Electronic Journal of Statistics
Vol. 7 (2013) 2947–2969
ISSN: 1935-7524
DOI: 10.1214/13-EJS864

Bayesian semi-parametric estimation of the long-memory parameter under FEXP-priors*

Willem Kruijer

Droevendaalsesteeg 1
Biometris, Building 107
6708 PB Wageningen
The Netherlands
e-mail: willem.kruijer@wur.nl

and

Judith Rousseau

3, Avenue Pierre Larousse
92245 Malakoff
France
e-mail: rousseau@ceremade.dauphine.fr

Abstract: In this paper we study the semi-parametric problem of the estimation of the long-memory parameter d in a Gaussian long-memory model. Considering a family of priors based on FEXP models, called FEXP priors in Rousseau et al. (2012), we derive concentration rates together with a Bernstein-von Mises theorem for the posterior distribution of d, under Sobolev regularity conditions on the short-memory part of the spectral density. Three different variations on the FEXP priors are studied. We prove that one of them leads to the minimax (up to a $\log n$ term) posterior concentration rate for d, under Sobolev conditions on the short memory part of the spectral density, while the other two lead to sub-optimal posterior concentration rates in d. Interestingly these results are contrary to those obtained in Rousseau et al. (2012) for the global estimation of the spectral density.

AMS 2000 subject classifications: Primary 62G20; secondary 62M15.
Keywords and phrases: Bayesian semi-parametric, Bernstein-von Mises, FEXP priors, Gaussian long-memory processes, rates of convergence.

Received April 2013.

*This work has been partially supported by the 2007–2010 ANR-07-BLAN-0237-01 "SP Bayes" grant.

†Most of this work was written when the first author was affiliated with the Université Paris Dauphine.

1. Introduction

Let X_t, $t \in \mathbb{Z}$, be a stationary Gaussian time series with zero mean and spectral density $f_o(x)$, $x \in [-\pi, \pi]$, which takes the form

$$|1 - e^{ix}|^{-2d_o} M_o(x), \qquad x \in [-\pi, \pi], \tag{1.1}$$

where $d_o \in (-\frac{1}{2}, \frac{1}{2})$ is called the long-memory parameter, and M_o is a positive continuous function that describes the short-memory behavior of the series. If d_o is positive, this makes the autocorrelation function $\rho(h)$ decay at rate $h^{-(1-2d_o)}$, and the time series is said to have long-memory. When $d_o = 0$, X_t has short memory, and the case $d_o < 0$ is referred to as intermediate or negative memory. Long memory time series models are used in a wide range of applications, such as hydrological or financial time series; see for example Beran (1994) or Robinson (1994). In parametric approaches, a finite dimensional model is used for the short memory part M_o; the most well known example is the ARFIMA(p,d,q) model. The asymptotic properties of maximum likelihood estimators (Dahlhaus (1989) or Lieberman et al. (2003)) and Bayesian estimators (Philippe and Rousseau (2002)) have been established in such models. When the sample size n tends to infinity these estimators are consistent and asymptotically normal with a convergence rate of order $\sqrt{n}$. However when the model for the short memory part is misspecified, the estimator for d can be inconsistent, calling for semi-parametric methods for the estimation of d. A key feature of semi-parametric estimators of the long-memory parameter is that they converge at a rate which depends on the smoothness of the short-memory part, and apart from the case where M_o is infinitely smooth, the convergence rate is smaller than $\sqrt{n}$. The estimation of the long-memory parameter d can thus be considered as a non-regular semi-parametric problem. In Moulines and Soulier (2003) (p. 274) it is shown that when f_o satisfies (1.4), the minimax rate for d is $n^{-\frac{2\beta-1}{4\beta}}$. There are frequentist estimators for d based on the periodogram that achieve this rate (see Hurvich et al. (2002) and Moulines and Soulier (2003)), even without the Gaussianity assumption.

Although Bayesian methods in long-memory models have been widely used (see for instance Ko et al. (2009), Jensen (2004) or Holan et al. (2009)), the literature on convergence properties of non- and semi-parametric estimators is sparse. Rousseau et al. (2012) (RCL hereafter) obtain consistency and rates for the L_2-norm of the log-spectral densities (Theorems 3.1 and 3.2), but for d they only show consistency (Corollary 1). No results exist on the posterior concentration rate on d, and thus on the convergence rates of Bayesian semi-parametric estimators of d. In this paper we aim to fill this gap for a specific family of semi-parametric priors. Obtaining theoretical results on the asymptotic behavior of posterior distributions in semi-parametric problems is difficult. A major difficulty comes from the necessity to control precisely the marginal likelihood of the parameter of interest which is an integral over an infinite dimensional parameter space. Only few results have been obtained so far, and these only deal with regular semi-parametric models, where a $\sqrt{n}$ convergence rate for the parameter of interest can be achieved, see Castillo (2012), Bickel and Kleijn (2012) and Rivoirard and Rousseau (2012). In regular semi-parametric models, a generic technique can be applied based on an expansion of the log-likelihood and the least favorable direction as developed in van der Vaart (1998). Here the model is irregular, in the sense that the convergence rate of an estimator of d depends on the smoothness of the nuisance parameter, which is the short memory part of the spectral density. Consequently, the approach for regular semi-parametric models is not feasible. In this paper we propose an alternative approach for studying irregular semi-parametric models. Although we only consider FEXP-models, we believe that this approach can be used in other contexts. More discussion on this is provided in Section 6.

We study Bayesian estimation of d within the FEXP-model (Beran (1993), Robinson (1995)), that contains densities of the form

$$f_{d,k,\theta}(x) = |1 - e^{ix}|^{-2d} \exp\left\{ \sum_{j=0}^{k} \theta_j \cos(jx) \right\}, \tag{1.2}$$

where $d \in (-\frac{1}{2}, \frac{1}{2})$, k is a nonnegative integer and $\theta \in \mathbb{R}^{k+1}$. The factor $\exp\{\sum_{j=0}^{k} \theta_j \cos(jx)\}$ models the function M_o in (1.1). In contrast to the original finite-dimensional FEXP-model (Beran (1993)), where k was supposed to be known, or at least bounded, f_o may have an infinite FEXP-expansion, and we allow k to increase with the number of observations to obtain approximations f that are increasingly close to f_o. Note that the case where the true spectral density satisfies $f_o = f_{d_o,k_o,\theta_o}$, is considered in Holan et al. (2009). In this paper we will pursue a fully Bayesian semi-parametric estimation of d, the short memory parameter being considered as an infinite-dimensional nuisance parameter.

The priors we consider have been implemented and used in particular by Holan et al. (2009) and Chopin et al. (2013), where in the latter case an efficient sequential monte carlo algorithm is proposed. Here we obtain results on the convergence rate and asymptotic distribution of the posterior distribution for d, which we summarize below in Section 1.1. These are to our knowledge the first of this kind in the Bayesian literature on semi-parametric time series.

1.1. Asymptotic framework and overview

For observations $X = (X_1, \ldots, X_n)$ from a Gaussian stationary time series with spectral density f, let $T_n(f)$ denote the associated covariance matrix and $l_n(f)$ denote the log-likelihood

$$l_n(f) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\det(T_n(f)) - \frac{1}{2}X^t T_n^{-1}(f)X.$$

We consider semi-parametric priors on f based on the FEXP-model defined by (1.2), inducing a parametrization of f in terms of (d, k, θ). Assuming priors π_d for d, and, independent of d, π_k for k and $\pi_{\theta|k}$ for $\theta|k$, we study the (marginal) posterior for d, given by

$$\Pi(d \in D|X) = \frac{\sum_{k=0}^{\infty} \pi_k(k) \int_D \int_{\mathbb{R}^{k+1}} e^{l_n(d,k,\theta)} d\pi_{\theta|k}(\theta) d\pi_d(d)}{\sum_{k=0}^{\infty} \pi_k(k) \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{\mathbb{R}^{k+1}} e^{l_n(d,k,\theta)} d\pi_{\theta|k}(\theta) d\pi_d(d)}. \tag{1.3}$$

The posterior mean or median can be taken as point-estimates for d, but we will focus on the posterior $\Pi(d|X)$ itself.

It is assumed that the true spectral density is of the form

$$\begin{aligned} f_o(x) &= |1 - e^{ix}|^{-2d_o} \exp\left\{ \sum_{j=0}^{\infty} \theta_{o,j} \cos(jx) \right\}, \\ \theta_o &\in \Theta(\beta, L_o) = \{\theta \in l_2(\mathbb{N}) : \sum_{j=0}^{\infty} \theta_j^2 (1+j)^{2\beta} \leq L_o\}, \end{aligned} \tag{1.4}$$

for some known $\beta > 1$.

In particular, we derive bounds on the rate at which $\Pi(d \in D|X)$ concentrates at d_o, together with a Bernstein-von Mises (BVM) property of this distribution. The posterior concentration rate for d is defined as the fastest sequence α_n converging to zero such that

$$\Pi(|d - d_o| < K\alpha_n|X) \xrightarrow{P_o} 1, \quad \text{for a given fixed K.} \tag{1.5}$$

In this paper we study three types of priors for model (1.2), namely priors **A**, **B** and **C** presented in Section 2.1. The priors are based on the sieve model defined by (1.2), with k increasing at the rate $(n/\log n)^{1/(2\beta)}$ (prior **A**), k increasing at rate $(n/\log n)^{1/(2\beta+1)}$ (prior **B**) or with random k (prior **C**). In Section 2 we show that the latter outperforms priors **B** and **C** for the estimation of the long-memory parameter d and leads to minimax posterior concentration rates (up to a $\log n$ term). We provide a lower bound for priors **B** and **C**, which shows that they lead to suboptimal posterior concentration rates for d (Theorem 2.2). This is not a unique phenomenon in (Bayesian) semi-parametric estimation and is encountered for instance in the estimation of a linear functional of the signal in white-noise models, see Li and Zhao (2002) or Arbel et al. (2013).

In addition, we derive a Bernstein-von Mises (hereafter denoted BVM) theorem for the posterior distribution of d (Theorem 2.1). The BVM property means that asymptotically the posterior distribution of d behaves like $\alpha_n^{-1}(d - \hat{d}) \sim \mathcal{N}(0,1)$, where $\hat{d}$ is an estimate whose frequentist distribution (associated to the parameter d) is $\mathcal{N}(d_o, \alpha_n^2)$. We prove such a property on the posterior distribution of d given $k = k_n$. In regular parametric long-memory models, the BVM property has been established by Philippe and Rousseau (2002). It is however much more difficult to establish BVM theorems in infinite dimensional setups, even for independent and identically distributed models; see for instance Freedman (1999), Castillo (2012) and Rivoirard and Rousseau (2012). In particular it has been proved that the BVM property may not be valid, even for reasonable priors. The BVM property is however very useful since it induces a strong connection between frequentist and Bayesian methods. In particular, it implies that Bayesian credible regions are asymptotically also frequentist confidence regions with the same nominal level. In section 2.3 we discuss this issue in more detail. It is unclear whether the above results can be extended to other long memory time series models. This generalization however seems difficult, as (to our knowledge) there are no BVM-theorems in the literature covering broad classes of semi-parametric models.

In section 3 we give a decomposition of $\Pi(d \in D|X)$ defined in (1.3), and obtain bounds for the terms in this decomposition in sections 3.2 and 3.3. Using these results we prove Theorems 2.1 and 2.2 in respectively sections 4 and 5. Conclusions are given in section 6. In the supplementary material (Kruijer and Rousseau (2013)) we give the proofs of the lemmas in section 3, as well as results on the derivatives of the log-likelihood. We conclude this introduction with an overview of the notation.

1.2. Notation

The m-dimensional identity matrix is denoted I_m. We write $|A|$ for the Frobenius or Hilbert-Schmidt norm of a matrix A, i.e. $|A| = \sqrt{\text{tr}AA^t}$, where A^t denotes the transpose of A. The operator or spectral norm is denoted $\|A\|^2 = \sup_{\|x\|=1} x^t A^t A x$. We also use $\|\cdot\|$ for the Euclidean norm on $\mathbb{R}^k$ or $l^2(\mathbb{N})$. The inner-product is denoted $|\cdot|$. We make frequent use of the relations

$$\begin{aligned} &|AB| = |BA| \leq \|A\| \cdot |B|, \quad \|AB\| \leq \|A\| \cdot \|B\|, \quad \|A\| \leq |A| \leq \sqrt{n}\|A\|, \\ &|\text{tr}(AB)| = |\text{tr}(BA)| \leq |A| \cdot |B|, \quad |x^t A x| \leq x^t x \|A\|, \end{aligned} \tag{1.6}$$

see Dahlhaus (1989), p. 1754. For any function $h \in L_1([-\pi, \pi])$, $T_n(h)$ is the matrix with entries $\int_{-\pi}^{\pi} e^{i|l-m|x} h(x)dx$, $l, m = 1, \dots, n$. For example, $T_n(f)$ is the covariance matrix of observations $X = (X_1, \dots, X_n)$ from a time series with spectral density f. If h is square integrable on $[-\pi, \pi]$ we denote

$$\|h\|_2 = \int_{-\pi}^{\pi} h^2(x)dx.$$

The loss $\sqrt{l}$ between spectral densities f and g is defined as

$$l(f, g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\log f(x) - \log g(x))^2 dx.$$

Unless stated otherwise, all expectations and probabilities are with respect to P_o, the law associated with the true spectral density f_o. To avoid ambiguous notation (e.g. θ_0 versus $\theta_{0,0}$) we write θ_o instead of θ_0. Related quantities such as f_o and d_o are also denoted with the o-subscript.

The symbols o_P and O_P have their usual meaning. We use boldface when they are uniform over a certain parameter range. Given a probability law P, a family of random variables $\{W_d\}_{d \in A}$ and a positive sequence a_n, $W_d = \mathbf{o_P}(a_n, A)$

means that

$$P\left(\sup_{d\in A}|W_d|/a_n > \epsilon\right) \to 0, (n\to\infty).$$

When the parameter set is clear from the context we simply write $\mathbf{o_P}(a_n)$. In a similar fashion, we write $\mathbf{o}(a_n)$ when the sequence is deterministic. In conjunction with the o_P and O_P notation we use the letters δ and ϵ as follows. When, for some $\tau > 0$ and a probability P we write $Z = O_P(n^{\tau+\epsilon})$, this means that $Z = O_P(n^{\tau+\epsilon})$ *for all* $\epsilon > 0$. When, on the other hand, $Z = O_P(n^{\tau-\delta})$, we mean that this is true for *some* $\delta > 0$. If the value of δ is of importance it is given a name, for example δ_1 in Lemma 3.4.

The true spectral density of the process is denoted f_o. We denote k-dimensional Sobolev-balls by

$$\Theta_k(\beta, L) = \left\{\theta \in \mathbb{R}^{k+1} : \sum_{j=0}^{k}\theta_j^2(1+j)^{2\beta} \le L\right\} \subset \mathbb{R}^{k+1}. \tag{1.7}$$

For any real number x, let x_+ denote $\max(0, x)$. The number r_k denotes the sum $\sum_{j\ge k+1} j^{-2}$. Let η be the sequence defined by $\eta_j = -2/j$, $j \ge 1$ and $\eta_0 = 0$. For an infinite sequence $u = (u_j)_{j\ge 0}$, let $u_{[k]}$ denote the vector of the first $k+1$ elements. In particular, $\eta_{[k]} = (\eta_0, \dots, \eta_k)$. The letter C denotes any generic constant independent of L_o and L, which are the constants appearing in the assumptions on f_o and the definition of the prior.

2. Main results

Before stating Theorems 2.1 and 2.2 in section 2.3, we state the assumptions on f_o and the prior, and give examples of priors satisfying these assumptions.

2.1. Assumptions on the prior and the true spectral density

We assume observations $X = (X_1, \dots, X_n)$ from a stationary Gaussian time series with law P_o, which is a zero mean Gaussian distribution, whose covariance structure is defined by a spectral density f_o satisfying (1.4), for known $\beta > 1$. It is assumed that for a small constant $t > 0$, $d_o \in [-\frac{1}{2}+t, \frac{1}{2}-t]$.

Assumptions on Π We consider different priors, and first state the assumptions that are common to all these priors. The prior on the space of spectral densities consists of independent priors π_d, π_k and, conditional on k, $\pi_{\theta|k}$. The prior for d has density π_d which is strictly positive on $[-\frac{1}{2}+t, \frac{1}{2}-t]$, the interval which is assumed to contain d_o, and zero elsewhere. The prior for θ given k has a density $\pi_{\theta|k}$ with respect to Lebesgue measure. This density satisfies condition $\mathrm{Hyp}(\mathcal{K}, \mathrm{c}_0, \beta, \mathrm{L_o})$, by which we mean that for a subset $\mathcal{K}$ of $\mathbb{N}$,

$$\min_{k\in\mathcal{K}} \inf_{\theta\in\Theta_k(\beta,L_o)} e^{c_0 k\log k}\pi_{\theta|k}(\theta) > 1,$$

where L_o is as in (1.4). The choice of $\mathcal{K}$ depends on the prior for k and $\theta|k$. We consider the following classes of priors.

- **Prior A:** k is deterministic and increasing at rate

$$k_n = \lfloor k_A (n/\log n)^{\frac{1}{2\beta}} \rfloor, \tag{2.1}$$

  for a constant $k_A > 0$. The prior density for $\theta|k$ satisfies $\mathrm{Hyp}(\{\mathrm{k_n}\}, \mathrm{c_0}, \beta - \frac{1}{2}, \mathrm{L_o})$ for some $c_0 > 0$ and has support $\Theta_k(\beta - \frac{1}{2}, L)$. In addition, for all $\theta, \theta' \in \Theta_k(\beta - \frac{1}{2}, L)$ such that $\|\theta - \theta'\| \leq L(n/\log n)^{-\frac{2\beta-1}{4\beta}}$,

$$\log \pi_{\theta|k}(\theta) - \log \pi_{\theta|k}(\theta') = h_k^t(\theta - \theta') + o(1), \tag{2.2}$$

  for vectors h_k satisfying $\|h_k\| \leq C(n/k)^{1-\rho_0}$, with constants $C, \rho_0 > 0$. Finally, it is assumed that L is sufficiently large compared to L_o.
- **Prior B:** k is deterministic and increasing at rate

$$k_n^{'} = \lfloor k_B (n/\log n)^{\frac{1}{1+2\beta}} \rfloor,$$

  where k_B is such that $k_n^{'} < k_n$ for all n. The prior for $\theta|k$ has density $\pi_{\theta|k}$ with respect to Lebesgue measure which satisfies condition $\mathrm{Hyp}(\{\mathrm{k_n^{'}}\}, \mathrm{c_0}, \beta, \mathrm{L_o})$ for some $c_0 > 0$ and is assumed to have support $\Theta_k(\beta, L)$. The density also satisfies

$$\log \pi_{\theta|k}(\theta) - \log \pi_{\theta|k}(\theta') = o(1),$$

  for all $\theta, \theta' \in \Theta_k(\beta, L)$ such that $\|\theta - \theta'\| \leq L(n/\log n)^{-\frac{\beta}{2\beta+1}}$. This condition is similar to (2.2), but with $h_k = 0$, and support $\Theta_k(\beta, L)$.
- **Prior C:** $k \sim \pi_k$ on $\mathbb{N}$ with $e^{-c_1 k \log k} \leq \pi_k(k) \leq e^{-c_2 k \log k}$ for k large enough, where $0 < c_1 < c_2 < +\infty$. There exists $\beta_s > 1$ such that for all $\beta \geq \beta_s$, the prior for $\theta|k$ has density $\pi_{\theta|k}$ with respect to Lebesgue measure which satisfies, for some $c_0 > 0$, condition $\mathrm{Hyp}(\{\mathrm{k} \leq \mathrm{k_0(n/\log n)^{1/(2\beta+1)}}\}, \mathrm{c_0}, \beta, \mathrm{L_o})$, for all $k_0 > 0$ as soon as n is large enough. It has support included in $\Theta_k(\beta, L)$ and satisfies

$$\log \pi_{\theta|k}(\theta) - \log \pi_{\theta|k}(\theta') = o(1),$$

  for all $\theta, \theta' \in \Theta_k(\beta, L)$ such that $\|\theta - \theta'\| \leq L(n/\log n)^{-\frac{\beta}{2\beta+1}}$.

Note that **prior A** is obtained when we take $\beta' = \beta - \frac{1}{2}$ in prior B.

2.2. Examples of priors

The Lipschitz conditions on $\log \pi_{\theta|k}$ considered for the three types of priors are satisfied for instance for the uniform prior on $\Theta_k(\beta - \frac{1}{2}, L)$ (resp. $\Theta_k(\beta, L)$), and for the truncated multivariate Gaussian prior, where, for some constants A and $\alpha > 0$,

$$\pi_{\theta|k}(\theta) \propto \mathbb{I}_{\Theta_k(\beta - \frac{1}{2}, L)}(\theta) \exp\left(-A \sum_{j=0}^{k} j^\alpha \theta_j^2\right),$$

with $\mathbb{I}_\Gamma$ denoting the indicator function of the set Γ. In the case of **Prior A**, the conditions on $\log \pi_{\theta|k}$ and h_k in (2.2) are satisfied for $\alpha < 4\beta - 2$. To see this, note that for all $\theta, \theta^{'} \in \Theta_h(\beta - 1/2, L)$, Hölder's inequality leads to

$$\sum_{j=0}^{k} j^\alpha |\theta_j^2 - (\theta_j^{'})^2| \leq L^{1/2} \|\theta - \theta'\| k^{\alpha-\beta+1/2} = o((n/k)^{1-\delta}).$$

In the case of **Prior B** and and **Prior C** we may choose $\alpha < 2\beta$, since for some positive k_0

$$\sum_{j=0}^{k} j^\alpha |\theta_j^2 - (\theta_j^{'})^2| \leq L^{1/2} \|\theta - \theta^{'}\| k^{\alpha-\beta} = o(1),$$

for all $k \leq k_0(n/\log n)^{1/(2\beta+1)}$ and all $\theta, \theta^{'} \in \Theta_k(\beta, L)$ such that $\|\theta - \theta^{'}\| \leq (n/\log n)^{-\beta/(2\beta+1)}$.

Also a truncated Laplace distribution is possible, in which case

$$\pi_{\theta|k}(\theta) \propto \mathbb{I}_{\Theta_k(\beta-\frac{1}{2},L)}(\theta) \exp\left(-a\sum_{j=0}^{k}|\theta_j|\right).$$

The condition on π_k in **Prior C** is satisfied for instance by Poisson distributions.

The restriction of the prior to Sobolev balls is required to obtain a proper concentration rate or even consistency of the posterior of the spectral density f itself, which is a necessary step in the proof of our results. This is discussed in more detail in section 3.1.

2.3. Convergence rates and BVM-results under different priors

Assuming a Poisson prior for k, RCL (Theorem 4.2) obtain a near-optimal convergence rate for $l(f, f_o)$. In Corollary 3.1 below, we show that the optimal rate for l implies that we have at least a suboptimal rate for $|d - d_o|$. Whether this can be improved to the optimal rate critically depends on the prior on k. By our first main result the answer is positive under **prior A**. The proof is given in section 4.

Theorem 2.1. *Under **prior A**, the posterior distribution has the asymptotic expansion*

$$\Pi\left[\sqrt{\frac{nr_{k_n}}{2}}(d - d_o - b_n(d_o)) \leq z|X\right] = \Phi(z) + o_{P_o}(1), \tag{2.3}$$

where, for $r_{k_n} = \sum_{j \geq k_n+1} \eta_j^2$ and some small enough $\delta > 0$,

$$b_n(d_o) = \frac{1}{r_{k_n}} \sum_{j=k_n+1}^{\infty} \eta_j \theta_{o,j} + Y_n + o(n^{-1/2-\delta} k_n^{1/2}), \quad Y_n = \frac{\sqrt{2}}{\sqrt{nr_{k_n}}} Z_n,$$

Z_n being a sequence of random variables converging weakly to a Gaussian variable with mean zero and variance 1.

Corollary 2.1. *Under **prior A**, the convergence rate for d is $\delta_n = (n/\log n)^{-\frac{2\beta-1}{4\beta}}$, i.e.*

$$\lim_{n\to\infty} E_0^n \left[\Pi(d : |d - d_o| > \delta_n|X)\right] = 0.$$

Equation (2.3) is a Bernstein-von Mises type of result: the posterior distribution is asymptotically normal, centered at a point $d_o + b_n(d_o)$, whose distribution is normal with mean d_o and variance $2/(nr_{k_n})$. The expressions for the posterior mean and variance give more insight in how the prior for k affects the posterior rate for d. The standard deviation of the limiting normal distribution (2.3) is $\sqrt{2/(nr_{k_n})} = O(n^{-\frac{2\beta-1}{4\beta}}(\log n)^{\frac{1}{4\beta}}) = o(\delta_n)$ and $b_n(d_o)$ equals

$$\frac{1}{r_{k_n}} \sum_{j=k_n+1}^{\infty} \eta_j \theta_{o,j} + O_{P_o}(r_{k_n}^{-\frac{1}{2}} n^{-\frac{1}{2}}) + o(n^{-\frac{1}{2}-\delta_1} k_n^{\frac{1}{2}}).$$

From the Cauchy-Schwarz inequality, the definition of η_j, k_n and r_{k_n} and the assumption on θ_o, it follows that

$$\frac{1}{r_{k_n}} \left| \sum_{j=k_n+1}^{\infty} \eta_j \theta_{o,j} \right| \le \frac{2}{r_{k_n}} \sqrt{\sum_{j>k_n} \theta_{o,j}^2 j^{2\beta}} \sqrt{\sum_{j>k_n} j^{-2\beta-2}} = o(k_n^{-\beta+\frac{1}{2}}). \tag{2.4}$$

Hence,

$$|b_n(d_o)| \le \delta_n, \tag{2.5}$$

and we obtain the δ_n-rate of Corollary 2.1. For smaller k, the standard deviation is smaller but the bias $b_n(d_o)$ is larger. In Theorem 2.2 below it is shown that this indeed leads to a suboptimal rate. Note that by allowing L to grow at a rate $\log\log n$, which is possible without modifying the result, the Bayesian procedure does not require a prior knowledge on L_o. Moreover, $\theta_{o,0}$ plays a special role since it is the variance of the residuals, we believe that we could assume a Gamma prior on $\theta_{o,0}$ and still obtain the same type of result. We have not pursued the computations here however.

An important consequence of the BVM-result is that posterior credible regions for d (HPD or equal-tails for instance) will also be asymptotic frequentist confidence regions. Consider for instance one-sided credible intervals for d defined by $P^\pi(d \le z_n(\alpha)|X) = \alpha$, so that $z_n(\alpha)$ is the α-th quantile of the posterior distribution of d. Equation (2.3) in Theorem 2.1 then implies that

$$z_n(\alpha) = d_o + b_n(d_o) + \sqrt{\frac{2k_n}{n}}\Phi^{-1}(\alpha)(1 + \mathbf{o_{P_o}}(1)).$$

As soon as $\sum_{j\ge k_n} j^{2\beta}\theta_{o,j}^2 = o((\log n)^{-1})$, we have that

$$z_n(\alpha) = d_o + \sqrt{2/(nr_{k_n})}Z_n + \sqrt{2/(nr_{k_n})}\Phi^{-1}(\alpha)(1 + \mathbf{o_{P_o}}(1))$$

and

$$P_o^n\left(d_o \le z_n(\alpha)\right) = P\left(Z_n \le \Phi^{-1}(\alpha)(1 + o(1))\right) = \alpha + o(1).$$

Similar computations can be made on equal-tail credible intervals or HPD regions for d.

Note that in this paper we assume that the smoothness β of f_o is greater than 1 instead of $1/2$, as is required in Moulines and Soulier (2003). This condition is used throughout the proof. Actually had we only assumed that $\beta > 3/2$, the proof of Theorem 2.1 would have been greatly simplified as many technicalities in the paper come from controlling terms when $1 < \beta \leq 3/2$. We do not believe that it is possible to weaken this constraint to $\beta > 1/2$ in our setup.

Our second main result states that if k is increasing at a slower rate than k_n, the posterior on d concentrates at a suboptimal rate. The proof is given in section 5.

Theorem 2.2. *Given $\beta > 5/2$, there exists $\theta_o \in \Theta(\beta, L_o)$ and a constant $k_v > 0$ such that under prior B and C defined above,*

$$\Pi(|d - d_o| > k_v w_n (\log n)^{-1} | X) \xrightarrow{P_o} 1.$$

with $w_n = C_w (n/\log n)^{-\frac{2\beta-1}{4\beta+2}}$ and $C_w = C_1 (L + L_o)^{\frac{1}{4\beta}} l_0^{\frac{2\beta-1}{2\beta}}$.

The constant C_w comes from the suboptimal rate for $|d - d_o|$ derived in Corollary 3.1. Theorem 2.2 is proved by considering the vector θ_o defined by $\theta_{o,j} = c_0 j^{-(\beta+\frac{1}{2})}(\log j)^{-1}$, for $j \geq 2$. This vector is close to the boundary of the Sobolev-ball $\Theta(\beta, L_o)$, in the sense that for all $\beta' > \beta$, $\sum_j j^{2\beta'} \theta_{o,j}^2 = +\infty$. The proof consists in showing that conditionally on k, the posterior distribution is asymptotically normal as in (2.3), with k replacing k_n, and that the posterior distribution concentrates on values of k smaller than $O(n^{1/(2\beta+1)})$, so that the bias $b_n(d_o)$ becomes of order $w_n(\log n)^{-1}$. The constraint $\beta > 5/2$ is used to simplify the computations and is not sharp.

It is interesting to note that similar to the frequentist approach, a key issue is a bias-variance trade-off, which is optimized when $k \sim n^{1/(2\beta)}$. This choice of k depends on the smoothness parameter β, and since it is not of the same order as the *optimal* values of k for the loss $l(f, f')$ on the spectral densities, the adaptive (near) minimax Bayesian nonparametric procedure proposed in Rousseau and Kruijer (2011) does not lead to optimal posterior concentration rate for d. While it is quite natural to obtain an adaptive (nearly) minimax Bayesian procedure under the loss $l(.,.)$ by choosing a random k, obtaining an adaptive minimax procedure for d remains an open problem, at least in a Bayesian context (by contrast, the frequentist estimators proposed in Iouditsky et al. (2001) are adaptive). This dichotomy is found in other semi-parametric Bayesian problems, see for instance Arbel et al. (2013) in the case of the white noise model or Rivoirard and Rousseau (2012) for BVM properties.

3. Decomposing the posterior for d

To prove Theorems 2.1 and 2.2 we need to take a closer look at (1.3), to understand how the integration over Θ_k affects the posterior for d. We develop

$\theta \to l_n(d,k,\theta)$ at a point $\bar{\theta}_{d,k}$ defined below and decompose the likelihood as

$$\exp\{l_n(d,k,\theta)\} = \exp\{l_n(d,k)\}\exp\{l_n(d,k,\theta) - l_n(d,k)\},$$

where $l_n(d,k)$ is short-hand notation for $l_n(d,k,\bar{\theta}_{d,k})$. Define

$$I_n(d,k) = \int_{\Theta_k} e^{l_n(d,k,\theta)-l_n(d,k)} d\pi_{\theta|k}(\theta), \tag{3.1}$$

where Θ_k is the generic notation for $\Theta_k(\beta - \frac{1}{2}, L)$ under **prior A** and $\Theta_k(\beta, L)$ for priors B and C. The posterior for d given in (1.3) can be written as

$$\Pi(d \in D|X) = \frac{\sum_{k=0}^{\infty} \pi_k(k) e^{l_n(d_o,k)} \int_D e^{l_n(d,k)-l_n(d_o,k)} I_n(d,k) d\pi_d(d)}{\sum_{k=0}^{\infty} \pi_k(k) e^{l_n(d_o,k)} \int_{-\frac{1}{2}+t}^{\frac{1}{2}-t} e^{l_n(d,k)-l_n(d_o,k)} I_n(d,k) d\pi_d(d)}. \tag{3.2}$$

The factor $\exp\{l_n(d,k) - l_n(d_o,k)\}$ is independent of θ, and will under certain conditions dominate the marginal likelihood. In section 3.2 we give a Taylor-approximation which, for given k, allows for a normal approximation to the marginal posterior. However, to obtain the convergence rates in Theorems 2.1 and 2.2, it also needs to be shown that the integrals $I_n(d,k)$ with respect to θ do not vary too much with d. This is the most difficult part of the proof of Theorem 2.1 and the argument is presented in section 3.3. Since Theorem 2.2 is essentially a counter-example and it is not aimed to be as general as Theorem 2.1, as far as the range of β is concerned, we can restrict attention to larger β's, i.e. $\beta > 5/2$, for which controlling $I_n(d,k)$ is much easier.

3.1. Preliminaries

First we define the point $\bar{\theta}_{d,k}$ at which we develop $\theta \to l_n(d,k,\theta)$. Since the function $\log(2 - 2\cos(x))$ has Fourier coefficients against $\cos jx$, $j \in \mathbb{N}$ equal to $0, 2, \frac{2}{2}, \frac{2}{3}, \ldots$, FEXP-spectral densities can be written as

$$|1 - e^{ix}|^{-2d} \exp\left\{\sum_{j=0}^{\infty} \theta_j \cos(jx)\right\} = \exp\left\{\sum_{j=0}^{\infty} (\theta_j + d\eta_j)\cos(jx)\right\}.$$

Given $f = f_{d,k,\theta}$ and $f' = f_{d',k',\theta'}$ we can therefore express the norm $l(f,f')$ in terms of $(\theta - \theta')$ and $(d - d')$:

$$l(f,f') = \frac{1}{2}\sum_{j=0}^{\infty}((\theta_j - \theta'_j) + \eta_j(d - d'))^2, \tag{3.3}$$

where θ_j and θ'_j are understood to be zero when j is larger than k respectively k'. Equation (3.3) implies that for given d and k, $l(f_o, f_{d,k,\theta})$ is minimized by

$$\bar{\theta}_{d,k} := \text{argmin}_{\theta \in \mathbb{R}^{k+1}} \sum_{j=0}^{\infty} (\theta_j - \theta_{o,j} + (d - d_o)\eta_j)^2 = \theta_{o[k]} + (d_o - d)\eta_{[k]}.$$

In particular, $\theta = \theta_{o[k]}$ minimizes $l(f_o, f_{d,k,\theta})$ only when $d = d_o$; when $d \neq d_o$ we need to add $(d_o - d)\eta_{[k]}$. The following lemma shows that an upper bound on $l(f_o, f_{d,k,\theta})$ leads to upper bounds on $|d - d_o|$ and $\|\theta - \theta_o\|$.

Lemma 3.1. *Let $\gamma \leq \beta$ and α_n a positive sequence tending to 0. There exist constants $C_1, C_2 > 0$ such that for all n, all $\theta \in \Theta_k(\gamma, L)$ and all $\theta_o \in \Theta_k(\beta, L_o)$, $l(f_o, f_{d,k,\theta}) \leq \alpha_n^2$ implies that*

$$|d - d_o| \leq C_1 (L + L_o)^{\frac{1}{4\gamma}} \alpha_n^{\frac{2\gamma-1}{2\gamma}}, \qquad \|\theta - \theta_o\| \leq C_2 (L + L_o)^{\frac{1}{4\gamma}} \alpha_n^{\frac{2\gamma-1}{2\gamma}}.$$

Proof. For all (d, k, θ) such that $l(f_{d,k,\theta}, f_o) \leq \alpha_n$, we have, using (3.3),

$$\begin{aligned} 2\alpha_n^2 \geq 2l(f_{d,k,\theta}, f_o) &= 2(\theta_{o,0} - \theta_0)^2 + \sum_{j \geq 1} \left((\theta_{o,j} - \theta_j) + \eta_j (d_o - d)\right)^2 \\ &\geq \left(\|\theta - \theta_o\| - |d - d_o| \|\eta\|\right)^2 . \end{aligned}$$

The inequalities remain true if we replace all sums over $j \geq 1$ by sums over $j \geq m_n$, for any nondecreasing sequence m_n. Since $\|(\eta_j 1_{j > m_n})_{j \geq 1}\|^2$ is of order m_n^{-1} and $\|((\theta - \theta_o)_j 1_{j > m_n})_{j \geq 1}\|^2 \leq m_n^{-2\gamma} \sum_{j > m_n} (1 + j)^{2\gamma} (\theta_j - \theta_{o,j})^2 < 2(L + L_o) m_n^{-2\gamma}$, setting $m_n = (\alpha_n / \sqrt{L + L_o})^{-\frac{1}{\gamma}}$ gives the desired rate for $|d - d_o|$ as well as for $\|\theta - \theta_o\|$. ☐

The convergence rate for $l(f_o, f_{d,k,\theta})$ required in Lemma 3.1 can be found in Rousseau and Kruijer (2011). For easy reference we restate it here. Compared to a similar result in RCL, the power of the $(\log n)$ factor is improved.

Lemma 3.2. *Under **prior A**, there exists a constant l_0 depending only on L_o and k_A (and not on L) such that*

$$\Pi((d, k, \theta) : l(f_{d,k,\theta}, f_o) \geq l_0^2 \delta_n^2 | X) \xrightarrow{P_o} 0,$$

*where $\delta_n = (n / \log n)^{-\frac{2\beta-1}{4\beta}}$. Under **priors B** and **C**, this statement holds with $\epsilon_n = (n / \log n)^{-\frac{\beta}{2\beta+1}}$ replacing δ_n.*

In the proof of Theorem 2.1 (resp. 2.2), this result allows us to restrict attention to the set of spectral densities f such that $l(f, f_o) \leq l_0^2 \delta_n^2$ (resp. $l_0^2 \epsilon_n^2$). In addition, by combination with Lemma 3.1 we can now deduce bounds on $|d - d_o|$ and $\|\theta - \bar{\theta}_{d,k}\|$. These bounds, although suboptimal, will be important in the sequel for obtaining the near-optimal rate in Theorem 2.1.

Corollary 3.1. *Under the result of Lemma 3.2 and **prior A**, we can apply Lemma 3.1 with $\alpha_n^2 = l_0^2 \delta_n^2$ and $\gamma = \beta - \frac{1}{2}$, and obtain*

$$\Pi_d(d : |d - d_o| \geq \bar{v}_n | X) \xrightarrow{P_o} 0, \quad \Pi(\|\theta - \bar{\theta}_{d,k}\| \geq 2 l_0 \delta_n | X) \xrightarrow{P_o} 0,$$

*where $\bar{v}_n = C_1 (L + L_o)^{\frac{1}{4\beta-2}} l_0^{\frac{2\beta-2}{2\beta-1}} (n / \log n)^{-\frac{\beta-1}{2\beta}}$. Under **priors B** and **C** we have $\gamma = \beta$; the rate for $|d - d_o|$ is then $w_n = C_w (n / \log n)^{-\frac{2\beta-1}{4\beta+2}}$ and the rate for $\|\theta - \bar{\theta}_{d,k}\|$ is $2 l_0 \epsilon_n$. The constant $C_w = C_1 (L + L_o)^{\frac{1}{4\beta}} l_0^{\frac{2\beta-1}{2\beta}}$ is as in Theorem 2.2.*

Proof. The rate for $|d-d_o|$ follows directly from Lemma 3.1. To obtain the rate for $\|\theta - \bar{\theta}_{d,k}\|$, let α_n denote either $l_0\delta_n$ (the rate for $l(f_o, f)$ under **prior A**) or $l_0\epsilon_n$ (the rate under **priors B** and **C**). Although Lemma 3.1 suggests that the Euclidean distance from θ_o to θ (contained in $\Theta_k(\beta, L)$ or $\Theta_k(\beta-\frac{1}{2}, L)$) may be larger than α_n, the distance from θ to $\bar{\theta}_{d,k}$ is certainly of order α_n. To see this, note that Lemma 3.2 implies the existence of d, k, θ in the model with $l(f_o, f_{d,k,\theta}) \leq \alpha_n^2$. From the definition of $\bar{\theta}_{d,k}$ it follows that $l(f_o, f_{d,k,\bar{\theta}_{d,k}}) \leq \alpha_n^2$. The triangle inequality gives $\|\theta - \bar{\theta}_{d,k}\|^2 = l(f_{d,k,\theta}, f_{d,k,\bar{\theta}_{d,k}}) \leq 4\alpha_n^2$. □

The rates $\bar{v}_n$ and w_n obtained in Corollary 3.1 are clearly suboptimal; their importance however lies in the fact that they narrow down the set for which we need to prove Theorems 2.1 and 2.2. To prove Theorem 2.2 for example it suffices to show that the posterior mass on $k_v w_n(\log n)^{-1} < |d-d_o| < w_n$ tends to zero. Note that the lower and the upper bound differ only by a factor $(\log n)$. Hence under priors B and C, the combination of Corollary 3.1 and Theorem 2.2 characterizes the posterior concentration rate (up to a $\log n$ term) for the given θ_o. Another consequence of Corollary 3.1 is that we may neglect the posterior mass on all (d, k, θ) for which $\|\theta - \bar{\theta}_{d,k}\|$ is larger than $2l_0\delta_n$ (under prior A) or $2l_0\epsilon_n$ (under priors B and C).

We conclude this section with a result on $\bar{\theta}_{d,k}$ and $\Theta_k(\beta, L)$. In the definition of $\bar{\theta}_{d,k}$ we minimize over $\mathbb{R}^{k+1}$, whereas the support of priors A–C is the Sobolev ball $\Theta_k(\beta, L)$ or $\Theta_k(\beta-\frac{1}{2}, L)$. Under the assumptions of Theorems 2.1 and 2.2 however, $\bar{\theta}_{d,k}$ is contained in $\Theta_k(\beta-\frac{1}{2}, L)$ and $\Theta_k(\beta, L)$, respectively. Also the l_2-ball of radius $2l_0\delta_n$ (or $2l_0\epsilon_n$) is contained in these Sobolev-balls.

Lemma 3.3. *Under the assumptions of Theorem 2.1, $B_k(\bar{\theta}_{d,k}, 2l_0\delta_n)$ is contained in $\Theta_k(\beta-\frac{1}{2}, L)$ for all $d \in [d_o - \bar{v}_n, d_o + \bar{v}_n]$, if L is large enough. In particular, $\bar{\theta}_{d,k} \in \Theta_k(\beta-\frac{1}{2}, L)$. Similarly, under the assumptions of Theorem 2.2, $B_k(\bar{\theta}_{d,k}, 2l_0\epsilon_n) \subset \Theta_k(\beta, L)$, for all $d \in [d_o - w_n, d_o + w_n]$.*

Proof. Since the constant l_0 is independent of L, $\theta \in B_k(\bar{\theta}_{d,k}, 2l_0\delta_n)$ implies that for n large enough,

$$\begin{aligned}\sum_{j=0}^{k} \theta_j^2 (j+1)^{2\beta-1} &\leq 2\sum_{j=0}^{k} (\theta - \bar{\theta}_{d,k})_j^2 (j+1)^{2\beta-1} + 2\sum_{j=0}^{k} (\bar{\theta}_{d,k})_j^2 (j+1)^{2\beta-1} \\ &\leq 8 l_0^2 (n/\log n)^{-\frac{2\beta-1}{2\beta}} (k_n+1)^{2\beta-1} + 4\sum_{j=0}^{k_n} \theta_{o,j}^2 (j+1)^{2\beta-1} \\ &\quad + 16(d-d_o)^2 \sum_{j=1}^{k_n} (j+1)^{2\beta-3}.\end{aligned}$$

The first two terms on the right only depend on L_o, and are smaller than $L/4$ when L is chosen sufficiently large. Because $\bar{v}_n = C_1(L+L_o)^{\frac{1}{4\beta-2}} l_0^{\frac{2\beta-2}{2\beta-1}} (n/\log n)^{-\frac{\beta-1}{2\beta}}$, the last term in the preceding display is at most

$$C_1^2 (L+L_o)^{\frac{1}{2\beta-1}} l_0^{\frac{4\beta-4}{2\beta-1}} (n/\log n)^{-\frac{\beta-1}{\beta}} k_A^{2\beta-2} (n/\log n)^{\frac{\beta-1}{\beta}},$$

which, since $\beta > 1$, is smaller than $L/2$ when L is large enough. We conclude that $B_k(\bar\theta_{d,k}, 2l_0\delta_n)$ is contained in $\Theta_k(\beta - \frac{1}{2}, L)$ provided L is chosen sufficiently large. The second statement can be proved similarly. □

3.2. A Taylor approximation for $l_n(d,k)$

Provided that the integrals $I_n(d,k)$ have negligible impact on the posterior for d, the conditional distribution of d given k will only depend on $\exp\{l_n(d,k) - l_n(d_o,k)\}$. Let $l_n^{(1)}(d,k)$, $l_n^{(2)}(d,k)$ denote the first two derivatives of the map $d \mapsto l_n(d,k)$. There exists a $\bar d$ between d and d_o such that

$$l_n(d,k) \quad = \quad l_n(d_o,k) + (d-d_o) l_n^{(1)}(d_o,k) + \frac{(d-d_o)^2}{2} l_n^{(2)}(\bar d,k). \qquad (3.4)$$

Defining

$$b_n(d) = -\frac{l_n^{(1)}(d_o,k)}{l_n^{(2)}(d,k)}, \qquad (3.5)$$

which is the b_n used in Theorem 2.1, we can rewrite (3.4) as

$$l_n(d,k) - l_n(d_o,k) = -\frac{1}{2}\frac{(l_n^{(1)}(d_o,k))^2}{l_n^{(2)}(\bar d,k)} + \frac{1}{2} l_n^{(2)}(\bar d,k)\left(d - d_o - b_n(\bar d)\right)^2. \qquad (3.6)$$

Note that each derivative $l_n^{(i)}(d,k)$, $i = 1,2$, can be decomposed into a centered quadratic form denoted $\mathcal{S}(l_n^{(i)}(d,k))$ and a deterministic term $\mathcal{D}(l_n^{(i)}(d,k))$. In the following lemma we give expressions for $l_n^{(1)}(d_o,k)$, $l_n^{(2)}(d,k)$ and b_n, making explicit their dependence on k and θ_o. Since $k_n^{'} \leq k_n$ and $w_n < \bar v_n$ (see Corollary 3.1) the result is valid for all priors under consideration. The proof is given in Section 4 of the supplementary material (Kruijer and Rousseau (2013)).

Lemma 3.4. *Given $\beta > 1$, let $\theta_o \in \Theta(\beta, L_o)$, there exists $\delta_1 > 0$ such that if $k \leq k_n$ and $|d - d_o| \leq \bar v_n$,*

$$\begin{aligned} l_n^{(1)}(d_o,k) &:= \mathcal{S}(l_n^{(1)}(d_o,k)) + \mathcal{D}(l_n^{(1)}(d_o,k)) \\ &= \mathcal{S}(l_n^{(1)}(d_o,k)) + \frac{n}{2}\sum_{j=k+1}^{\infty} \eta_j \theta_{o,j} + \mathbf{o}(n^{\epsilon}(k^{-\beta+3/2} + n^{-1/(2\beta)})), \\ l_n^{(2)}(d,k) &= l_n^{(2)}(d_o,k)\left(1 + \mathbf{o_{P_o}}\left(k^{1/2} n^{-1/2-\epsilon} + |d-d_o| n^{\epsilon}\right)\right) \\ &= -\frac{1}{2} n r_k \left(1 + \mathbf{o_{P_o}}(n^{-\delta_1})\right), \end{aligned}$$

where $\mathcal{S}(l_n^{(1)}(d_o,k))$ is a centered quadratic form with variance

$$Var(\mathcal{S}(l_n^{(1)}(d_o,k))) = \frac{n}{2}\left(\sum_{j>k}\eta_j^2\right)(1+o(1)) = \frac{n r_k}{2}(1+o(1)) = O(nk^{-1}).$$

Consequently,

$$b_n(d) = -\frac{l_n^{(1)}(d_o,k)}{l_n^{(2)}(d,k)} = \frac{1}{r_k}\left(\sum_{j=k+1}^{\infty}\eta_j\theta_{o,j}\right)(1+\mathbf{o}_{\mathbf{P_o}}(n^{-\delta})) + \frac{2\mathcal{S}(l_n^{(1)}(d_o,k))(1+\mathbf{o}_{\mathbf{P_o}}(n^{-\delta}))}{nr_k} + \mathbf{o}_{\mathbf{P_o}}(n^{\epsilon-1}k^{-\beta+5/2}+n^{\epsilon-1}), \tag{3.7}$$

with

$$\frac{2\mathcal{S}(l_n^{(1)}(d_o,k))}{nr_k} = \mathbf{O}_{\mathbf{P_o}}(n^{-\frac{1}{2}}k^{\frac{1}{2}}).$$

Remark 3.1. Recall from (2.4) that $r_k^{-1}\sum_{j=k+1}^{\infty}\eta_j\theta_{o,j}$ is $O(k^{-\beta+1/2})$. The term $2\mathcal{S}(l_n^{(1)}(d_o,k))/(nr_k)$ is $O_{P_o}(k^{-\beta+1/2})$ whenever $k\sim n^{1/(2\beta)}$, which is the case under all priors under consideration.

Substituting the above results on $l_n^{(1)}$, $l_n^{(2)}$ and b_n in (3.6), we can give the following informal argument leading to Theorems 2.1 and Theorem 2.2. If we consider k to be fixed and $I_n(d,k)$ constant in d, then (3.6) implies that the posterior distribution for d is asymptotically normal with mean $d_o+b_n(d_o)$ and variance of order k/n.

3.3. Integration of the short memory parameter

A key ingredient in the proofs of both Theorems 2.1 and 2.2 is the control of the integral $I_n(d,k)$ appearing in (1.3), which is proved to be almost constant in d compared to $\exp\{l_n(d,k)-l_n(d_o,k)\}$. In Lemma 3.5 below we prove this to be the case under the assumptions of Theorems 2.1 and 2.2. For the case of Theorem 2.2 this is fairly simple: the conditional posterior distribution of θ given (d,k) can be proved to be asymptotically Gaussian by a Laplace-approximation. For smaller β and larger k the control is technically more demanding. In both cases the proof is based on the following Taylor expansion of $l_n(d,k,\theta)$ around $\bar\theta_{d,k}$:

$$l_n(d,k,\theta)-l_n(d,k) = \sum_{j=1}^{J}\frac{(\theta-\bar\theta_{d,k})^{(j)}\nabla^j l_n(d,k)}{j!} + R_{J+1,d}(\theta), \tag{3.8}$$

where

$$(\theta-\bar\theta_{d,k})^{(j)}\nabla^j l_n(d,k) = \sum_{l_1,\dots,l_j=0}^{k}(\theta-\bar\theta_{d,k})_{l_1}\dots(\theta-\bar\theta_{d,k})_{l_j}\frac{\partial^j l_n(d,k,\bar\theta_{d,k})}{\partial\theta_{l_1}\dots\partial\theta_{l_j}},$$

$$R_{J+1,d}(\theta) = \frac{1}{(J+1)!}\sum_{l_1,\dots,l_{J+1}=0}^{k}(\theta-\bar\theta_{d,k})_{l_1}\dots(\theta-\bar\theta_{d,k})_{l_{J+1}}\frac{\partial^{J+1} l_n(d,k,\tilde\theta)}{\partial\theta_{l_1}\dots\partial\theta_{l_{J+1}}}. \tag{3.9}$$

The above expressions are used to derive the following lemma, which gives control of the term $I_n(d,k)$.

Lemma 3.5. *Under the conditions of Theorem 2.1, the integral $I_n(d,k)$ defined in (3.1) equals*

$$I_n(d_o,k)\exp\left\{\mathbf{o_{P_o}}(1)+\mathbf{o_{P_o}}\left(|d-d_o|n^{\frac{1}{2}-\delta_2}k^{-\frac{1}{2}}\right)+\mathbf{o_{P_o}}\left((d-d_o)^2n^{1-\delta_2}k^{-1}\right)\right\},$$

for some $\delta_2>0$. Under the conditions of Theorem 2.2,

$$I_n(d,k)=I_n(d_o,k)\exp\{\mathbf{o_{P_o}}(1)\}.$$

The proof is given in Sections 5 and 6 of the supplementary material (Kruijer and Rousseau (2013)), and relies on expressions for the derivatives $\nabla^j l_n$. Lemma 3.5 should be seen in relation to Lemma 3.4 and the expressions for $\Pi(d|X)$ and $l_n(d,k)-l_n(d_o,k)$ in equations (3.2) and (3.4). Lemma 3.5 then shows that the dependence on the integrals $I_n(d,k)$ on d is asymptotically negligible with respect to $l_n(d,k)-l_n(d_o,k)$. This is made rigorous in the following section.

4. Proof of Theorem 2.1

By Lemma 3.2 we may assume posterior convergence of $l(f_o,f_{d,k,\theta})$ at rate $l_0^2\delta_n^2$, and, by Corollary 3.1, also convergence of $|d-d_o|$ at rate $\bar{v}_n$. By Lemma 3.3, we may restrict the integration over θ to $B_k(\bar{\theta}_{d,k},2l_0\delta_n)$ and in the definition of $I_n(d,k)$ in (3.1) the integration in θ can be restricted to $B_k(\bar{\theta}_{d,k},2l_0\delta_n)$. To simplify notations we still denote $I_n(d,k)$ the integral over the restricted space. Let $\Gamma_n(z)=\{d:\sqrt{\frac{nr_k}{2}}(d-d_o-b_n(d_o))\le z\}$. Under **prior A**, it suffices to show that for $k=k_n$,

$$\begin{aligned}\frac{N_n}{D_n}&:=\frac{\int_{\Gamma_n(z)}e^{l_n(d,k)-l_n(d_o,k)}\int_{B_k(\bar{\theta}_{d,k},2l_0\delta_n)}e^{l_n(d,k,\theta)-l_n(d,k)}d\pi_{\theta|k}(\theta)d\pi_d(d)}{\int_{|d-d_o|<\bar{v}_n}e^{l_n(d,k)-l_n(d_o,k)}\int_{B_k(\bar{\theta}_{d,k},2l_0\delta_n)}e^{l_n(d,k,\theta)-l_n(d,k)}d\pi_{\theta|k}(\theta)d\pi_d(d)}\\&=\frac{\int_{\Gamma_n(z)}\exp\{l_n(d,k)-l_n(d_o,k)+\log I_n(d,k)\}d\pi_d(d)}{\int_{|d-d_o|<\bar{v}_n}\exp\{l_n(d,k)-l_n(d_o,k)+\log I_n(d,k)\}d\pi_d(d)}=\Phi(z)+\mathbf{o_{P_o}}(1).\end{aligned}\tag{4.1}$$

Using the results for $l_n(d,k)-l_n(d_o,k)$ and $I_n(d,k)$ given by Lemmas 3.4 and 3.5, we show that for $A_n\subset\mathbb{R}^n$ defined below such that $P_o^n(A_n)\to 1$,

$$\frac{N_n}{D_n}\le\Phi(z)+o(1),\quad \frac{N_n}{D_n}\ge\Phi(z)+o(1),\quad \forall X\in A_n.\tag{4.2}$$

Since $P_o^n(A_n)\to 1$ this implies the last equality in (4.1).

Note that Lemmas 3.4 and 3.5 also hold for all $\delta_1'<\delta_1$ and $\delta_2'<\delta_2$. In the remainder of the proof, let $0<\delta\le\min(\delta_1,\delta_2)$. For notational simplicity, let $\mathcal{D}=\mathcal{D}(l_n^{(1)}(d_o,k))$, the deterministic part of $l_n^{(1)}(d_o,k)$. For a sufficiently large constant C_1 and arbitrary $\epsilon_1>0$, let A_n be the set of $X\in\mathbb{R}^n$ such that

$$\left.\begin{array}{c}|\log I_n(d,k)-\log I_n(d_o,k)|\le\epsilon_1+(d-d_o)^2k^{-1}n^{1-\delta}+|d-d_o|k^{-\frac{1}{2}}n^{\frac{1}{2}-\delta}\\ \left|l_n^{(1)}(d_o,k)-\mathcal{D}\right|\le C_1n^{\frac{1}{2}}k^{-\frac{1}{2}}\sqrt{\log n},\qquad \left|l_n^{(2)}(d,k)+\frac{1}{2}nr_k\right|\le n^{1-\delta}k^{-1}\end{array}\right\}$$

for all $|d - d_o| \leq \bar{v}_n$. Since $k = k_n$ and $\beta > 1$, Lemmas 3.4 and 3.5 imply that $P_o^n(A_n^c) \to 0$. We prove the first inequality in (4.2); the second one can be obtained in the same way. Using (3.4) and the definition of A_n, it follows that for all $X \in A_n$,

$$\begin{aligned} l_n(d,k) - l_n(d_o,k) + \log I_n(d,k) - \log I_n(d_o,k) &\leq \epsilon_1 + (d-d_o)^2 n^{1-\delta} k^{-1} \\ &\quad + |d-d_o| n^{\frac{1}{2}-\delta} k^{-\frac{1}{2}} + (d-d_o) l_n^{(1)}(d_o,k) - \frac{nr_k}{4}(d-d_o)^2(1-n^{-\delta}) \\ \leq 2\epsilon_1 - \frac{nr_k}{4}\left(1-2n^{-\delta}\right)\left(d - d_o - \frac{2 l_n^{(1)}(d_o,k)}{(1-2n^{-\delta})\, nr_k}\right)^2 & \\ + |d-d_o|\, n^{\frac{1}{2}-\delta} k^{-\frac{1}{2}} + \frac{(l_n^{(1)}(d_o,k))^2}{(1-2n^{-\delta})\, nr_k} & \\ \leq 3\epsilon_1 - \frac{nr_k}{4}\left(1-2n^{-\delta}\right)\left(d - d_o - \frac{b_n(d_o,k)}{1-2n^{-\delta}}\right)^2 & \\ + \left|d - d_o - \frac{b_n(d_o,k)}{1-2n^{-\delta}}\right| n^{\frac{1}{2}-\delta} k^{-\frac{1}{2}} + \frac{(l_n^{(1)}(d_o,k))^2}{(1-2n^{-\delta})\, nr_k}. & \end{aligned} \tag{4.3}$$

The third inequality follows from (2.5) and Remark 3.1, by which $b_n(d_o) = O(k^{-\beta+\frac{1}{2}}) = O(\delta_n)$. This implies that $|b_n(d_o)| k^{-\frac{1}{2}} n^{\frac{1}{2}-\delta} < \epsilon_1$, again for large enough n. Similar to the preceding display, we have the lower-bound

$$\begin{aligned} l_n(d,k) - l_n(d_o,k) &+ \log I_n(d,k) - \log I_n(d_o,k) \\ &\geq -3\epsilon_1 - \frac{nr_k}{4}(1+2n^{-\delta})\left(d - d_o - \frac{b_n(d_o,k)}{(1+2n^{-\delta})}\right)^2 \\ &\quad - \left|d - d_o - \frac{b_n(d_o,k)}{(1+2n^{-\delta})}\right| k^{-\frac{1}{2}} n^{\frac{1}{2}-\delta} + \frac{(l_n^{(1)}(d_o,k))^2}{(1+2n^{-\delta}) nr_k}. \end{aligned} \tag{4.4}$$

Note that

$$\exp\left\{\frac{(l_n^{(1)}(d_o,k))^2}{(1-2n^{-\delta}) nr_k} - \frac{(l_n^{(1)}(d_o,k))^2}{(1+2n^{-\delta}) nr_k}\right\} = \exp\{o(1)\}, \tag{4.5}$$

which follows from the expression for $l_n^{(1)}(d_o,k)$ in Lemma 3.4, the definition of A_n and the assumption that $X \in A_n$. Therefore, substituting (4.3) in N_n and (4.4) in D_n, the terms $\frac{(l_n^{(1)}(d_o,k))^2}{4nr_k}$ cancel out and by (4.5) we can neglect the difference between $\frac{(l_n^{(1)}(d_o,k))^2}{(1\pm 2n^{-\delta})nr_k}$ and $\frac{(l_n^{(1)}(d_o,k))^2}{nr_k}$.

To conclude the proof that $N_n/D_n \leq \Phi(z) + o(1)$ for each $X \in A_n$, we make the change of variables

$$u = \sqrt{\frac{nr_k}{2}(1 \pm 2n^{-\delta})}\left(d - d_o - \frac{b_n(d_o)}{1 \pm 2n^{-\delta}}\right),$$

where we take $+$ in the lower bound for D_n and $-$ in the upper-bound for N_n. Using once more that $b_n(d_o) = O(\delta_n)$, we find that for large enough n,

$|u| \leq \frac{\bar{v}_n}{4}\sqrt{nr_k}$ implies $|d - d_o| \leq \bar{v}_n$. Hence we may integrate over $|u| \leq \frac{\bar{v}_n}{4}\sqrt{nr_k}$ in the lower-bound for D_n. In the upper-bound for N_n we may integrate over $u \leq z + \epsilon_1$.

Combining (4.3)-(4.5), it follows that for all ϵ_1 and all $X \in A_n$,

$$\begin{aligned}\frac{N_n}{D_n} &\leq e^{7\epsilon_1}\left(\frac{1+2n^{-\delta}}{1-2n^{-\delta}}\right)^{\frac{1}{2}} \frac{\int_{u<z+\epsilon_1}\exp\{-\frac{1}{2}u^2 + Cn^{-\delta}|u|\}du}{\int_{|u|\leq \frac{\bar{v}_n}{4}\sqrt{nr_k}}\exp\{-\frac{1}{2}u^2 - Cn^{-\delta}|u|\}du} \\ &\leq e^{8\epsilon_1}\frac{\int_{u<z+\epsilon_1}\exp\{-\frac{1}{2}u^2 + Cn^{-\delta}|u|\}du}{\int_{|u|\leq \frac{\bar{v}_n}{4}\sqrt{nr_k}}\exp\{-\frac{1}{2}u^2 - Cn^{-\delta}|u|\}du} \to \Phi(z+\epsilon_1)e^{8\epsilon_1},\end{aligned}$$

since $\bar{v}_n\sqrt{nr_k}$ goes to infinity. Similarly we prove that for all ϵ_1, $N_n/D_n \geq \Phi(z-\epsilon_1)e^{-8\epsilon_1}$, when n is large enough, which terminates the proof of Theorem 2.1.

5. Proof of Theorem 2.2

Let $\beta > 5/2$ and $\theta_{o,j} = c_0 j^{-(\beta+\frac{1}{2})}(\log j)^{-1}$. When the constant c_0 is chosen small enough, $\theta_o \in \Theta(\beta, L_o)$. In view of Corollary 3.1, the posterior mass on the events $\{(d,k,\theta) : \|\theta - \bar{\theta}_{d,k}\| \geq 2l_0\epsilon_n\}$ and $\{(d,k,\theta) : |d - d_o| \geq w_n\}$ tends to zero in probability, and may be neglected. Moreover Lemma 3.1 applied to $\gamma = \beta$ and $\alpha_n = (n/\log n)^{-\beta/(2\beta+1)}$ implies that with posterior probability going to 1, $\|\theta - \theta_0\| \lesssim (n/\log n)^{-(\beta-1/2)/(2\beta+1)}$. We first consider the case of **Prior C**. Within the $(k+1)$-dimensional FEXP-model, $\|\theta - \theta_o\|$ is minimized by setting $\theta_j = \theta_{o,j}$ $(j = 0, \dots, k)$, and for this choice of θ we have

$$\|\theta - \theta_o\|^2 = \sum_{l>k}\theta_{o,l}^2 \gtrsim k^{-2\beta}(\log k)^{-2}.$$

Consequently, the fact that $\|\theta - \theta_0\| \lesssim (n/\log n)^{-(\beta-1/2)/(2\beta+1)}$ implies that $k > k_n'' := k_C(n/\log n)^{(\beta-1/2)/(\beta(2\beta+1))}(\log n)^{-1/\beta}$, for some constant $k_C > 0$. We conclude that

$$\Pi\left(k \leq k_n''|X\right) = \mathbf{o_{P_o}}(1),$$

and we can restrict our attention to $k > k_n''$. For a positive constant k_v, we decompose $\Pi_d(|d - d_o| \leq k_v w_n(\log n)^{-1}, k > k_n''|X)$ as

$$\begin{aligned}&\sum_{m>k_n''}\Pi(|d - d_o| \leq k_v w_n(\log n)^{-1}, k = m|X) \\ &\quad = \sum_{m>k_n''}\Pi(k = m|X)\Pi_m(|d - d_o| \leq k_v w_n(\log n)^{-1}|X),\end{aligned}$$

where $\Pi_m(|d - d_o| \leq k_v w_n(\log n)^{-1}|X)$ is the posterior for d within the FEXP-model of dimension $m+1$, i.e. $\Pi_m(|d - d_o| \leq k_v w_n(\log n)^{-1}|X) := \Pi(|d - d_o| \leq k_v w_n(\log n)^{-1}|k = m, X)$. From Theorem 4.2 of RCL (see Appendix C of RCL), since

$$\pi_k(k \geq B(n/\log n)^{1/(2\beta+1)}) \leq e^{-cBn\epsilon_n^2},$$

for n large enough and all $B > 0$ large enough, setting $k_n^{(3)} = \lfloor B(n/\log n)^{1/(2\beta+1)}\rfloor + 1$, we obtain that for B large enough

$$\Pi\left(k \geq k_n^{(3)}|X\right) = o_p(1).$$

To prove Theorem 2.2 it now suffices to show that

$$E_0^n \max_{k_n'' \leq k \leq k_n^{(3)}} \Pi_k(|d-d_o| \leq k_v w_n (\log n)^{-1}|X) \stackrel{P_o}{\to} 0, \tag{5.1}$$

In the remainder we prove (5.1). For every $k \leq k_n^{(3)}$ we can write, using the notation of (4.1),

$$\begin{aligned}\Pi_k(|d-d_o| < k_v w_n(\log n)^{-1}|X) &\leq \frac{N_{n,k}}{D_{n,k}} \\ := \frac{\int_{|d-d_o|<k_v w_n(\log n)^{-1}} \exp\{l_n(d,k) - l_n(d_o,k) + \log I_n(d,k)\} d\pi_d(d)}{\int_{|d-d_o|<w_n} \exp\{l_n(d,k) - l_n(d_o,k) + \log I_n(d,k)\} d\pi_d(d)}.\end{aligned} \tag{5.2}$$

Let $\delta_2 > 0$ and A_n be the set of $X \in \mathbb{R}^n$ such that for all $k_n'' \leq k \leq k_n^{(3)}$,

$$\left.\begin{aligned} \sup_{|d-d_o|\leq \bar{v}_n} |\log I_n(d,k) - \log I_n(d_o,k)| \leq \epsilon_1, \\ \left|l_n^{(1)}(d_o,k) - \mathcal{D}(l_n^{(1)}(d_o,k))\right| \leq C_1 n^{\frac{1}{2}} k^{-\frac{1}{2}}\sqrt{\log n}, \\ \sup_{|d-d_o|\leq w_n} \left|l_n^{(2)}(d,k) - l_n^{(2)}(d_o,k)\right| \leq \epsilon_1 n^{1+\epsilon_1} r_k w_n. \end{aligned}\right\}$$

Compared to the definition of A_n in the proof of Theorem 2.1, the constraints on $l_n^{(2)}(d,k)$ and I_n are different. For the latter, recall from Lemma 3.5 that $\log I_n(d,k) = \log I_n(d_o,k) + \mathbf{o_{P_o}}(1)$, uniformly over $d \in (d_o - w_n, d_o + w_n)$. As in the proof of Theorem 2.1, it now follows from Lemmas 3.4 and 3.5 that $P_o^n(A_n^c) \to 0$. We can write

$$E_0^n\left[\max_{k_n''\leq k\leq k_n^{(3)}} \frac{N_{n,k}}{D_{n,k}}\right] \leq P_o^n(A_n^c) + E_0^n\left[\max_{k_n''\leq k\leq k_n^{(3)}} \frac{N_{n,k}}{D_{n,k}} 1_{A_n}\right],$$

and bound $N_{n,k}/D_{n,k}$ pointwise for $X \in A_n$. If ϵ_1 is small enough,

$$\frac{(l_n^{(1)}(d_o,k))^2}{2|l_n^{(2)}(d_o,k)|} = O_{p_o}(k^{-2\beta}n) = \mathbf{o_{P_o}}(n^{-\epsilon_1}/w_n),$$

uniformly over $k \in (k_n'', k_n^{(3)})$. Hence, analogous to (4.3) and (4.4), we find that for some $\delta_2 > 0$ and for all $X \in A_n$,

$$\begin{aligned} &l_n(d,k) - l_n(d_o,k) + \log I_n(d,k) - \log I_n(d_o,k) \\ &\quad \leq 2\epsilon_1 - \frac{|l_n^{(2)}(d_o,k)|(1-n^{-\delta_2})}{2}(d-d_o-b_n(d_o))^2 + \frac{(l_n^{(1)}(d_o,k))^2}{2|l_n^{(2)}(d_o,k)|} \\ &l_n(d,k) - l_n(d_o,k) + \log I_n(d,k) - \log I_n(d_o,k) \\ &\quad \geq -2\epsilon_1 - \frac{|l_n^{(2)}(d_o,k)|(1+n^{-\delta_2})}{2}(d-d_o-b_n(d_o))^2 + \frac{(l_n^{(1)}(d_o,k))^2}{2|l_n^{(2)}(d_o,k)|}, \end{aligned}$$

when n is large enough since $k > k_n''$. This follows from the definition of $b_n(d_o)$ in (3.5).

We now lower-bound $b_n(d_o)$ by bounding the terms on the right in (3.7) in Lemma 3.4. Since $\theta_{oj} = c_0 j^{-(\beta+1/2)}(\log j)^{-1}$, it follows that

$$r_k^{-1}\sum_{j>k} j^{-1}\theta_{o,j} = c_0 r_k^{-1}\sum_{j>k} j^{-\beta-\frac{3}{2}}/(\log j) \geq ck^{-\beta+\frac{1}{2}}(\log k)^{-1},$$

for some $c > 0$. Since $X \in A_n$, $2\mathcal{S}(l_n^{(1)}(d_o,k))/(nr_k) \leq 2C_1\sqrt{k/n}\sqrt{\log n}$. Since $k \leq k_n^{(3)}$, this bound is $o(k^{-\beta+\frac{1}{2}}(\log k)^{-1})$. The last term in (3.7) is $o(n^{\epsilon-1})$ when $\beta > 5/2$, and hence this term is also $o(k^{-\beta-\frac{1}{2}}(\log k)^{-1})$. Therefore, the last two terms in (3.7) are negligible with respect to $r_k^{-1}\sum_{j>k} j^{-1}\theta_{o,j}$. We deduce that $b_n(d_o) \geq ck^{-\beta+\frac{1}{2}}(\log k)^{-1} \geq cn^{-(2\beta-1)/(4\beta+2)}(\log n)^{-(2\beta+3)/(4\beta+2)}$ for n large enough.

Consequently, when the constant k_v is chosen sufficiently small, $\sqrt{nr_{k_n^{(3)}}}(b_n(d_o) - k_v w_n(\log n)^{-1}) \geq (c-k_v)n^{1/(4\beta+2)}(\log n)^{-(\beta+1)/(2\beta+1)} := z_n \to \infty$. We now substitute the above bounds on $l_n(d,k) - l_n(d_o,k) + \log I_n(d,k) - \log I_n(d_o,k)$ in the right hand side of (5.2), make the change of variables $u = d - d_o - b_n(d_o)$ and obtain uniformly over $k \in (k_n'', k_n^{(3)})$

$$\begin{aligned}\frac{N_{n,k}}{D_{n,k}} &\leq e^{5\epsilon_1}\frac{\int_{u\leq -k_v w_n(\log n)^{-1}-b_n(d_o)}\exp\{-\frac{nr_k u^2}{4}\}du}{\int_{|u|<w_n/2}\exp\{-\frac{nr_k u^2(1-n^{-\delta_2})}{4}\}du}\\ &\leq e^{5\epsilon_1}\frac{\int_{v>z_n(1-n^{-\delta_2})}\exp\{-\frac{v^2}{2}\}dv}{\int_{|v|<w_n\sqrt{nr_k/8}}\exp\{-\frac{v^2}{2}\}dv} = o(1).\end{aligned}$$

The case of **Prior B** corresponds to considering $k_n'' = k_n^{(3)}$ and is dealt with in the same way. This achieves the proof of Theorem 2.2.

6. Conclusion

In this paper we have derived conditions leading to a BVM type of result for the long memory parameter $d \in (-\frac{1}{2}, \frac{1}{2})$ of a stationary Gaussian process, for the class of FEXP-priors. The result implies in particular that asymptotically credible intervals for d have good frequentist coverage. To our knowledge this is the first result on Bernstein-von Mises theorems in non -regular semi-parametric models and we believe that the approach we have considered can be used in other semi-parametric non-regular models, such as models for extremes, and with other families of priors hence completing (not exhaustively) the recent works of Castillo (2012) and Bickel and Kleijn (2012).

Recently Hoffmann et al. (2013) have obtained a generic lower bound on the expectation under the true model P^o of posterior probability of complements of shrinking neighbourhoods in the form : given some loss function ℓ and some rate ϵ_n

$$E^o\left[\Pi\left(\ell(f_{d,\theta}, f_o) \leq \epsilon_n | X\right)\right] \geq e^{-Cn\phi_n^2}$$

$$\phi_n = \inf_{f_o} \inf\{\| \log f_{d,\theta} - \log f_o\|_2; \ell(f_{d,\theta}, f_o) \geq \epsilon_n, (d, \theta)\}.$$

In this paper we are interested in $\ell(f_{d,\theta}, f_o) = |d - d_o|$ and d, θ, together with d_o, θ_o vary in $[-t, t] \times \{\sum_{j\geq 1} \theta_j^2 (1+j)^{2\beta} \leq L\}$, in which case, using the same computations as in Section 3.1, Lemma 3.1, we have that $\phi_n \lesssim \sqrt{\log n/n}$. As discussed in Hoffmann et al. (2013) this implies that the approach based on tests proposed by Ghosal et al. (2000) and Ghosal and van der Vaart (2007) leads to suboptimal concentration rates and cannot be applied. It thus become necessary to have a much more involved analysis of the posterior distribution, explaining the difficulty in studying the frequentist properties of non-regular Bayesian semi-parametric approaches.

A by-product of our results is that the *most natural prior* (Prior C) from a Bayesian perspective, which is also the prior leading to adaptive minimax rates under the loss function l on f, leads to sub-optimal estimators in terms of d. Prior A leads to optimal estimators for d however it is not adaptive, contrary to the frequentist estimator proposed by Iouditsky et al. (2001). An interesting direction for future work would be to define an adaptive minimax estimation procedure for d. We believe this can be done using a more flexible basis for the expansion of $\log f$ such as wavelet bases and most of the computations considered here, but is beyond the scope of this paper.

Acknowledgements

This work was supported by the 800-20072010 grant ANR-07-BLAN-0237-01 SP Bayes.

Supplementary Material

Results on integrals, convergence of traces and the Hölder constants of various functions

(doi: 10.1214/13-EJS864SUPP; .pdf). This supplement also contains the proofs of Lemmas 3.4, 3.5.

References

Arbel, J., Gayraud, G., and Rousseau, J. (2013). Bayesian optimal adaptive estimation using a sieve prior. *Scandinavian Journal of Statistics*, 40(3):549–570.

Beran, J. (1993). Fitting long-memory models by generalized linear regression. *Biometrika*, 80(4):817–822. MR1282790

Beran, J. (1994). *Statistics for long-memory processes*, volume 61 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York. MR1304490

Bickel, P. J. and Kleijn, B. J. K. (2012). The semiparametric Bernstein-von Mises theorem. *Annals of Statistics*, 40:206–237. MR3013185

Castillo, I. (2012). A semiparametric Bernstein-von Mises theorem for Gaussian process priors. *Probability Theory and Related Fields*, 152(1–2):53–99. MR2875753

Chopin, N., Rousseau, J., and Liseo, B. (2013). Computational aspects of Bayesian spectral density estimation. *Journal of Computational and Graphical Statistics*, 22(3):533–557.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.*, 17(4):1749–1766. MR1026311

Freedman, D. (1999). On the Bernstein-von Mises theorem with infinite-dimensional parameters. *Ann. Statist.*, 27(4):1119–1140. MR1740119

Ghosal, S., Ghosh, J., and van der Vaart, A. (2000). Convergence rates of posterior distributions. *The Annals of Statistics*, 28(2):500–531. MR1790007

Ghosal, S. and van der Vaart, A. (2007). Convergence rates of posterior distributions for noniid observations. *The Annals of Statistics*, 35(1):192–225. MR2332274

Hoffmann, M., Rousseau, J., and Schmidt-Hieber, J. (2013). On adaptive posterior concentration rates. *ArXiv e-prints*.

Holan, S., McElroy, T., and Chakraborty, S. (2009). A Bayesian approach to estimating the long memory parameter. *Bayesian Anal.*, 4(1):159–190. MR2486243

Hurvich, C. M., Moulines, E., and Soulier, P. (2002). The FEXP estimator for potentially non-stationary linear time series. *Stochastic Process. Appl.*, 97(2):307–340. MR1875337

Iouditsky, A., Moulines, E., and Soulier, P. (2001). Adaptive estimation of the fractional differencing coefficient. *Bernoulli*, 7(5):699–731. MR1867079

Jensen, M. J. (2004). Semiparametric Bayesian inference of long-memory stochastic volatility models. *J. Time Ser. Anal.*, 25(6):895–922. MR2099644

Ko, K., Qu, L., and Vannucci, M. (2009). Wavelet-based Bayesian estimation of partially linear regression models with long memory errors. *Statist. Sinica*, 19(4):1463–1478. MR2589192

Kruijer, W. and Rousseau, J. (2013). Supplement to "Bayesian semi-parametric estimation of the long-memory parameter under FEXP-priors." DOI:10.1214/13-EJS864SUPP.

Li, X. and Zhao, L. H. (2002). Bayesian nonparametric point estimation under a conjugate prior. *Stat. Probab. Lett.*, 1(4):23–30. MR1900337

Lieberman, O., Rousseau, J., and Zucker, D. M. (2003). Valid asymptotic expansions for the maximum likelihood estimator of the parameter of a stationary, Gaussian, strongly dependent process. *Ann. Statist.*, 31(2):586–612. Dedicated to the memory of Herbert E. Robbins. MR1983543

Moulines, E. and Soulier, P. (2003). Semiparametric spectral estimation for fractional processes. In *Theory and applications of long-range dependence*, pages 251–301. Birkhäuser Boston, Boston, MA. MR1956053

Philippe, A. and Rousseau, J. (2002). Non-informative priors in the case of Gaussian long-memory processes. *Bernoulli*, 8(4):451–473. MR1914698

Rivoirard, V. and Rousseau, J. (2012). Bernstein-von Mises theorem for linear functionals of the density. *Ann. Stat.*, 40(3):1489–1523. MR3015033

Robinson, P. M. (1994). Time series with strong dependence. In *Advances in econometrics, Sixth World Congress, Vol. I* (Barcelona, 1990), volume 23 of *Econom. Soc. Monogr.*, pages 47–95. Cambridge Univ. Press, Cambridge. MR1278267

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.*, 23(5):1630–1661. MR1370301

Rousseau, J., Chopin, N., and Liseo, B. (2012). Bayesian nonparametric estimation of the spectral density of a long memory Gaussian process. *Annals of Statistics*, 40:964–995. MR2985940

Rousseau, J. and Kruijer, W. (2011). Adaptive Bayesian Estimation of a spectral density. `http://hal.archives-ouvertes.fr/hal-00641485`

van der Vaart, A. W. (1998). *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. MR1652247