}(\delta_n),$$

where

$$g_s(z) = \frac{s_0^2(z) - s^2(z)}{s^2(z)}, \quad \mathcal{G}_{s_0, \tilde{\varepsilon}_n} = \{g_s : \|s - s_0\|_\infty < \tilde{\varepsilon}_n\},$$

and the notation resembles the one in [1], so that the arguments become more transparent. Indeed, it suffices to note that by Assumption 2.3 we have $f_s(z_i) = g_s(z_i) + O(n^{-1})$ (the remainder term is of order $1/n$ uniformly in s), whence

$$\sup_{f_s \in \mathcal{F}_{s_0, \tilde{\varepsilon}_n}} \left| \frac{1}{n} \sum_{i=1}^n \mathcal{W}_i f_s(z_i) \right| \leq \sup_{g_s \in \mathcal{G}_{s_0, \tilde{\varepsilon}_n}} \left| \frac{1}{n} \sum_{i=1}^n \mathcal{W}_i g_s(z_i) \right| + O_{\mathbb{P}_{s_0}} \left(\frac{1}{n} \right).$$

In order to apply Corollary 8.8 from [1], we need to verify its conditions, and in particular we need to check formulae (8.23)–(8.29) there. This involves somewhat lengthy computations. Firstly, we need to find a constant R_n , such that $\sup_{g_s \in \mathcal{G}_{s_0, \tilde{\varepsilon}_n}} \|g_s\|_{Q_n}^2 \leq R_n^2$. Here $Q_n = n^{-1} \sum_{i=1}^n \delta_{z_i}$ is the empirical measure associated with the points z_i and $\|g_s\|_{Q_n}^2 = n^{-1} \sum_{i=1}^n g_s^2(z_i)$. Now, $\|g_s\|_{Q_n}^2 \leq 4\mathcal{K}^2 \tilde{\varepsilon}_n^2 / \kappa^4$ for $g_s \in \mathcal{G}_{s_0, \tilde{\varepsilon}_n}$, and thus it suffices to take $R_n = 2\mathcal{K} \tilde{\varepsilon}_n / \kappa^2$. Next, set $K_1 = 3$. Using the rough bound $|e^x - 1 - x| \leq x^2 e^{|x|}$, we get that

$$2K_1^2 \mathbb{E}_{s_0} \left[e^{|\mathcal{W}_i|/K_1} - 1 - \frac{|\mathcal{W}_i|}{K_1} \right] \leq 2\mathbb{E}_{s_0} \left[\mathcal{W}_i^2 e^{|\mathcal{W}_i|/3} \right] < \infty.$$

Let $\sigma_0^2 = 2\mathbb{E}_{s_0} \left[\mathcal{W}_i^2 e^{|\mathcal{W}_i|/3} \right]$ and note that this quantity is finite. With these K_1 and σ_0 , (8.23) in [1] will be satisfied. Next we need to find a constant K_2 , such that the inequality $\sup_{g_s \in \mathcal{G}_{s_0, \tilde{\varepsilon}_n}} \|g_s\|_\infty \leq K_2$ holds. One can take $K_2 = 2\mathcal{K} \tilde{\varepsilon}_n / \kappa^2$, and this verifies (8.24) in [1]. We take $C_1 = 3$, set $K = 4K_1 K_2$, and note that for

all n large enough, $\delta_n \leq C_1 2R_n^2 \sigma_0^2 / K$ and $\delta_n \leq 8\sqrt{2}R_n \sigma_0$ hold, because $\tilde{\varepsilon}_n \rightarrow 0$. This choice of C_1 and K thus yields (8.25)–(8.27) in [1]. Next let $C_0 = 2C$, where C is a universal constant as in Corollary 8.8 in [1]. This choice of C_0 yields (8.29) in [1]. It remains to check (8.28) in [1], *i.e.*

$$\sqrt{n}\delta_n \geq C_0 \left(\int_0^{\sqrt{2}R_n \sigma_0} H_B^{1/2} \left(\frac{u}{\sqrt{2}\sigma_0}, \mathcal{G}_{s_0, \tilde{\varepsilon}_n}, Q_n \right) du \vee \sqrt{2}R_n \sigma_0 \right), \quad (\text{A.1})$$

where $H_B(\delta, \mathcal{G}_{s_0, \tilde{\varepsilon}_n}, Q_n)$ is the δ -entropy with bracketing of $\mathcal{G}_{s_0, \tilde{\varepsilon}_n}$ for the $L_2(Q_n)$ -metric (see Def. 2.2 in [1]), and $a \vee b$ denotes the maximum of two numbers a and b . By Lemma 2.1 in [1] and Problem 2.5 there, $H_B(\delta, \mathcal{G}_{s_0, \tilde{\varepsilon}_n}, Q_n) \leq H_\infty(\delta/2, \mathcal{G}_{s_0, \tilde{\varepsilon}_n})$, where $H_\infty(\delta, \mathcal{G}_{s_0, \tilde{\varepsilon}_n})$ is the δ -entropy of $\mathcal{G}_{s_0, \tilde{\varepsilon}_n}$ for the supremum norm (see Def. 2.3 in [1]). Lemma 3.9 in [1] implies that there exists a constant $A_1 > 0$, such that $H_\infty(\delta, \mathcal{G}_{s_0, \tilde{\varepsilon}_n}) \leq A_1 \delta^{-1}$ for all $\delta > 0$ (the fact that the matrix Σ_{Q_n} from the statement of that lemma is non-singular, can be shown by a minor variation of an argument from the proof of Lem. 1.4 in [15]). Hence

$$\int_0^{\sqrt{2}R_n \sigma_0} H_B^{1/2} \left(\frac{u}{\sqrt{2}\sigma_0}, \mathcal{G}_{s_0, \tilde{\varepsilon}_n}, Q_n \right) du \leq \sqrt{A_1} \int_0^{\sqrt{2}R_n \sigma_0} \left(\frac{u}{\sqrt{22}\sigma_0} \right)^{-1/2} du \leq 4\sigma_0 \sqrt{A_1 R_n} \asymp \sqrt{\tilde{\varepsilon}_n}.$$

Since $\tilde{\varepsilon}_n \rightarrow 0$, the right-hand side of (A.1) is of order $\sqrt{\tilde{\varepsilon}_n}$, and then $\tilde{\varepsilon}_n \asymp n^{-1/3} \log n$ is enough to ensure that (A.1), or equivalently, formula (8.28) in [1], holds for all n large enough. This completes verification of the conditions in Corollary 8.8 in [1]. As a result, *cf.* formula (8.30) in [1], for all n large enough we get the bound

$$\mathbb{P}_{s_0} \left(\sup_{g \in \mathcal{G}_{s_0, \tilde{\varepsilon}_n}} \left| \frac{1}{n} \sum_{i=1}^n \mathcal{W}_i g(z_i) \right| \geq \delta_n \right) \leq C \exp \left(- \frac{n\delta_n^2}{C^2(C_1 + 1)2R_n^2 \sigma_0^2} \right).$$

The right-hand side of this expression converges to zero as $n \rightarrow \infty$, because $n\tilde{\varepsilon}_n^2 \rightarrow \infty$. This completes the proof of the lemma. \square

Lemma A.2. *Let the conditions of Theorem 2.4 hold, assume that n is large enough and let $s \in V_{s_0, \tilde{\varepsilon}_n}$. Then*

$$\begin{aligned} \frac{1}{2} \frac{1}{n} \sum_{i=1}^n \{ \log(1 + f_s(z_i)) - f_s(z_i) \} &\geq -\frac{1}{2} \int_0^1 \frac{(s_0^2(u) - s^2(u))^2}{s^4(u)} du + O\left(\frac{1}{n}\right) \\ &\geq -\frac{2\mathcal{K}^2}{\kappa^4} \tilde{\varepsilon}_n^2 + O\left(\frac{1}{n}\right), \end{aligned}$$

where the remainder term is of order n^{-1} uniformly in s .

Proof. By the elementary inequality $|\log(1+t) - t| \leq t^2$ that is valid for $|t| < 1/2$, we have for all n large enough and uniformly in $s \in V_{s_0, \tilde{\varepsilon}_n}$ that

$$|\log(1 + f_s(z_i)) - f_s(z_i)| \leq f_s^2(z_i).$$

Hence

$$\log(1 + f_s(z_i)) - f_s(z_i) \geq -f_s^2(z_i),$$

and therefore

$$\frac{1}{2} \frac{1}{n} \sum_{i=1}^n \{ \log(1 + f_s(z_i)) - f_s(z_i) \} \geq -\frac{1}{2} \frac{1}{n} \sum_{i=1}^n f_s^2(z_i).$$

The statement of the lemma now follows by a simple computation employing Assumption 2.3 and the Riemann sum approximation of the integral, yielding that for all n large enough,

$$\begin{aligned} -\frac{1}{2} \frac{1}{n} \sum_{i=1}^n f_s^2(z_i) &= -\frac{1}{2} \int_0^1 \frac{(s_0^2(u) - s^2(u))^2}{s^4(u)} du + O\left(\frac{1}{n}\right) \\ &\geq -\frac{2\mathcal{K}^2}{\kappa^4} \varepsilon_n^2 + O\left(\frac{1}{n}\right), \end{aligned}$$

where the remainder term is of order n^{-1} uniformly in s . \square

Lemma A.3. *Let the conditions of Theorem 2.4 hold and let $\varepsilon_n \asymp n^{-1/3} \log n$. Denote $\sigma_0^2 = 2\mathbb{E}_{s_0} [\mathcal{W}_i^2 e^{|\mathcal{W}_i|/3}]$. There exists a constant $\tilde{c}_0 > 0$, such that $\tilde{c}_0 \leq \mathcal{K}^4 \sigma_0 (\sigma_0 \wedge 4) / \kappa^4$, another constant c_1 , such that $c_1 < \tilde{c}_0 \kappa^2 / (2\mathcal{K}^4)$, and a universal constant $C > 0$, for which the inequality*

$$\mathbb{P}_{s_0} \left(\sup_{s \in U_{s_0, \varepsilon_n}^c} \prod_{i=1}^n \frac{p_{i,n,s}(Y_{i,n})}{p_{i,n,0}(Y_{i,n})} \geq \exp(-c_1 n \varepsilon_n^2) \right) \leq C M_{\varepsilon_n} \exp\left(-\frac{(\tilde{c}_0 \kappa^2 / (2\mathcal{K}^4) - c_1)^2}{8C^2(4\mathcal{K}^2/\kappa^4 + 1)\sigma_0^2} n \varepsilon_n^2\right)$$

holds for all n large enough. Here $a \wedge b$ denotes the minimum of two numbers a and b . In particular, as $n \rightarrow \infty$, the right-hand side of the above display converges to zero.

Proof. As in the proof of Lemma A.1, we will use empirical process theory to establish the result. We use the convention that the supremum over the empty set is equal to zero. By Assumption 2.3, we have $\|s - s_0\|_2^2 \leq 4\mathcal{K}^2$. Hence, using the definition of M_{ε_n} and A_{j,ε_n} at the beginning of this appendix, we can write

$$\mathbb{P}_{s_0} \left(\sup_{s \in U_{s_0, \varepsilon_n}^c} \prod_{i=1}^n \frac{p_{i,n,s}(Y_{i,n})}{p_{i,n,0}(Y_{i,n})} \geq \exp(-c_1 n \varepsilon_n^2) \right) = \sum_{j=0}^{M_{\varepsilon_n}} \mathbb{P}_{s_0} \left(\sup_{s \in A_{j,\varepsilon_n}} \prod_{i=1}^n \frac{p_{i,n,s}(Y_{i,n})}{p_{i,n,0}(Y_{i,n})} \geq \exp(-c_1 n \varepsilon_n^2) \right).$$

We will individually bound the summands on the right-hand side of the above display, thereby obtaining a bound on its left-hand side, and will show that this upper bound converges to zero as $n \rightarrow \infty$.

Using Lemma A.4 ahead (note that the constant \tilde{c}_0 in its statement can be taken arbitrarily small) and recalling the definition of $Z_{i,n,s}(Y_{i,n})$, A_{j,ε_n} and B_{j,ε_n} at the beginning of this appendix, we obtain that for all n large enough

$$\begin{aligned} \mathbb{P}_{s_0} \left(\sup_{s \in A_{j,\varepsilon_n}} \prod_{i=1}^n \frac{p_{i,n,s}(Y_{i,n})}{p_{i,n,0}(Y_{i,n})} \geq \exp(-c_1 n \varepsilon_n^2) \right) &\leq \mathbb{P}_{s_0} \left(\sup_{s \in A_{j,\varepsilon_n}} \exp\left(\sum_{i=1}^n \{Z_{i,n,s}(Y_{i,n}) - \mathbb{E}_{s_0}[Z_{i,n,s}(Y_{i,n})]\}\right) \right) \\ &\geq \exp\left(2^j n \varepsilon_n^2 \left(\frac{\tilde{c}_0 \kappa^2}{\mathcal{K}^4} - \frac{\tilde{C}_0}{2^j n \varepsilon_n^2} - \frac{c_1}{2^j}\right)\right) \\ &\leq \mathbb{P}_{s_0} \left(\sup_{s \in B_{j,\varepsilon_n}} \exp\left(\sum_{i=1}^n \{Z_{i,n,s}(Y_{i,n}) - \mathbb{E}_{s_0}[Z_{i,n,s}(Y_{i,n})]\}\right) \right) \\ &\geq \exp\left(2^j n \varepsilon_n^2 \left(\frac{\tilde{c}_0 \kappa^2}{\mathcal{K}^4} - \frac{\tilde{C}_0}{2^j \varepsilon_n^2 n} - \frac{c_1}{2^j}\right)\right) \\ &\leq \mathbb{P}_{s_0} \left(\sup_{s \in B_{j,\varepsilon_n}} \left| \frac{1}{n} \sum_{i=1}^n \mathcal{W}_i f_s(z_i) \right| \geq \delta_n \right), \end{aligned} \tag{A.2}$$

where we have set

$$\delta_n = \bar{\delta} 2^{j+1} \varepsilon_n^2 = \left(\frac{\tilde{c}_0 \kappa^2}{\mathcal{K}^4} - \frac{\tilde{C}_0}{2^j \varepsilon_n^2 n} - \frac{c_1}{2^j} \right) 2^{j+1} \varepsilon_n^2. \tag{A.3}$$

Positivity of $\bar{\delta}$ for n large enough is a consequence of the assumptions in the statement of the lemma. We want to apply Corollary 8.8 from [1] to the last term in (A.2). In order to do so, we need to verify its conditions, which can be done using arguments similar to those from the proof of Lemma A.1 in this Appendix. We first need to find a constant R_n , such that $\sup_{s \in B_{j,\varepsilon_n}} \|f_s\|_{Q_n} \leq R_n$. We have for all n large enough and all $j = 0, 1, \dots, M_{\varepsilon_n}$,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\int_{z_i}^{z_{i+1}} [s_0^2(u) - s^2(u)] du}{\int_{z_i}^{z_{i+1}} s^2(u) du} \right\}^2 &= \int_0^1 \frac{(s_0^2(u) - s^2(u))^2}{s^4(u)} du + \left[\frac{1}{n} \sum_{i=1}^n \left\{ \frac{\int_{z_i}^{z_{i+1}} [s_0^2(u) - s^2(u)] du}{\int_{z_i}^{z_{i+1}} s^2(u) du} \right\}^2 \right. \\ &\quad \left. - \int_0^1 \frac{(s_0^2(u) - s^2(u))^2}{s^4(u)} du \right] \\ &\leq \left(\frac{4\mathcal{K}^2}{\kappa^4} + 1 \right) 2^{j+1} \varepsilon_n^2, \end{aligned}$$

where we used Assumption 2.3, definition of B_{j,ε_n} and the assumption that $\varepsilon_n \asymp n^{-1/3} \log n$ to see the last inequality. We can thus take

$$R_n = \left\{ \frac{4\mathcal{K}^2}{\kappa^4} + 1 \right\}^{1/2} 2^{(j+1)/2} \varepsilon_n.$$

Next, define the constants K_1, C, C_0 and C_1 as in the proof of Lemma A.1. Since $\|f_s\|_\infty \leq 2\mathcal{K}^2/\kappa^2$, we can take $K_2 = 2\mathcal{K}^2/\kappa^2$. We also set $K = 4K_1K_2$. We want that the inequalities $\delta_n \leq C_1 2R_n^2 \sigma_0^2 / K$, $\delta_n \leq 8\sqrt{2}R_n \sigma_0$ and

$$\sqrt{n} \delta_n \geq C_0 \left(\int_0^{\sqrt{2}R_n \sigma_0} H_B^{1/2} \left(\frac{u}{\sqrt{2}\sigma_0}, B_{j,\varepsilon_n}, Q_n \right) du \vee \sqrt{2}R_n \sigma_0 \right) \quad (\text{A.4})$$

hold. It is not difficult to check by a direct computation that the first two of these inequalities hold with δ_n as in (A.3) and \tilde{c}_0 and c_1 as in the statement of the lemma. Verification of (A.4), on the other hand, requires some additional arguments. In order to check (A.4), we need to show that for all n large enough and all $j = 0, 1, \dots, M_{\varepsilon_n}$, the inequalities $n\delta_n^2 \geq C_0^2 2R_n^2 \sigma_0^2$ and

$$n\delta_n^2 \geq C_0^2 \left(\int_0^{\sqrt{2}R_n \sigma_0} H_B^{1/2} \left(\frac{u}{\sqrt{2}\sigma_0}, B_{j,\varepsilon_n}, Q_n \right) du \right)^2 \quad (\text{A.5})$$

hold. It is easy to see that the first of these two inequalities follows from the fact that $n\varepsilon_n^2 \rightarrow \infty$. As far as the second one is concerned, we note that for all $\delta > 0$ for some constant $A > 0$,

$$H_B(\delta, B_{j,\varepsilon_n}, Q_n) \leq \frac{A}{\delta},$$

where we have used the fact that $B_{j,\varepsilon_n} \subseteq \mathcal{X}$, as well as Lemma 2.1 and Theorem 2.4 from [1]. Therefore,

$$\begin{aligned} \int_0^{\sqrt{2}R_n \sigma_0} H_B^{1/2} \left(\frac{u}{\sqrt{2}\sigma_0}, B_{j,\varepsilon_n}, Q_n \right) du &\leq \sqrt{A} \int_0^{\sqrt{2}R_n \sigma_0} \left(\frac{u}{\sqrt{2}2\sigma_0} \right)^{-1/2} du \\ &= 4\sqrt{AR_n} \sigma_0. \end{aligned}$$

Since

$$n\bar{\delta}^2 2^{2(j+1)} \varepsilon_n^4 \geq 16C_0^2 A \sigma_0^2 \left(\frac{4\mathcal{K}^2}{\kappa^4} + 1 \right) 2^{(j+1)/2} \varepsilon_n$$

for all n large enough and all $j = 0, 1, \dots, M_{\varepsilon_n}$ (this follows from the assumption that $\varepsilon_n \asymp n^{-1/3} \log n$), we get that (A.5), and hence (A.4) too, hold. Thus all the assumptions from Corollary 8.8 in [1] are satisfied.

As a result, the inequality (8.30) from Corollary 8.8 combined with formula (A.2) and some further bounding gives that

$$\mathbb{P}_{s_0} \left(\sup_{s \in A_{j, \varepsilon_n}} \prod_{i=1}^n \frac{p_{i,n,s}(Y_{i,n})}{p_{i,n,s}(Y_{i,n})} \geq \exp(-c_1 n \varepsilon_n^2) \right) \leq C \exp \left(-\frac{(\tilde{c}_0 \kappa^2 / (2\mathcal{K}^4) - c_1)^2}{8C^2 \sigma_0^2 (4\mathcal{K}^2 / \kappa^4 + 1)} n \varepsilon_n^2 \right)$$

holds for all n large enough and all $j = 0, 1, \dots, M_{\varepsilon_n}$. The statement of the lemma is an easy consequence of this bound, the fact that $M_{\varepsilon_n} \asymp \log_2(1/\varepsilon_n)$ for $\varepsilon_n \rightarrow 0$ and the fact that $\varepsilon_n \asymp n^{-1/3} \log n$. \square

Lemma A.4. *Under the same conditions as in Lemma A.3, there exist two constants $\tilde{c}_0 > 0$ and $\tilde{C}_0 > 0$, such that for all n large enough and all $s \in A_{j, \varepsilon_n}$, $j = 0, 1, \dots, M_{\varepsilon_n}$, we have*

$$\sum_{i=1}^n \mathbb{E}_{s_0} [Z_{i,n,s}(Y_{i,n})] \leq -\frac{\tilde{c}_0 \kappa^2}{\mathcal{K}^4} 2^j \varepsilon_n^2 n + \tilde{C}_0.$$

Proof. Thanks to Assumption 2.3, using the differentiability of the integrands, we have

$$\begin{aligned} & \int_{z_i}^{z_{i+1}} \frac{(s^2(u) - s_0^2(u))^2}{s^4(u)} du - \frac{1}{n} \left\{ \frac{\int_{z_i}^{z_{i+1}} [s_0^2(u) - s^2(u)] du}{\int_{z_i}^{z_{i+1}} s^2(u) du} \right\}^2 \\ &= \frac{1}{n} \frac{(s^2(z_i) - s_0^2(z_i))^2}{s^4(z_i)} + O\left(\frac{1}{n^2}\right) - \frac{1}{n} \left\{ \frac{[s_0^2(z_i) - s^2(z_i)] + O(n^{-1})}{s^2(z_i) + O(n^{-1})} \right\}^2 = O\left(\frac{1}{n}\right), \end{aligned}$$

where the last term is of order $1/n$ uniformly in s . Note that

$$\begin{aligned} \mathbb{E}_{s_0} [Z_{i,n,s}(Y_{i,n})] &= \frac{1}{2} \log \left(1 + \frac{\int_{z_i}^{z_{i+1}} [s_0^2(u) - s^2(u)] du}{\int_{z_i}^{z_{i+1}} s^2(u) du} \right) \\ &\quad - \frac{1}{2} \frac{\int_{z_i}^{z_{i+1}} [s_0^2(u) - s^2(u)] du}{\int_{z_i}^{z_{i+1}} s^2(u) du}. \end{aligned}$$

A standard argument shows that for any fixed constant $\overline{C}_0 > 0$, there exists another constant $\tilde{c}_0 > 0$, such that for $-1 \leq x < \overline{C}_0$, the inequality $\log(1+x) - x \leq -\tilde{c}_0 x^2$ holds. Therefore, for all n large enough,

$$\begin{aligned} \sum_{i=1}^n \mathbb{E}_{s_0} [Z_{i,n,s}(Y_{i,n})] &\leq -\frac{\tilde{c}_0 n}{2} \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\int_{z_i}^{z_{i+1}} [s_0^2(u) - s^2(u)] du}{\int_{z_i}^{z_{i+1}} s^2(u) du} \right\}^2 \\ &= -\frac{\tilde{c}_0 n}{2} \int_0^1 \frac{(s^2(u) - s_0^2(u))^2}{s^4(u)} du + O(1) \\ &\leq -\frac{\tilde{c}_0 \kappa^2}{\mathcal{K}^4} 2^j \varepsilon_n^2 n + \tilde{C}_0, \end{aligned}$$

where we used Assumption 2.3 and the definition of A_{j, ε_n} . Here $\tilde{C}_0 > 0$ is some constant independent of a particular s and n , by the argument at the beginning of this proof. This completes the proof of the lemma. \square

Acknowledgements. The authors would like to thank the Associate Editor and the referees for their remarks and suggestions. The research leading to the results in this paper has received funding from the European Research Council under ERC Grant Agreement 320637.

REFERENCES

- [1] S.A. van de Geer, Applications of Empirical Process Theory. Vol. 6 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge (2000).
- [2] V. Genon-Catalot, C. Laredo and D. Picard, Nonparametric estimation of the diffusion coefficient by wavelets methods. *Scand. J. Statist.* **19** (1992) 317–335.
- [3] S. Ghosal, J.K. Ghosh and A.W. van der Vaart, Convergence rates of posterior distributions. *Ann. Statist.* **28** (2000) 500–531.
- [4] S. Ghosal and A.W. van der Vaart, Convergence rates of posterior distributions for non-i.i.d. observations. *Ann. Statist.* **35** (2007) 192–223.
- [5] S. Ghosal, J.K. Ghosh and R.V. Ramamoorthi, Non-informative priors via sieves and packing numbers. Advances in Statistical Decision Theory and Applications, *Stat. Ind. Technol.* Birkhäuser Boston, Boston, MA (1997) 119–132.
- [6] S. Ghosal, J.K. Ghosh and R.V. Ramamoorthi, Consistency issues in Bayesian nonparametrics. Asymptotics, Nonparametrics, and Time Series. Vol. 158 of *Statist. Textbooks Monogr.* Dekker, New York (1999) 639–667.
- [7] S. Gugushvili and P. Spreij, Non-parametric Bayesian estimation of a dispersion coefficient of the stochastic differential equation. *ESAIM: PS* **18** (2014) 332–341.
- [8] M. Hoffmann, Minimax estimation of the diffusion coefficient through irregular samplings. *Statist. Probab. Lett.* **32** (1997) 11–24.
- [9] I. Karatzas and S.E. Shreve, Brownian Motion and Stochastic Calculus. Vol. 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York (1988).
- [10] B. Kleijn and A.W. van der Vaart, Misspecification in infinite-dimensional Bayesian statistics, *Ann. Statist.* **34** (2006) 837–877.
- [11] R. Nickl and J. Söhl, Nonparametric Bayesian posterior contraction rates for discretely observed scalar diffusions. Preprint [arXiv:1510.05526](https://arxiv.org/abs/1510.05526) [math.ST] (2015).
- [12] C.E. Rasmussen and C.K.I. Williams, Gaussian Processes for Machine Learning. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA (2006).
- [13] X. Shen and L. Wasserman, Rates of convergence of posterior distributions. *Ann. Statist.* **29** (2001) 687–714.
- [14] P. Soulier, Nonparametric estimation of the diffusion coefficient of a diffusion process. *Stochastic Anal. Appl.* **16** (1998) 185–200.
- [15] A.B. Tsybakov, Introduction to Nonparametric Estimation. *Springer Series in Statistics*. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. Springer, New York (2009).
- [16] A.W. van der Vaart and J.H. van Zanten, Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.* **36** (2008) 1435–1463.
- [17] J. van Waaij and H. van Zanten, Gaussian process methods for one-dimensional diffusions: optimal rates and adaptation. Preprint [arXiv:1506.00515](https://arxiv.org/abs/1506.00515) [math.ST] (2015).
- [18] L. Wasserman, Asymptotic properties of nonparametric Bayesian procedures. In Practical Nonparametric and Semiparametric Bayesian Statistics. Vol. 133 of Lect. Notes Stat. Springer, New York (1998) 293–304.
- [19] W.H. Wong and X. Shen, Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.* **23** (1995) 339–362.