Bernoulli **12**(5), 2006, 863–888

Convergence rates of posterior distributions for Brownian semimartingale models

F.H. VAN DER MEULEN[1], A.W. VAN DER VAART[2,*] and J.H. VAN ZANTEN[2,**]

[1]*Institute for Business and Industrial Statistics of the University of Amsterdam, IBIS, UvA, Plantage Muidergracht 24, 1018 TV Amsterdam, Netherlands. E-mail: meulen@science.uva.nl*
[2]*Department of Mathematics, Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, Netherlands. E-mail: *aad@cs.vu.nl; **harry@cs.vu.nl*

We consider the asymptotic behaviour of posterior distributions based on continuous observations from a Brownian semimartingale model. We present a general result that bounds the posterior rate of convergence in terms of the complexity of the model and the amount of prior mass given to balls centred around the true parameter. This result is illustrated for three special cases of the model: the Gaussian white noise model, the perturbed dynamical system and the ergodic diffusion model. Some examples for specific priors are discussed as well.

Keywords: Bayesian estimation; continuous semimartingale; Dirichlet process; Hellinger distance; infinite-dimensional model; rate of convergence; wavelets

1. Introduction

Suppose that we observe the stochastic process $X^n = (X_t^n, 0 \leqslant t \leqslant T_n)$ defined through the stochastic differential equation (SDE)

$$dX_t^n = \beta^{\theta,n}(t, X^n)dt + \sigma^n(t, X^n)dB_t^n, \quad t \in [0, T_n], \qquad X_0^n = X_0, \tag{1.1}$$

where B^n is a standard Brownian motion. Based on a realization of X^n, we wish to make inference on the parameter θ that determines the shape of the 'drift coefficient' $\beta^{\theta,n}$. The 'diffusion coefficient' σ^n is considered known, as it can be determined without error from continuous observation of the process. The natural number $n \in \mathbb{N}$ serves as an indexing parameter for our asymptotic set-up, in which n tends to infinity. The endpoint T_n of the observational interval may be fixed or tend to infinity. This *Brownian semimartingale model* contains the *diffusion model* as the special case in which $\beta^{\theta,n}$ and σ^n are measurable functions of the process X_t^n at time t only.

To set this up more formally, we assume that $\beta^{\theta,n}$ and σ^n are measurable functions that satisfy regularity conditions that ensure that the SDE (1.1) has a unique weak solution. We then let $P^{\theta,n}$ be the induced distribution on the Borel sets of the space $C[0, T_n]$ of continuous functions on $[0, T_n]$ of a solution $X^n = (X_t^n, 0 \leqslant t \leqslant T_n)$, and consider the statistical experiment $(P^{\theta,n} : \theta \in \Theta^n)$ for a given parameter set Θ^n. We are mostly

1350–7265

interested in the case where the parameter set Θ^n is infinite-dimensional, but our results also apply to parametric models.

The Bayesian approach to statistical inference consists of putting a prior distribution Π^n on the parameter set Θ^n and making inference based on the posterior distribution $\Pi^n(\cdot|X^n)$. The latter is the conditional distribution of the parameter θ given the observation X^n if the measures $P^{\theta,n}$ are considered the conditional distributions of X^n given the parameter θ. In this paper we adopt the Bayesian framework to define the posterior distribution, but study the properties of the posterior distribution from a frequentist point of view. This entails that we assume that the observation X^n is generated from a measure $P^{\theta_0,n}$ in the model, where the value θ_0 is referred to as the 'true value' of the parameter.

We are interested in the asymptotic behaviour of the posterior distributions, as $n \to \infty$. If the priors Π^n do not exclude θ_0 as a possible value of θ, then we may expect posterior consistency, meaning that the sequence of random measures $\Pi^n(\cdot|X^n)$ converges weakly to the degenerate measure at θ_0. In this paper we are interested in the rate of this convergence, measured by the size of the largest shrinking balls around θ_0 that contain 'most' of the posterior probability. Our main result is a characterization of this rate through a measure of the amount of prior mass near θ_0 and a measure of the complexity of the parameter set Θ^n relative to the SDE model.

Earlier work on versions of this problem includes Ibragimov and Has'minskii (1981), Kutoyants (1994, 2004), Zhao (2000), Shen and Wasserman (2001) and Ghosal and van der Vaart (2004). The last paper relates the problem to general Bayesian inference, and we refer to this paper for further references and a overview of the literature on Bayesian asymptotics. Ghosh and Ramamoorthi (2003) give many examples of prior distributions in nonparametric models, and discuss consistency. Results on non-Bayesian methods can be found in Prakasa Rao (1999) and Kutoyants (1984).

Versions of the *parametric* Brownian semimartingale model, in which the process $\beta^{\theta,n}$ depends smoothly on a Euclidean parameter, have been studied in detail. The *Gaussian white noise model*, in which the drift coefficient is a deterministic function of time, the diffusion coefficient is a sequence of constants tending to zero and the observational interval is fixed, is well understood, also from a Bayesian point of view. Results on parametric Bayesian estimation are summarized in Ibragimov and Has'minskii (1981: Theorem II.5.1); they prove asymptotic normality and efficiency for Bayes estimators under various loss functions under conditions that imply local asymptotic normality of the statistical models. The rate of convergence in this case is equal to the size of the drift constants σ^n. The *perturbed dynamical system* is an extension of the white noise model, which allows the drift coefficient to depend on the solution X_t^n in addition to t. This model is treated in depth in the book by Kutoyants (1994). Under natural regularity conditions these models are locally asymptotically normal, and Bayes estimators typically converge at rate σ^n and are asymptotically normal (Kutoyants 1994: Theorem 2.2.3). Results on non-standard situations, such as model misspecification or non-regular parametrizations, can be found in that book too. In the *ergodic diffusion model* both the drift and diffusion coefficients may depend on the solution X_t^n, but they are assumed to have a form independent of n. The asymptotics here is on the endpoint T_n of the observational interval, which tends to infinity. Again these models are locally asymptotically normal under natural conditions, with scaling rate $\sqrt{T_n}$. Results on these models are derived in Kutoyants (2004).

Much less is known about the *nonparametric* Brownian semimartingale model, except for the very special case of the Gaussian white noise model. The Gaussian white noise model has been studied from many perspectives, and in the Bayesian set-up with many priors (see Zhao 2000; Shen and Wasserman 2001). It was put in a more general framework of models that are not independently and identically distributed (i.i.d.) in Ghosal and van der Vaart (2004: Section 5). Unfortunately, the general Brownian semimartingale model is much more complicated. The main focus of the present paper is on this general model.

A key difficulty of the general Brownian semimartingale model is that the *Hellinger semimetric* is, in general, a random process rather than a true semimetric. The square of the Hellinger semimetric h_n is given by

$$h_n^2(\theta, \theta_0) = \int_0^{T_n} \left(\frac{\beta^{\theta,n} - \beta^{\theta_0,n}}{\sigma^n} \right)^2 (t, X^n)\mathrm{d}t.$$

It is the natural semimetric to use, as the log-likelihood process (with respect to $P^{\theta_0,n}$) of the model can be written as $M - \frac{1}{2}[M]$ for a certain continuous local martingale M and the squared Hellinger semimetric $h_n^2(\theta, \theta_0) = [M]_{T_n}$ is the quadratic variation of this martingale M.

The best possible rate of convergence is of course determined by the likelihood process of the model, and in a more technical way by the existence of appropriate tests of the true parameter versus balls of alternatives. The martingale representation of the log-likelihood and Bernstein's inequality allow such tests to be constructed relative to the Hellinger semimetric. Unfortunately, the randomness of this semimetric causes complications that preclude straightforward extension of the Ghosal and van der Vaart (2004) result and motivate the present paper. In part we follow ideas from van Zanten (2005), who considers convergence rates for the maximum likelihood estimator of the Brownian semimartingale model.

Our main theorem (Theorem 2.1) bounds the posterior rate of convergence in terms of the complexity of the model and the amount of prior mass given to balls centred around the true parameter. In the statement of the theorem, the distance from θ to the true parameter θ_0 is measured by the Hellinger semimetric, but it is often possible to translate this result in terms of a deterministic semimetric d_n. We illustrate the usefulness of our main result by three classes of examples of SDEs: the Gaussian white noise model, the perturbed dynamical system and the ergodic diffusion model. Explicit calculations using a variety of priors are included. Priors based on series expansions yield a rate of posterior convergence that is within the minimax rate of estimation up to a logarithmic factor provided the tuning constants (truncation level, prior spread of the coefficients) satisfy broad inequalities. A natural prior on montone functions based on the Dirichlet process is also nearly optimal, without the need for tuning. These results indicate that certain natural priors may work well, although 'most' priors can be expected to give suboptimal results. The logarithmic factors in the results mentioned may be due to our method of proof. We also include an example of an artificial prior that attains the minimax rate.

In the examples, we treat in detail the case where the random Hellinger metric converges at a deterministic rate to a non-random limiting metric. The latter is not crucial for our results to be applicable, however. In certain null recurrent or transient diffusion models the Hellinger metric converges, at a deterministic rate, to a random limit (see, for instance, Dietz and Kutoyants 2003; Höpfner and Kutoyants 2003; Loukianova and Loukianov 2005). Our general result applies in such situations as well – see also van Zanten (2005: Section 5.4) for the maximum likelihood estimator case.

The organization of the paper is as follows. In Section 2 we present our main result. We specialize this result to three classes of SDEs in Section 3. The proof of the main result and technical complements are deferred to Section 4.

2. Main result

For $n \in \mathbb{N}$, given numbers $T_n > 0$, and each θ in an arbitrary set Θ^n, let $\beta^{\theta,n}$ and σ^n be measurable and non-anticipative functions on $[0, T_n] \times C[0, T_n]$ such that the SDE (1.1) possesses a unique weak solution $X^n = (X_t^n, t \in [0, T_n])$. Here B^n is a standard Brownian motion. Denote the distribution of the process X^n on the Borel sets $\mathcal{C}^n$ of the space $C[0, T_n]$ by $P^{\theta,n}$. The parameter value $\theta_0 \in \Theta^n$, which may also depend on n, will refer to the 'true value' of the parameter: throughout we consider the distribution of X^n under the assumption that X^n satisfies the SDE with θ_0 instead of θ.

Under regularity conditions the measures $P^{\theta,n}$ are equivalent and possess densities

$$p^{\theta,n}(X^n) = \exp\left(\int_0^{T_n}\left(\frac{\beta^{\theta,n}}{(\sigma^n)^2}\right)(t, X^n)\mathrm{d}X_t^n - \frac{1}{2}\int_0^{T_n}\left(\frac{\beta^{\theta,n}}{\sigma^n}\right)^2(t, X^n)\mathrm{d}t\right) \tag{2.1}$$

relative to a common dominating measure. The following conditions are necessary and sufficient for this to be true, and are assumed throughout the paper:

- There exists a standard filtered probability space $(\Omega^n, \mathcal{F}^n, (\mathcal{F}_t^n, t \geqslant 0), \mathrm{Pr}^n)$ and a parameter value θ_0 on which the SDE (1.1) with θ_0 substituted for θ possesses a solution $X^n = (X_t^n, t \in [0, T_n])$.
- This solution satisfies $\int_0^{T_n}(\beta^{\theta,n}/\sigma^n)^2(t, X^n)\mathrm{d}t < \infty$ Pr^n-almost surely and

$$\mathrm{E}^n \exp\left(\int_0^{T_n}\left(\frac{\beta^{\theta,n} - \beta^{\theta_0,n}}{\sigma^n}\right)(t, X^n)\mathrm{d}B_t^n - \frac{1}{2}\int_0^{T_n}\left(\frac{\beta^{\theta,n} - \beta^{\theta_0,n}}{\sigma^n}\right)^2(t, X^n)\mathrm{d}t\right) = 1.$$

The necessity of these conditions is clear (note that the exponential in the second condition is the quotient $p^{\theta,n}/p^{\theta_0,n}(X^n)$), and the sufficiency follows readily with the help of Girsanov's theorem. There are several approaches in the literature to verifying the first condition under more concrete conditions on the drift and diffusions functions. The second condition is generally hardest to verify. Liptser and Shiryayev (1977) discuss this issue at length and provide elementary sufficient conditions. We defer a discussion of their results to the special examples in the next section.

We assume that the parameter set Θ^n is equipped with some σ-field $\mathcal{B}^n$ and that, for all

n, the map $(x, \theta) \mapsto p^{\theta,n}(x)$ is jointly measurable relative to $\mathcal{C}^n \times \mathcal{B}^n$. Then, given a prior distribution Π^n, a probability distribution on $(\Theta^n, \mathcal{B}^n)$, the *posterior distribution* can be defined by

$$\Pi^n(B|X^n) = \frac{\int_B p^{\theta,n}(X^n)\mathrm{d}\Pi^n(\theta)}{\int_{\Theta^n} p^{\theta,n}(X^n)\mathrm{d}\Pi^n(\theta)}, \qquad B \in \mathcal{B}^n. \tag{2.2}$$

Because the measures $P^{\theta,n}$ are equivalent (by assumption), the expression on the right is well defined with probability one and, apart from definition on a null set, gives a Markov kernel. In the Bayesian set-up it is the conditional distribution of the parameter given X^n, but in this paper we take (2.2) as a definition of the kernel on the left, and study its behaviour under the measures $P^{\theta_0,n}$.

Under mild conditions $\Pi^n(B|X^n) \to 1$ in $P^{\theta_0,n}$-probability as $n \to \infty$ for any fixed 'neighbourhood' B of θ_0. We are interested in the maximal rate at which we can shrink balls around θ_0, while still capturing almost all posterior mass. This can be formalized using a semimetric d_n on the parameter set Θ^n by saying that the sequence of posterior distributions converges to θ_0 (at least) at rate μ_n if for every sequence $M_n \to \infty$,

$$P^{\theta_0,n}\Pi^n(\theta \in \Theta^n : d_n(\theta, \theta_0) \geqslant M_n\mu_n|X^n) \to 0.$$

The posterior rate of convergence reveals the size of Bayesian credibility regions (central regions of mass $1 - \alpha$ in the posterior distribution). It also implies the same rate for a variety of derived point estimators, such as the posterior mode and (under some conditions) the posterior mean.

Our main result is formulated in terms of three semimetrics h_n, d_n and $\bar{d}_n$ on the parameter set Θ^n. The first is the *Hellinger semimetric* h_n given by

$$h_n^2(\theta, \psi) := \int_0^{T_n} \left(\frac{\beta^{\theta,n} - \beta^{\psi,n}}{\sigma^n}\right)^2 (t, X^n)\mathrm{d}t, \qquad \theta, \psi \in \Theta. \tag{2.3}$$

The Hellinger semimetric is random, unlike the other two semimetrics d_n and $\bar{d}_n$ we shall employ, which are ordinary semimetrics. They are related to the Hellinger semimetric through the following assumption. Let μ_n be the desired rate of convergence, a sequence of positive numbers.

Assumption 2.1. *For every $\gamma > 0$ there exist positive constants $c = c_\gamma$, $C = C_\gamma$ and a non-negative constant $D = D_\gamma$ such that*

$$\liminf_{n\to\infty} P^{\theta_0,n}(c d_n(\theta, \theta_0) \leqslant h_n(\theta, \theta_0), \forall \theta \in \Theta^n \text{ with } h_n(\theta, \theta_0) \geqslant D\mu_n,$$

$$\text{and } h_n(\theta, \psi) \leqslant C\bar{d}_n(\theta, \psi), \forall \theta, \psi \in \Theta^n \text{ with } h_n(\theta, \psi) \geqslant D\mu_n) \geqslant 1 - \gamma.$$

The ε*-covering number* of a set A for a semimetric ρ, denoted by $N(\varepsilon, A, \rho)$, is defined as the minimal number of ρ-balls of radius ε needed to cover the set A. The logarithm of the covering number is referred to as the entropy.

Our main theorem poses two conditions: the first (2.4), measures the complexity of the model by the so-called *local Kolmogorov entropy* or *Le Cam dimension*; the second, (2.5), requires that the prior puts sufficient mass close to the true parameter value θ_0. Denote by $B^n(\theta_0, \varepsilon)$ and $\bar{B}^n(\theta_0, \varepsilon)$ the balls of d_n- and $\bar{d}_n$-radius ε around θ_0.

Theorem 2.1. *Let μ_n be a sequence of positive numbers that is bounded away from zero. Suppose Assumption 2.1 is satisfied by the sequence μ_n and that, for every $a > 0$, there exists a constant $g(a) < \infty$ such that*

$$\sup_{\mu > \mu_n} \log N(a\mu, B^n(\theta_0, \mu), \bar{d}_n) \leqslant \mu_n^2 g(a). \tag{2.4}$$

Furthermore, assume that for every $\xi > 0$ there exists an integer J such that, for $j \geqslant J$,

$$\frac{\Pi^n(B^n(\theta_0, j\mu_n))}{\Pi^n(\bar{B}^n(\theta_0, \mu_n))} \leqslant e^{\xi \mu_n^2 j^2}. \tag{2.5}$$

Then for every $M_n \to \infty$, we have that

$$P^{\theta_0, n}\Pi^n(\theta \in \Theta^n : h_n(\theta, \theta_0) \geqslant M_n\mu_n | X^n) \to 0. \tag{2.6}$$

If $\inf_{\gamma > 0} c_\gamma / C_\gamma \geqslant a_0 > 0$, *then the entropy condition* (2.4) *needs to hold for $a = a_0/8$ only. If* $\inf_{\gamma>0} c_\gamma \geqslant c_0 > 0$, *then the prior mass condition* (2.5) *needs to hold for $\xi = c_0^2/9216$ only.*

The proof of the theorem is deferred to Section 4. The assertion of the theorem remains true if h_n in (2.6) is replaced by the lower semimetric d_n.

In our examples the semimetrics satisfy $d_n = c_n d$ and $\bar{d}_n = c_n \bar{d}$, for a sequence of positive numbers c_n and fixed semimetrics d and $\bar{d}$. Scaling properties of entropies and neighbourhoods then yield a rate of convergence $\mu_n = c_n \varepsilon_n$ (with respect to d_n) for ε_n satisfying the bounds

$$\sup_{\varepsilon > \varepsilon_n} \log N(a\varepsilon, B(\theta_0, \varepsilon), \bar{d}) \leqslant c_n^2 \varepsilon_n^2 g(a). \tag{2.7}$$

$$\frac{\Pi^n(B(\theta_0, j\varepsilon_n))}{\Pi^n(\bar{B}(\theta_0, \varepsilon_n))} \leqslant e^{\xi c_n^2 \varepsilon_n^2 j^2}. \tag{2.8}$$

Here $B(\theta_0, \varepsilon)$ and $\bar{B}(\theta_0, \varepsilon)$ are the balls of radius ε around θ_0 for the fixed semimetrics d and $\bar{d}$, respectively. These two equations replace (2.4) and (2.5) in the preceding theorem. It is then still assumed that Assumption 2.1 holds, with $\mu_n = c_n \varepsilon_n$, $d_n = c_n d$ and $\bar{d}_n = c_n \bar{d}$.

The prior mass conditions (2.5) and (2.8) concern the relative amount of prior mass close to θ_0 (denominator) and farther from θ_0 (numerator). Because the numerator is trivially bounded above by 1, (2.5) is implied by the condition

$$\Pi^n(\bar{B}^n(\theta_0, \mu_n)) \geqslant e^{-\mu_n^2}. \tag{2.9}$$

This is a lower bound on the prior mass close to θ_0.

The entropy condition (2.4) is sometimes restrictive, because it treats the parameter set in a uniform way, irrespective of the prior mass. The presence of a subset of parameters with large entropy but small prior mass typically does not affect the rate of convergence. The following lemma allows such situations to be dealt with. We first remark that the preceding theorem remains true if the prior measures Π^n are supported on larger parameter sets $\overline{\Theta}^n \supset \Theta^n$, where the balls $B^n(\theta_0, \varepsilon) = \{\theta \in \Theta^n : d_n(\theta, \theta_0) \leq \varepsilon\}$ are still defined to be subsets of the smaller set Θ^n and the assertion (2.6) remains unchanged. Thus the entropy (2.4) is measured only within Θ^n, but the assertion also only concerns the posterior within Θ^n. (The posterior distribution is now defined by (2.2) with Θ^n replaced by $\overline{\Theta}^n$, for measurable sets $B \subset \overline{\Theta}^n$.) The following lemma, whose proof is given in Section 4.4, allows this to be complemented with a result for parameter values in $\overline{\Theta}^n \backslash \Theta^n$. It shows that sets $\overline{\Theta}^n \backslash \Theta^n$ with very small prior measure automatically have negligible posterior measure, and hence can be ignored.

Lemma 2.2. *If, for every $\gamma > 0$,*

$$\frac{\Pi^n(\overline{\Theta}^n \backslash \Theta^n)}{\Pi^n(\bar{B}^n(\theta_0, \mu_n))} = o(\mathrm{e}^{-(C_\gamma \vee D_\gamma)^2 \mu_n^2}), \tag{2.10}$$

then

$$P^{\theta_0, n}[\Pi^n(\overline{\Theta}^n \backslash \Theta^n | X^n)] \to 0, \qquad n \to \infty.$$

The proof is given in Section 4.4.

3. Special cases

In this section we consider a number of special cases of the Brownian semimartingale model. We give examples of priors and derive the rate of convergence according to our main theorem.

3.1. Signal in white noise

In the signal in white noise model we observe the process X^n satisfying

$$\mathrm{d}X_t^n = \theta_0(t)\mathrm{d}t + \sigma_n \mathrm{d}B_t, \quad t \leq T, \qquad X_0^n = x_0.$$

We observe the process X^n up to a fixed endpoint T. The 'noise level' σ_n is a deterministic sequence of positive numbers that tends to zero as $n \to \infty$. The parameter θ_0 is a deterministic function that belongs to a subset Θ of $L^2[0, T]$. Write $\|\cdot\|$ for the $L^2[0, T]$ norm.

In this case the Hellinger semimetric is non-random and given by

$$h_n(\theta, \psi) = \frac{1}{\sigma_n}\|\theta - \psi\|.$$

It follows that Assumption 2.1 holds with $\gamma = 0$, $c = C = 1$, $D = 0$ and $d_n = \bar{d}_n = h_n$. Theorem 2.1 then yields the following theorem.

Theorem 3.1. *Let ε_n be a sequence of positive numbers such that ε_n/σ_n is bounded away from zero. Suppose that there exists a constant $K < \infty$ such that*

$$\sup_{\varepsilon > \varepsilon_n} \log N(\varepsilon/8, \{\theta \in \Theta^n : \|\theta - \theta_0\| < \varepsilon\}, \|\cdot\|) \leqslant K(\varepsilon_n/\sigma_n)^2.$$

and assume that there exists an integer J such that, for $j \geqslant J$,

$$\frac{\Pi^n(\theta \in \Theta^n : \|\theta - \theta_0\| < j\varepsilon_n)}{\Pi^n(\theta \in \Theta^n : \|\theta - \theta_0\| < \varepsilon_n)} \leqslant \mathrm{e}^{j^2(\varepsilon_n/\sigma_n)^2/9216}. \tag{3.1}$$

Then for any sequence $M_n \to \infty$ we have

$$P^{\theta_0,n}\Pi^n(\theta \in \Theta^n : \|\theta - \theta_0\| \geqslant M_n\varepsilon_n | X^n) \to 0. \tag{3.2}$$

For $\sigma_n = \mathrm{n}^{-1/2}$, we recover Theorem 6 in Ghosal and van der Vaart (2004), who also give examples of priors. Note that the conditions are purely in terms of the L_2 distance.

3.2. Perturbed dynamical system

The 'perturbed dynamical system' is described by the SDE

$$\mathrm{d}X_t^n = \theta_0(X_t^n)\mathrm{d}t + \sigma_n \,\mathrm{d}B_t^n, \qquad t \leqslant T,\, X_0^n = x_0.$$

The 'noise level' σ_n is a sequence of positive constants that tends to zero. We observe the process X^n up to a fixed time T. The parameter θ_0 belongs to a class of functions Θ on the real line.

Under natural conditions, as $n \to \infty$ the processes X^n will tend to the solution $t \mapsto x_t$ of the unperturbed ordinary differential equation (ODE)

$$\mathrm{d}x_t = \theta_0(x_t)\mathrm{d}t.$$

For instance, if θ_0 is Lipschitz, then the Gronwall inequality (Karatzas and Shreve 1991: 287–288) implies that

$$\sup_{0 \leqslant t \leqslant T} |X_t^n - x_t| = O_{P^{\theta_0},n}(\sigma_n).$$

It follows that the process X^n will with probability tending to one take its values in a neighbourhood of the range of the deterministic function $t \mapsto x_t$, and hence in a compact set. The nature of the functions θ in the parameter set Θ therefore matters only through their restrictions to a compact set, and the semimetrics and entropies may be interpreted accordingly.

The convergence of the processes X^n is also the key to finding appropriate semimetrics d_n and $\bar{d}_n$. The Hellinger semimetric h_n is given by

$$h_n(\theta, \psi) = \frac{1}{\sigma_n}\sqrt{\int_0^T (\theta(X_t^n) - \psi(X_t^n))^2 \,\mathrm{d}t}.$$

The convergence of X^n to the solution $t \mapsto x_t$ of the corresponding ODE suggests that

$$\sigma_n^2 h_n^2(\theta, \theta_0) \to d^2(\theta, \theta_0),$$

for

$$d(\theta, \theta_0) = \sqrt{\int_0^T (\theta(x_t) - \theta_0(x_t))^2 \, \mathrm{d}t}, \tag{3.3}$$

We choose $(1/\sigma_n)$ times the semimetric d as both the lower semimetric d_n and upper semimetric $\bar{d}_n$ in the application of our main theorem. Typically, the solution of the ODE will be sufficiently regular to ensure that the semimetric d is equivalent to the L_2 semimetric on the range $\{x_t : t \in [0, T]\}$ of this solution. Of course, the semimetric d is always bounded above by the uniform norm on a neighbourhood of the range $\{x_t : t \in [0, T]\}$ of the solution to the ODE, and hence we may use the uniform metric as well.

That the approximation d/σ_n of h_n satisfies Assumption 2.1 is made precise under a Lipschitz condition in the following theorem.

Theorem 3.2. *Let ε_n be a sequence of positive numbers such that ε_n/σ_n is bounded away from zero. Assume that*

$$\sup_{\theta\in\Theta} \sup_x |\theta(x)| < \infty, \qquad \sup_{\theta\in\Theta} \sup_{x\neq y} \frac{|\theta(x) - \theta(y)|}{|x - y|} < \infty. \tag{3.4}$$

Suppose there exists a constant $K < \infty$ such that

$$\sup_{\varepsilon>\varepsilon_n} \log N(\varepsilon/24, \{\theta \in \Theta^n : d(\theta, \theta_0) < \varepsilon\}, d) \leqslant K(\varepsilon_n/\sigma_n)^2, \tag{3.5}$$

where d is given in (3.3). *Furthermore, assume there exists an integer J such that, for $j \geqslant J$,*

$$\frac{\Pi^n(\theta \in \Theta^n : d(\theta, \theta_0) < j\varepsilon_n)}{\Pi^n(\theta \in \Theta^n : d(\theta, \theta_0) < \varepsilon_n)} \leqslant \mathrm{e}^{\varepsilon_n^2 j^2/(20\,736\sigma_n^2)}. \tag{3.6}$$

Then for every $M_n \to \infty$, we have, as $M_n \to \infty$,

$$P^{\theta_0,n}\Pi^n(\theta \to \Theta^n : d(\theta, \theta_0) \geqslant M_n\varepsilon_n|X^n) \to 0. \tag{3.7}$$

Proof. Under the Lipschitz condition (3.4) the Gronwall inequality mentioned previously shows that

$$\sup_{\theta,\psi\in\Theta} |\sigma_n h_n(\theta, \psi) - d(\theta, \psi)| = O_{P^{\theta_0},n}(\sigma_n) \tag{3.8}$$

(see the proof of Proposition 5.2 in van Zanten 2005). Using Lemma 4.3 from the appendix with $\varepsilon = \frac{1}{2}$, we see that Assumption 2.1 is fulfilled for $c = \frac{2}{3}$, $C = 2$ and $d_n = \bar{d}_n = (1/\sigma_n)d$. The theorem now follows from Theorem 2.1. □

3.2.1. Discrete priors

The current standard for *α-regular functions* on an interval $[-M, M] \subset \mathbb{R}$ is the Besov space $B^\alpha_{p,\infty}$ of functions $\theta : [-M, M] \to \mathbb{R}$ with

$$\|\theta\|^\alpha_{p,\infty} := \|\theta\|_p + \sup_{t>0} \frac{1}{t^\alpha} \sup_{0<h<t} \|\Delta^{\overline{\alpha}}_h \theta\|_p < \infty.$$

Here $\|\cdot\|_p$ is the L_p norm with respect to Lebesgue measure, $\overline{\alpha}$ is an integer strictly bigger than α, and $\Delta^{\overline{\alpha}}_h$ is the $\overline{\alpha}$th difference operator, defined recursively by $\Delta^r_h = \Delta^{r-1}_h \Delta_h$ and $\Delta_h \theta(x) = \theta(x+h) - \theta(x)$ (Devore and Lorentz 1993). This Besov space contains in particular all functions that are $\overline{\alpha}$ times differentiable with bounded $\overline{\alpha}$th derivative. See also Definition 9.2 (p. 104) and Corollary 9.1 (p. 123) in Härdle *et al.* (1998).

For $p > 1/\alpha$ the entropy of the unit ball of the Besov space $B^\alpha_{p,\infty}$ for the uniform norm is of the order $(1/\varepsilon)^{1/\alpha}$ (Birgé and Massart 2000; Kerkyacharian and Picard 2004).

We choose a multiple of this unit ball as parameter set Θ, and define a prior Π^n by choosing for given numbers $\varepsilon_n > 0$ a minimal $\varepsilon_n/2$-net over Θ for the uniform norm and defining Π^n to be the discrete uniform measure on this finite set of functions. If N^n is the number of points in the support of this prior, then $\log N^n$ is of order $(1/\varepsilon_n)^{1/\alpha}$. A uniform neighbourhood of radius ε_n around some $\theta_0 \in \Theta$ contains at least one point of the support, and hence has prior mass at least $1/N^n$.

It follows that the entropy and prior mass conditions (3.5) and (3.6) are satisfied if

$$(1/\varepsilon_n)^{1/\alpha} \leqslant K(\varepsilon_n/\sigma_n)^2,$$

$$\exp(-(1/\varepsilon_n)^{1/\alpha}) \geqslant \mathrm{e}^{-\varepsilon_n^2/\sigma_n^2}.$$

(Bound the numerator of (3.6) by one.) This is satisfied for $\varepsilon_n = \sigma_n^{2\alpha/(2\alpha+1)}$. If the parameters are also uniformly Lipschitz, then the rate of convergence relative to the semimetrics $\sigma_n h_n$ or d is $\sigma_n^{2\alpha/(2\alpha+1)}$.

3.2.2. Priors based on wavelet expansions

Consider as parameter space Θ the set of all functions $\theta : [-M, M] \to \mathbb{R}$ with a bounded αth derivative, for some given natural number α. This parameter set is contained in the Besov space $B^\alpha_{\infty\infty}$ and therefore we can represent every parameter θ in a suitable orthonormal wavelet basis $(\psi_{j,k} : j \in \mathbb{N}, k = 1, \ldots, 2^j)$ in the form

$$\theta(x) = \sum_{j=1}^{\infty} \sum_{k=1}^{2^j} \theta_{j,k} \psi_{j,k}(x),$$

where the Fourier coefficients $\theta_{j,k}$ satisfy

$$\|\theta\|^\alpha_{\infty,\infty} := \sup_j 2^{j\alpha} 2^{j/2} \max_k |\theta_{j,k}| < \infty.$$

A prior on Θ can be defined structurally as

$$\theta \overset{d}{=} \sum_{j=1}^{J} \sum_{k=1}^{2^j} \delta_j Z_{j,k} \psi_{j,k},$$

where $J = J_n$ is chosen dependent on n at a rate to be determined later, δ_j are constants, and $(Z_{j,k} : j \in \mathbb{N},\ k = 1, \ldots, 2^j)$ are i.i.d. standard normal random variables.

We shall show that if $2^{J_n} \sim \sigma_n^{-2/(2\alpha+1)}$ and $\delta_j = 2^{-j/2}$, then the Bayesian rate of convergence relative to the semimetrics h_n/σ_n or d is equal to $\sigma_n^{2\alpha/(2\alpha+1)}$ up to a logarithmic factor. The logarithmic factor is possibly a defect from our proof. The rate $\sigma_n^{2\alpha/(2\alpha+1)}$ is known to be the sharp estimation rate for non-Bayesian procedures, and hence also cannot be improved in the present context.

We derive the rate from Theorem 3.2, setting Θ^n equal to the set of functions $\theta = \sum_{j=1}^{J} \sum_k \theta_{j,k} \psi_{j,k}$ with coefficients $\theta_{j,k}$ bounded absolutely by $M_{j,n} := \delta_j 2^{J/2} a_n$ for $k = 1, \ldots, 2^j$ and $\{a_n\}$ a sequence of positive numbers. Then

$$\Pi^n(\Theta \backslash \Theta^n) = \Pr(\exists j,\ k : |\delta_j Z_{j,k}| > M_{j,n}) \leqslant \sum_{j=1}^{J} 2^j 2(1 - \Phi(M_{j,n}/\delta_j))$$

$$\leqslant \sum_{j=1}^{J} 2^{j+1} \mathrm{e}^{-M_{j,n}^2/2\delta_j^2} \leqslant 2^{J+2} \mathrm{e}^{-2^J a_n^2/2}.$$

We may then use Lemma 2.2 to show that (by an appropriate choice of the numbers $\{a_n\}$) the posterior mass within $\Theta \backslash \Theta^n$ is negligible, and concentrate on the posterior mass inside Θ^n.

The uniform norm of a function θ in the Besov space $B^\alpha_{\infty\infty}$ is equivalent to the norm

$$\|\theta\|_\infty = \sum_{j=1}^{\infty} 2^{j/2} \max_k |\theta_{j,k}|$$

on the Fourier coefficients of the function. If the true parameter θ_0 is contained in $B^\alpha_{\infty\infty}$, then the uniform distance between θ_0 and its projection $\theta_0^J := \sum_{j=1}^{J} \sum_k \theta_{0;j,k} \psi_{j,k}$ on the space spanned by the wavelets of resolution up to J satisfies

$$\|\theta_0 - \theta_0^J\|_\infty = \sum_{j>J} 2^{j/2} \max_k |\theta_{0;j,k}| \leqslant \sum_{j>J} \|\theta_0\|^\alpha_{\infty\infty} 2^{-j\alpha} \leqslant 2^{-J\alpha} \|\theta_0\|^\alpha_{\infty\infty}.$$

See also Section 9.5 of Härdle *et al.* (1998), in particular formulae (9.34) and (9.35). By the triangle inequality it follows that for $2^{-J\alpha}\|\theta_0\|^\alpha_{\infty\infty} < \varepsilon_n$,

$$\begin{aligned}\Pi^n(\theta\in\Theta^n:\|\theta-\theta_0\|_\infty\leqslant 2\varepsilon_n)&\geqslant \Pr\Bigl(\sum_{j=1}^J 2^{j/2}\max_k|\delta_j Z_{j,k}-\theta_{0;j,k}|\leqslant\varepsilon_n\Bigr)\\&\geqslant\prod_{j=1}^J\Pr\Bigl(2^{j/2}\max_k|\delta_j Z_{j,k}-\theta_{0;j,k}|\leqslant\frac{\varepsilon_n}{J}\Bigr)\\&\geqslant\prod_{j=1}^J\prod_k\Bigl[\mathrm{e}^{-\theta_{0;j,k}^2/\delta_j^2}\frac{1}{\sqrt{2}}\Pr\Bigl(\frac{1}{\sqrt{2}}|Z_{j,k}|\leqslant\frac{\varepsilon_n}{J2^{j/2}\delta_j}\Bigr)\Bigr].\end{aligned}$$

In the last step we use the fact that the $N(\theta, 1)$ density is bounded below by $\mathrm{e}^{-\theta^2}/\sqrt{2}$ times the $N(0,\frac{1}{2})$ density, so that $\Pr(|Z-\theta|\leqslant\varepsilon)\geqslant(\mathrm{e}^{-\theta^2}/\sqrt{2})\Pr(|Z|/\sqrt{2}\leqslant\varepsilon)$. For $\varepsilon_n/(J2^{j/2}\delta_j)$ bounded above, the right-hand side is bounded below by, for some positive constant C,

$$\begin{aligned}&C^{2^J}\exp\Bigl(-\sum_{j=1}^J\sum_k\frac{\theta_{0;j,k}^2}{\delta_j^2}\Bigr)\prod_{j=1}^J\Bigl(\frac{\varepsilon_n}{J2^{j/2}\delta_j}\Bigr)^{2^j}\\&\geqslant C^{2^J}\exp\Bigl(-\sum_{j=1}^J\frac{2^{-2j\alpha}}{\delta_j^2}(\|\theta_0\|_{\infty,\infty}^\alpha)^2\Bigr)\exp\Bigl(-\sum_{j=1}^J 2^j\log\Bigl(\frac{J2^{j/2}\delta_j}{\varepsilon_n}\Bigr)\Bigr).\end{aligned}$$

We shall use these estimates to verify the prior mass condition (3.6).

To compute the entropy of Θ^n we choose, for each fixed j, a minimal $2^{(j/2-J)}M_{j,n}\varepsilon/a_n$-net over the interval $[-M_{j,n}, M_{j,n}]^{2^j}\subset\mathbb{R}^{2^j}$ for the maximum norm on $\mathbb{R}^{2^j}$, and form a net over Θ^n by forming arrays $\theta=(\theta_{j,k})$ with the coefficients $(\theta_{j,1},\ldots,\theta_{j,2^j})$ at each level $j\in\{1,\ldots,J\}$ chosen equal to an arbitrary element of the net over $[-M_{j,n}, M_{j,n}]^{2^j}$, and $\theta_{j,k}=0$ for $j>J$. The logarithm of the total number of points θ is bounded by

$$\log\prod_{j=1}^J\Bigl(\frac{3M_{j,n}a_n}{2^{(j/2-J)}M_{j,n}\varepsilon}\Bigr)^{2^j}\leqslant\sum_{j=1}^J 2^j\Bigl(\log\frac{3a_n}{\varepsilon}+(J-j/2)\Bigr)\lesssim 2^J\Bigl(\log\frac{1}{\varepsilon}+\log a_n+J\Bigr).$$

The uniform distance of an arbitrary point $\theta\in\Theta^n$ to the net is bounded above by

$$\sum_{j=1}^J 2^{j/2}2^{(j/2-J)}M_{j,n}\varepsilon/a_n=\varepsilon 2^{-J/2}\sum_{j=1}^J 2^j\delta_j.$$

If the right-hand side is bounded by ε, then it follows that the ε-entropy of Θ^n for the uniform norm is bounded above by $2^J(\log(1/\varepsilon)+J)$.

Combining the foregoing with Lemma 2.2 and Theorem 3.2, we see that the rate of convergence relative to the semimetrics $\sigma_n h_n$ or d is equal to ε_n if the following inequalities are satisfied:

$$\sum_{j=1}^{J} 2^j \log\left(\frac{J2^{j/2}\delta_j}{\varepsilon_n}\right) + \sum_{j=1}^{J} \frac{2^{-2j\alpha}}{\delta_j^2} \lesssim \frac{\varepsilon_n^2}{\sigma_n^2},$$

$$\frac{\varepsilon_n}{J2^{j/2}\delta_j} \lesssim 1,$$

$$2^{-J\alpha} \lesssim \varepsilon_n,$$

$$2^J\left(\log\frac{1}{\varepsilon_n} + \log a_n + J\right) \lesssim K\frac{\varepsilon_n^2}{\sigma_n^2},$$

$$2^{-J/2}\sum_{j=1}^{J} 2^j\delta_j \lesssim 1,$$

where $\lesssim$ denotes inequality up to a fixed positive multiplicative constant.

The first three conditions ensure that the prior-mass condition is satisfied, whereas the fourth and the fifth conditions take care of the entropy condition. It can be verified that the above inequalities are satisfied for $2^J \sim \sigma_n^{-2/(2\alpha+1)}$ and $\varepsilon_n = \sigma_n^{2\alpha/(2\alpha+1)}\log(1/\sigma_n)$ if $a_n = (\log\sigma_n)^2$.

3.3. Ergodic diffusion

In this subsection we consider the SDE

$$\mathrm{d}X_t = \theta_0(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t, \qquad t \leqslant T_n,$$

for a given measurable function σ. Under regularity conditions (see Karatzas and Shreve 1991: Section 5.5), this equation generates a strong Markov process on an interval $I \subseteq \mathbb{R}$, with scale function s_{θ_0} given by

$$s_{\theta_0}(x) = \int_{x_0}^{x} \exp\left(-2\int_{x_0}^{y}\frac{\theta_0(z)}{\sigma^2(z)}\mathrm{d}z\right)\mathrm{d}y$$

(x_0 is an arbitrary, but fixed point in the state space) and speed measure

$$m_{\theta_0}(\mathrm{d}x) = \frac{\mathrm{d}x}{s'_{\theta_0}(x)\sigma^2(x)}.$$

We assume that m_{θ_0} has finite total mass, $m_{\theta_0}(I) < \infty$. Then the diffusion is ergodic, and the normalized speed measure $\mu_0 = m_{\theta_0}/m_{\theta_0}(I)$ is the unique invariant probability measure. For simplicity, the initial law of the diffusion is supposed to be degenerate at some point $x \in I$. The endpoint T_n of the observation interval is assumed to tend to infinity as $n \to \infty$. The parameter set Θ is a collection of real functions on the interval I.

In this model the square of the Hellinger semimetric h_n in (2.3) is given by

$$h_n^2(\theta, \psi) = \int_0^{T_n}\left(\frac{\theta(X_t) - \psi(X_t)}{\sigma(X_t)}\right)^2 \mathrm{d}t.$$

Using the occupation time formula $\int_0^t f(X_s)\mathrm{d}s = \int_I f\, l_t \mathrm{d}m_0$, we can rewrite this semimetric in terms of the diffusion local time $(l_t(x),\ t \geqslant 0,\ x \in I)$ of the process X relative to its speed measure m_{θ_0} (see Itô and McKean 1965) as

$$h_n^2(\theta, \psi) = \int_I \left(\frac{\theta(x) - \psi(x)}{\sigma(x)}\right)^2 l_{T_n}(x)\mathrm{d}m_{\theta_0}(x).$$

An immediate consequence is that for any interval $I^* \subset I$,

$$\inf_{x\in I^*} l_{T_n}(x)\left\|\frac{\theta-\psi}{\sigma}\mathbf{1}_{I^*}\right\|^2_{L^2(m_{\theta_0})} \leqslant h_n^2(\theta,\psi) \leqslant \sup_{x\in I} l_{T_n}(x)\left\|\frac{\theta-\psi}{\sigma}\right\|^2_{L^2(m_{\theta_0})}. \qquad (3.9)$$

Because the infimum and supremum over the scaled local time $(1/T_n)l_{T_n}$ are appropriately bounded away from zero and infinity (see the proof below), we can choose $\sqrt{T_n}$ times the L_2 metrics appearing on the left and right of this display as the semimetrics d_n and $\bar{d}_n$ in the application of our main theorem.

This leads to the following theorem.

Theorem 3.3. *Let ε_n be a sequence of positive numbers such that $T_n\varepsilon_n^2$ is bounded away from zero. Let I^* be a compact subinterval of I. Suppose that for every $a > 0$ there exists a constant $K < \infty$ such that*

$$\sup_{\varepsilon>\varepsilon_n} \log N(a\varepsilon, \{\theta\in\Theta : \|(\theta-\theta_0)\mathbf{1}_{I^*}/\sigma\|_{L_2(\mu_0)} < \varepsilon\}, L_2(\mu_0)) \leqslant KT_n\varepsilon_n^2. \qquad (3.10)$$

Furthermore, assume that for every $\xi > 0$ there is a constant J such that for $j \geqslant J$,

$$\frac{\Pi^n(\theta\in\Theta : \|(\theta-\theta_0)\mathbf{1}_{I^*}/\sigma\|_{L_2(\mu_0)} < j\varepsilon_n)}{\Pi^n(\theta\in\Theta : \|(\theta-\theta_0)/\sigma\|_{L_2(\mu_0)} < \varepsilon_n)} \leqslant \mathrm{e}^{\xi T_n\varepsilon_n^2 j^2}. \qquad (3.11)$$

Then for every $M_n \to \infty$, we have that

$$P^{\theta_0,n}\Pi^n(\theta\in\Theta^n : \|(\theta-\theta_0)\mathbf{1}_{I^*}/\sigma\|_{L_2(\mu_0)} \geqslant M_n\varepsilon_n | X^n) \to 0. \qquad (3.12)$$

Proof. The assertion follows from Theorem 2.1 once it has been established that Assumption 2.1 is satisfied for $d_n := \sqrt{T_n}d$ and $\bar{d}_n := \sqrt{T_n}\bar{d}$, where d and $\bar{d}$ are the L_2 metrics appearing on the left and right of (3.9).

Now, according to Theorems 3.1 and 3.2 of van Zanten (2003) we have, with $M = m_{\theta_0}(I)$, that

$$\sup_{x\in I} l_{T_n}(x) = O_P(T_n),$$

$$\sup_{x\in I^*}\left|\frac{1}{T_n}\, l_{T_n}(x) - \frac{1}{M}\right| \xrightarrow{\mathrm{p}} 0.$$

Hence, for $\gamma > 0$ there exists a constant $C = C_\gamma > 0$ such that

$$P^{\theta_0,n}\left(\frac{1}{T_n}\sup_{x\in I} l_{T_n}(x) \leqslant C\right) \geqslant 1-\gamma,$$

and we have that

$$P^{\theta_0,n}\left(\inf_{x\in I^*}\frac{1}{T_n}\, l_{T_n}(x) \geqslant \frac{1}{2M}\right) \geqslant P^{\theta_0,n}\left(\sup_{x\in I^*}\left|\frac{1}{T_n}\, l_{T_n}(x) - \frac{1}{M}\right| \leqslant \frac{1}{2M}\right) \to 1.$$

Therefore, the events $U^n = \{1/(2M) \leqslant (1/T_n) l_{T_n}(x) \leqslant C\ \forall x \in I^*\}$ have probability satisfying $\liminf_{n\to\infty} P^{\theta_0,n}(U^n) \geqslant 1-\gamma$, and on U^n we have $1/(2M)d_n^2 \leqslant h_n^2 \leqslant C\bar{d}_n^2$ for all θ, $\psi \in \Theta^n$. Thus Assumption 2.1 is satisfied with $\mu_n = 1$. □

From a modelling perspective the most interesting case is that the state space I of the diffusion is a bounded open interval. Then we shall never observe the full state space in finite time, as the sample paths $t \mapsto X_t$ are continuous functions with range strictly within the state space. A model will specify the parameters $\theta : I \to \mathbb{R}$ on an interval containing the range of the observed sample path. (Note that correspondingly the preceding theorem gives consistency of the estimator on compact subintervals of the state space only.) These parameters should also be specified so that the resulting diffusion equation possesses an ergodic solution that remains within the interval. The most interesting (and simplest) case is that the diffusion function σ is strictly positive on the state space I and tends to zero at its boundaries, so that the diffusion part of the differentials $\mathrm{d}X_t$ become negligible as the sample path $t \mapsto X_t$ approaches the boundary. The drift parameters θ should then be positive near the left boundary of I and negative near the right boundary, so that the differentials $\mathrm{d}X_t$ become positive and negative at these two boundaries, thus deflecting the sample path near the boundaries of the state space.

Following Liptser and Shiryayev (1977), we give conditions that make the foregoing precise and ensure that the conditions at the beginning of Section 2 are satisfied. Then, we discuss examples of prior distributions. For simplicity of notation we take the state space equal to the open unit interval $I = (0, 1)$. We assume that the drift function $\sigma : (0, 1) \mapsto \mathbb{R}$ is strictly positive and Lipschitz, with, for some numbers p, $q \geqslant 0$,

$$\sigma(x) \sim x^{1+p}, \quad \text{as } x \downarrow 0, \qquad \sigma(x) \sim (1-x)^{1+q}, \quad \text{as } x \uparrow 1,$$

where $f \sim g$ denotes that the quotient of the functions f and g tends to a strictly positive finite constant. Then the diffusion equation

$$\mathrm{d}X_t = \theta(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t, \quad t \leqslant T_n, \qquad X_0 = x_0$$

possesses a unique strong solution X for any initial value $x_0 \in (0, 1)$ for any Lipschitz function $\theta : (0, 1) \to \mathbb{R}$ that is positive and bounded away from zero in a neighbourhood of 0 and negative and bounded away from zero in a neighbourhood of 1. The corresponding scale function s_θ can be seen to satisfy $s_\theta(x) \to -\infty$ as $x \downarrow 0$ and $s_\theta(x) \to \infty$ as $x \uparrow 1$ and hence maps I onto $\mathbb{R}$ (Proposition 5.22(a) in Karatzas and Shreve 1991). It follows that the diffusion X is recurrent on the state space I with speed measure m_θ that has a continuous density, which is bounded by

$$\frac{C_1}{x^{2+2p}}\,\mathrm{e}^{-C_2 x^{-1-2p}} \quad \text{and} \quad \frac{C_1}{(1-x)^{2+2q}}\,\mathrm{e}^{-C_3(1-x)^{-1-2q}}$$

near 0 and 1, respectively. Here C_1, C_2 and C_3 are positive constants. In particular, the speed measure m_θ is finite, so that the diffusion is positive recurrent and ergodic. We also have that $\int_0^1 \sigma^{-2}(x)\mathrm{d}m_\theta(x) < \infty$, so that

$$\int_0^{T_n} \left(\frac{\vartheta}{\sigma}\right)^2 (X_t)\mathrm{d}t \leqslant \sup_x L^{\vartheta}_{T_n}(x) \int_0^t \left(\frac{\vartheta}{\sigma}\right)^2 (x)\mathrm{d}m_\theta(x) < \infty,$$

for any bounded function $\vartheta : (0, 1) \to \mathbb{R}$. According to Theorems 7.19 and 7.20 of Liptser and Shiryayev (1977), the induced distributions $P^{\theta,n}$ on the Borel sets of $C[0, T_n]$ of the solutions are equivalent, and their likelihood process is given by (2.1).

Thus for a diffusion function σ as given we obtain a valid statistical model for the parameter set Θ equal to the set of Lipschitz functions $\theta : [0, 1] \to \mathbb{R}$ that are positive and bounded away from zero near 0, and negative and bounded away from zero near 1. In the following sections we discuss examples of priors on this parameter set.

3.3.1. Monotone drift functions

Let the parameter set Θ be the set of all monotone Lipschitz functions $\theta : [0, 1] \to \mathbb{R}$ with $\theta(0) > 0$ and $\theta(1) < 0$. Given a finite measure α with a continuous positive density on (0, 1) and a positive integer L, we define a prior on this parameter set through the following steps:

- $(D(1/L), D(2/L) - D(1/L), \ldots, D(1) - D(1 - 1/L))$ is Dirichlet distributed on the unit simplex in $\mathbb{R}^L$ with parameter vector $(\alpha(0, 1/L], \alpha(1/L, 2/L], \ldots, \alpha(1 - 1/L, 1))$.
- D is extended to a function $D : (0, 1) \to [0, 1]$ by setting $D(0) = 0$, $D(1) = 1$, and linearly interpolating on the intervals $((j - 1)/L, j/L]$.
- U and V are independent random variables, both independent of D. U is uniformly distributed on (0, 1) and V has a distribution on $[0, \infty)$ with bounded, strictly positive density such that $P(V \geqslant v) \leqslant \mathrm{e}^{-\mathrm{e}^v}$ for large values of v.
- $\theta \overset{d}{=} \mathrm{VU} - VD$.

Thus D is a random distribution function on (0, 1) that is reflected $(-D)$ shifted up to cross the horizontal axis at a random location $(U - D)$ and finally scaled by multiplication with V.

We shall now show that for any $\theta_0 \in \Theta$ the rate of convergence relative to the L_2 metric on a compact subinterval $I^* \subset I$ is at least $T_n^{-1/3} \log T_n$. The rate $T_n^{-1/3}$ is known to be the minimax rate of estimation for this problem, and hence our natural prior yields a posterior which concentrates at a nearly optimal frequentist rate.

We apply Theorem 3.3 with Θ^n equal to $\{\theta \in \Theta : \|\theta\|_\infty \leqslant K_n\}$ for $K_n \sim (\log T_n)^2$. Because a function $\mathrm{VU} - VD$ decreases from VU at 0 to $V(U - 1)$ at 1, its absolute value can take values larger than K only if $V \geqslant K$. Consequently, for n sufficiently large,

$$\Pi^n(\Theta \backslash \Theta^n) \leqslant \Pr(V \geqslant K_n) \leqslant \mathrm{e}^{-\mathrm{e}^{K_n}}.$$

With the help of Lemma 2.2 we shall be able to discard this part of the prior.

The set Θ^n consists of monotone functions $\theta : [0, 1] \to [-K_n, K_n]$. The measure Q_0 defined by $dQ_0(x) = \sigma^{-2}(x)d\mu_0(x)$ is finite. Therefore the ε-entropy of Θ^n relative to the $L_2(Q)$ semimetric is bounded above by a multiple of K_n/ε (van der Vaart and Wellner 1996).

To lower-bound the prior mass of a neighbourhood of θ_0 we first note that, by the triangle inequality, with $D_0 = (\theta_0(0) - \theta_0)/(\theta_0(0) - \theta_0(1))$,

$$\|VU - VD - \theta_0\|_\infty \leqslant |VU - \theta_0(0)| + \|D_0 - D\|_\infty(\theta_0(0) - \theta_0(1)) + |\theta_0(0) - \theta_0(1) - V|.$$

Here $\theta_0(0)$ and $\theta_0(0) - \theta_0(1)$ are positive numbers by assumption, and hence the probability of the intersection of the events that $|VU - \theta_0(0)| < \varepsilon$ and $|\theta_0(0) - \theta_0(1) - V| < \varepsilon$ is of order ε^2 as $\varepsilon \downarrow 0$. By Lemma 3 in Ghosal and van der Vaart (2003) we also have that, for $J\varepsilon \leqslant 1$ and positive constants c and C,

$$\Pr\left(\sum_{j=1}^{L}\left|D\left(\frac{j-1}{L}, \frac{j}{L}\right] - p_j\right| < \varepsilon\right) \geqslant Ce^{-cL\log(1/\varepsilon)},$$

uniformly in $(p_1, \ldots, P_L)$ in the unit simplex. The function D_0 is the cumulative distribution of a probability distribution on (0, 1) and is Lipschitz. It can be seen that

$$\|D_0 - D\|_\infty \leqslant \|D_0\|_{\mathrm{Lip}}\frac{1}{L} + \sum_{j=1}^{L}\left|D\left(\frac{j-1}{L}, \frac{j}{L}\right] - D_0\left(\frac{j-1}{L}, \frac{j}{L}\right]\right|.$$

Here, the Lipschitz norm of a function f is defined by $\|f\|_{\mathrm{Lip}} = \sup_{x\neq y}|f(x) - f(y)|/|x - y|$. Combining these facts, it follows that

$$\begin{aligned}
&\Pi^n(\theta \in \Theta : \|\theta - \theta_0\|_\infty \leqslant 3\varepsilon)\\
&\quad \geqslant \Pr(|VU - \theta_0(0)| < \varepsilon, |\theta_0(0) - \theta_0(1) - V| < \varepsilon)\Pr(\|D_0 - D\|_\infty < \varepsilon)\\
&\quad \gtrsim \varepsilon^2 e^{-cJ\log(1/\varepsilon)}.
\end{aligned}$$

If we choose $J \sim T_n^{1/3}\log T_n$, $K_n = (\log T_n)^2$ and $\varepsilon_n \sim T_n^{-1/3}\log T_n$, then the entropy and prior mass conditions are satisfied.

3.3.2. Parametric models

Consider the ergodic diffusion model with the drift function taking a parametric form. We shall denote the parameter again by θ and write the drift function in the form β^θ. Thus the process X satisfies the SDE

$$dX_t = \beta^\theta(X_t)dt + \sigma(X_t)dB_t,$$

for a given measurable function σ.

Let the parameter θ range over a subset of k-dimensional Euclidean space $(\mathbb{R}^k, \|\cdot\|)$, and assume that there exist functions $\underline{\beta}$ and $\bar{\beta}$ satisfying

$$0 < \int_{I^*} \left(\frac{\underline{\beta}}{\sigma}\right)^2 \mathrm{d}\mu_0(x), \qquad \int_I \left(\frac{\bar{\beta}}{\sigma}\right)^2 \mathrm{d}\mu_0(x) < \infty,$$

and such that, for all $x \in I$ and all $\theta, \psi \in \Theta$,

$$\underline{\beta}(x)\|\theta - \psi\| \leqslant |\beta^\theta(x) - \beta^\psi(x)| \leqslant \bar{\beta}(x)\|\theta - \psi\|.$$

For our purpose it suffices that the first inequality be satisfied for $x \in I^* \subseteq I$.

Under this assumption the entropy and prior mass conditions of Theorem 3.3 can be expressed in terms of corresponding conditions with respect to Euclidean distance, and we obtain the following corollary.

Corollary 3.4. *Let the prior Π^n be independent of n and possess a Lebesgue density that is bounded and bounded away from zero on a neighbourhood of θ_0. Let functions $\underline{\beta}$ and $\bar{\beta}$ exist as in the foregoing. Then for every $M_n \to \infty$, we have, as $n \to \infty$,*

$$P^{\theta_0, n}\Pi^n\left(\theta \to \Theta^n : \|\theta - \theta_0\| \geqslant M_n/\sqrt{T_n}|X^n\right) \to 0. \tag{3.13}$$

Proof. The assumptions imply the existence of positive constants L, U such that

$$L\|\theta - \psi\| \leqslant \left\|\frac{\theta - \psi}{\sigma}\mathbf{1}_{I^*}\right\|_{L_2(\mu_0)} \leqslant \left\|\frac{\theta - \psi}{\sigma}\right\|_{L_2(\mu_0)} \leqslant U\|\theta - \psi\|.$$

These inequalities allow the calculations for Theorem 3.3 to be carried out using Euclidean balls and distances.

First, the bounds imply that the left-hand side of (3.10) is bounded above by

$$\sup_{\varepsilon > \varepsilon_n} \log N(a\varepsilon/U, \{\theta : \|\theta - \theta_0\| \leqslant \varepsilon/L\}, \|\cdot\|) \leqslant k \log\left(\frac{5\varepsilon/L}{a\varepsilon/U}\right),$$

(Pollard 1990: Lemma 4.1) which is bounded above by a constant, independently of ε.

Secondly, the comparison of norms shows that the quotient on the left-hand side of (3.11) is bounded above by

$$\frac{\Pi^n(\theta \to \Theta^n : \|\theta - \theta_0\| \leqslant j\varepsilon/L)}{\Pi^n(\theta \to \Theta^n : \|\theta - \theta_0\| \leqslant \varepsilon/U)} \leqslant \frac{M}{m}\left(\frac{j\varepsilon/L}{\varepsilon/U}\right)^k = \frac{M}{m}\left(\frac{jU}{L}\right)^k,$$

where m and M are lower and upper bounds on the density of the prior. □

The rate of convergence $T_n^{-1/2}$ is sharp and was previously obtained by Kutoyants (2004).

4. Proofs

4.1. Proof of Theorem 2.1

For given μ_n and $M_n \to \infty$, denote by U^n the random set

$$U^n = \{\theta \in \Theta^n : h_n(\theta, \theta_0) \geqslant M_n \mu_n\}.$$

For given positive constants c, C, D, define events

$$\bar{A}_{n,C,D} := \{\omega : h_n(\theta, \psi)(\omega) \leqslant C\bar{d}_n(\theta, \psi), \forall \theta, \psi \in \Theta^n \text{ with } h_n(\theta, \psi)(\omega) \geqslant D\mu_n\},$$

$$A_{n,c,D} := \{\omega : h_n(\theta, \theta_0)(\omega) \geqslant c d_n(\theta, \theta_0)), \forall \theta \in \Theta^n \text{ with } h_n(\theta, \theta_0)(\omega) \geqslant D\mu_n\},$$

According to Assumption 2.1, there exist positive constants c, C, D such that the events $A_{n,c,D} \cap \bar{A}_{n,C,D}$ have probability arbitrarily close to one as $n \to \infty$. It therefore suffices to show that the sequence $P^{\theta_0,n}\Pi^n(U^n|X^n)1_{A_{n,c,D}\cap\bar{A}_{n,C,D}}$ tends to zero for fixed positive constants c, C, D. Furthermore, if the constants c_γ, C_γ in Assumption 2.1 satisfy $\inf_{\gamma>0} c_\gamma / C_\gamma \geqslant a_0 > 0$ and/or $\inf_{\gamma>0} c_\gamma \geqslant c_0 > 0$, then it suffices to consider c, C, D satisfying these restrictions only.

In Lemma 4.1 we construct test functions $\varphi^n : \Omega \to [0, 1]$ that are consistent for the null hypothesis $H_0 : \theta = \theta_0$, that is, $P^{\theta_0,n}\varphi^n \to 0$ as $n \to \infty$. Since $1 = \varphi^n + (1 - \varphi^n)$, we can bound

$$P^{\theta_0,n}\Pi^n(U^n|X^n)\mathbf{1}_{A_{n,c,D}\cap\bar{A}_{n,C,D}} \leqslant P^{\theta_0,n}\varphi^n + P^{\theta_0,n}\Pi^n(U^n|X^n)(1 - \varphi^n)\mathbf{1}_{A_{n,c,D}\cap\bar{A}_{n,C,D}}. \tag{4.1}$$

Here the first term on the right tends to zero by consistency, and hence it suffices to concentrate on the second term. We rewrite the posterior distribution (2.2) as

$$\Pi^n(B|X^n) = \frac{\int_B p^{\theta,n}/p^{\theta_0,n}(X^n)\mathrm{d}\Pi^n(\theta)}{\int_{\Theta^n} p^{\theta,n}/p^{\theta_0,n}(X^n)\mathrm{d}\Pi^n(\theta)}, \qquad B \in \mathcal{B}^n. \tag{4.2}$$

The set of interest is the union $U^n = \cup_{i \geqslant M_n} \Theta_i^n$ of the random rings defined by

$$\Theta_i^n = \{\theta \in \Theta^n : i\mu_n \leqslant h_n(\theta, \theta_0) < (i+1)\mu_n\}, \qquad i \in \mathbb{N}.$$

Therefore, we can bound the second term on the right in (4.1) by

$$\sum_{i \geqslant M_n} P^{\theta_0,n}\left[\frac{\int_{\Theta_i^n} p^{\theta,n}/p^{\theta_0,n}\,\mathrm{d}\Pi^n(\theta)}{\int p^{\theta,n}/p^{\theta_0,n}\,\mathrm{d}\Pi^n(\theta)}(1 - \varphi^n)\mathbf{1}_{A_{n,c,D}\cap\bar{A}_{n,C,D}}\right]. \tag{4.3}$$

The main part of the proof is to construct the test functions in such a way that the terms in this sum are small. Here we bound the denominator from below by a constant, and use Fubini's theorem to bound

$$P^{\theta_0,n}\int_{\Theta_i^n}\frac{p^{\theta,n}}{p^{\theta_0,n}}\,\mathrm{d}\Pi^n(\theta)(1-\varphi^n)\mathbf{1}_{A_{n,c,D}\cap\bar{A}_{n,C,D}}$$
$$\leqslant\int P^{\theta,n}\mathbf{1}_{\{\theta\in\Theta_i^n\}}(1-\varphi^n)\mathbf{1}_{A_{n,c,D}\cap\bar{A}_{n,C,D}}\,\mathrm{d}\Pi^n(\theta).$$

The following two lemmas assert the existence of appropriate test functions φ^n and give the lower bound on the denominator.

Lemma 4.1. *If condition* (2.4) *holds, then for every positive constant* μ_n, *c, C, D and sufficiently large integer I there exists a test* φ^n *(depending on* μ_n, *c, C, D, I) such that*

$$P^{\theta_0,n}\varphi^n\leqslant\exp\left(\mu_n^2 g\left(\frac{c}{8C}\right)\right)\sum_{i\geqslant I}\mathrm{e}^{-i^2\mu_n^2/512} \tag{4.4}$$

and, for all $i\geqslant I$,

$$P^{\theta,n}(1-\varphi^n)\mathbf{1}_{\{\theta\in\Theta_i^n\}}\mathbf{1}_{A_{n,c,D}\cap\bar{A}_{n,C,D}}\leqslant\mathrm{e}^{-i^2\mu_n^2/1152}. \tag{4.5}$$

Lemma 4.2. *For every* $\varepsilon>0$ *and* $K>0$,

$$P^{\theta_0,n}\left(\int\frac{p^{\theta,n}}{p^{\theta_0,n}}\,\mathrm{d}\Pi^n(\theta)\leqslant\exp\left(-\left(\frac{1}{2}(C^2\varepsilon^2\vee D^2\mu_n^2)+K\varepsilon^2\right)\right)\Pi^n(\bar{B}^n(\theta_0,\varepsilon)),\ \bar{A}_{n,C,D}\right)$$
$$\leqslant\exp\left(-\frac{K^2\varepsilon^4}{2(C^2\varepsilon^2\vee D^2\mu_n^2)}\right).$$

The proofs of these lemmas are deferred to Sections 4.2 and 4.3. We first proceed with the proof of the main theorem. Choose $I=M_n\to\infty$ and let φ^n be tests as in Lemma 4.1.

Since $g(c/8C)<\infty$, assertion (4.4) of Lemma 4.1 implies that $P^{\theta_0,n}\varphi^n\to 0$ if μ_n is bounded away from zero and $I=I_n\to\infty$.

By Lemma 4.2, applied with $\varepsilon=\mu_n$, expression (4.3) can be bounded by

$$\sum_{i\geqslant M_n}P^{\theta_0,n}\frac{\int_{\Theta_i^n}p^{\theta,n}/p^{\theta_0,n}\,\mathrm{d}\Pi^n(\theta)}{\exp\left(\left(-\frac{1}{2}(C\vee D)^2+K\right)\mu_n^2\right)\Pi^n(\bar{B}^n(\theta_0,\mu_n))}(1-\varphi^n)\mathbf{1}_{A_{n,c,D}\cap\bar{A}_{n,C,D}}$$
$$+\exp\left(-\frac{K^2\mu_n^2}{2(C\vee D)^2}\right).$$

The second term can be made arbitrarily small by choice of K. The first term can be handled using Fubini's theorem as in (4.4), and inequality (4.5). Here, since $\Theta_i^n(\omega)\subset B^n(\theta_0,2i\mu_n/c)$ if $\omega\in A_{n,c,D}$ and $i\geqslant D\vee 2$, we may restrict the integral to the (non-random) set $B^n(\theta_0,2i\mu_n/c)$. Thus, for n sufficiently large, we obtain the bound

$$\sum_{i \geq M_n} \frac{\Pi^n(B^n(\theta_0, 2\mu_n i/c))}{\Pi^n(\bar{B}^n(\theta_0, \mu_n))} \exp\left(\left(\frac{1}{2}(C \vee D)^2 + K\right)\mu_n^2 - \frac{i^2 \mu_n^2}{1152}\right).$$

Taking $\xi = c^2/(8 \cdot 1152)$ in condition (2.5), we see that the latter is, for sufficiently large n, bounded by

$$\sum_{i \geq M_n} \exp\left\{\left(\frac{1}{2}(C \vee D)^2 + K\right)\mu_n^2 - \frac{i^2 \mu_n^2}{2 \times 1152}\right\},$$

which tends to zero, as $M_n \to \infty$, for any fixed C, D, K. This concludes the proof of the main theorem.

4.2. Proof of Lemma 4.1

The proof is based on the following version of Bernstein's inequality: if M is a continuous local martingale vanishing at 0 with quadratic variation process $[M]$, then, for any stopping time T and all x, $L > 0$,

$$P\left(\sup_{0 \leq t \leq T} |M_t| \geq x, [M]_T \leq L\right) \leq \mathrm{e}^{-x^2/(2L)}$$

(see, for instance, Revuz and Yor 1999: 153–154). We shall apply this inequality to two local martingales derived from the log-likelihood.

First (cf. (2.1)), the log-likelihood ratio process can be written as

$$\ell(\theta) := \log \frac{p^{\theta,n}}{p^{\theta_0,n}}(X^n) = M_{T_n}^{\theta,n} - \tfrac{1}{2}[M^{\theta,n}]_{T_n},$$

where $M^{\theta,n}$ is the $P^{\theta_0,n}$-local martingale

$$M_t^{\theta,n} = \int_0^t \left(\frac{\beta_s^{\theta,n} - \beta_s^{\theta_0,n}}{\sigma_s^n}\right) \mathrm{d}B_s^n, \qquad t \geq 0,\ \theta \in \Theta, \tag{4.6}$$

for B^n a Brownian motion under $P^{\theta_0,n}$. The quadratic variation of $M^{\theta,n}$ at T_n is precisely the square Hellinger semidistance $h_n^2(\theta, \theta_0) = [M^{\theta,n}]_{T_n}$.

Under $P^{\theta,n}$ the process $M^{\theta,n}$ is not a local martingale. However, by Girsanov's theorem the process

$$B_t^{\theta,n} = B_t^n - \int_0^t \left(\frac{\beta_s^{\theta,n} - \beta_s^{\theta_0,n}}{\sigma_s^n}\right) \mathrm{d}s$$

is a $P^{\theta,n}$-Brownian motion, and we can write

$$\ell(\theta_1) = Z_{T_n}^{\theta_1,\theta,n} + \tfrac{1}{2}[Z^{\theta_1,\theta,n}]_{T_n} + \int_0^{T_n} \left(\frac{\beta_t^{\theta_1,n} - \beta_t^{\theta_0,n}}{\sigma_t^n}\right)\left(\frac{\beta_t^{\theta,n} - \beta_t^{\theta_1,n}}{\sigma_t^n}\right) \mathrm{d}t, \tag{4.7}$$

for the $P^{\theta,n}$-local martingale $Z^{\theta_1,\theta,n}$ defined by

$$Z_t^{\theta_1,\theta,n} = \int_0^t \left(\frac{\beta_s^{\theta_1,n} - \beta_s^{\theta_0,n}}{\sigma_s^n}\right) \mathrm{d}B_s^{\theta,n}, \qquad \theta \in \Theta^n.$$

The quadratic variation of the process $Z^{\theta_1,\theta,n}$ at T_n is again equal to the squared Hellinger semidistance $h_n^2(\theta_1, \theta_0) = [Z^{\theta_1,\theta,n}]_{T_n}$. (The process $Z^{\theta_1,\theta_0 n}$ is equal to the process $M^{\theta_1,n}$ introduced earlier.)

For fixed natural numbers i and n, let $\theta_1, \ldots, \theta_N \in \Theta^n$ be a minimal $\mu_n i/(4C)$-net for $\bar{d}_n$ over the set $B^n(\theta_0, 2i\mu_n/c)$. For sufficiently large i we have $2i\mu_n/c \geqslant \mu_n$, and hence by condition (2.4) the number of points in the net is bounded by

$$N \leqslant N\left(\frac{\mu_n i}{4C}, B^n\left(\theta_0, \frac{2i\mu_n}{c}\right), \bar{d}_n\right) \leqslant \exp\left(\mu_n^2 g\left(\frac{c}{8C}\right)\right). \tag{4.8}$$

Define for each $i \in \mathbb{N}$ a deterministic map $\tau_{ni} : \Theta^n \to \{\theta_1, \ldots, \theta_N\}$ by mapping each $\theta \in B^n(\theta_0, 2i\mu_n/c)$ into a closest point of the net and mapping each other $\theta \in \Theta^n$ in an arbitrary point of the net. For each $\theta \in \Theta^n$ and $i \in \mathbb{N}$ define a test by

$$\varphi_i^{\theta,n} := \mathbf{1}\{\ell(\theta) > 0,\, i\mu_n/2 < h_n(\theta, \theta_0) < 2i\mu_n\},$$

and set

$$\varphi^n := \sup_{i \geqslant I} \sup_{\theta \in \tau_{ni}(\Theta^n)} \varphi_i^{\theta,n},$$

We shall show that the latter tests satisfy (4.4) and (4.5) if I is sufficiently large.

The error of the first kind (4.4) of these tests satisfies

$$P^{\theta_0,n}\varphi^n \leqslant \sum_{i \geqslant I} \sum_{\theta \in \tau_{ni}(\Theta^n)} P^{\theta_0,n}\varphi_i^{\theta,n} \leqslant \left(\sup_{i \geqslant I} \#\tau_{ni}(\Theta^n)\right) \sum_{i \geqslant I} \max_{\theta \in \tau_{ni}(\Theta^n)} P^{\theta_0,n}\varphi_i^{\theta,n}.$$

Here the cardinality of the sets $\tau_{ni}(\Theta^n)$ is bounded above in (4.8). The probabilities on the right of the last display can be bounded with the help of Bernstein's inequality

$$\begin{aligned} P^{\theta_0,n}\varphi_i^{\theta,n} &= P^{\theta_0,n}\left(M_{T_n}^{\theta,n} - \tfrac{1}{2}h_n^2(\theta, \theta_0) > 0,\, i\mu_n/2 < h_n(\theta, \theta_0) < 2i\mu_n\right) \\ &\leqslant P^{\theta_0,n}\left(M_{T_n}^{\theta,n} > \tfrac{1}{2}(i\mu_n/2)^2,\, [M^{\theta,n}]_{T_n} < (2i\mu_n)^2\right) \leqslant \mathrm{e}^{-i^2\mu_n^2/512}, \end{aligned}$$

uniformly in $\theta \to \Theta^n$. Inserting this bound and the bound (4.8) in the preceding display, we obtain (4.4).

The expectation in (4.5) is restricted to the intersection of the events $\bar{A}_{n,C,D} \cap A_{n,c,D}$ and $\theta \in \Theta_i^n$. By construction of the net $\theta_1, \ldots, \theta_N$,

$$\bar{d}_n(\theta, \tau_{ni}(\theta)) \leqslant \frac{\mu_n i}{4C}, \qquad \text{if } \theta \in B^n\left(\theta_0, \frac{2i\mu_n}{c}\right).$$

We have $\Theta_i^n(\omega) \subset B^n(\theta_0, 2i\mu_n/c)$ if $\omega \in A_{n,c,D}$ and $i \geqslant D \vee 2$. Furthermore, if $\omega \in \bar{A}_{n,C,D}$, then either $h_n(\theta, \tau_{ni}(\theta)) \leqslant D\mu_n$ or the Hellinger semimetric is bounded above by $C\bar{d}_n$. It follows that for $i \geqslant \mathrm{I} \geqslant 4D$, if $\omega \in \bar{A}_{n,C,D} \cap A_{n,c,D}$ and $\theta \in \Theta_i^n(\omega)$, then

$$h_n(\theta, \tau_{ni}(\theta)) \leqslant \frac{\mu_n i}{4}. \tag{4.9}$$

By the triangle inequality it then follows that

$$\frac{3i\mu_n}{4} \leqslant h_n(\theta_0, \tau_{ni}(\theta)) \leqslant (i + 1 + \tfrac{1}{4}i)\mu_n < 2\mu_n i, \qquad i \geqslant 2. \tag{4.10}$$

Therefore, if $\omega \in A_{n,c,D} \cap \bar{A}_{n,C,D}$ and $\theta \in \Theta_{n,i}(\omega)$,

$$1 - \varphi^n \leqslant 1 - \varphi_i^{\tau_{ni}(\theta),n} \leqslant \mathbf{1}\{\ell(\tau_{ni}(\theta)) \leqslant 0\}.$$

We write the log-likelihood ratio $\ell(\tau_{ni}(\theta))$ in terms of the process $Z^{\tau_{ni}(\theta),\theta,n}$ as in (4.7), where by the Cauchy–Schwarz inequality the inner product in (4.7) can be bounded as

$$\begin{aligned}\left|\int_0^{T_n}\left(\frac{\beta_t^{\tau_{ni}(\theta),n} - \beta_t^{\theta_0,n}}{\sigma_t^n}\right)\left(\frac{\beta_t^{\theta,n} - \beta_t^{\tau_{ni}(\theta),n}}{\sigma_t^n}\right)\mathrm{d}t\right| &\leqslant h_n(\tau_{ni}(\theta), \theta_0)h_n(\theta, \tau_{ni}(\theta)) \\ &\leqslant \frac{1}{3}\,h_n^2(\tau_{ni}(\theta), \theta_0),\end{aligned}$$

for $\omega \in A_{n,c,D} \cap \bar{A}_{n,C,D}$ and $\theta \in \Theta_{n,i}(\omega)$, since $h_n(\theta, \tau_{ni}(\theta)) \leqslant \mu_n i/4 \leqslant h_n(\theta_0, \tau_{ni}(\theta))/3$ on this event, by (4.9) and (4.10). It follows that the variable $\ell(\tau_{ni}(\theta))$ is bounded below by $Z_{T_n}^{\tau_{ni}(\theta),\theta,n} + [Z^{\tau_{ni}(\theta),\theta,n}]_{T_n}/6$, and therefore

$$\begin{aligned}&P^{\theta,n}(1-\varphi^n)\mathbf{1}_{\{\theta\in\Theta_i^n\}}\mathbf{1}_{A_{n,c,D}\cap\bar{A}_{n,C,D}} \\ &\quad\leqslant P^{\theta,n}(Z_{T_n}^{\tau_{ni}(\theta),\theta,n} + [Z^{\tau_{ni}(\theta),\theta,n}]_{T_n}/6 \leqslant 0, \{\theta \in \Theta_i^n\}) \\ &\quad\leqslant P^{\theta,n}\left(|Z_{T_n}^{\tau_{ni}(\theta),\theta,n}| \geqslant \frac{1}{12}\mu_n^2 i^2, [Z^{\tau_{ni}(\theta),\theta,n}]_{T_n} \leqslant 4\mu_n^2 i^2\right) \\ &\quad\leqslant \mathrm{e}^{-i^2\mu_n^2/1152},\end{aligned}$$

by Bernstein's inequality.

4.3. Proof of Lemma 4.2

Let $\tilde{\Pi}^n$ be equal to the measure Π^n restricted and renormalized to be a probability measure on $\bar{B}^n(\theta_0, \varepsilon)$. By Jensen's inequality, with $M^{\theta,n}$ the local martingale in (4.6),

$$\begin{aligned}\log\int\frac{p^{\theta,n}}{p^{\theta_0,n}}\frac{\mathrm{d}\Pi^n(\theta)}{\Pi^n(\bar{B}^n(\theta_0,\varepsilon))} &\geqslant \int\log\frac{p^{\theta,n}}{p^{\theta_0,n}}\,\mathrm{d}\tilde{\Pi}^n(\theta) \\ &= \int(M_T^{\theta,n} - \tfrac{1}{2}h_n^2(\theta,\theta_0))\mathrm{d}\tilde{\Pi}^n(\theta) \\ &\geqslant Z_T^n - \tfrac{1}{2}(C^2\varepsilon^2 \vee D^2\mu_n^2),\end{aligned} \tag{4.11}$$

on $\bar{A}_{n,C,D}$, where the process Z^n is defined by

$$Z_t^n := \int M_t^{\theta,n}\, \mathrm{d}\tilde{\Pi}^n(\theta) = \int_0^t \int \left(\frac{\beta_s^{\theta,n} - \beta_s^{\theta_0,n}}{\sigma_s^n} \right) \mathrm{d}\tilde{\Pi}^n(\theta) \mathrm{d}B_s^{\theta_0}.$$

The last equality follows from the stochastic Fubini theorem (see Theorem 64 of Chapter IV in Protter 2004). The process Z^n is a continuous local martingale with respect to $P^{\theta_0,n}$ with quadratic variation process

$$[Z^n]_t = \int_0^t \left(\int \left(\frac{\beta_s^{\theta,n} - \beta_s^{\theta_0,n}}{\sigma_s^n} \right) \mathrm{d}\tilde{\Pi}^n(\theta) \right)^2 \mathrm{d}s,$$

By Jensen's inequality and Fubini's theorem,

$$[Z^n]_T \leqslant \int_0^T \int \left(\frac{\beta_t^{\theta,n} - \beta_t^{\theta_0,n}}{\sigma_t^n} \right)^2 \mathrm{d}\tilde{\Pi}^n(\theta)\mathrm{d}t = \int h_n^2(\theta, \theta_0)\mathrm{d}\tilde{\Pi}^n(\theta).$$

Thus $[Z^n]_T \leqslant C^2\varepsilon^2 \vee D^2\mu_n^2$ on the event $\bar{A}_{n,C,D}$. In view of (4.11) the probability in the lemma is bounded by

$$P^{\theta_0,n}(Z_T^n \leqslant -K\varepsilon^2, [Z^n]_T \leqslant C^2\varepsilon^2 \vee D^2\mu_n^2) \leqslant \exp\left(-\frac{K^2\varepsilon^4}{2(C^2\varepsilon^2 \vee D^2\mu_n^2)} \right),$$

by Bernstein's inequality for continuous local martingales.

4.4. Proof of Lemma 2.2

By Fubini's theorem and the fact that $P^{\theta_0,n}(p^{\theta,n}/p^{\theta_0,n}) \leqslant 1$,

$$P^{\theta_0,n}\left[\int_{\bar{\Theta}^n \setminus \Theta^n} \frac{p^{\theta,n}}{p^{\theta_0,n}}\, \mathrm{d}\Pi^n(\theta) \right] \leqslant \Pi^n(\bar{\Theta}^n \setminus \Theta^n).$$

By Lemma 4.2 with $\varepsilon = \mu_n$ and arbitrary $K > 0$ on the event $\bar{A}_{n,C,D}$ the denominator of the posterior distribution is bounded below by $\exp((-\frac{1}{2}(C \vee D)^2 + K)\mu_n^2)\Pi^n(\bar{B}^n(\theta_0, \mu_n))$ with probability at least $1 - \exp(-K^2\mu_n^2/2(C \vee D^2)$. Choosing $C = C_\gamma$, $D = D_\gamma$, and combining this with the previous display, we obtain

$$\begin{aligned} P^{\theta_0,n}[\Pi^n(\bar{\Theta}^n \setminus \Theta^n | X^n)\mathbf{1}_{\bar{A}_{n,C,D}}] &\leqslant \frac{\Pi^n(\bar{\Theta}^n \setminus \Theta^n)\exp((-\frac{1}{2}(C \vee D)^2 + K)\mu_n^2)}{\Pi^n(\bar{B}^n(\theta_0, \mu_n))} + \exp\left(-\frac{K^2\mu_n^2}{2(C \vee D)^2} \right) \\ &\leqslant o(1)\exp((-\tfrac{1}{2}(C \vee D)^2 + K)\mu_n^2) + \exp\left(-\frac{K^2\mu_n^2}{2(C \vee D)^2} \right), \end{aligned}$$

by assumption (2.10).

If $\mu_n \to \infty$, then we choose $K < (C \vee D)^2/2$, and both terms on the right tend to zero. If μ_n remains bounded, then so is the factor $\exp(-(C \vee D)^2/2 + K)\mu_n^2)$ and hence the first term on the right tends to zero for any fixed K. Furthermore, the second term on the right can be made arbitrarily small by choosing large K in this case.

Thus we have proved the assertion of the lemma on the event $\bar{A}_{n,C,D}$ for each $\gamma > 0$.

This suffices, since the probability of this event can be made arbitrarily large by choice of γ.

4.5. A technical result

The following lemma is helpful in checking Assumption 2.1. It gives a sufficient condition for Assumption 2.1 with $\mu_n = 1$ (and hence also $\mu_n \to \infty$).

Lemma 4.3. *If h_n and d_n are random semimetrics on a set Θ^n with*

$$\sup_{\theta,\psi\in\Theta^n} |h_n(\theta, \psi) - d_n(\theta, \psi)| = O_{P^{\theta_0,n}}(1), \tag{4.12}$$

then for all $\gamma > 0$ there exists a positive constant L such that, for all $\varepsilon \in (0, 1)$,

$$\liminf_{n\to\infty} P^{\theta_0,n}\left(\frac{1}{2-\varepsilon} d_n(\theta, \psi) \leqslant h_n(\theta, \psi) \leqslant \frac{1}{\varepsilon} d_n(\theta, \psi) \right.$$
$$\left. \text{for all } \theta, \psi \in \Theta^n \text{ with } h_n(\theta, \psi) \geqslant \frac{L}{1-\varepsilon}\right) \geqslant 1 - \gamma.$$

Proof. For any $\gamma > 0$ there exists a constant $L_\gamma < \infty$ such that, on an event with probability at least $1 - \gamma$,

$$h_n(\theta, \psi) - L_\gamma \leqslant d_n(\theta, \psi) \leqslant h_n(\theta, \psi) + L_\gamma, \qquad \forall\theta, \psi \in \Theta^n.$$

If $h_n(\theta, \psi) \geqslant L_\gamma/(1-\varepsilon)$, then on the same event

$$\varepsilon h_n(\theta, \psi) \leqslant d_n(\theta, \psi) \leqslant (2-\varepsilon)h_n(\theta, \psi), \qquad \forall\theta, \psi \in \Theta^n.$$

This is the same event as in the assertion of the lemma. □

References

Birgé, L. and Massart, P. (2000) An adaptive compression algorithm in Besov spaces. *Constr. Approx.*, **16**, 1–36.

Devore, R.A. and Lorentz, G.G. (1993) *Constructive Approximation*. Berlin: Springer-Verlag.

Dietz, H.M. and Kutoyants, Yu.A. (2003) Parameter estimation for some non-recurrent solutions of SDE. *Statist. Decisions*, **21**(1), 29–45.

Ghosal, S. and van der Vaart, A.W. (2003) Posterior convergence rates of Dirichlet mixtures of normal distributions for smooth densities. Preprint 2003-12. http://www.math.vu.nl/stochastics/publications.php (accessed 24 February 2006).

Ghosal, S. and van der Vaart, A.W. (2004) Convergence rates of posterior distributions for noniid observations. Preprint 2003-17. http://www.math.vu.nl/stochastics/publications.php (accessed 24 February 2006). *Ann. Statist.* To appear.

Ghosh J.K. and Ramamoorthi, R.V. (2003) *Bayesian Nonparametrics*. New York: Springer-Verlag.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998) *Wavelets, Approximation, and Statistical Applications*. Lectures Notes in Statist. 129. New York: Springer-Verlag.

Höpfner, R. and Kutoyants, Yu.A. (2003) On a problem of statistical inference in null recurrent diffusions. *Statist. Inference Stochastic Process.*, **6**(1), 25–42.

Ibragimov, I.A. and Has'minskii R.Z. (1981) *Statistical Estimation: Asymptotic Theory*. New York: Springer-Verlag.

Itô, K. and McKean, H.P., Jr. (1965) *Diffusion Processes and Their Sample Paths*. Berlin: Springer-Verlag.

Karatzas, I. and Shreve, S.E. (1991) *Brownian Motion and Stochastic Calculus*, 2nd edn. New York: Springer-Verlag.

Kerkyacharian, G. and Picard, D, (2004) Entropy, universal coding, approximation and bases properties. *Constr. Approx.*, **20**, 1–37.

Kutoyants, Yu.A. (1984) *Parameter Estimation for Stochastic Processes*. Berlin: Heldermann Verlag.

Kutoyants, Yu.A. (1994) *Identification of Dynamical Systems with Small Noise*. Dordrecht: Kluwer Academic.

Kutoyants, Yu.A. (2004) *Statistical Inference for Ergodic Diffusion Processes*. New York: Springer-Verlag.

Liptser, R.S. and Shiryayev, A.N. (1977) *Statistics of Random Processes I*. New York: Springer-Verlag.

Loukianova, D. and Loukianov, O. (2005) Deterministic equivalents of additive functionals of recurrent diffusions and drift estimation. Preprint.

Pollard, D. (1990) *Empirical Processes: Theory and Applications*, NSF-CBMS Reg. Conf. Ser. Probab. Statist. 2. Hayward, CA: Institute of Mathematical Statistics; Alexandria, VA: American Statistical Association.

Protter, P. (2004) *Stochastic Integration and Differential Equations*, 2nd edn. Berlin: Springer-Verlag.

Prakasa Rao, B.L.S. (1999) *Statistical Inference for Diffusion Type Processes*. London: Edward Arnold.

Revuz, D. and Yor, M. (1999) *Continuous Martingales and Brownian Motion*, 3rd edn. Berlin: Springer-Verlag.

Shen, X. and Wasserman, L. (2001) Rates of convergence of posterior distributions. *Ann. Statist.*, **29**, 687–714.

van der Vaart, A.W. and Wellner, J.A. (1996) *Weak Convergence and Empirical Processes with Applications to Statistics*. New York: Springer-Verlag.

van Zanten, J.H. (2003) On uniform laws of large numbers for ergodic diffusions and consistency of estimators. *Statist. Inference Stochastic Process.*, **6**(2), 199–213.

van Zanten, J.H. (2005) On the rate of convergence of the maximum likelihood estimator in Brownian semimartingale models. *Bernoulli*, **11**, 643–664.

Zhao, L.H. (2000) Bayesian aspects of some nonparametric problems. *Ann. Statist.*, **28**, 532–552.

Received April 2005 and revised February 2006