Bernstein von Mises Theorems for Gaussian Regression with increasing number of regressors

Dominique Bontemps

September 8, 2010

Abstract

This paper brings a contribution to the Bayesian theory of nonparametric and semiparametric estimation. We are interested in the asymptotic normality of the posterior distribution in Gaussian linear regression models when the number of regressors increases with the sample size. Two kinds of Bernstein-von Mises Theorems are obtained in this framework: nonparametric theorems for the parameter itself, and semiparametric theorems for functionals of the parameter. We apply them to the Gaussian sequence model and to the regression of functions in Sobolev and C^α classes, in which we get the minimax convergence rates. Adaptivity is reached for the Bayesian estimators of functionals in our applications.

Contents

1 Introduction

To estimate a parameter of interest in a statistical model, a Bayesian puts a prior distribution on it and looks at the posterior distribution, given the observations. A Bernstein-von Mises Theorem is a result stating that under adequate conditions the posterior distribution is asymptotically normal, centered at the maximum likelihood estimator (MLE) of the model used, with a variance equal to the asymptotic frequentist variance of the MLE.

Such an asymptotic posterior normality is important because it allows to construct approximate credible regions, based on the posterior distribution, which keep good frequentist properties. In particular it is difficult to build frequentist confidence regions in complex models, while the Monte-Carlo Markov chain algorithms (MCMC) make more feasible the construction of Bayesian confidence regions — however Bernstein-von Mises Theorems are difficult to derive in complex models.

For parametric models, the Bernstein-von Mises Theorem is a well-known result, for which we refer to [18]. In nonparametric models (where the parameter space is infinite-dimensional or growing), and semiparametric models (when the parameter of interest is a finite-dimensional functional of the complete infinite-dimensional parameter), there are still relatively few asymptotic normality results. [6] gives negative results, and we recall some positive ones below. However many recent papers deal with the convergence rate of posterior distributions in various settings, which is linked with the model complexity: we refer to [9, 16] as early representatives of this school.

Nonparametric Bernstein-von Mises Theorems have been developed for models based on a sieve approximation, where the dimension of the parameter grows with the sample size. In particular two situations have been studied: regression models in [7]; exponential models in [8], [4], and [2] (this last one deals with the discrete case, when the observations follow some unknown infinite multinomial distribution).

In semiparametric frameworks the asymptotic normality has been obtained in several situations. [12] and [11] study the nonparametric right-censoring model and the proportional hazard model. [3] obtains Bernstein-von Mises Theorems for Gaussian process priors, in the semiparametric framework where the unknown quantity is (θ, f), with θ the parameter of interest and f an infinite-dimensional nuisance parameter. It clarifies a preceding paper [15], which considers also the more general framework where the quantity of interest is a finite-dimensional function $g(f)$ of the infinite-dimensional parameter f of the model. [14] obtains the Bernstein-von Mises Theorem for linear functionals of the density of the observations, in the context of a sieve approximation; they achieve also the frequentist minimax estimation rate for densities in specific regularity classes with a deterministic (non-adaptive) value of the cutoff k_n.

In the current paper we obtain nonparametric and semiparametric Bernstein-von Mises Theorems in a Gaussian regression framework with an increasing number of regressors.

Our nonparametric results cover the case of a specific Gaussian prior, and the case of more generic smooth priors. They are said nonparametric because we use sieve priors and the dimension of the parameter grows. These results improve on the preceding ones by [7] which did not suppose the normality of the errors but imposed other conditions, in particular on the growth rate of the number of regressors. We apply them to the periodic Sobolev classes and to regularity classes $C^\alpha[0,1]$ in the context of the regression model (using respectively trigonometric polynomials and splines as regressors), as well as to the Gaussian sequence model. In all these situations we get the asymptotic normality of the posterior in addition to the minimax convergence rates, with appropriate (non-adaptive) choices of the prior. We also show that for some priors known to reach this convergence rate, the Bernstein-von Mises Theorem does not hold. We derive also semiparametric Bernstein-von Mises Theorems for linear and nonlinear functionals of the parameter. The linear case is an immediate corollary of the nonparametric theorems and do not need any additional condition. We apply these results to the periodic Sobolev classes to estimate a linear functional and the L^2 norm of the regression function f if enough smoothness is present, and in both cases we are able to build an adaptive Bayesian estimator which achieves the minimax convergence rate whatever the unknown parameter of the class is, in addition to the asymptotic normality.

The paper is organized as follows. We present the framework in section 2. Section 3 states the nonparametric Bernstein-von Mises Theorems, for Gaussian or non-Gaussian priors. In section 4 we expound the semiparametric Bernstein-von Mises Theorems for linear and non-linear functionals of the parameter. Then we consider in section 5 applications to the Gaussian sequence model, and to the regression of a function in a Sobolev and $C^\alpha[0,1]$ class. In section 6 the nonparametric and semiparametric Bernstein-von Mises Theorems are proved. Eventually the Appendix contains various technical tools used in the main analysis.

2 Framework

We consider a Gaussian linear regression framework. For any $n \geq 1$, our observation $Y = (Y_1, \dots, Y_n) \in \mathbb{R}^n$ is a Gaussian random vector

$$Y = F + \varepsilon \tag{1}$$

where the vector of errors $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n) \sim \mathcal{N}(0, \sigma_n^2 I_n)$ is centered normal and the mean vector F belongs to $\mathbb{R}^n$. The observations Y_i and the variance σ_n^2 of the errors may depend on n, but σ_n^2 is known. Fix a (sequence of) mean vector(s) F_0. We denote by P_{F_0} the probability distribution of a random variable following $\mathcal{N}(F_0, \sigma_n^2 I_n)$, and E the associated expectation.

Let $\phi_1, \dots, \phi_{k_n}$ a collection of k_n linearly independent regressors in $\mathbb{R}^n$, where $k_n \leq n$ grows with n. We gather these regressors in the $n \times k_n$-matrix Φ of rank k_n, and we denote $\langle \phi \rangle$ their linear span. $\langle \phi \rangle$ is the misspecified model in which the Bernstein-von Mises Theorems will be stated. It can be parametrized as $\langle\phi\rangle = \left\{\Phi\theta : \theta = (\theta_1, \dots, \theta_{k_n}) \in \mathbb{R}^{k_n}\right\}$. We denote by P_θ the probability distribution of a random variable following $\mathcal{N}(\Phi\theta, \sigma_n^2 I_n)$, and E_θ the associated expectation.

As examples, we present three different frameworks, each one with its own collection of regressors. In section 5 the Bernstein-von Mises Theorems are applied to each one of these frameworks.

1. **The Gaussian sequence model.**
   Our first application concerns the Gaussian sequence model, which is also equivalent to the white noise model (see [13, ch. 4] for instance). We consider the infinite dimensional setting

$$Y_j = \theta_j^0 + \frac{1}{\sqrt{n}}\xi_j, \qquad j \geq 1 \tag{2}$$

   where the random variables $\xi_j, j \geq 1$ are independant and have distribution $\mathcal{N}(0,1)$. Projecting on the first k_n coordinates with $k_n \leq n$, we retrieve our model (1) with $\theta_0 = (\theta_j^0)_{1\leq j\leq k_n}$, $\sigma_n = 1/\sqrt{n}$, and $\Phi^T\Phi = I_{k_n}$.

2. **Regression of a function in a Sobolev class.**
   Let $f : [0,1] \to \mathbb{R}$ be a function in $\mathbb{L}^2([0,1])$. We observe realizations of random variables

$$Y_i = f(i/n) + \varepsilon_i \tag{3}$$

   for $1 \leq i \leq n$, where the errors ε_i are iid $\mathcal{N}(0, \sigma_n^2)$ and σ_n does not depend on n.

   We denote by $(\varphi_j)_{j\geq 1}$ the Fourier basis

$$\begin{aligned} \varphi_1 &\equiv 1 \\ \varphi_{2m}(x) &= \sqrt{2}\cos(2\pi m x) \qquad \forall m \geq 1 \\ \varphi_{2m+1}(x) &= \sqrt{2}\sin(2\pi m x) \qquad \forall m \geq 1 \end{aligned} \tag{4}$$

   For the regression on Fourier's basis we choose a regular design $x_i = i/n$ for $1 \leq i \leq n$. This gives the collection of regressors $\phi_j = (\varphi_j(i/n))_{1\leq i\leq n}$, $1 \leq j \leq k_n$.

   In practice we suppose that f belongs to one of the Sobolev classes:

   Definition 1. *Let $\alpha > 0$ and $L > 0$. Let $(\varphi_j)_{j\geq 1}$ denote the Fourier basis (4). We define the Sobolev class $\mathcal{W}(\alpha, L)$ as the collection of all functions $f = \sum_{j=1}^{\infty} \theta_j\varphi_j$ in $\mathbb{L}^2([0,1])$ such that $\theta = (\theta_j)_{j\geq 1}$ is an element of the ellipsoid of $\ell^2(\mathbb{N})$*

$$\Theta(\alpha, L) = \left\{\theta \in \ell^2(\mathbb{N}) : \sum_{j=1}^{\infty} a_j^2\theta_j^2 \leq \frac{L^2}{\pi^{2\alpha}}\right\}$$

   where

$$a_j = \begin{cases} j^\alpha & \text{if } j \text{ is even;} \\ (j-1)^\alpha & \text{if } j \text{ is odd.} \end{cases} \tag{5}$$

3. **Regression of a function in $C^\alpha[0,1]$.**
Fix a regularity $\alpha > 0$, and consider a function $f \in C^\alpha[0,1]$. This means that f is α_0 times continuously differentiable with $\|f\|_\alpha < \infty$, α_0 being the greatest integer less than α and the seminorm being defined by

$$\|f\|_\alpha = \sup_{x \neq x'} \frac{\left|f^{(\alpha_0)}(x) - f^{(\alpha_0)}(x')\right|}{|x - x'|^{\alpha - \alpha_0}}.$$

Consider a design $\left(x_i^{(n)}\right)_{n \geq 1, 1 \leq i \leq n}$, not necessarily uniform. Here F_0 is the vector $\left(f\left(x_i^{(n)}\right)\right)_{1 \leq i \leq n}$. Once again we suppose that $\sigma_n = \sigma$ does not depend on n.

Fix an integer $q \geq \alpha$, and let $K = k_n + 1 - q$. Partition the interval $(0,\ 1]$ into K subintervals $((j-1)/K,\ j/K]$ for $1 \leq j \leq K$. We want to perform the regression of f in the space of splines of order q defined on that partition, and use the B-splines basis $(B_j)_{1 \leq j \leq k_n}$ (see [5] for instance). Our collection of regressors is $\phi_j = \left(B_j\left(x_i^{(n)}\right)\right)_{1 \leq i \leq n}$, for $1 \leq j \leq k_n$.

For any value of $n \geq 1$, let W be a prior distribution on F, with support included in $\langle\phi\rangle$. Equivalently, W is induced by a probability distribution $\widetilde{W}$ on θ by the application $\theta \mapsto \Phi\theta$. P^W denotes the marginal distribution of Y under prior W, and $W(dG(F)|Y)$ denotes the posterior distribution of a functional $G(F)$. Note that everything depends on n — W for instance is a distribution on $\mathbb{R}^n$ — even if we do not use n as index to simplify our notations.
W is a sieve prior. Such priors are specially well adapted for increasing dimension frameworks; they also make clear the relations between the parametric and nonparametric results. On the other hand the question of the choice of the cutoff k_n arises.
The exact parametrization by θ and the corresponding collection of regressors $\phi_1, \ldots, \phi_{k_n}$ are somehow arbitrary: what matters is the posterior distribution of F and this depends on $\langle\phi\rangle$, which is characterized by the matrix $\Sigma = \Phi(\Phi^T\Phi)^{-1}\Phi^T$ of the orthogonal projection onto $\langle\phi\rangle$. In practice it is difficult to dissociate $\langle\phi\rangle$ and the collection $\phi_1, \ldots, \phi_{k_n}$, but we have chosen to emphasize W and F over $\widetilde{W}$ and θ.

In the model $\langle\phi\rangle$, the MLE of F_0 is the orthogonal projection $Y_{\langle\phi\rangle}$ of Y; so $Y_{\langle\phi\rangle} = \Sigma Y$. We set $\theta_Y = (\Phi^T\Phi)^{-1}\Phi^T Y$ its associated parameter. Let also $F_{\langle\phi\rangle} = \Phi\theta_0$ be the projection of F_0 on $\langle\phi\rangle$, with $\theta_0 = (\Phi^T\Phi)^{-1}\Phi^T F_0$. Even if $\langle\phi\rangle$ contains the support of the prior distribution W, we do not suppose that F_0 belongs to $\langle\phi\rangle$, and this improves on some previous results. θ_0 has not to be seen as some "true" parameter.

Although the MLE is naturally defined *in the sieve* $\langle\phi\rangle$, it heavily depends on the choice of $\langle\phi\rangle$. Therefore the Bernstein-von Mises Theorems we establish depends on the choice of the sieve the prior distribution is built on. This is true in particular in a maybe more veiled way for our semiparametric results, in which the centering point is a plug-in estimator based on the MLE defined on $\langle\phi\rangle$. In nonparametric models constructed on an infinite dimensional parameter, there is no definition of a MLE; what should be the natural centering for a Bernstein-von Mises Theorem in such situations is not clear.

To conclude this section, the following immediate frequentist result gives the distribution of $Y_{\langle\phi\rangle}$ under P_{F_0}:

$$Y_{\langle\phi\rangle} \sim \mathcal{N}\left(F_{\langle\phi\rangle}, \sigma_n^2 \Sigma\right).$$

3 Nonparametric Bernstein-von Mises Theorems

The proofs of our nonparametric results are delayed to section 6.

3.1 With Gaussian priors

We consider here a centered, normal prior distribution W which is isotropic on $\langle\phi\rangle$, so that $W = \mathcal{N}\left(0, \tau_n^2 \Sigma\right)$ for some sequence τ_n. Essentially the only assumption needed in this case is that the prior becomes flat enough as n grows. $\|Q - Q'\|_{\mathrm{TV}}$ denotes the total variation norm between two probability distributions Q and Q'.

Theorem 1. *Assume that $\sigma_n = o(\tau_n)$, $\|F_0\| = o(\tau_n^2/\sigma_n)$ and $k_n = o(\tau_n^4/\sigma_n^4)$. Then*

$$E\left\|W(dF|Y) - \mathcal{N}\left(Y_{\langle\phi\rangle}, \sigma_n^2 \Sigma\right)\right\|_{\mathrm{TV}} \to 0 \text{ as } n \to \infty.$$

Since the support of W is included in $\langle\phi\rangle$, we can equivalently state

$$E\left\|\widetilde{W}(d\theta|Y) - \mathcal{N}\left(\theta_Y, \sigma_n^2 (\Phi^T \Phi)^{-1}\right)\right\|_{\mathrm{TV}} \to 0 \text{ as } n \to \infty.$$

Theorem 1 does not deal with the modeling bias introduced by taking a prior restricted to $\langle\phi\rangle$. This is an important question in nonparametric statistics, and k_n has to be chosen in order to achieve the bias-variance tradeoff. In most cases this bias has already been studied in frequentist papers on sieve approximation.

As an example, let us consider an usual regression framework with $F_0 = (f(x_i))_{1\le i\le n}$, where f is some function and $(x_i)_{1\le i\le n}$ some design. If σ_n does not depend on n, both conditions $\|F_0\| = o(\tau_n^2/\sigma_n)$ and $k_n = o(\tau_n^4/\sigma_n^4)$ are verified for instance if f is bounded and $n^{1/4} = o(\tau_n)$. These conditions can be read in the other way: τ_n^4 must be large enough with respect to $\|F_0\|$ and k_n.

3.2 With smooth priors

We consider now more general priors. We get an abstract result, but with powerful applications.

Theorem 2. *Suppose that W is induced by a distribution on θ admitting a density $w(\theta)$ with respect to Lebesgue measure. If there exists a sequence $(M_n)_{n\ge 1}$ such that*

1. $$\sup_{h^T \Phi^T \Phi h \le \sigma_n^2 M_n,\, g^T \Phi^T \Phi g \le \sigma_n^2 M_n} \frac{w(\theta_0 + h)}{w(\theta_0 + g)} \to 1 \text{ as } n \to \infty.$$
2. $k_n \ln k_n = o(M_n)$
3. $$\max\left(0, \ln\left(\frac{\sqrt{\det(\Phi^T \Phi)}}{\sigma_n^{k_n} w(\theta_0)}\right)\right) = o(M_n)$$

Then

$$E \left\| W(dF|Y) - \mathcal{N}\left(Y_{\langle\phi\rangle}, \sigma_n^2 \Sigma\right) \right\|_{\mathrm{TV}} \to 0 \ \textit{as}\ n \to \infty.$$

Since the support of W is included in $\langle\phi\rangle$, we can equivalently state

$$E \left\| \widetilde{W}(d\theta|Y) - \mathcal{N}\left(\theta_Y, \sigma_n^2 (\Phi^T\Phi)^{-1}\right) \right\|_{\mathrm{TV}} \to 0 \ \textit{as}\ n \to \infty.$$

With Condition 1 we ask for a sufficiently flat prior in a given neighborhood of θ_0. By Conditions 2 and 3 we insure that this neighborhood has enough prior weight. This kind of assumptions is quite common in the literature dealing with the concentration of posterior distributions. These assumptions are needed together in order to get the Gaussian shape of the posterior distribution. Several of our applications illustrate that priors known to induce the posterior minimax convergence rate may not be flat enough to get the Gaussian shape with the asymptotic variance $\sigma_n^2\Sigma$.

Our main applications, to the Gaussian sequence model, and to the regression model using trigonometric polynomials and splines, are developed in section 5. We now present two remarks about the parametric case and the comparison with the pioneer work of Ghosal [7].

The parametric case. Consider the regression of a function f defined on $[0,1]$, with a fixed number k of regressors. Set a design $(x_i^{(n)})_{n\geq 1, 1\leq i\leq n}$, with $x_i^{(n)} \in [(i-1)/n, i/n]$ for any $n \geq 1$, and $F_0 = \left(f(x_i^{(n)})\right)_{1\leq i\leq n}$. Choose a finite number of piecewise continuous and linearly independent regressors $(\varphi_j)_{1\leq j\leq k}$ on $[0,1]$, and set $\phi_j = \left(\varphi_j\left(x_i^{(n)}\right)\right)_{1\leq i\leq n}$ for $1 \leq j \leq k$. f, $k_n = k$, $\sigma_n = \sigma$, and W do not depend on n.

We would like to compare Theorem 2 with the usual Bernstein-von Mises Theorem for parametric models, applied to such a regression framework. In that setting, let us suppose that w is continuous and positive, and that f is bounded. Then Condition 1 becomes $M_n = o(n)$, while Condition 3 reduces to $\ln n = o(M_n)$. Clearly, there exist such sequences $(M_n)_{n\geq 1}$, and Theorem 2 applies.

The rescaling by $\sqrt{n}$ of the Bernstein-von Mises Theorem for parametric models is here hidden in the asymptotic posterior variance $\sigma^2(\Phi^T\Phi)^{-1}$ of the parameter θ. Indeed, $(1/n)\ \Phi^T\Phi$ is a Riemann sum, and converges towards the Gramian matrix of the collection $(\varphi_j)_{1\leq j\leq k}$ in $\mathbb{L}^2([0,1])$.

Proof. We have $\|\Phi\theta_0\| \leq \|F_0\| \leq \sqrt{n}\|f\|_\infty$, and $\|\theta_0\|^2 \leq \left\|(\Phi^T\Phi)^{-1}\right\| \|\Phi\theta_0\|^2 \leq \left\|n(\Phi^T\Phi)^{-1}\right\| \|f\|_\infty^2$. $(1/n)\ \Phi^T\Phi$ converges towards the Gramian matrix of the collection $(\varphi_j)_{1\leq j\leq k}$ in $\mathbb{L}^2([0,1])$, and its smallest eigenvalue is lower bounded for n large enough. Therefore θ_0 is bounded, and we can consider it lies in some compact set on which w is uniformly continuous and lower bounded by a positive constant. The rest follows. □

Comparison with Ghosal's conditions. The Bernstein-von Mises Theorem in a regression setting when the number of parameters goes to infinity has been first studied by Ghosal [7] as an early step in the development of frequentist nonparametric Bayesian theory. In his paper the errors ε_i are not supposed to be Gaussian. Under the Gaussianity assumption, we get improved results. In particular our condition for the prior smoothness is simpler, and the growth rate of the dimension k_n is much less constrained.

- [7] does not admit a modeling bias between F_0 and $\Phi\theta_0$. In the present work the normality of the errors permits to take $F_0 \neq \Phi\theta_0$ without any cost, as it appears in the core of the proof (Lemma 7).
- In [7] σ_n is constant, which does not allow the application to the Gaussian sequence model.
- At last, [7] restricts the growth of the dimension k_n to $k_n^4 \ln k_n = o(n)$ (see below). It is then not possible to obtain the applications to the Gaussian sequence model or to the regression model for Sobolev or C^α classes.

Let $\delta_n^2 = \|(\Phi^T\Phi)^{-1}\|$ be the operator norm of $(\Phi^T\Phi)^{-1}$ for the ℓ^2 metric, and let η_n^2 be the maximal value on the diagonal of Σ. With our notations, the remaining assumptions of [7] become

(A3) There exists $\eta_0 > 0$ such that $w(\theta_0) > \eta_0^{k_n}$. Moreover

$$|\ln w(\theta) - \ln w(\theta_0)| \leq L_n(C)\|\theta - \theta_0\|, \tag{6}$$

whenever $\|\theta - \theta_0\| \leq C\delta_n k_n \sqrt{\ln k_n}$, where the Lipschitz constant $L_n(C)$ is subject to some growth restriction (see assumption A4).

(A4)

$$\forall C > 0, L_n(C)\delta_n k_n \sqrt{\ln k_n} \to 0 \quad \text{and} \quad \eta_n k_n^{3/2} \sqrt{\ln k_n} \to 0. \tag{7}$$

Further the design satisfies a condition on the trace of $\Phi^T\Phi$:

$$\mathrm{tr}(\Phi^T\Phi) = O(nk_n). \tag{8}$$

Since Σ is an orthogonal projection matrix on a k_n-dimensional space, $\mathrm{tr}(\Sigma) = k_n$ and $\eta_n^2 \geq k_n/n$. Consequently the last part of (7) entails $k_n^4 \ln k_n = o(n)$.

If we add the normality of the errors and a slight technical condition $\ln n = o(k_n \ln k_n)$, these assumptions entail ours. Indeed, set $M_n = C^2 k_n^2 \ln k_n$ for some arbitrary value of C. Our condition 2 is immediate. Condition 1 is got from (6) and the first part of (7). The beginning of (A3) entails $-\ln w(\theta_0) = O(k_n) = o(M_n)$. Using the concavity of the ln function and (8), we get $\ln\det(\Phi^T\Phi) \leq k_n \ln \mathrm{tr}(\Phi^T\Phi) - k_n \ln k_n = O(k_n \ln n) = o(M_n)$. Therefore our condition 3 holds.

4 Semiparametric Bernstein-von Mises Theorems

We consider two kinds of functionals of F: linear and non-linear ones. These results can be easily adapted to functionals of θ, using the maps $\theta \mapsto \Phi\theta$ and $F \mapsto (\Phi^T\Phi)^{-1}\Phi^T F$.

4.1 The linear case

For linear functionals of F, we have the following corollary:

Corollary 1. *Let $p \geq 1$ fixed, and G be a $\mathbb{R}^p \times \mathbb{R}^n$-matrix. Suppose that the conditions of either Theorem 1 or Theorem 2 are verified. Then*

$$E\left\|W(d(GF)|Y) - \mathcal{N}\left(GY_{\langle\phi\rangle}, \sigma_n^2 G\Sigma G^T\right)\right\|_{\mathrm{TV}} \to 0 \text{ as } n \to \infty.$$

Further, the distribution of $GY_{\langle\phi\rangle}$ is $\mathcal{N}\left(GF_{\langle\phi\rangle}, \sigma_n^2 G\Sigma G^T\right)$.

Corollary 1 is just a linear transform of the preceding Theorems, and of the distribution of $Y_{\langle\phi\rangle}$.

An example of application is given in subsection 5.2, in the context of the regression on Fourier's basis.

4.2 The nonlinear case

Let $p \geq 1$ fixed, and $G : \mathbb{R}^n \mapsto \mathbb{R}^p$ be a twice continuously differentiable function. For $F \in \mathbb{R}^n$, let $\dot{G}_F$ denote the Jacobian matrix of G at F, and $D_F^2 G(\cdot,\cdot)$ the second derivative of G, as a bilinear function on $\mathbb{R}^n$. For any $F \in \langle\phi\rangle$ and $a > 0$, let

$$B_F(a) = \sup_{h\in\langle\phi\rangle : \|h\|^2 \leq \sigma_n^2 a} \sup_{0\leq t\leq 1} \left\|D_{F+th}^2 G(h,h)\right\|. \tag{9}$$

where $\|\cdot\|$ denotes the Euclidean norm of $\mathbb{R}^p$.

We also consider the following nonnegative symmetric matrix

$$\Gamma_F = \sigma_n^2 \dot{G}_F \Sigma \dot{G}_F^T. \tag{10}$$

In the following, $\|\Gamma_F^{-1}\|$ denotes the Euclidean operator norm of Γ_F^{-1}, which is also the inverse of the smallest eigenvalue of Γ_F.

Let $\mathcal{I}$ be the collection of all intervals in $\mathbb{R}$, and for any $I \in \mathcal{I}$, let $\psi(I) = P(Z \in I)$, where Z is a $\mathcal{N}(0,1)$ random variable.

Theorem 3. *Let $G : \mathbb{R}^n \mapsto \mathbb{R}^p$ be a twice continuously differentiable function, and let Γ_F be as just defined. Suppose that $\Gamma_{F_{\langle\phi\rangle}}$ is nonsingular, and that there exists a sequence $(M_n)_{n\geq 1}$ such that $k_n = o(M_n)$ and*

$$B_{F_{\langle\phi\rangle}}^2(M_n) = o\left(\left\|\Gamma_{F_{\langle\phi\rangle}}^{-1}\right\|^{-1}\right). \tag{11}$$

Suppose further that the conditions of either Theorem 1 or Theorem 2 are verified. Then, for any $b \in \mathbb{R}^p$,

$$E\left[\sup_{I\in\mathcal{I}}\left|W\left(\frac{b^T\left(G(F) - G(Y_{\langle\phi\rangle})\right)}{\sqrt{b^T\Gamma_{F_{\langle\phi\rangle}} b}} \in I \,\middle|\, Y\right) - \psi(I)\right|\right] \to 0 \text{ as } n \to \infty.$$

Under the same conditions,

$$\sup_{I\in\mathcal{I}}\left|P\left(\frac{b^T\left(G(Y_{\langle\phi\rangle}) - G(F_{\langle\phi\rangle})\right)}{\sqrt{b^T\Gamma_{F_{\langle\phi\rangle}} b}} \in I\right) - \psi(I)\right| \to 0 \text{ as } n \to \infty. \tag{12}$$

$\sup_{I\in\mathcal{I}} |Q(I) - Q'(I)|$ is the Levy-Prokhorov distance between two distributions Q and Q' on $\mathbb{R}$. The Levy-Prokhorov distance metricizes the convergence in distribution. So, when $p = 1$ (12) says that the Levy-Prokhorov distance between the distribution of $\frac{b^T\left(G(Y_{\langle\phi\rangle}) - G(F_{\langle\phi\rangle})\right)}{\sqrt{b^T\Gamma_{F_{\langle\phi\rangle}} b}}$ and $\mathcal{N}(0,1)$ goes to 0 in mean.

An application of Theorem 3 is given in subsection 5.2, in the context of the regression on Fourier's basis. The proof is delayed to subsection 6.3.

5 Applications

We present now the three applications announced in section 2. The models studied and the collections of regressors used have been defined there.

5.1 The Gaussian sequence model

We consider the model (2). Here the MLE is only the projection $\theta_Y = (Y_j)_{1\le j\le k_n}$.

The nonparametric case corresponds to the estimation of θ^0. Under the assumption that θ^0 is in some regularity class, we obtain a Bernstein-von Mises Theorem with the posterior convergence rate already obtained in previous works. On the contrary, for some priors known to achieve this rate, the centering point and the asymptotic variance of the posterior distribution do not fit with the ones expected in a Bernstein-von Mises Theorem. We also look at the semiparametric estimation of the squared ℓ^2 norm of θ^0.

5.1.1 The nonparametric estimation of θ^0

Proposition 1. *Suppose that $\sum_{j=1}^{k_n} (\theta_j^0)^2$ is bounded. This is verified in particular when θ^0 is an element of $\ell^2(\mathbb{N})$ non depending on n. With a prior $\widetilde{W} = \mathcal{N}\left(0, \tau_n^2 I_{k_n}\right)$ such that $n^{-1/4} = o(\tau_n)$, we have whatever $k_n \le n$,*

$$E\left\|\widetilde{W}(d\theta|Y) - \mathcal{N}\left(\theta_Y, \frac{1}{n} I_{k_n}\right)\right\|_{\mathrm{TV}} \to 0 \text{ as } n \to \infty.$$

and the convergence rate of θ towards *θ_0 is $\sqrt{\frac{k_n}{n}}$: for every $\lambda_n \to \infty$,*

$$E\left[\widetilde{W}\left(\|\theta - \theta_0\| \ge \lambda_n \sqrt{\frac{k_n}{n}} \,\middle|\, Y\right)\right] \to 0.$$

Proof. The beginning is an immediate corollary of Theorem 1. For the convergence rate, let $\lambda_n \to \infty$. Since $\theta_Y - \theta_0 \sim \mathcal{N}\left(0, \frac{1}{n} I_{k_n}\right)$,

$$P\left(\|\theta_Y - \theta_0\| \ge \frac{\lambda_n}{2}\sqrt{\frac{k_n}{n}}\right) \to 0.$$

In the same way

$$\begin{aligned} E\left[\widetilde{W}\left(\|\theta - \theta_Y\| \ge \frac{\lambda_n}{2}\sqrt{\frac{k_n}{n}}\right)\right] &\le E\left\|\widetilde{W}(d\theta|Y) - \mathcal{N}\left(\theta_Y, \frac{1}{n} I_{k_n}\right)\right\|_{\mathrm{TV}} \\ &\quad + \mathcal{N}\left(0, \frac{1}{n} I_{k_n}\right)\left(\left\{h : \|h\| \le \frac{\lambda_n}{2}\sqrt{\frac{k_n}{n}}\right\}\right) \\ &\to 0. \end{aligned}$$

Therefore

$$E\left[\widetilde{W}\left(\|\theta - \theta_0\| \ge \lambda_n \sqrt{\frac{k_n}{n}}\right)\right] \to 0.$$

□

However in such a general setting we have no information about the bias between θ^0 and its projection θ_0. Several authors add the assumption that the true parameter belongs to a Sobolev class of regularity $\alpha > 0$, defined by the relation $\sum_{j=1}^{\infty} |\theta_j^0|^2 j^{2\alpha} < \infty$. In this setting we show that for some priors the induced posterior may achieve the nonparametric convergence rate but with a centering point and a variance different from what is expected in the Bernstein-von Mises Theorem. Then we exhibit priors for which both the Bernstein-von Mises Theorem and the nonparametric convergence rate hold.

From now on, we suppose that $\sum_{j=1}^{\infty} |\theta_j^0|^2 j^{2\alpha} < \infty$. In this setting [10, §7.6] considers a prior $\widetilde{W}$ such that θ_1, θ_2,... are independent, and θ_j is normally distributed with variance σ^2_{j,k_n}. Further, the variances are supposed to verify

$$c/k_n \leq \min\{\sigma^2_{j,k_n} j^{2\alpha} : 1 \leq j \leq k_n\} \leq C/k_n \tag{13}$$

for some positive constants c and C. Suppose that $\alpha \geq 1/2$ and there exists constants C_1 and C_2 such that $C_1 n^{1/(1+2\alpha)} \leq k_n \leq C_2 n^{1/(1+2\alpha)}$. Then [10, Theorem 11] proved that the posterior converges at the rate $n^{-\alpha/(1+2\alpha)}$.

In order to get $n^{-1} I_{k_n}$ as asymptotic variance, we need more stringent conditions on k_n, or a flatter prior. As a counterexample consider, for $k_n \approx n^{1/(1+2\alpha)}$, the following choices of σ_{j,k_n}:

$$\sigma^2_{j,k_n} = \begin{cases} k_n^{-1} & \text{if } 1 \leq j \leq k_n/2, \\ 2^{2\alpha}/n & \text{if } j > k_n/2. \end{cases}$$

Then $\min\{\sigma^2_{j,k_n} j^{2\alpha} : 1 \leq j \leq k_n\} \approx k_n^{-1}$, and [10, Theorem 11] applies.
In this case we can perform an explicit calculus of the posterior distribution, similar to the one made in the proof of Theorem 1. The coordinates are independent, and

$$\widetilde{W}(d\theta_j|Y) = \mathcal{N}\left(\frac{\sigma^2_{j,k_n}}{\sigma_n^2 + \sigma^2_{j,k_n}} Y_j, \frac{\sigma_n^2 \sigma^2_{j,k_n}}{\sigma_n^2 + \sigma^2_{j,k_n}}\right).$$

For $j > k_n/2$, $\frac{\sigma^2_{j,k_n}}{\sigma_n^2+\sigma^2_{j,k_n}} = \frac{4^\alpha}{1+4^\alpha}$, and therefore $\left\|\widetilde{W}(d\theta_j|Y) - \mathcal{N}\left(Y_j, \sigma_n^2\right)\right\|_{\mathrm{TV}}$ is bounded away from 0.

On the contrary with an isotropic prior, flat in all directions, we obtain the centering point and the asymptotic variance we expected, and the same convergence rate as previously.

Proposition 2. *Suppose that θ^0 belongs to the Sobolev class of regularity $\alpha > 0$. Choose a prior $\widetilde{W} = \mathcal{N}\left(0, \tau_n^2 I_{k_n}\right)$ such that $n^{-1/4} = o(\tau_n)$, which insures the asymptotic normality of the posterior distribution as in Proposition 1.*
If further $k_n \approx n^{1/(1+2\alpha)}$, then the convergence rate of θ towards θ_0 and towards θ^0 is $n^{-\alpha/(1+2\alpha)}$: for every $\lambda_n \to \infty$,

$$E\left[\widetilde{W}\left(\|\theta - \theta^0\| \geq \lambda_n n^{-\alpha/(1+2\alpha)} \middle| Y\right)\right] \to 0.$$

Proof. We consider θ and θ_0 as elements of $\ell^2(\mathbb{N})$ by setting $\theta_j = \theta_{0,j} = 0$ for $j \geq k_n + 1$. The convergence rate towards θ_0 has already been established in Proposition 1. Since $\theta_{0,j} = \theta_j^0$ for $1 \leq j \leq k_n$, $\|\theta^0 - \theta_0\| \leq k_n^{-\alpha} \sqrt{\sum_{j=k_n+1}^{\infty} (\theta_j^0)^2 j^{2\alpha}} = O\left(k_n^{-\alpha}\right)$. Therefore the convergence rate of θ towards θ^0 is also $n^{-\alpha/(1+2\alpha)}$. □

5.1.2 Semiparametric theorem for the ℓ^2 norm of θ^0

We still consider the same prior distribution as before, but now we look at the posterior distribution of $\|\theta\|^2$. To get the asymptotic normality with variance $n^{-1/2}$, we just need $k_n = o(\sqrt{n})$. To control the bias term we need $\alpha > 1/2$, and in this case we get an adaptive Bayesian estimator.

Proposition 3. *Let $\alpha > 1/2$ and suppose that θ^0 belongs to the Sobolev class of regularity α. Choose a prior $\widetilde{W} = \mathcal{N}\left(0, \tau_n^2 I_{k_n}\right)$ such that $n^{-1/4} = o(\tau_n)$. Then, for any choice of k_n such that $k_n = o(\sqrt{n})$ and $\sqrt{n} = o(k_n^{2\alpha})$,*

$$E\left[\sup_{I\in\mathcal{I}}\left|\widetilde{W}\left(\frac{\sqrt{n}\left(\|\theta\|^2 - \|\theta_Y\|^2\right)}{2\|\theta^0\|} \in I \,\middle|\, Y\right) - \psi(I)\right|\right] \to 0 \text{ as } n \to \infty$$

and $\dfrac{\sqrt{n}\left(\|\theta_Y\|^2 - \|\theta_0\|^2\right)}{2\|\theta^0\|} \to \mathcal{N}(0,1)$ *in distribution, as $n \to \infty$. Further, the bias is negligible with respect to the square root of the variance:*

$$\frac{\sqrt{n}\left(\|\theta_0\|^2 - \|\theta^0\|^2\right)}{2\|\theta^0\|} = o(1).$$

In particular the choice $k_n = \sqrt{n/\ln n}$ is adaptive in α.

Proof. The conditions of Theorem 1 are fulfilled, as in Proposition 1.

Here $G(\theta) = \theta^T\theta$, $\dot{G}_\theta = 2\theta^T$ and $\ddot{G}_\theta = 2I_{k_n}$. Therefore $B_{\theta_0}(M_n) = 2M_n/n$, while $\Gamma_{\theta_0} = 4\|\theta_0\|^2/n$.

Let us choose $(M_n)_{n\geq 1}$ such that $k_n = o(M_n)$ and $M_n = o(\sqrt{n})$. Such sequences exist and fulfill the conditions of Theorem 3.

Since $\|\theta_0\|^2 \to \|\theta^0\|^2$, we can substitute the variance Γ_{θ_0} by $4\|\theta^0\|^2/n$ and get the two asymptotic normality results.

Eventually $\|\theta^0\|^2 - \|\theta_0\|^2 = \|\theta^0 - \theta_0\|^2 = O\left(k_n^{-2\alpha}\right)$, as in the proof of Proposition 2. If $\sqrt{n} = o(k_n^{2\alpha})$, we get $\sqrt{n}\left(\|\theta_0\|^2 - \|\theta^0\|^2\right) = o(1)$. □

5.2 Regression on Fourier's basis

Now we consider the regression model (3) with a function f in a Sobolev class $\mathcal{W}(\alpha, L)$, and use Fourier's basis (4). For any $\theta \in \mathbb{R}^{k_n}$, we define $f_\theta = \sum_{j=1}^{k_n}\theta_j\varphi_j$. We also denote by $\theta^0 \in \ell^2(\mathbb{N})$ the sequence of Fourier's coefficients of f: $f = \sum_{j=1}^{\infty}\theta_j^0\varphi_j$.

The following useful Lemma about our collection of regressors can be found for instance in [17] (we slightly modified it to take into account the case n even):

Lemma 1. *Suppose either that n is odd and $k_n \leq n$, or n is even and $k_n \leq n-1$. Consider the collection $(\phi_j)_{1\leq j\leq k_n}$ defined before, and Φ the associated matrix. Then*

$$\Phi^T\Phi = nI_{k_n}.$$

This makes the regression on Fourier's basis very close to the Gaussian sequence model, and the result we obtain are similar.

We consider first the nonparametric estimation of f in a Sobolev class, for which we get a Bernstein-von Mises Theorem and the frequentist minimax $n^{-\alpha/(1+2\alpha)}$ posterior convergence rate for the L^2 norm.

Then we consider two semiparametric settings: the estimation of a linear functional of f, and the estimation of the L^2 norm of f. We get the adaptive $\sqrt{n}$ convergence rate for any $\alpha > 1/2$.

5.2.1 Nonparametric Bernstein-von Mises Theorem in Sobolev classes

Proposition 4. *Suppose that f belongs to some Sobolev class $\mathcal{W}(\alpha, L)$ for $L > 0$ and $\alpha > 1/2$. Let $k_n \approx n^{1/(1+2\alpha)}$ and $\widetilde{W} = \mathcal{N}(0, \gamma_n I_{k_n})$ be the prior on θ, for a sequence $(\gamma_n)_{n\geq 1}$ such that $1/\sqrt{n} = o(\gamma_n)$. Then*

$$E\left\|\widetilde{W}(d\theta|Y) - \mathcal{N}\left(\theta_Y, \frac{\sigma^2}{n} I_{k_n}\right)\right\|_{\mathrm{TV}} \to 0 \text{ as } n \to \infty$$

and the convergence rate relative to the euclidean norm for f_θ is $n^{-\alpha/(1+2\alpha)}$: for every $\lambda_n \to \infty$,

$$E\left[\widetilde{W}\left(\|f_\theta - f\| \geq \lambda_n n^{-\alpha/(1+2\alpha)} \middle| Y\right)\right] \to 0.$$

Proof. The conditions of Theorem 1 are fulfilled: with $\tau_n^2 = n\gamma_n$, we have $n = o(\tau_n^4)$. The first assertion follows.

Because of the orthogonal nature of Fourier's basis, $\|f_\theta - f\| = \|\theta - \theta^0\|$ in $\ell^2(\mathbb{N})$. We use the decomposition $\|\theta - \theta^0\|^2 \leq \|\theta - \theta_0\|^2 + \|\theta_0 - \theta^0\|^2$. In the same way as in the proof of Proposition 1, for any $\lambda_n \to \infty$,

$$E\left[\widetilde{W}\left(\|\theta - \theta_0\| \geq \lambda_n \sqrt{\frac{k_n}{n}}\right)\right] \to 0.$$

Going back to Definition 1, we have

$$\|\theta_0 - \theta^0\|^2 = \sum_{j=k_n+1}^{\infty} (\theta_j^0)^2 \leq k_n^{-2\alpha} \sum_{j=k_n+1}^{\infty} a_j^{2\alpha} (\theta_j^0)^2 = O(k_n^{-2\alpha}).$$

This permits to get

$$E\left[\widetilde{W}\left(\|\theta - \theta^0\| \geq \lambda_n n^{-\alpha/(1+2\alpha)} \middle| Y\right)\right] \to 0.$$

□

5.2.2 Linear functionals of f

Let $g : [0,1] \to \mathbb{R}$ be a function in $\mathbb{L}^2([0,1])$. We want to estimate $\mathcal{F}(f) = \int_0^1 fg$, and we approximate it by

$$\frac{1}{n}\sum_{i=1}^{n} g(i/n) f(i/n) = GF_0$$

where $G = (g(i/n)/n)_{1\leq i\leq n}^T$. The plug-in MLE estimator of GF_0 in the misspecified model $\langle\phi\rangle$ is $GY_{\langle\phi\rangle}$. More generally, we consider the functional $F \mapsto GF$.

The following result is adaptive, in the sense that the same choice $k_n = \lfloor n/\ln n \rfloor$ entails the convergence rate $n^{-1/2}$ for all values of $\alpha > 1/2$.

Proposition 5. *Suppose f is bounded, and let W be the prior induced by the $\mathcal{N}(0, \gamma_n\, I_{k_n})$ distribution on θ, for a sequence $(\gamma_n)_{n\geq 1}$ such that $1/\sqrt{n} = o(\gamma_n)$. Then*

1. $$E\left\|W(d(GF)|Y) - \mathcal{N}\left(GY_{\langle\phi\rangle}, \sigma^2 G\Sigma G^T\right)\right\|_{\mathrm{TV}} \to 0$$
   and the distribution of $GY_{\langle\phi\rangle}$ is $\mathcal{N}\left(GF_{\langle\phi\rangle}, \sigma^2 G\Sigma G^T\right)$.

2. *Suppose further that f and g belong to some Sobolev class $\mathcal{W}(\alpha, L)$ for $L > 0$ and $\alpha > 1/2$. Then $G\Sigma G^T \sim \frac{1}{n}\int_0^1 g^2$,*
   $$E\left\|W\left(d\frac{\sqrt{n}(GF - GY_{\langle\phi\rangle})}{\sigma\sqrt{\int_0^1 g^2}}\middle| Y\right) - \mathcal{N}(0,1)\right\|_{\mathrm{TV}} \to 0,$$
   and $\dfrac{\sqrt{n}(GY_{\langle\phi\rangle} - GF_{\langle\phi\rangle})}{\sigma\sqrt{\int_0^1 g^2}} \to \mathcal{N}(0,1)$ in distribution, as $n \to \infty$.

3. *Suppose that f and g belong to some Sobolev class $\mathcal{W}(\alpha, L)$ for $L > 0$ and $\alpha > 1/2$, and suppose further that k_n is large enough so that $n = o(k_n^{2\alpha})$. Then the bias is negligible with respect to the square root of the variance:*
   $$\frac{\sqrt{n}\left(GF_{\langle\phi\rangle} - \mathcal{F}(f)\right)}{\sigma\sqrt{\int_0^1 g^2}} = o(1).$$

Before the proof we give two lemmas, proved in Appendix B, about the error terms of the approximation of a Sobolev class by a sieve build on Fourier's basis, and of the approximation of an integral by a Riemann sum.

Lemma 2. *Let $\alpha > 1/2$ and $L > 0$. We suppose n odd or $k_n < n$. If $f \in \mathcal{W}(\alpha, L)$,*
$$\|F_0 - F_{\langle\phi\rangle}\| \leq (1 + o(1))\frac{\sqrt{2}L}{\pi^\alpha}\frac{\sqrt{n}}{k_n^\alpha}.$$
Further, $\|F_0\| \sim \sqrt{n\int_0^1 f^2}$ and $\|F_0 - F_{\langle\phi\rangle}\| = O(k_n^{-\alpha}\|F_0\|)$.

Lemma 3. *Let two functions $f \in \mathcal{W}(\alpha, L)$ and $g \in \mathcal{W}(\alpha', L')$ for some $\alpha, \alpha' > 1/2$ and two positive numbers L and L'. Then*
$$\left|\frac{1}{n}\sum_{i=1}^n f(i/n)g(i/n) - \int_0^1 fg\right| = O\left(n^{-\inf(\alpha,\alpha')}\right).$$

Proof of Proposition 5. 1. The first assertion is just Corollary 1. The conditions of Theorem 1 are fulfilled, as in the proof of Proposition 4.

2. If $g \in \mathcal{W}(\alpha, L)$ for $L > 0$ and $\alpha > 1/2$, $G\Sigma G^T = \|\Sigma G^T\|^2 \sim \|G^T\|^2$ by Lemma 2. In the meantime $\|G^T\|^2 = \frac{1}{n^2}\sum_{i=1}^n g^2(x_i) \sim \frac{1}{n}\int_0^1 g^2$ by Lemma 3. So $G\Sigma G^T \sim \frac{1}{n}\int_0^1 g^2$, and the variance in the formulas of Corollary 1 can be substituted with $\frac{1}{n}\int_0^1 g^2$.

3. We decompose the bias into two terms, $|GF_0 - \mathcal{F}(f)|$ and $|GF_{\langle\phi\rangle} - GF_0|$, and show that both are $o(n^{-1/2})$. The first term is controlled by Lemma 3. For the last one, $|GF_{\langle\phi\rangle} - GF_0| \leq \|G^T\|\,\|F_{\langle\phi\rangle} - F_0\|$. $\|G^T\| = O(n^{-1/2})$, $\|F_{\langle\phi\rangle} - F_0\| = O(k_n^{-\alpha}\|F_0\|)$ by Lemma 2, and $\|F_0\| = O(\sqrt{n})$. We conclude thanks to the assumption $n = o(k_n^{2\alpha})$.

□

5.2.3 L^2-norm of f

Suppose that we want to estimate $\mathcal{F}(f) = \int_0^1 f^2$. We can consider the plug-in MLE estimator

$$G(Y_{\langle\phi\rangle}) = \frac{1}{n}\left\|Y_{\langle\phi\rangle}\right\|^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{k_n}\theta_{Y,j}\varphi_j(i/n)\right)^2.$$

More generally we define, for any $F \in \mathbb{R}^n$,

$$G(F) = \frac{1}{n}\|F\|^2. \tag{14}$$

With a Gaussian prior, we obtain the following result, which is also adaptive: the same $k_n = \lfloor\sqrt{n}/\ln n\rfloor$ is suitable whatever $\alpha > 1/2$.

Proposition 6. *Let $G(F) = \|F\|^2/n$. Suppose that $f \in \mathcal{W}(\alpha, L)$ for some $L > 0$ and $\alpha > 1/2$. Let W be the prior induced by the $\mathcal{N}(0, \gamma_n\, I_{k_n})$ distribution on θ, for a sequence $(\gamma_n)_{n\geq 1}$ such that $1/\sqrt{n} = o(\gamma_n)$. The sequence $(k_n)_{n\geq 1}$ can be chosen such that $k_n = o(\sqrt{n})$ and $\sqrt{n} = o(k_n^{2\alpha})$, and with such a choice,*

$$E\left[\sup_{I\in\mathcal{I}}\left|W\left(\frac{\sqrt{n}\left(G(F) - G(Y_{\langle\phi\rangle})\right)}{2\sigma\sqrt{\mathcal{F}(f)}} \in I \,\middle|\, Y\right) - \psi(I)\right|\right] \to 0 \text{ as } n \to \infty$$

and $\dfrac{\sqrt{n}\left(G(Y_{\langle\phi\rangle}) - G(F_{\langle\phi\rangle})\right)}{2\sigma\sqrt{\mathcal{F}(f)}} \to \mathcal{N}(0,1)$ *in distribution, as $n \to \infty$. Further, the bias is negligible with respect to the square root of the variance:*

$$\frac{\sqrt{n}\left(G(F_{\langle\phi\rangle}) - \mathcal{F}(f)\right)}{2\sigma\sqrt{\mathcal{F}(f)}} = o(1).$$

A similar corollary can be stated for a non-Gaussian prior.

Proof. First, let us note that the conditions of Theorem 1 are fulfilled, as in the proof of Proposition 4. Lemma 10 in Appendix B insures that f is bounded.

In this setting $\dot{G}_F = (2/n)\,F^T$ and $D_F^2 G(h,h) = (2/n)\,\|h\|^2$ for any $F \in \mathbb{R}^n$ and any $h \in \mathbb{R}^n$. Therefore $B_F(a) = 2\sigma^2 a/n$, and $\Gamma_F = 4(\sigma^2/n^2)\,\|F\|^2$. By Lemma 2, $\|F_{\langle\phi\rangle}\|^2 \sim \|F_0\|^2 \sim n\,\mathcal{F}(f)$. Thus $\Gamma_{F_{\langle\phi\rangle}} = 4(1+o(1))\mathcal{F}(f)/n$.

Let us choose $(M_n)_{n\geq 1}$ such that $k_n = o(M_n)$ and $M_n = o(\sqrt{n})$. Such sequences exist and fulfill the conditions of Theorem 3. We can substitute the variance $\Gamma_{F_{\langle\phi\rangle}}$ by $4\mathcal{F}(f)/n$ and get the two asymptotic normality results.

Let us now consider the bias term.

$$\mathcal{F}(f) - G(F_{\langle\phi\rangle}) \leq \frac{\|F_0\|^2 - \|F_{\langle\phi\rangle}\|^2}{n} + \left(\int_0^1 f^2 - \frac{1}{n}\sum_{i=1}^{n} f^2(i/n)\right)$$

We use Lemma 2 to control $\|F_0\|^2 - \|F_{\langle\phi\rangle}\|^2$, and Lemma 3 for the other term:

$$\left|\mathcal{F}(f) - G(F_{\langle\phi\rangle})\right| = O\left(k_n^{-2\alpha}\right) + O\left(n^{-\alpha}\right).$$

This is a $o(1/\sqrt{n})$ under the assumptions of Corollary 6. □

5.3 Regression on splines

Here we consider the regression model for functions in $C^\alpha[0,1]$ with $\alpha > 0$, using splines. The problem has been set in section 2. We first develop further the framework and the assumptions used here, and recall the previous result of [10, §7.7.1] which obtains the posterior concentration at the frequentist minimax rate. Then we present two Bernstein-von Mises Theorems: the first one with the same prior as [10] but a stronger condition on k_n (or equivalently on α); the second one with a flatter prior, for which we retrieve the minimax convergence rate in addition to the asymptotic Gaussianity of the posterior distribution.

For any $\theta \in \mathbb{R}^{k_n}$, we define $f_\theta = \sum_{j=1}^{k_n} \theta_j B_j$. The B-splines basis has the following approximation property: for any $\alpha > 0$, there exist $C_\alpha > 0$ such that, if $f \in C^\alpha[0,1]$, there exists $\theta^\infty \in \mathbb{R}^{k_n}$ verifying

$$\|f - f_{\theta^\infty}\|_\infty \leq C_\alpha k_n^{-\alpha} \|f\|_\alpha. \tag{15}$$

We need the design $\left(x_i^{(n)}\right)_{n\geq 1, 1\leq i\leq n}$ to be sufficiently regular but, as stressed in [10], the spacial separation property of B-splines permits to express the precise condition in terms of the covariance matrix $\Phi^T\Phi$. We suppose that there exist positive constants C_1 and C_2 such that, as n increases, whatever $\theta \in \mathbb{R}^{k_n}$,

$$C_1 \frac{n}{k_n}\|\theta\|^2 \leq \theta^T\Phi^T\Phi\theta \leq C_2 \frac{n}{k_n}\|\theta\|^2. \tag{16}$$

A norm $\|f\|_n = \sqrt{\frac{1}{n}\sum_{i=1}^n |f(x_i)|^2}$ is associated to the design. Note that $\sqrt{n}\|f_\theta\|_n = \|\Phi\theta\|$ if $\theta \in \mathbb{R}^{k_n}$. Under condition (16) we have a relation between $\|\cdot\|_n$ and the euclidean norm on the parameter space: for every θ_1 and θ_2

$$C_1\|\theta_1 - \theta_2\| \leq \sqrt{k_n}\,\|f_{\theta_1} - f_{\theta_2}\|_n \leq C_2\|\theta_1 - \theta_2\|.$$

With these conditions [10, Theorem 12] gets the posterior concentration at the minimax rate. Take $\alpha \geq 1/2$, let $\widetilde{W} = \mathcal{N}(0, I_{k_n})$ be the prior on the spline coefficients, and suppose there exists constants C_3 and C_4 such that $C_3 n^{1/(1+2\alpha)} \leq k_n \leq C_4 n^{1/(1+2\alpha)}$. Then the posterior concentrates at the minimax rate $n^{-\alpha/(1+2\alpha)}$ relative to $\|\cdot\|_n$: for every $\lambda_n \to \infty$,

$$E\left[\widetilde{W}\left(\|f_\theta - f\|_n \geq \lambda_n n^{-\alpha/(1+2\alpha)}\,\Big|\, Y\right)\right] \to 0.$$

This is equivalent to a convergence rate $n^{\frac{1-2\alpha}{2(1+2\alpha)}}$ relative to the euclidean norm for θ:

$$E\left[\widetilde{W}\left(\|\theta - \theta_0\| \geq \lambda_n n^{\frac{1-2\alpha}{2(1+2\alpha)}}\,\Big|\, Y\right)\right] \to 0.$$

Indeed (15) and the projection property entail

$$\|f_{\theta_0} - f\|_n \leq \|f_{\theta^\infty} - f\|_n \leq \|f_{\theta^\infty} - f\|_\infty \leq C_\alpha\|f\|_\alpha k_n^{-\alpha}.$$

With modified assumptions we get also the Bernstein-von Mises Theorem in two different settings. First, with the same prior as [10]:

Proposition 7. *Assume that f is bounded, $k_n = o\left(\left(\frac{n}{\ln n}\right)^{1/3}\right)$, and (16) holds. Let $\widetilde{W} = \mathcal{N}(0, I_{k_n})$ be the prior on the spline coefficients. Then*

$$E\left\|\widetilde{W}(d\theta|Y) - \mathcal{N}\left(\theta_Y, \sigma^2(\Phi^T\Phi)^{-1}\right)\right\|_{\mathrm{TV}} \to 0 \text{ as } n \to \infty$$

and the convergence rate relative to the euclidean norm for θ is $\frac{k_n}{\sqrt{n}}$.

We need $\alpha > 1$ to get the Gaussian shape with the same convergence rate as in [10]. The conditions of Proposition 7 are verified in particular if there exists constants C_3 and C_4 such that $C_3 n^{1/(1+2\alpha)} \leq k_n \leq C_4 n^{1/(1+2\alpha)}$. In this case the convergence rate for θ is $n^{\frac{1-2\alpha}{2(1+2\alpha)}}$.

Proof. We apply Theorem 2. We can choose M_n such that $k_n \ln n = o(M_n)$ and $M_n = o\left(\frac{n}{k_n^2}\right)$. Assumption 2 is then trivially verified.

From (16) we get $\left\|\Phi^T\Phi\right\| \leq C_2 \frac{n}{k_n}$ and $\left\|(\Phi^T\Phi)^{-1}\right\| \leq C_1^{-1}\frac{k_n}{n}$. We have also $\ln\det(\Phi^T\Phi) \leq k_n \ln C_2 + k_n \ln\left(\frac{n}{k_n}\right) = O(k_n \ln n) = o(M_n)$. Since $\theta_0 = \Phi(\Phi^T\Phi)^{-1}F_0$,

$$\|\theta_0\|^2 \leq \frac{k_n}{C_1 n}\|F_0\|^2 \leq \frac{\|f\|_\infty}{C_1}k_n.$$

Therefore $-\ln w(\theta_0) = O(1) + \frac{1}{2}\|\theta_0\|^2 = O(k_n) = o(M_n)$, and assumption 3 holds.

Let $h \in \mathbb{R}^{k_n}$ such that $h^T\Phi^T\Phi h \leq \sigma^2 M_n$. We have $\|h\|^2 \leq \left\|(\Phi^T\Phi)^{-1}\right\| \|\Phi h\|^2 \leq \frac{\sigma^2 k_n M_n}{C_1 n} = o\left(k_n^{-1}\right)$. Therefore

$$\sup_{h^T\Phi^T\Phi h \leq \sigma^2 M_n}\left|\ln\frac{w(\theta_0+h)}{w(\theta_0)}\right| \leq \sup_{h^T\Phi^T\Phi h \leq \sigma^2 M_n}\frac{\|h\|^2 + 2\|h\|\|\theta_0\|}{2} = o(1) \quad (17)$$

and assumption 1 follows.

Let us now prove the convergence rate. Let $\lambda_n \to \infty$. Then

$$P\left(\|\theta_Y - \theta_0\| \geq \frac{\lambda_n k_n}{2\sqrt{n}}\right) \leq P\left(\|\Phi(\theta_Y - \theta_0)\|^2 \geq \frac{C_1\lambda_n^2 k_n}{4}\right) \to 0$$

since $\|\Phi(\theta_Y - \theta_0)\|^2 \sim \sigma^2\chi^2(k_n)$. In the same way

$$\begin{aligned} E\left[\widetilde{W}\left(\|\theta - \theta_Y\| \geq \frac{\lambda_n k_n}{2\sqrt{n}}\right)\right] &\leq E\left\|\widetilde{W}(d\theta|Y) - \mathcal{N}\left(\theta_Y, \sigma^2(\Phi^T\Phi)^{-1}\right)\right\|_{\mathrm{TV}} \\ &\quad + \mathcal{N}\left(0, \sigma^2(\Phi^T\Phi)^{-1}\right)\left(\left\{h:\ \|h\| \leq \frac{\lambda_n k_n}{2\sqrt{n}}\right\}\right) \\ &\to 0. \end{aligned}$$

Therefore

$$E\left[\widetilde{W}\left(\|\theta - \theta_0\| \geq \frac{\lambda_n k_n}{\sqrt{n}}\right)\right] \to 0.$$

□

The situation is similar to the one we encountered with the Gaussian sequence model. To get the Bernstein-von Mises Theorem with the same convergence rate as [10] for $\alpha \leq 1$, we need a flatter prior:

Proposition 8. *Assume that f is bounded and (16) holds. Let $\widetilde{W} = \mathcal{N}\left(0, \tau_n^2 I_{k_n}\right)$ be the prior on the spline coefficients, with the sequence τ_n verifying*

$$\frac{k_n^2 \ln n}{n} = o(\tau_n^2) \qquad \text{and} \qquad \frac{k_n^3 \ln n}{n} = o(\tau_n^4).$$

Then

$$E\left\|\widetilde{W}(d\theta|Y) - \mathcal{N}\left(\theta_Y, \sigma^2(\Phi^T\Phi)^{-1}\right)\right\|_{\mathrm{TV}} \to 0 \text{ as } n \to \infty$$

and the convergence rate relative to the euclidean norm for θ is $\frac{k_n}{\sqrt{n}}$.

When $\alpha > 0$ and k_n is of order $n^{1/(1+2\alpha)}$, the conditions reduce to $n^{\frac{2-2\alpha}{1+2\alpha}} \ln n = o(\tau_n^4)$. So we retrieve the convergence rate of [10] in addition to the Gaussian shape with the same k_n, even for $\alpha \leq 1$, but with a different prior.

Proof. The proof is essentially the same as for Proposition 7. M_n can be chosen such as $k_n \ln n = o(M_n)$, $M_n = o\left(\frac{n\tau_n^2}{k_n}\right)$, and $M_n = o\left(\frac{n\tau_n^4}{k_n^2}\right)$. These last two conditions are the ones needed to obtain the same upper bounds as in (17). □

6 Proofs

6.1 Proof of Theorem 1

In the present setting all distributions are explicit and admit densities with respect to the corresponding Lebesgue measure. We decompose any $y \in \mathbb{R}^n$ in two orthogonal components $y = \Phi\theta_y + y'$, with $\Phi^T y' = 0$. Then

$$\begin{aligned}
dP_\theta(y) &= c_1 \exp\left\{-\frac{1}{2\sigma_n^2}\left(\|\Phi\theta\|^2 + \|\Phi\theta_y\|^2 + \|y'\|^2 - 2\theta^T\Phi^T\Phi\theta_y\right)\right\} \\
d\widetilde{W}(\theta) &= c_2 \exp\left\{-\frac{1}{2\tau_n^2}\|\Phi\theta\|^2\right\} \\
dP_\theta(y)\, d\widetilde{W}(\theta) &= c_1 c_2 \exp\left\{-\frac{\sigma_n^2+\tau_n^2}{2\sigma_n^2\tau_n^2}\left\|\Phi\left(\theta - \frac{\tau_n^2}{\sigma_n^2+\tau_n^2}\theta_y\right)\right\|^2\right. \\
&\qquad \left. -\frac{1}{2(\sigma_n^2+\tau_n^2)}\|\Phi\theta_y\|^2 - \frac{1}{2\sigma_n^2}\|y'\|^2\right\}
\end{aligned}$$

where $c_1 = (2\pi)^{-n/2}\sigma_n^{-n}$ and $c2 = (2\pi)^{-k_n/2}\tau_n^{-k_n}\det(\Phi^T\Phi)^{-1}$.

Using the Bayes rule, we get the density of $\widetilde{W}(d\theta|Y)$, in which we recognize the normal distribution

$$\widetilde{W}(d\theta|Y) = \mathcal{N}\left(\frac{\tau_n^2}{\sigma_n^2+\tau_n^2}\theta_Y, \frac{\sigma_n^2\tau_n^2}{\sigma_n^2+\tau_n^2}(\Phi^T\Phi)^{-1}\right). \tag{18}$$

At that point, we have got an exact expression of $\widetilde{W}(d\theta|Y)$, but nor the centering nor the variance correspond to the limit distribution given in Theorem 1. Therefore we make use of the triangle inequality, with intermediate distribution $Q = \mathcal{N}\left(\frac{\tau_n^2}{\sigma_n^2+\tau_n^2}\theta_Y, \sigma_n^2(\Phi^T\Phi)^{-1}\right)$. We first deal with the change in the variance.

Let $\alpha_n = \frac{\tau_n}{\sigma_n}\sqrt{\ln\left(1+\frac{\sigma_n^2}{\tau_n^2}\right)}$, and f and g be respectively the density functions of $\mathcal{N}(0, I_{k_n})$ and $\mathcal{N}\left(0, \frac{\tau_n^2}{\sigma_n^2+\tau_n^2} I_{k_n}\right)$. Let U be a random variable following the chi-square distribution with k_n degrees of freedom $\chi^2(k_n)$. Then

$$\begin{aligned}
\left\|\widetilde{W}(d\theta|Y) - Q\right\|_{\mathrm{TV}} &= \left\|\mathcal{N}(0, I_{k_n}) - \mathcal{N}\left(0, \frac{\tau_n^2}{\sigma_n^2+\tau_n^2} I_{k_n}\right)\right\|_{\mathrm{TV}} \\
&= \int_{\mathbb{R}^{k_n}} (g-f)_+ = \int_{\|x\|\le\sqrt{k_n}\alpha_n} g(x) - f(x)\, d^n x \\
&= P(U \le k_n\alpha_n^2) - P\left(U \le \frac{\sigma_n^2+\tau_n^2}{\tau_n^2} k_n \alpha_n^2\right) \\
&= P\left(\alpha_n^2 \le \frac{U}{k_n} \le \frac{\sigma_n^2+\tau_n^2}{\tau_n^2}\alpha_n^2\right).
\end{aligned}$$

As n goes to infinity, $\frac{U}{k_n}$ converges towards $\mathcal{N}(0,1)$ in distribution. Since $\sigma_n = o(\tau_n)$, both $\frac{\sigma_n^2+\tau_n^2}{\tau_n^2}$ and α_n go to 1. As a consequence, $\left\|\widetilde{W}(d\theta|Y) - Q\right\|_{\mathrm{TV}}$ goes to zero as n goes to infinity.

Let us now deal with the centering term.

Lemma 4. *Let U be a standard normal random variable, let $k \ge 1$, and let $Z \in \mathbb{R}^k$. Then*

$$\|\mathcal{N}(0, I_k) - \mathcal{N}(Z, I_k)\|_{\mathrm{TV}} = P(|U| \le \|Z\|/2) \le \|Z\|/\sqrt{2\pi}.$$

Proof. Let g be the density of $\mathcal{N}(0, I_k)$. Then

$$\begin{aligned}
\|\mathcal{N}(0, I_k) - \mathcal{N}(Z, I_k)\|_{\mathrm{TV}} &= \int_{\mathbb{R}^k} (g(x) - g(x-Z))_+\, d^k x \\
&= \int_{\{2x^TZ \le \|Z\|^2\}} (g(x) - g(x-Z))\, d^k x \\
&= P(U \le \|Z\|/2) - P(U + \|Z\| \le \|Z\|/2) \\
&\le \|Z\|/\sqrt{2\pi}.
\end{aligned}$$

The last line comes from the density of $\mathcal{N}(0,1)$ being bounded by $1/\sqrt{2\pi}$. ☐

Let $\sqrt{\Phi^T\Phi}$ be a square root of the matrix $\Phi^T\Phi$. Then

$$\begin{aligned}
\left\|\mathcal{N}\left(\theta_Y, \sigma_n^2(\Phi^T\Phi)^{-1}\right) - Q\right\|_{\mathrm{TV}} &= \left\|\mathcal{N}(0, I_{k_n}) - \mathcal{N}\left(\frac{\sigma_n}{\tau_n^2+\sigma_n^2}\sqrt{\Phi^T\Phi}\theta_Y, I_{k_n}\right)\right\|_{\mathrm{TV}} \\
&\le \frac{1}{\sqrt{2\pi}} \frac{\sigma_n}{(\tau_n^2+\sigma_n^2)} \|\Phi\theta_Y\| \\
&\le \frac{1}{\sqrt{2\pi}} \frac{\sigma_n}{(\tau_n^2+\sigma_n^2)} \left(\|F_0\| + \sqrt{\varepsilon^T\Sigma\varepsilon}\right).
\end{aligned}$$

$\varepsilon^T\Sigma\varepsilon$ is a random variable following $\sigma_n^2\chi^2(k_n)$ distribution. Therefore

$$E\left\|\mathcal{N}\left(\theta_Y, \sigma_n^2(\Phi^T\Phi)^{-1}\right) - Q\right\|_{\mathrm{TV}} \le \frac{1}{\sqrt{2\pi}} \frac{\sigma_n}{\tau_n^2+\sigma_n^2}\left(\|F_0\| + \sigma_n\sqrt{k_n}\right).$$

This goes to zero under the assumptions of Theorem 1.

To conclude the proof, let us just note that we deduce the results on $W(dF|Y)$ from the ones on $\widetilde{W}(d\theta|Y)$, by the linear relation $F = \Phi\theta$.

6.2 Proof of Theorem 2.

We make the proof for $\widetilde{W}(d\theta|Y)$. Then the result for $W(dF|Y)$ is immediate. Our method is adapted from [2].

For $M > 0$, consider the ellipsoid

$$\mathcal{E}_{\theta_0,\Phi}(M) = \left\{\theta \in \mathbb{R}^{k_n} : (\theta-\theta_0)^T\Phi^T\Phi(\theta-\theta_0) \le \sigma_n^2 M\right\}. \tag{19}$$

To any probability measure P on $\mathbb{R}^{k_n}$, we associate the probability

$$P^M = \frac{P\left(\cdot \cap \mathcal{E}_{\theta_0,\Phi}(M)\right)}{P\left(\mathcal{E}_{\theta_0,\Phi}(M)\right)} \tag{20}$$

with support in $\mathcal{E}_{\theta_0,\Phi}(M)$. It can be easily checked that

$$\left\|P - P^M\right\|_{\mathrm{TV}} = P\left(\mathcal{E}^c_{\theta_0,\Phi}(M)\right). \tag{21}$$

Then the calculus is divided in three parts, M_n being used as a threshold to truncate the queues of the probability distributions. Gathered, these lemmas give Theorem 2.

Lemma 5. *If $k_n < 4M_n$, then*

$$E\left\|\mathcal{N}\left(\theta_Y, \sigma_n^2(\Phi^T\Phi)^{-1}\right) - \mathcal{N}^{M_n}\left(\theta_Y, \sigma_n^2(\Phi^T\Phi)^{-1}\right)\right\|_{\mathrm{TV}} \le 2e^{-\frac{\left(\sqrt{M_n}-2\sqrt{k_n}\right)^2}{8}}.$$

If $k_n = o(M_n)$, for n large enough, this bound can be replaced by $\exp(-M_n/9)$.

Proof. Two cases occur, depending on whether θ_Y is near or far from θ_0:

$$\begin{aligned}\left\|\mathcal{N}\left(\theta_Y, \sigma_n^2(\Phi^T\Phi)^{-1}\right) - \mathcal{N}^{M_n}\left(\theta_Y, \sigma_n^2(\Phi^T\Phi)^{-1}\right)\right\|_{\mathrm{TV}} \\ &= \mathcal{N}\left(\theta_Y, \sigma_n^2(\Phi^T\Phi)^{-1}\right)\left(\mathcal{E}^c_{\theta_0,\Phi}(M_n)\right) \\ &\le \mathbb{1}_{(\theta_Y-\theta_0)^T\Phi^T\Phi\,(\theta_Y-\theta_0) > \sigma_n^2 M_n/4} \\ &\quad + \mathcal{N}\left(\theta_0, \sigma_n^2(\Phi^T\Phi)^{-1}\right)\left(\mathcal{E}^c_{\theta_0,\Phi}(M_n/4)\right)\end{aligned}$$

Let U a random variable following the $\chi^2(k_n)$ distribution. Taking the expectation in the last line we get

$$E\left\|\mathcal{N}\left(\theta_Y, \sigma_n^2(\Phi^T\Phi)^{-1}\right) - \mathcal{N}^{M_n}\left(\theta_Y, \sigma_n^2(\Phi^T\Phi)^{-1}\right)\right\|_{\mathrm{TV}} \le 2P(U > M_n/4).$$

To conclude we use Cirelson's inequality [13]:

$$P(\sqrt{U} > \sqrt{k_n} + \sqrt{2x}) \le \exp(-x) \tag{22}$$

□

Lemma 6. *If* $\sup_{h^T\Phi^T\Phi h \le \sigma_n^2 M_n,\, g^T\Phi^T\Phi g \le \sigma_n^2 M_n} \frac{w(\theta_0+h)}{w(\theta_0+g)} \to 1$ *as $n \to \infty$, then*

$$E\left\|\widetilde{W}^{M_n}(d\theta|Y) - \mathcal{N}^{M_n}\left(\theta_Y, \sigma_n^2(\Phi^T\Phi)^{-1}\right)\right\|_{\mathrm{TV}} \to 0 \text{ as } n \to \infty.$$

Proof. Let us first note that, for every θ and τ in $\mathbb{R}^{k_n}$, for every $Y \in \mathbb{R}^n$,

$$\frac{dP_\theta(Y)}{dP_\tau(Y)} = \exp\left\{\frac{-\|\Phi\theta\|^2 + \|\Phi\tau\|^2 - 2Y^T\Phi(\tau-\theta)}{2\sigma_n^2}\right\} = \frac{d\mathcal{N}\left(\theta_Y, \sigma_n^2(\Phi^T\Phi)^{-1}\right)(\theta)}{d\mathcal{N}\left(\theta_Y, \sigma_n^2(\Phi^T\Phi)^{-1}\right)(\tau)}. \tag{23}$$

In the following we mainly use the convexity of $x \mapsto (1-x)_+$. We abbreviate $\mathcal{N}^{M_n}\left(\theta_Y, \sigma_n^2(\Phi^T\Phi)^{-1}\right)$ into $\mathcal{N}^{M_n}$. Then

$$\begin{aligned}
&\left\|\widetilde{W}^{M_n}(d\theta|Y) - \mathcal{N}^{M_n}\right\|_{\mathrm{TV}} \\
&\quad= \int \left(1 - \frac{d\mathcal{N}^{M_n}(\theta)}{d\widetilde{W}^{M_n}(\theta|Y)}\right)_+ d\widetilde{W}^{M_n}(\theta|Y) \\
&\quad= \int \left(1 - \frac{d\mathcal{N}^{M_n}(\theta)\int \frac{w(\tau)}{d\mathcal{N}^{M_n}(\tau)}\, dP_\tau(Y)\, d\mathcal{N}^{M_n}(\tau)}{w(\theta)\, dP_\theta(Y)}\right)_+ d\widetilde{W}^{M_n}(\theta|Y) \\
&\quad\leq \int\int \left(1 - \frac{w(\tau)\, d\mathcal{N}^{M_n}(\theta)\, dP_\tau(Y)}{w(\theta)\, d\mathcal{N}^{M_n}(\tau)\, dP_\theta(Y)}\right)_+ d\mathcal{N}^{M_n}(\tau)\, d\widetilde{W}^{M_n}(\theta|Y) \\
&\quad= \int\int \left(1 - \frac{w(\tau)}{w(\theta)}\right)_+ d\mathcal{N}^{M_n}(\tau)\, d\widetilde{W}^{M_n}(\theta|Y) \\
&\quad\leq 1 - \inf_{h^T\Phi^T\Phi h \leq \sigma_n^2 M_n,\, g^T\Phi^T\Phi g \leq \sigma_n^2 M_n} \frac{w(\theta_0+h)}{w(\theta_0+g)}.
\end{aligned}$$

□

Proposition 9 (Posterior concentration)**.** *Suppose that Condition 1, Condition 2, and Condition 3 of Theorem 2 hold. Then*

$$\begin{aligned}
E\left\|\widetilde{W}(d\theta|Y) - \widetilde{W}^{M_n}(d\theta|Y)\right\|_{\mathrm{TV}} &= E\left[\widetilde{W}\left(\mathcal{E}^C_{\theta_0,\Phi}(M_n)\middle|\, Y\right)\right] \\
&\to 0 \text{ as } n \to \infty.
\end{aligned}$$

Proposition 9 is proved in Appendix A, using the important following Lemma.

Lemma 7. *Let $a \in \mathbb{R}^n$ such that $\Phi^T a = 0$. Then, for any $y \in \mathbb{R}^n$, $W(\cdot|Y = y) = W(\cdot|Y = y + a)$.*

Lemma 7 states that the distribution $W(\cdot|Y)$ is invariant by any translation of Y orthogonal to $\langle\phi\rangle$. As a consequence, proving Proposition 9 in the case $F_0 = \Phi\theta_0$ is enough.

Proof. We decompose any $y \in \mathbb{R}^n$ in two orthogonal components $y = \Phi\theta_y + y'$, with $\Phi^T y' = 0$. The density of P_θ is equal to

$$\begin{aligned}
dP_\theta(y) &= \frac{1}{\left(\sigma_n\sqrt{2\pi}\right)^n}\exp\left\{-\frac{\|\Phi\theta_y + y' - \Phi\theta\|^2}{2\sigma_n^2}\right\} \\
&= \frac{1}{\left(\sigma_n\sqrt{2\pi}\right)^n}\exp\left\{-\frac{\|y'\|^2}{2\sigma_n^2}\right\}\exp\left\{-\frac{\|\Phi\theta_y - \Phi\theta\|^2}{2\sigma_n^2}\right\}.
\end{aligned}$$

On the same way,

$$dP_\theta(y+a) = \frac{1}{\left(\sigma_n\sqrt{2\pi}\right)^n}\exp\left\{-\frac{\|y'+a\|^2}{2\sigma_n^2}\right\}\exp\left\{-\frac{\|\Phi\theta_y-\Phi\theta\|^2}{2\sigma_n^2}\right\}$$
$$= \exp\left\{-\frac{\|y'+a\|^2-\|y'\|^2}{2\sigma_n^2}\right\}dP_\theta(y).$$

Therefore

$$dP^W(y+a) = \int dP_\theta(y+a)w(\theta)\,d\theta$$
$$= \exp\left\{-\frac{\|y'+a\|^2-\|y'\|^2}{2\sigma_n^2}\right\}dP^W(y)$$

and

$$\widetilde{W}(d\theta|Y=y+a) = \frac{dP_\theta(y+a)w(\theta)\,d\theta}{dP^W(y+a)}$$
$$= \widetilde{W}(d\theta|Y=y).$$

□

6.3 Proof of Theorem 3.

Let us consider the following Taylor expansion:

$$G(F)-G(Y_{\langle\phi\rangle}) = \dot{G}_{F_{\langle\phi\rangle}}(F-Y_{\langle\phi\rangle})$$
$$+\frac{1}{2}\int_0^1(1-t)D^2_{F_{\langle\phi\rangle}+t\,(F-F_{\langle\phi\rangle})}G(F-F_{\langle\phi\rangle},F-F_{\langle\phi\rangle})\,dt$$
$$-\frac{1}{2}\int_0^1(1-t)D^2_{F_{\langle\phi\rangle}+t\,(Y_{\langle\phi\rangle}-F_{\langle\phi\rangle})}G(Y_{\langle\phi\rangle}-F_{\langle\phi\rangle},Y_{\langle\phi\rangle}-F_{\langle\phi\rangle})\,dt.$$

Suppose that $F\in\langle\phi\rangle$, $\|F-F_{\langle\phi\rangle}\|^2\le\sigma_n^2M_n$, and $\|Y_{\langle\phi\rangle}-F_{\langle\phi\rangle}\|^2\le\sigma_n^2M_n$. Then, for any $b\in\mathbb{R}^p$,

$$\left|b^T\left(G(F)-G(Y_{\langle\phi\rangle})-\dot{G}_{F_{\langle\phi\rangle}}(F-Y_{\langle\phi\rangle})\right)\right|\le\|b\|B_{F_{\langle\phi\rangle}}(M_n).$$

On the other hand, $\sqrt{b^T\Gamma_{F_{\langle\phi\rangle}}b}\ge\sqrt{\left\|\Gamma^{-1}_{F_{\langle\phi\rangle}}\right\|^{-1}}\|b\|$. Moreover

$$\left\|W\left(d\frac{b^T\dot{G}_{F_{\langle\phi\rangle}}(F-Y_{\langle\phi\rangle})}{\sqrt{b^T\Gamma_{F_{\langle\phi\rangle}}b}}\middle|Y\right)-\mathcal{N}(0,1)\right\|_{\mathrm{TV}}\le\left\|W(dF|Y)-\mathcal{N}\left(Y_{\langle\phi\rangle},\sigma_n^2\Sigma\right)\right\|_{\mathrm{TV}}.$$

Let $\eta_n=\sqrt{\left\|\Gamma^{-1}_{F_{\langle\phi\rangle}}\right\|}B_{F_{\langle\phi\rangle}}(M_n)$, which tends to 0 by hypothesis. Let also

$$I_{\eta_n}=\{x\in\mathbb{R}:\exists x'\in I,|x-x'|\le\eta_n\}.$$

Note that $\psi\left(I_{\eta_n}\right) \leq \psi(I) + \sqrt{\frac{2}{\pi}}\eta_n$.

Gathering all this information, we can get the upper bound

$$
\begin{aligned}
W\left(\frac{b^T\left(G(F) - G(Y_{\langle\phi\rangle})\right)}{\sqrt{b^T \Gamma_{F_{\langle\phi\rangle}} b}} \in I \,\middle|\, Y \right) &\leq \psi(I) + \sqrt{\frac{2}{\pi}}\eta_n \\
&\quad + \left\| W(dF|Y) - \mathcal{N}\left(Y_{\langle\phi\rangle}, \sigma_n^2 \Sigma\right) \right\|_{\mathrm{TV}} \\
&\quad + \mathbb{1}_{\|Y_{\langle\phi\rangle} - F_{\langle\phi\rangle}\|^2 > \sigma_n^2 M_n} \\
&\quad + W\left(\|F - F_{\langle\phi\rangle}\|^2 > \sigma_n^2 M_n \,\middle|\, Y \right).
\end{aligned}
$$

A lower bound is obtained in the same way. Taking the expectation,

$$
\begin{aligned}
E\left| W\left(\frac{b^T\left(G(F) - G(Y_{\langle\phi\rangle})\right)}{\sqrt{b^T \Gamma_{F_{\langle\phi\rangle}} b}} \in I \,\middle|\, Y \right) - \psi(I) \right| \\
\leq o(1) + P(\|Y_{\langle\phi\rangle} - F_{\langle\phi\rangle}\|^2 > \sigma_n^2 M_n) \\
+ E\left[W\left(\|F - F_{\langle\phi\rangle}\|^2 > \sigma_n^2 M_n \,\middle|\, Y \right) \right].
\end{aligned}
$$

But $\|Y_{\langle\phi\rangle} - F_{\langle\phi\rangle}\|^2$ follows the $\sigma_n^2 \chi^2(k_n)$ distribution, and since $k_n = o(M_n)$,

$$
P(\|Y_{\langle\phi\rangle} - F_{\langle\phi\rangle}\|^2 > \sigma_n^2 M_n) = o(1).
$$

To conclude the proof of the Bayesian part of Theorem 3, we use the following:

Lemma 8. *Suppose that the conditions of either Theorem 1 or Theorem 2 are verified. Then*

$$
E\left[W\left(\|F - F_{\langle\phi\rangle}\|^2 > \sigma_n^2 M_n \,\middle|\, Y \right) \right] \to 0 \text{ as } n \to \infty.
$$

Proof. For smooth priors, this is an immediate corollary of Proposition 9. Let us suppose we are under the conditions of Theorem 1.

Let Z be a $\mathcal{N}\left(0, \frac{\sigma_n^2 \tau_n^2}{\sigma_n^2 + \tau_n^2} (\Phi^T \Phi)^{-1}\right)$ random vector in $\mathbb{R}^n$ independent on Y, and U a random variable following $\chi^2(k_n)$. Using (18), we get

$$
\begin{aligned}
&W\left(\|F - F_{\langle\phi\rangle}\|^2 > \sigma_n^2 M_n \,\middle|\, Y \right) \\
&\qquad = P\left(\left\| Z + \frac{\tau_n^2}{\sigma_n^2 + \tau_n^2} Y_{\langle\phi\rangle} - F_{\langle\phi\rangle} \right\|^2 > \sigma_n^2 M_n \right) \\
&\qquad \leq P\left(\|Z\| > \sigma_n \sqrt{M_n} - \left\| \frac{\tau_n^2}{\sigma_n^2 + \tau_n^2} Y_{\langle\phi\rangle} - F_{\langle\phi\rangle} \right\| \right) \\
&\qquad \leq \begin{cases} 1 & \text{if } \left\| \frac{\tau_n^2}{\sigma_n^2 + \tau_n^2} Y_{\langle\phi\rangle} - F_{\langle\phi\rangle} \right\| > \frac{2\sigma_n \sqrt{M_n}}{3} \\ P\left(U > \frac{\sigma_n^2 + \tau_n^2}{\tau_n^2} \frac{M_n}{9} \right) & \text{otherwise.} \end{cases}
\end{aligned}
$$

Since $k_n = o(M_n)$, $P(U > M_n/9) = o(1)$. On the other hand,

$$
\begin{aligned}
\left\| \frac{\tau_n^2}{\sigma_n^2 + \tau_n^2} Y_{\langle\phi\rangle} - F_{\langle\phi\rangle} \right\| &= \left\| \Sigma \left(\frac{\tau_n^2}{\sigma_n^2 + \tau_n^2} \varepsilon + \frac{\sigma_n^2}{\sigma_n^2 + \tau_n^2} F_0 \right) \right\| \\
&\leq \|\Sigma \varepsilon\| + \frac{\sigma_n}{\sqrt{\sigma_n^2 + \tau_n^2}} \|F_0\|
\end{aligned}
$$

Since $\|F_0\| = o(\tau_n^2/\sigma_n)$, $\frac{\sigma_n^2\|F_0\|^2}{\sigma_n^2+\tau_n^2} = o(1) < \frac{M_n}{9}$ for n large enough. $\|\Sigma\varepsilon\|^2$ is a $\chi^2(k_n)$ variable. Therefore, for n large enough,

$$E\left[W\left(\|F - F_{\langle\phi\rangle}\|^2 > \sigma_n^2 M_n \middle| Y\right)\right] \leq 2P(U > M_n/9) = o(1).$$

□

The frequentist assertion (12) is proved in a similar way from Taylor's expansion

$$G(Y_{\langle\phi\rangle}) - G(F_{\langle\phi\rangle}) = \dot{G}_{F_{\langle\phi\rangle}}(Y_{\langle\phi\rangle} - F_{\langle\phi\rangle}) + \frac{1}{2}\int_0^1 (1-t) D^2_{F_{\langle\phi\rangle}+t\,(Y_{\langle\phi\rangle}-F_{\langle\phi\rangle})} G(Y_{\langle\phi\rangle} - F_{\langle\phi\rangle}, Y_{\langle\phi\rangle} - F_{\langle\phi\rangle})\,dt.$$

Acknowledgment

The author would like to thank I. Castillo and E. Gassiat for valuable discussions and suggestions.

References

[1] Dominique Bontemps. Universal coding on infinite alphabets: Exponentially decreasing envelopes. http://fr.arxiv.org/pdf/0806.0562v2.

[2] Stéphane Boucheron and Elisabeth Gassiat. A Bernstein-von Mises theorem for discrete probability distributions. *Electron. J. Statist.*, 3:114–148, 2009.

[3] Ismaël Castillo. A semi-parametric Bernstein-von Mises theorem. http://www.proba.jussieu.fr/~castillo/bvm.pdf.

[4] Bertrand Clarke and Subhashis Ghosal. Posterior normality and reference priors for exponential families with increasing dimension. http://www4.stat.ncsu.edu/~sghosal/papers/HighDimRefPr.pdf.

[5] Carl de Boor. *A practical guide to splines*. Springer-Verlag, New York, 1978.

[6] David Freedman. Wald lecture: On the Bernstein-von Mises theorem with infinite-dimensional parameters. *Ann. Statist.*, 27(4):1119–1141, 1999.

[7] Subhashis Ghosal. Asymptotic normality of posterior distributions in high-dimensional linear models. *Bernoulli*, 5(2):315–331, 1999.

[8] Subhashis Ghosal. Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. *Journal of Multivariate Analysis*, 74:49–68, 2000.

[9] Subhashis Ghosal, Jayanta K. Ghosh, and Aad W. van der Vaart. Convergence rates of posterior distributions. *Ann. Statist.*, 28(2):500–531, 2000.

[10] Subhashis Ghosal and Aad W. van der Vaart. Convergence rates of posterior distributions for noniid observations. *Ann. Statist.*, 35(1):192–223, 2007.

[11] Yongdai Kim. The Bernstein-von Mises theorem for the proportional hazard model. *Ann. Statist.*, 34(4):1678–1700, 2006.

[12] Yongdai Kim and Jaeyong Lee. A Bernstein–von Mises theorem in the nonparametric right-censoring model. *Ann. Statist.*, 32(4):1492–1512, 2004.

[13] Pascal Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2007.

[14] Vincent Rivoirard and Judith Rousseau. Bernstein von Mises Theorem for linear functionals of the density. http://arxiv.org/abs/0908.4167.

[15] Xiaotong Shen. Asymptotic normality of semiparametric and nonparametric posterior distributions. *J. Amer. Statist. Assoc.*, 97(457):222–235, 2002.

[16] Xiaotong Shen and Larry Wasserman. Rates of convergence of posterior distributions. *Ann. Statist.*, 29(3):687–714, 2001.

[17] Alexandre B. Tsybakov. *Introduction à l'estimation non-paramétrique*. Mathématiques et Applications. Springer-Verlag, Berlin, 2004.

[18] Aad W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical ans Probabilistic Mathematics. CUP, Cambridge, 1998.

A Posterior Consistency

Here we prove Proposition 9. Lemma 7 allows us to suppose $F_0 = \Phi\theta_0$. Let U a random variable following the $\chi^2(k_n)$ distribution. Proceeding as in [2, 18], we introduce a test

$$T_n = \mathbb{1}_{(\theta_Y-\theta_0)^T\Phi^T\Phi\,(\theta_Y-\theta_0)>\sigma_n^2 M_n/4}. \tag{24}$$

Note that $ET_n = P(U > M_n/4) = o(1)$. Then

$$E\left[\widetilde{W}\left(\mathcal{E}^C_{\theta_0,\Phi}(M_n)\middle|\,Y\right)\right] \le ET_n + E\left[(1-T_n)\widetilde{W}\left(\mathcal{E}^C_{\theta_0,\Phi}(M_n)\middle|\,Y\right)\right].$$

Next, let $(r_n)_{n\ge1}$ be a sequence of positive numbers such that r_n goes to 0 and $-\ln(r_n) = o(M_n/k_n)$ as n goes to infinity. We replace the distribution P_{θ_0} by the mixture distribution $P^W_{\theta_0,r_n}$ with density

$$dP^W_{\theta_0,r_n}(y) = \int_{\mathcal{E}_{\theta_0,\Phi}(r_n)} dP_\theta(y)\widetilde{W}^{r_n}(d\theta). \tag{25}$$

where $\widetilde{W}^{r_n}$ is the rescaled restriction of $\widetilde{W}$ to $\mathcal{E}_{\theta_0,\Phi}(r_n)$, as in (20). The following Lemma illustrates the link between P_{θ_0} and $P^W_{\theta_0,r_n}$.

Lemma 9. *Using the preceding notations,*

$$\left\|P^W_{\theta_0,r_n} - P_{\theta_0}\right\|_{\mathrm{TV}} \le \sqrt{\frac{r_n}{2\pi}} = o(1).$$

Proof. We use convexity, and Lemma 4 since $P_\theta = \mathcal{N}\left(\Phi\theta, \sigma_n^2 I_n\right)$:

$$\left\|P^W_{\theta_0,r_n} - P_{\theta_0}\right\|_{\mathrm{TV}} \leq \sup_{h^T\Phi^T\Phi h \leq \sigma_n^2 r_n} \|P_{\theta_0+h} - P_{\theta_0}\|_{\mathrm{TV}} \leq \frac{\sqrt{r_n}}{\sqrt{2\pi}}.$$

□

At that point, the Bayes rule and the Fubini Theorem give

$$\begin{aligned}
&E^W_{\theta_0,r_n}\left[(1-T_n)\widetilde{W}\left(\mathcal{E}^C_{\theta_0,\Phi}(M_n)\middle|\, Y\right)\right] \\
&= \frac{1}{\widetilde{W}(\mathcal{E}(r_n))}\int_{\mathcal{E}(r_n)}\left(\int_{\mathbb{R}^n}\left[(1-T_n)\int_{\mathcal{E}^C(M_n)}\frac{dP_\tau(Y)w(\tau)\,d\tau}{\int_{\mathbb{R}^{k_n}} dP_\eta(Y)w(\eta)\,d\eta}\right] dP_\theta(Y)\right) w(\theta)\,d\theta \\
&= \frac{1}{\widetilde{W}(\mathcal{E}(r_n))}\int_{\mathcal{E}^C(M_n)} E_\tau\left[(1-T_n)\widetilde{W}(\mathcal{E}(r_n)|Y)\right] w(\tau)\,d\tau \\
&\leq \frac{1}{\widetilde{W}\left(\mathcal{E}_{\theta_0,\Phi}(r_n)\right)}\sup_{h^T\Phi^T\Phi h > \sigma_n^2 M_n} E_{\theta_0+h}(1-T_n) \\
&\leq \frac{1}{\widetilde{W}\left(\mathcal{E}_{\theta_0,\Phi}(r_n)\right)}\sup_{h^T\Phi^T\Phi h > \sigma_n^2 M_n} P_{\theta_0+h}\left(\|\Phi\left(\theta_Y-\theta_0-h\right)\|^2 > \sigma_n^2 M_n/4\right) \\
&= \frac{P(U > M_n/4)}{\widetilde{W}\left(\mathcal{E}_{\theta_0,\Phi}(r_n)\right)}.
\end{aligned}$$

Let $B_k(0,1)$ be the unit ball in $\mathbb{R}^k$. We make use of the following relation (see for instance [1, Lemma 2])

$$-\ln\mathrm{vol}\left(B_k(0,1)\right) = \ln\frac{\Gamma(1+k/2)}{\pi^{k/2}} \underset{k\to\infty}{\sim} \frac{k}{2}\ln k$$

together with a control on the volume of the ellipsoid $\mathcal{E}_{\theta_0,\Phi}(r_n)$

$$\widetilde{W}\left(\mathcal{E}_{\theta_0,\Phi}(r_n)\right) \geq \left(\inf_{h^T\Phi^T\Phi h\leq\sigma_n^2 r_n}\frac{w(\theta_0+h)}{w(\theta_0)}\right)\frac{\sigma_n^{k_n} w(\theta_0)}{\sqrt{\det(\Phi^T\Phi)}} r_n^{k_n/2}\,\mathrm{vol}\left(B_{k_n}(0,1)\right).$$

Next we can use Cirelson's inequality (22) as in Lemma 5 and get, for n large enough,

$$\begin{aligned}
&\ln\left(E^W_{\theta_0,r_n}\left[(1-T_n)\widetilde{W}\left(\mathcal{E}_{\theta_0,\Phi}(M_n)\middle|\, Y\right)\right]\right) \\
&\qquad\leq \ln\left(\frac{\sqrt{\det(\Phi^T\Phi)}}{\sigma_n^{k_n} w(\theta_0)}\right) - \frac{M_n}{9} - \frac{k_n}{2}\ln(r_n) - \ln\mathrm{vol}\left(B_{k_n}(0,1)\right) + o(1) \\
&\qquad\sim -\frac{M_n}{9}
\end{aligned}$$

which goes to minus infinity as n goes to infinity.

B Sobolev classes

We begin with a simple lemma, then we prove Lemma 2 and Lemma 3

Lemma 10. *Let $\alpha > 1/2$, $L > 0$, and $\theta \in \Theta(\alpha, L)$. Then*

$$\sum_{j=1}^{\infty} |\theta_j| < \infty.$$

As a consequence, f is the uniform limit of the series $\sum_{j=1}^{\infty} \theta_j \varphi_j$ and f is continuous.

Proof of Lemma 10. We have a simple control on the sum of the coefficients

$$\sum_{j=2}^{\infty} |\theta_j| \leq \sqrt{\sum_{j\geq 2} a_j^{-2}} \sqrt{\sum_{j\geq 2} a_j^2 \theta_j^2} \leq \frac{L}{\pi^\alpha} \sqrt{\sum_{j\geq 1} j^{-2\alpha}} < \infty.$$

Since all functions φ_j are continuous and bounded by $\sqrt{2}$, the other points follow. □

Proof of Lemma 2. $F_{\langle\phi\rangle}$ is the orthogonal projection of F_0 on the convex span of the first k_n vectors of the orthogonal basis $(\phi_j)_{1\leq j\leq n}$ of $\mathbb{R}^n$. So

$$\|F_0 - F_{\langle\phi\rangle}\|^2 = \sum_{j=k_n+1}^{n} (F_0^T \phi_j)^2 = n \sum_{j=k_n+1}^{n} \left(\frac{1}{n} \sum_{i=1}^{n} f(i/n)\varphi_j(i/n) \right)^2.$$

Following [17], we set $\zeta_j = \frac{1}{n} \sum_{i=1}^{n} f(i/n)\varphi_j(i/n) - \theta_j^0$ for $1 \leq j \leq n$. Then

$$\|F_0 - F_{\langle\phi\rangle}\|^2 = n \sum_{j=k_n+1}^{n} \left(\zeta_j + \theta_j^0\right)^2 \leq 2n \left(\sum_{j=1}^{n} \zeta_j^2 + \sum_{j=k_n+1}^{n} (\theta_j^0)^2 \right).$$

Using Lemma 1, for any $1 \leq j \leq n$,

$$\begin{aligned}
\zeta_j &= \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{m=1}^{\infty} \theta_m^0 \varphi_m(i/n) \right) \varphi_j(i/n) - \theta_j^0 \\
&= \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{m=n+1}^{\infty} \theta_m^0 \varphi_m(i/n) \right) \varphi_j(i/n).
\end{aligned}$$

So, using Lemma 1 again,

$$\sum_{j=1}^{n} \zeta_j^2 \leq \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{m=n+1}^{\infty} \theta_m^0 \varphi_m(i/n) \right)^2.$$

We recognize a Riemann sum of the function $\left(\sum_{m=n+1}^{\infty} \theta_m^0 \varphi_m\right)^2$, which is continuous according to Lemma 10. Therefore

$$\sum_{j=1}^{n} \zeta_j^2 \leq (1 + o(1)) \int_0^1 \left(\sum_{m=n+1}^{\infty} \theta_m^0 \varphi_m \right)^2 = \sum_{m=n+1}^{\infty} (\theta_m^0)^2$$

and

$$\begin{aligned}
\|F_0 - F_{\langle\phi\rangle}\|^2 &\le (2n + o(n)) \sum_{m=k_n+1}^{\infty} (\theta^0_m)^2 \\
&\le \frac{2n + o(n)}{a^2_{k_n+1}} \sum_{m=k_n+1}^{\infty} a^2_m (\theta^0_m)^2 \\
&\le \frac{2L^2}{\pi^{2\alpha}} \frac{n + o(n)}{k_n^{2\alpha}}.
\end{aligned}$$

On the other hand, f is continuous, and $(1/n)\,\|F_0\|^2$ is a Riemann sum of f^2. Therefore $(1/n)\,\|F_0\|^2$ goes to $\int_0^1 f^2$ as n goes to infinity. □

Proof of Lemma 3. Let $(\theta'_j)_{j\ge1}$ the Fourier coefficients of g. As in the previous proof, we set $\zeta_j = \frac{1}{n}\sum_{i=1}^n f(i/n)\varphi_j(i/n) - \theta^0_j$ and $\zeta'_j = \frac{1}{n}\sum_{i=1}^n g(i/n)\varphi_j(i/n) - \theta'_j$ for $1 \le j \le n$. We have $F_0 = \sum_{j=1}^n (\zeta_j + \theta^0_j)\phi_j$, so

$$\frac{1}{n}\sum_{i=1}^n f(i/n)g(i/n) = \sum_{j=1}^n (\zeta_j + \theta^0_j)(\zeta'_j + \theta'_j).$$

In the meantime

$$\int_0^1 fg = \sum_{j=1}^{\infty} \theta^0_j \theta'_j.$$

So

$$\begin{aligned}
\left| \frac{1}{n}\sum_{i=1}^n f(i/n)g(i/n) - \int_0^1 fg \right| &= \left| \sum_{j=1}^n \zeta_j\zeta'_j + \sum_{j=1}^n \zeta_j\theta'_j + \sum_{j=1}^n \theta^0_j\zeta'_j - \sum_{j=n+1}^{\infty} \theta^0_j\theta'_j \right| \\
&\le \sqrt{\sum_{j=1}^n \zeta_j^2}\sqrt{\sum_{j=1}^n {\zeta'_j}^2} + \sqrt{\sum_{j=1}^n \zeta_j^2}\sqrt{\sum_{j=1}^n {\theta'_j}^2} \\
&\quad + \sqrt{\sum_{j=1}^n {\zeta'_j}^2}\sqrt{\sum_{j=1}^n (\theta^0_j)^2} + \sqrt{\sum_{j=n+1}^{\infty} (\theta^0_j)^2}\sqrt{\sum_{j=n+1}^{\infty} {\theta'_j}^2}.
\end{aligned}$$

As in the proof of Lemma 2, we have

$$\sum_{j=1}^n \zeta_j^2 \sim \sum_{j=n+1}^{\infty} (\theta^0_j)^2 \le \frac{L^2}{\pi^{2\alpha} n^{2\alpha}}$$

and on the other hand,

$$\sum_{j=1}^n (\theta^0_j)^2 \le \int_0^1 f^2 = \|f\|^2.$$

Thus

$$\begin{aligned}
\left| \frac{1}{n}\sum_{i=1}^n f(i/n)g(i/n) - \int_0^1 fg \right| &\le (1 + o(1)) \left(\frac{LL'}{\pi^{\alpha+\alpha'} n^{\alpha+\alpha'}} + \frac{L'\|f\|}{\pi^{\alpha'} n^{\alpha'}} + \frac{L\|g\|}{\pi^{\alpha} n^{\alpha}} \right) \\
&= O\left(n^{-\inf(\alpha,\alpha')} \right).
\end{aligned}$$

□