

BERNSTEIN-VON MISES THEOREM FOR LINEAR FUNCTIONALS OF THE DENSITY

BY VINCENT RIVOIRARD*,

Université Paris Sud and Ecole Normale Supérieure

E-MAIL: Vincent.Rivoirard@math.u-psud.fr

BY JUDITH ROUSSEAU*,

Université Paris Dauphine and CREST-ENSAE

E-MAIL: rousseau@ceremade.dauphine.fr

In this paper, we study the asymptotic posterior distribution of linear functionals of the density. In particular, we give general conditions to obtain a semi-parametric version of the Bernstein-von Mises theorem. We then apply this general result to non-parametric priors based on infinite dimensional exponential families. As a byproduct, we also derive adaptive non-parametric rates of concentration of the posterior distributions under these families of priors on the class of Sobolev and Besov spaces.

1. Introduction. The Bernstein-von Mises property, in Bayesian analysis, concerns the asymptotic form of the posterior distribution of a quantity of interest, and more specifically it corresponds to the asymptotic normality of the posterior distribution centered at some kind of maximum likelihood estimator with the variance being equal to the asymptotic frequentist variance of the centering point. Such results are well known in parametric frameworks, see for instance [16] where general conditions are given. This is an important property for both practical and theoretical reasons. In particular the asymptotic normality of the posterior distributions allows us to construct approximate credible regions and the duality between the behavior of the posterior distribution and the frequentist distribution of the asymptotic centering point of the posterior implies that credible regions will have also good frequentist properties. These results are given in many Bayesian textbooks see for instance [19] or [1].

In a frequentist perspective the Bernstein-von Mises property enables the construction of confidence regions since under this property a Bayesian credi-

*This work has been partially supported by the ANR-SP Bayes grant

AMS 2000 subject classifications: Primary 62G07, 62G20

Keywords and phrases: Bayesian non-parametric, rates of convergence, Bernstein - von Mises, adaptive estimation

ble region will be asymptotically a frequentist confidence region as well. This is even more important in complex models, since in such models the construction of confidence regions can be difficult whereas the Markov Chain Monte Carlo algorithms usually make the construction of a Bayesian credible region feasible. However the more complex the model the harder it is to derive Bernstein-von Mises theorems. In infinite dimensional setups, the mechanisms are even more complex.

Semi-parametric and non-parametric models are widely popular both from a theoretical and practical perspective and have been used by frequentists as well as Bayesians although their theoretical asymptotic properties have been mainly studied in the frequentist literature. The use of Bayesian non-parametric or semi-parametric approaches is more recent and has been made possible mainly by the development of algorithms such as Markov Chain Monte-Carlo algorithms but has grown rapidly over the past decade.

However, there is still little work on asymptotic properties of Bayesian procedures in semi-parametric models or even in non-parametric models. Most of existing works on the asymptotic posterior distributions deal with consistency or rates of concentration of the posterior. In other words it consists in controlling objects of the form $\mathbb{P}^\pi [U_n | X^n]$ where $\mathbb{P}^\pi [\cdot | X^n]$ denotes the posterior distribution given a n vector of observations X^n and U_n denotes either a fixed neighborhood (consistency) or a sequence of shrinking neighborhoods (rates of concentration). As remarked by [6] consistency is an important condition since it is not possible to construct subjective prior in a non-parametric framework. Obtaining concentration rates of the posterior helps to understand the impact of the choice of a specific prior and allows for a comparison between priors to some extent. However, to obtain a Bernstein-von Mises theorem it is necessary not only to bound $\mathbb{P}^\pi [U_n | X^n]$ but to determine an equivalent of $\mathbb{P}^\pi [U_n | X^n]$ for some specific types of sets U_n . This difficulty explains that there is up to now hardly any work on Bernstein-von Mises theorems in infinite dimensional models. The most well known results are negative results and are given in [7]. Some positive results are provided by [8] on the asymptotic normality of the posterior distribution of the parameter in an exponential family with increasing number of parameters. In a discrete setting, [2] derive Bernstein-von Mises results, in particular satisfied by Dirichlet priors. Nice positive results are obtained in [14] and [15], however they rely heavily on a conjugacy property and on the fact that their priors put mass one on discrete probabilities which makes the comparison with the empirical distribution more tractable.

In a semi-parametric framework, where the parameter can be separated into a finite dimensional parameter of interest and infinite dimensional nuisance

sance parameter, [3] obtains interesting conditions leading to a Bernstein-von Mises theorem on the parameter of interest, clarifying an earlier work of [20].

In this paper we are interested in studying the existence of a Bernstein-von Mises property in semi-parametric models where the parameter of interest is a functional of the density of the observations. The estimation of functionals of infinite dimensional parameters such as the cumulative distribution function at a specific point, is a widely studied problem both in the frequentist literature and in the Bayesian literature. There is a vast literature on the rates of convergence and on the asymptotic distribution of frequentist estimates of functionals of unknown curves and of finite dimensional functionals of curves in particular, see for instance [23] for an excellent presentation of a general theory on such problems.

One of the most common functionals considered in the literature is the cumulative distribution function calculated at a given point, say $F(x_0)$. The empirical cumulative distribution function is a natural frequentist estimator and its asymptotic distribution is Gaussian with mean $F(x_0)$ and variance $F(x_0)(1 - F(x_0))/n$.

The Bayesian counterpart of this estimator is the one derived from a Dirichlet process prior and it is well known to be asymptotically equivalent to $F_n(x_0)$, see for instance [11]. This result is obtained by using the conjugate nature of the Dirichlet prior, leading to an explicit posterior distribution. Other frequentist estimators, based on density estimates such as kernel estimators have also been studied in the frequentist literature. Hence a natural question arises. Can we generalize the Bernstein-von Mises theorem of the Dirichlet estimator to other Bayesian estimators? What happens if the prior has support on distributions absolutely continuous with respect to the Lebesgue measure?

In this paper we provide an answer to these questions by establishing conditions under which a Bernstein-von Mises theorem can be obtained for linear functional of the density of f such as $F(x_0)$. We also study cases where the asymptotic posterior distribution of the functional is not asymptotically Gaussian but is asymptotically a mixture of Gaussian distributions with different centering points.

1.1. *Notation and aim.* In this paper, we assume that, given a distribution \mathbb{P} with a compactly supported density f with respect to the Lebesgue measure, X_1, \dots, X_n are independent and identically distributed according to \mathbb{P} . We set $X^n = (X_1, \dots, X_n)$ and denote F the cumulative distribution function associated with f . Without loss of generality we assume that for

any i , $X_i \in [0, 1]$ and we set

$$\mathcal{F} = \left\{ f : [0, 1] \rightarrow \mathbb{R}^+ \text{ s.t. } \int_0^1 f(x)dx = 1 \right\}.$$

We denote $\ell_n(f)$ the log-likelihood associated with the density f and if f is parametrized by a finite dimensional parameter θ , we set $\ell_n(\theta) = \ell_n(f_\theta)$. For any integrable function g , we set $F(g) = \int_0^1 f(u)g(u)du$. We denote by $\langle \cdot, \cdot \rangle_f$ the inner product and by $\|\cdot\|_f$ the associated norm in

$$\mathbb{L}_2(F) = \left\{ g \text{ s.t. } \int g^2(x)f(x)dx < +\infty \right\}.$$

We also consider the classical inner product in $\mathbb{L}_2[0, 1]$, denoted $\langle \cdot, \cdot \rangle_2$, and $\|\cdot\|_2$, the associated norm. The Kullback-Leibler divergence and the Hellinger distance between two densities f and f' will be respectively denoted $K(f, f')$ and $h(f, f')$. We recall that

$$K(f, f') = F(\log(f/f')), \quad h(f, f') = \left[\int \left(\sqrt{f(x)} - \sqrt{f'(x)} \right)^2 dx \right]^{1/2}.$$

In the sequel, we shall also use

$$V(f, f') = F\left((\log(f/f'))^2\right).$$

Let \mathbb{P}_0 be the true distribution of the observations X_i whose density and cumulative distribution function are respectively denoted f_0 and F_0 . We consider usual notation on empirical processes, namely

$$P_n(g) = \frac{1}{n} \sum_{i=1}^n g(X_i), \quad G_n(g) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [g(X_i) - F_0(g)]$$

and F_n is the empirical distribution function. Now, we simply denote $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ instead of $\langle \cdot, \cdot \rangle_{f_0}$ and $\|\cdot\|_{f_0}$ respectively.

For any given $\psi \in \mathbb{L}_\infty[0, 1]$, we consider Ψ the functional on \mathcal{M} , the set of finite measures on $[0, 1]$, defined by

$$(1.1) \quad \Psi(\mu) = \int \psi d\mu, \quad \mu \in \mathcal{M}.$$

In particular, we have

$$\Psi(P_n) = P_n(\psi) = \frac{\sum_{i=1}^n \psi(X_i)}{n}.$$

Most of the time, to simplify notation when μ is absolutely continuous with respect to the Lebesgue measure with $g = \frac{d\mu}{dx}$, we use $\Psi(g)$ instead of $\Psi(\mu)$. A typical example of such functionals is given by

$$\Psi_{x_0}(f) = F(x_0) = \int \mathbb{1}_{x \leq x_0} f(x) dx, \quad x_0 \in \mathbb{R}.$$

Now, we consider a prior π on the set \mathcal{F} . The aim of this paper is to study the posterior distribution of $\Psi(f)$ and to derive conditions under which

$$(1.2) \quad \mathbb{P}^\pi [\sqrt{n}(\Psi(f) - \Psi(P_n)) \leq z | X^n] \rightarrow \Phi_{V_0}(z) \quad \text{in } \mathbb{P}_0\text{-probability,}$$

where V_0 is the variance of $\sqrt{n}\Psi(P_n)$ under \mathbb{P}_0 and $\Phi_{V_0}(z)$ is the cumulative distribution function of a centered Gaussian random variable with variance V_0 . Note that under this duality between the Bayesian and the frequentist behaviors, highest posterior credible regions for $\Psi(f)$ (such as equal tail or one-sided intervals) have also the correct asymptotic frequentist coverage.

In this paper we propose general conditions leading to (1.2) and we study in detail the special case of infinite dimensional exponential families as described in the following section.

1.2. *Infinite dimensional exponential families based on Fourier and wavelet expansions.* Fourier and wavelet bases are the dictionaries from which we build exponential families in the sequel. We recall that Fourier bases constitute unconditional bases of periodized Sobolev spaces W^γ where γ is the smoothness parameter. Wavelet expansions of any periodized function h take the following form:

$$h(x) = \theta_{-10} \mathbb{1}_{[0,1]}(x) + \sum_{j=0}^{+\infty} \sum_{k=0}^{2^j-1} \theta_{jk} \varphi_{jk}(x), \quad x \in [0, 1]$$

where $\theta_{-10} = \int_0^1 h(x) dx$ and $\theta_{jk} = \int_0^1 h(x) \varphi_{jk}(x) dx$. We recall that the functions φ_{jk} are obtained by periodizing dilations and translations of a mother wavelet φ that can be assumed to be compactly supported. Under standard properties of φ involving its regularity and its vanishing moments (see Lemma 4.1), wavelet bases constitute unconditional bases of Besov spaces $\mathcal{B}_{p,q}^\gamma$ for $1 \leq p, q \leq +\infty$ and $\gamma > \max\left(0, \frac{1}{p} - \frac{1}{2}\right)$. We refer the reader to [12] for a good review of wavelets and Besov spaces. We just mention that the scale of Besov spaces includes Sobolev spaces: $W^\gamma = \mathcal{B}_{2,2}^\gamma$. In the sequel, to shorten notation, the considered orthonormal basis will be denoted $\Phi = (\phi_\lambda)_{\lambda \in \mathbb{N}}$, where $\phi_0 = \mathbb{1}_{[0,1]}$ and

- for the Fourier basis, if $\lambda \geq 1$,

$$\phi_{2\lambda-1}(x) = \sqrt{2} \sin(2\pi\lambda x), \quad \phi_{2\lambda}(x) = \sqrt{2} \cos(2\pi\lambda x),$$

- for the wavelet basis, if $\lambda = 2^j + k$, with $j \in \mathbb{N}$ and $k \in \{0, \dots, 2^j - 1\}$,

$$\phi_\lambda = \varphi_{jk}.$$

Now, the decomposition of each periodized function $h \in \mathbb{L}_2[0, 1]$ on $(\phi_\lambda)_{\lambda \in \mathbb{N}}$ is written as follows:

$$h(x) = \sum_{\lambda \in \mathbb{N}} \theta_\lambda \phi_\lambda(x), \quad x \in [0, 1],$$

where $\theta_\lambda = \int_0^1 h(x) \phi_\lambda(x) dx$. We denote $\|\cdot\|_\gamma$ and $\|\cdot\|_{\gamma,p,q}$ the norms associated with W^γ and $\mathcal{B}_{p,q}^\gamma$ respectively.

We use such expansions to build non-parametric priors on \mathcal{F} in the following way: For any $k \in \mathbb{N}^*$, we set

$$\mathcal{F}_k = \left\{ f_\theta = \exp \left(\sum_{\lambda=1}^k \theta_\lambda \phi_\lambda - c(\theta) \right) \text{ s.t. } \theta \in \mathbb{R}^k \right\},$$

where

$$(1.3) \quad c(\theta) = \log \left(\int_0^1 \exp \left(\sum_{\lambda=1}^k \theta_\lambda \phi_\lambda(x) \right) dx \right).$$

So, we define a prior π on the set $\mathcal{F}_\infty = \cup_k \mathcal{F}_k \subset \mathcal{F}$ by defining a prior p on \mathbb{N}^* and then, once k is chosen, we fix a prior π_k on \mathcal{F}_k . Such priors are often considered in the Bayesian non-parametric literature. See for instance [21]. The special case of log-spline priors has been studied by [9] and [13], whereas the prior considered by [24] is based on Legendre polynomials. For the wavelet case, [13] considered the special case of the Haar basis.

We now specify the class of priors π on these models.

DEFINITION 1.1. *Given $\beta > 1/2$, the prior p on k satisfies one of the following conditions:*

[Case (PH)] *There exist two positive constants c_1 and c_2 such that for any $k \in \mathbb{N}^*$,*

$$(1.4) \quad \exp(-c_1 k L(k)) \leq p(k) \leq \exp(-c_2 k L(k)),$$

where L is the function that can be either $L(x) = 1$ or $L(x) = \log(x)$.

[Case (D)] If $k_n^* = n^{1/(2\beta+1)}$,

$$p(k) = \delta_{k_n^*}(k),$$

where $\delta_{k_n^*}$ denotes the Dirac mass at the point k_n^* .

Conditionally on k the prior π_k on \mathcal{F}_k is defined by

$$\frac{\theta_\lambda}{\sqrt{\tau_\lambda}} \stackrel{iid}{\sim} g, \quad \tau_\lambda = \tau_0 \lambda^{-2\beta} \quad 1 \leq \lambda \leq k,$$

where τ_0 is a positive constant and g is a continuous density on \mathbb{R} such that for any x ,

$$A_* \exp(-\tilde{c}_* |x|^{p_*}) \leq g(x) \leq B_* \exp(-c_* |x|^{p_*}),$$

where p_* , A_* , B_* , \tilde{c}_* and c_* are positive constants.

Observe that the prior is not necessarily Gaussian since we allow for densities g to have different tails. In the Dirac case (D), the prior on k is non random. For the case (PH), $L(x) = \log(x)$ typically corresponds to a Poisson prior on k and the case $L(x) = 1$ typically corresponds to hypergeometric priors.

1.3. *Organization of the paper.* After emphasizing difficulties raised by natural heuristics for proving Bernstein-von Mises theorems in non-parametric setups (see Section 2.1), Theorem 2.1 of Section 2.2 gives the asymptotic posterior distribution of $\Psi(f)$ which can be either Gaussian or a mixture of Gaussian distributions, when the prior is based on infinite dimensional exponential families. Corollary 2.2 illustrates positive results with respect to our purpose, but Proposition 2.1 shows that some bad phenomenons may happen. Theorem 2.1 is derived from Theorem 2.2, a more general result established in Section 2.3. These theorems depend on concentration rates established in Theorem 3.1. Since our purpose is not to focus on such results, Theorem 3.1 is postponed in Section 3. Proofs of the results are given in Section 4.

2. Bernstein-von Mises theorems.

2.1. *Some heuristics for proving Bernstein-von Mises theorems.* We first introduce some notions that are useful in the study of asymptotic properties of semi-parametric models. More details can be found, for instance, in [23].

As in Chapter 25 of [23], the usual way to study the asymptotic behavior of semi-parametric models is to consider local 1-dimensional differentiable paths around the true parameter f_0 , namely submodels of the form: $u \rightarrow f_{u,s}^*$ for $0 < u < u_0$, for some $u_0 > 0$ such that for each path there exists a measurable function s called the score function for the submodel $\{f_{u,s}^* \text{ s.t. } 0 < u < u_0\}$ at $u = 0$ satisfying

$$(2.1) \quad \lim_{u \rightarrow 0} \int_{\mathbb{R}} \left(\frac{f_{u,s}^{*1/2}(x) - f_0^{1/2}(x)}{u} - \frac{1}{2}s(x)f_0^{1/2}(x) \right)^2 dx = 0.$$

We denote by \mathcal{F}_{f_0} the tangent set, i.e. the collection of score functions s associated with these differentiable paths. Using (2.1), \mathcal{F}_{f_0} can be identified with a subset of $\{s \in \mathbb{L}_2(F_0) \text{ s.t. } F_0(s) = 0\}$. For instance, when considering all probability laws, the most usual collection of differentiable paths is given by

$$(2.2) \quad f_{u,s}^*(x) = d(u)f_0(x)e^{us(x)}$$

with $\|s\|_{\infty} < \infty$ and d such that $d(0) = 1$ and $d'(0) = 0$. In this case, s is the score function. Note that as explained in [23], the collection of differentiable paths of the form $f_{u,s}^*(x) = 2d(u)f_0(x)(1 + \exp(-2us(x)))^{-1}$ (with previous conditions on d), leads to the tangent space given by $\{s \in \mathbb{L}_2(F_0) \text{ s.t. } F_0(s) = 0\}$.

Consider a functional Ψ associated to a function $\psi \in \mathbb{L}_{\infty}[0, 1]$, as defined in Section 1.1, then for any differentiable path $u \rightarrow f_{u,s}^*$ with score function s ,

$$\begin{aligned} \frac{\Psi(f_{u,s}^*) - \Psi(f_0)}{u} &= \int \psi(x)s(x)f_0(x)dx + \int \frac{(f_{u,s}^{*1/2}(x) - f_0^{1/2}(x))^2}{u} \psi(x)dx \\ &+ 2 \int \psi(x) \left(\frac{f_{u,s}^{*1/2}(x) - f_0^{1/2}(x)}{u} - \frac{1}{2}s(x)f_0^{1/2}(x) \right) f_0^{1/2}(x)dx \\ &= \langle \psi, s \rangle + o(1). \end{aligned}$$

Then, we can define the efficient influence function $\tilde{\psi}$ belonging to $\overline{\text{lin}}(\mathcal{F}_{f_0})$ (the closure of the linear space generated by \mathcal{F}_{f_0}) that satisfies for any $s \in \mathcal{F}_{f_0}$,

$$\int \tilde{\psi}(x)s(x)f_0(x)dx = \int \psi(x)s(x)f_0(x)dx.$$

This implies:

$$(2.3) \quad \lim_{u \rightarrow 0} \frac{\Psi(f_{u,s}^*) - \Psi(f_0)}{u} = \langle \tilde{\psi}, s \rangle.$$

The efficient influence function, which is a key notion to characterize asymptotically efficient estimators (see Section 25.3 of [23]), will play an important role for our purpose. To shed lights on these notions, we consider the following three examples:

EXAMPLE 2.1. *As in Section 1.1, for fixed $x_0 \in \mathbb{R}$, consider for any density function f whose cdf is F ,*

$$\Psi_{x_0}(f) = \int \mathbb{1}_{x \leq x_0} f(x) dx = F(x_0)$$

so that in this case, if \mathcal{F}_{f_0} is the subspace of $\mathbb{L}_2(F_0)$ of functions s satisfying $F_0(s) = 0$ then $\tilde{\psi}(x) = \mathbb{1}_{x \leq x_0} - F_0(x_0)$.

EXAMPLE 2.2. *More generally, for any measurable set A consider $\psi(x) = \mathbb{1}_{x \in A}$. For any density function f*

$$\Psi_A(f) = \int \mathbb{1}_{x \in A} f(x) dx$$

satisfies the above conditions and $\tilde{\psi}(x) = \mathbb{1}_{x \in A} - \int_A f_0(x) dx$.

EXAMPLE 2.3. *If f_0 has bounded support, say on $[0, 1]$, then the functional*

$$\Psi(f) = \mathbb{E}_f[X_1] = \int_0^1 x f(x) dx$$

satisfies the above conditions. Then, $\psi(x) = x$ and $\tilde{\psi}(x) = x - \mathbb{E}_{f_0}[X_1]$.

In this framework, the Bernstein-von Mises theorem could be derived from the convergence of the following Laplace transform defined for any $t \in \mathbb{R}$ by

$$\begin{aligned} L_n(t) &= \mathbb{E}^\pi[\exp(t\sqrt{n}(\Psi(f) - \Psi(P_n))) | X^n] \\ &= \frac{\int \exp(t\sqrt{n}(\Psi(f) - \Psi(P_n)) + \ell_n(f) - \ell_n(f_0)) d\pi(f)}{\int \exp(\ell_n(f) - \ell_n(f_0)) d\pi(f)}. \end{aligned}$$

Now, let us set $f_{s,n}^{**} = f_{u,s}^*$ if $u = n^{-\frac{1}{2}}$. We have:

$$\begin{aligned} \sqrt{n} \left(\Psi(f_{s,n}^{**}) - \Psi(P_n) \right) &= \sqrt{n} \int \psi(x) (f_{s,n}^{**}(x) - f_0(x)) dx - G_n(\tilde{\psi}) \\ &= \Delta_n(s) + \langle \tilde{\psi}, s \rangle - G_n(\tilde{\psi}), \end{aligned}$$

with

$$\Delta_n(s) = \sqrt{n} \left(\Psi(f_{s,n}^{**}) - \Psi(f_0) \right) - \langle \tilde{\psi}, s \rangle.$$

Furthermore,

$$\ell_n(f_{s,n}^{**}) - \ell_n(f_0) = R_n(s) + G_n(s) - \frac{F_0(s^2)}{2},$$

with

$$R_n(s) = nP_n \left(\log \left(\frac{f_{s,n}^{**}}{f_0} \right) \right) - G_n(s) + \frac{F_0(s^2)}{2}.$$

So,

$$\begin{aligned} t\sqrt{n} \left(\Psi(f_{s,n}^{**}) - \Psi(P_n) \right) + \ell_n(f_{s,n}^{**}) - \ell_n(f_0) \\ = R_n(s) - \frac{F_0(s^2)}{2} + G_n(s - t\tilde{\psi}) + t\Delta_n(s) + t \langle \tilde{\psi}, s \rangle \\ = R_n(s - t\tilde{\psi}) + G_n(s - t\tilde{\psi}) - \frac{F_0((s - t\tilde{\psi})^2)}{2} + \frac{t^2 F_0(\tilde{\psi}^2)}{2} + U_n(s), \end{aligned}$$

with

$$U_n(s) = t\Delta_n(s) + R_n(s) - R_n(s - t\tilde{\psi}).$$

Lemma 25.14 of [23] shows that under (2.1), $R_n(s) = o(1)$ and (2.3) yields $\Delta_n(s) = o(1)$ for a fixed s . It is not enough however to derive a Bernstein-von Mises theorem. Nonetheless if we can choose a prior distribution π adapted to the previous framework to obtain uniformly $U_n = o(1)$ and

$$\sqrt{n} \left(\Psi(f_{s,n}^{**}) - \Psi(f) \right) + \ell_n(f_{s,n}^{**}) - \ell_n(f) = o(1)$$

then

$$\begin{aligned} L_n(t) &= \exp \left(\frac{t^2 F_0(\tilde{\psi}^2)}{2} \right) \frac{\int e^{R_n(s-t\tilde{\psi})+G_n(s-t\tilde{\psi})-\frac{F_0((s-t\tilde{\psi})^2)}{2}} d\pi(f)}{\int e^{R_n(s)+G_n(s)-\frac{F_0(s^2)}{2}} d\pi(f)} (1 + o(1)) \\ (2.4) \quad &= \exp \left(\frac{t^2 F_0(\tilde{\psi}^2)}{2} \right) (1 + o(1)) \end{aligned}$$

if

$$\begin{aligned} \frac{\int e^{R_n(s-t\tilde{\psi})+G_n(s-t\tilde{\psi})-\frac{F_0((s-t\tilde{\psi})^2)}{2}} d\pi(f)}{\int e^{R_n(s)+G_n(s)-\frac{F_0(s^2)}{2}} d\pi(f)} &= \frac{\int \exp(\ell_n(f) - \ell_n(f_0)) d\pi(f_{s+t\tilde{\psi}})}{\int \exp(\ell_n(f) - \ell_n(f_0)) d\pi(f)} \\ (2.5) \quad &= 1 + o(1). \end{aligned}$$

However $s \mapsto U_n(s)$ is not uniformly bounded on \mathcal{F}_{f_0} . We thus consider an alternative approach, which uses however some of the ideas described above

but allows for a better control of a term similar to $U_n(s)$ but of a slightly different nature. A condition similar to (2.5) will still be required. Since it may be hard to prove and even to handle this condition in many setups, we first focus, in Section 2.2, on a specific family of non-parametric priors.

In the sequel, we consider a functional Ψ as defined in (1.1) associated with the function $\psi \in \mathbb{L}_\infty[0, 1]$ and we set

$$(2.6) \quad \tilde{\psi}(x) = \psi(x) - F_0(\psi).$$

Note that this notation is coherent with the definition of the influence function associated with the tangent set $\{s \in \mathbb{L}_2(F_0) \text{ s.t. } F_0(s) = 0\}$.

2.2. Bernstein-von Mises in infinite dimensional exponential families. In this section, we consider the non-parametric models (priors) defined in Section 1.2. Assume that f_0 is 1-periodic and $f_0 \in \mathcal{F}_\infty$. Let $\Phi = (\phi_\lambda)_{\lambda \in \mathbb{N}}$ be one of the bases introduced in Section 1.2, then there exists a sequence $\theta_0 = (\theta_{0\lambda})_{\lambda \in \mathbb{N}^*}$ such that

$$f_0(x) = \exp \left(\sum_{\lambda \in \mathbb{N}^*} \theta_{0\lambda} \phi_\lambda(x) - c(\theta_0) \right).$$

We denote $\Pi_{f_0, k}$ the projection operator on the vector space generated by $(\phi_\lambda)_{0 \leq \lambda \leq k}$ for the scalar product $\langle \cdot, \cdot \rangle_{f_0}$ and

$$\Delta_{\psi, k} = \psi - \Pi_{f_0, k} \psi = \tilde{\psi} - \Pi_{f_0, k} \tilde{\psi},$$

where $\tilde{\psi}$ is defined in (2.6). We expand the functions $\tilde{\psi}$ and $\Pi_{f_0, k} \tilde{\psi}$ on Φ :

$$\tilde{\psi}(x) = \sum_{\lambda \in \mathbb{N}} \tilde{\psi}_\lambda \phi_\lambda(x), \quad \Pi_{f_0, k} \tilde{\psi}(x) = \sum_{\lambda=0}^k \tilde{\psi}_{\Pi, \lambda} \phi_\lambda(x), \quad x \in [0, 1]$$

so that $(\tilde{\psi}_\lambda)_{\lambda \in \mathbb{N}}$ and $(\tilde{\psi}_{\Pi, \lambda})_{\lambda \leq k}$ denote the sequences of coefficients of the expansions of the functions $\tilde{\psi}$ and $\Pi_{f_0, k} \tilde{\psi}$ respectively. We finally note:

$$\tilde{\psi}_{\Pi}^{[k]} = (\tilde{\psi}_{\Pi, 1}, \dots, \tilde{\psi}_{\Pi, k}).$$

Now, we consider the sequence $(\epsilon_n)_n$ decreasing to zero defined in Theorem 3.1 (see Section 3). We use the sequence $L(n)$ introduced in Definition 1.1 for the case (PH) and, in the sequel, we set $L(n) = 1$ in the case (D) by convention. Using Definition 1.1, for all $a > 0$, there exists a constant

$l_0 > 0$ large enough so that $\mathbb{P}_p \left(k > \frac{l_0 n \epsilon_n^2}{L(n)} \right) \leq e^{-an \epsilon_n^2}$. From [9], it implies that there exists $c > 0$ and l_0 large enough such that

$$\mathbb{P}_0 \left[\mathbb{P}^\pi \left(k > \frac{l_0 n \epsilon_n^2}{L(n)} \middle| X^n \right) \leq e^{-cn \epsilon_n^2} \right] = 1 + o(1).$$

Define $l_n = l_0 n \epsilon_n^2 / L(n)$ in the case (PH). In the case (D) we set $l_n = k_n^*$. We have the following result.

THEOREM 2.1. *We assume that $\|\log(f_0)\|_\infty < \infty$ and $\log(f_0) \in \mathcal{B}_{p,q}^\gamma$ with $p \geq 2$, $1 \leq q \leq \infty$ and $\gamma > 1/2$ is such that*

$$\beta < 1/2 + \gamma \quad \text{if } p_* \leq 2 \quad \text{and} \quad \beta < \gamma + 1/p_* \quad \text{if } p_* > 2.$$

Let us also assume that the prior is defined as in Definition 1.1 and assume that for all $t \in \mathbb{R}$, for all $1 \leq k \leq l_n$, uniformly on

$$\left\{ \theta \in \mathbb{R}^k \quad \text{s.t.} \quad \sum_{\lambda=1}^k (\theta_\lambda - \theta_{0\lambda})^2 \leq \frac{(\log n)^3}{L(n)^2} \epsilon_n^2 \right\},$$

we have:

$$(2.7) \quad \frac{\pi_k(\theta)}{\pi_k \left(\theta - \frac{t \tilde{\psi}_{\frac{I}{n}}^{[k]}}{\sqrt{n}} \right)} = 1 + o(1)$$

and

$$(2.8) \quad \sup_{k \leq l_n} \left\{ \left\| \sum_{\lambda > k} \tilde{\psi}_\lambda \phi_\lambda \right\|_\infty + \sqrt{k} \left\| \sum_{\lambda > k} \tilde{\psi}_\lambda \phi_\lambda \right\|_2 \right\} = o \left(\frac{(\log n)^{-3}}{\sqrt{n} \epsilon_n^2} \right)$$

(replace $k \leq l_n$ with $k = l_n$ in the case (D)). Then, for all $z \in \mathbb{R}$,

$$(2.9) \quad \mathbb{P}^\pi \left[\sqrt{n} (\Psi(f) - \Psi(P_n)) \leq z | X^n \right] = \sum_k p(k | X^n) \Phi_{V_{0k}}(z + \mu_{n,k}) + o_{\mathbb{P}_0}(1),$$

where

$$\begin{aligned} - V_{0k} &= F_0(\tilde{\psi}^2) - F_0(\Delta_{\psi,k}^2), \\ - \mu_{n,k} &= \sqrt{n} F_0 \left(\Delta_{\psi,k} \sum_{\lambda \geq k+1} \theta_{0\lambda} \phi_\lambda \right) + G_n(\Delta_{\psi,k}). \end{aligned}$$

In the case (D), if

$$(2.10) \quad \sum_{\lambda > k_n^*} \tilde{\psi}_\lambda^2 = o \left(n^{\frac{2\gamma}{2\beta+1}-1} \right)$$

then

$$(2.11) \quad \mathbb{P}^\pi [\sqrt{n}(\Psi(f) - \Psi(P_n)) \leq z|X^n] = \Phi_{V_0}(z) + o_{\mathbb{P}_0}(1),$$

where $V_0 = F_0(\tilde{\psi}^2)$.

Note that condition (2.10) is satisfied if $\gamma > \beta + 1/2$ or if $\gamma > \beta$ and ψ is a piecewise constant function, or a smooth function like a continuously differentiable function in the case of the Fourier basis. The proof of Theorem 2.1 is given in Section 4.3. This result is a consequence of Theorem 2.2 depending on three assumptions (A1), (A2) and (A3). More precisely, Conditions (A1) and (A2) are verified using Theorem 3.1 so that (2.7) and (2.8) are required to prove assumption (A3). Condition (2.7) corresponds to the heuristics we have given in Section 2.1 and connects (A3) to a change of parametrization (see Section 2.3). Condition (2.8) requires some minimal smoothness on ψ through the decay to zero of its coefficients. These two extra conditions are rather mild as will be shown in the few examples below, so that quite generally, the posterior distribution of $\sqrt{n}(\Psi(f) - \Psi(P_n))$ is asymptotically a mixture of Gaussian distributions with variances $V_0 - F_0(\Delta_{\psi,k}^2)$ and mean values $-\mu_{n,k}$ with weights $p(k|X^n)$. To obtain an asymptotic Gaussian distribution with mean zero and variance V_0 it is necessary for $\mu_{n,k}$ to be small whenever $p(k|X^n)$ is not. This is satisfied in the case of a prior of type (D). In full generality, we have not proved that priors of type (PH) cannot lead to this result. Nevertheless we give below a counter-example for which the Bernstein-von Mises property is not satisfied in the case (PH) and we believe that in most cases, the asymptotic posterior distribution is either not a Gaussian distribution or it does not have the correct mean or variance. We also give a counter-example where the asymptotic normality with correct mean and variance is not satisfied in the case of a prior of type (D) when $\gamma < \beta$. We now discuss condition (2.7) in three different examples. For the sake of simplicity, we only consider the case $p = q = 2$.

COROLLARY 2.1. *Assume that $\log(f_0) \in W^\gamma$. Condition (2.7) is satisfied in the following cases:*

- g is the standard **Gaussian density** and $\gamma > \beta - 1/4$ for the case (PH), $\gamma > \beta - 1/2$ for the case (D).
- g is the **Laplace density** $g(x) \propto e^{-|x|}$ and $\gamma > \beta$ for the case (PH), $\gamma, \beta > 1/2$ for the case (D).
- g is a **Student density** $g(x) \propto (1 + x^2/d)^{-(d+1)/2}$ under the same conditions as for the Gaussian density.

Corollary 2.1 holds for any bounded function ψ . For the special case $\psi(x) = \mathbb{1}_{x \leq x_0}$, conditions on γ and β can be relaxed. In particular, in the case (PH), if g is the Laplace density, (2.7) is satisfied as soon as $\gamma > \beta - 1/2$. By choosing $1/2 < \beta \leq 1$, this is satisfied for any $\gamma > 1/2$ as imposed by Theorem 2.1. Note that in the case (PH), Theorem 3.1 implies that the posterior distribution concentrates with the adaptive minimax rate up to a logarithmic term, so that choosing β close to $1/2$ is not restrictive.

Interestingly, Theorem 2.1 shows that sieve models (increasing sequence of parametric models) have a mixed behavior between parametric and non-parametric models. Indeed if the posterior distribution puts most of its mass on k 's large enough, the posterior distribution has a Bernstein-von Mises property centered at the empirical (non-parametric MLE) estimator with the correct variance. On the contrary, if the posterior probability of small k 's is positive, then the posterior distribution is neither asymptotically Gaussian with the right centering, nor with the right variance. An extreme case corresponds to the situation where $F_0(\Delta_{\psi,k}^2) \neq o(1)$ under the posterior distribution, which is equivalent to

$$\exists k_0, \quad \text{s.t. } \forall \epsilon > 0 \quad \liminf_{n \rightarrow \infty} \mathbb{P}_0^n [\mathbb{P}^\pi [k_0 | X^n] > \epsilon] > 0.$$

For each fixed k , if $\inf_{\theta \in \mathbb{R}^k} K(f_0, f_\theta) > 0$, since the model is regular, there exists $c > 0$ such that $\mathbb{P}_0 [\mathbb{P}^\pi [k | X^n] > e^{-nc}] \rightarrow 1$. Therefore, $F_0(\Delta_{\psi,k}^2) \neq o(1)$ under the posterior distribution if there exists k_0 such that $\inf_{\theta \in \mathbb{R}^{k_0}} K(f_0, f_\theta) > 0$, i.e. if there exists $\theta_0 \in \mathbb{R}^{k_0}$ such that $f_0 = f_{\theta_0}$. In that case it can be proved that $\mathbb{P}^\pi [k_0 | X^n] = 1 + o_{\mathbb{P}_0}(1)$, see [4], and the posterior distribution of $\Psi(f)$ is asymptotically Gaussian with mean $\Psi(f_{\hat{\theta}_{k_0}})$, where $\hat{\theta}_{k_0}$ is the maximum likelihood estimator in \mathcal{F}_{k_0} , and the variance is the asymptotic variance of $\Psi(f_{\hat{\theta}_{k_0}})$. However, even if $\Delta_{\psi,k} = o(1)$, the posterior distribution might not satisfy the non-parametric Bernstein-von Mises property with the correct centering. See below for an illustration of these facts.

We illustrate this issue in the special case of the cumulative distribution function calculated at a given point x_0 : $\psi(x) = \mathbb{1}_{x \leq x_0}$. We recall that the variance of $G_n(\psi)$ under \mathbb{P}_0 is equal to $V_0 = F_0(x_0)(1 - F_0(x_0))$. We consider the case of the Fourier basis (the case of wavelet bases can be handled in the same way). Straightforward computations lead to the following result.

COROLLARY 2.2. *Assume that ψ is a piecewise constant function. Consider the prior defined in Section 1.2 in the case (D) with g being the Gaussian or the Laplace density. Then if $f_0 \in W^\gamma$, with $\gamma \geq \beta > 1/2$, the posterior distribution of $\sqrt{n}(F(x_0) - F_n(x_0))$ is asymptotically Gaussian with*

mean 0 and variance V_0 . If g is the Student density and if $\gamma \geq \beta > 1$, the same result holds.

We now illustrate the fact that when k is random, the Bernstein-von Mises property may be not valid.

PROPOSITION 2.1. *Let*

$$f_0(x) = \exp \left(\sum_{\lambda \geq k_0} \theta_{0\lambda} \phi_\lambda(x) - c(\theta_0) \right)$$

where k_0 is fixed and for any λ , $\theta_{0,2\lambda+1} = 0$ and

$$\theta_{0,2\lambda} = \frac{\sin(2\pi\lambda x_0)}{\lambda^{\gamma+1/2} \sqrt{\log \lambda \log \log \lambda}}.$$

Consider the prior defined in Section 1.2 with g being the Gaussian or the Laplace density but the prior p is now the Poisson distribution with parameter $\nu > 0$. If k_0 is large enough, there exists x_0 such that the posterior distribution of $\sqrt{n}(F(x_0) - F_n(x_0))$ is not asymptotically Gaussian with mean 0 and variance $F_0(x_0)(1 - F_0(x_0))$.

Actually, we prove that the asymptotic posterior distribution of $F(x_0) - F_n(x_0)$ is a mixture of Gaussian distributions with means $\mu_{n,k}$ and variance $F_0(x_0)(1 - F_0(x_0))/n$ and the support of the posterior distribution of k is included in $\{m \in \mathbb{N}^* \text{ s.t. } m \leq ck_n\}$ where c is a constant and k_n is defined in (4.23). Furthermore, we show that for all $k \leq ck_n$, $|\mu_{n,k}| \geq C\sqrt{\log n}$ for some positive constant C .

2.3. *Bernstein-von Mises theorem: general case.* To prove Theorem 2.1, we use a general result stated in this section. The subsequent theorem may deserve interest in its own right and can be used for other families of priors.

For each density function f , we define h such that for any x ,

$$h(x) = \sqrt{n} \log \left(\frac{f(x)}{f_0(x)} \right) \quad \text{or equivalently} \quad f(x) = f_0(x) \exp \left(\frac{h(x)}{\sqrt{n}} \right).$$

For the sake of clarity, we sometime write f_h instead of f and h_f instead of h to emphasize the relationship between f and h . Note that in this context h is not the score function since $F_0(h) \neq 0$. Then we consider the following assumptions.

(A1) The posterior distribution concentrates around f_0 . More precisely, there exists $u_n = o(1)$ such that if $A_{u_n}^1 = \{f \in \mathcal{F} \text{ s.t. } V(f_0, f) \leq u_n^2\}$ the posterior distribution of $A_{u_n}^1$ satisfies

$$\mathbb{P}^\pi \left\{ A_{u_n}^1 | X^n \right\} = 1 + o_{\mathbb{P}_0}(1).$$

(A2) The posterior distribution of the subset $A_n \subset A_{u_n}^1$ of densities such that

$$(2.12) \quad \int \left| \log \left(\frac{f(x)}{f_0(x)} \right) \right|^3 (f_0(x) + f(x)) dx = o(1)$$

satisfies

$$\mathbb{P}^\pi [A_n | X^n] = 1 + o_{\mathbb{P}_0}(1).$$

(A3) Let

$$R_n(h) = \sqrt{n}F_0(h) + \frac{F_0(h^2)}{2}$$

and for any x , for any t ,

$$\bar{\psi}_{t,n}(x) = \tilde{\psi}(x) + \frac{\sqrt{n}}{t} \log \left(F_0 \left[\exp \left(\frac{h}{\sqrt{n}} - \frac{t\tilde{\psi}}{\sqrt{n}} \right) \right] \right).$$

We have for any t ,

$$(2.13) \quad \frac{\int_{A_n} \exp \left(-\frac{F_0((h_f - t\bar{\psi}_{t,n})^2)}{2} + G_n(h_f - t\bar{\psi}_{t,n}) + R_n(h_f - t\bar{\psi}_{t,n}) \right) d\pi(f)}{\int_{A_n} \exp \left(-\frac{F_0(h_f^2)}{2} + G_n(h_f) + R_n(h_f) \right) d\pi(f)} = 1 + o_{\mathbb{P}_0}(1).$$

Before stating our general result, let us discuss these assumptions. Condition (A1) concerns concentration rates of the posterior distribution and there exists now a large literature on such results. See for instance [22] or [9] for general results. The difficulty here comes from the use of V instead of the Hellinger or the \mathbb{L}_1 distances. However since u_n does not need to be optimal, deriving rates in terms of V from those in terms of the Hellinger distance is often not a problem (see below).

Condition (A2) is a refinement of (A1) but can often be derived from (A1) as illustrated in the case of exponential families.

The main difficulty comes from condition (A3). Roughly speaking, the reason for (A3) can be glimpsed in the heuristic arguments given in Section 2.1, where computations made under the very strong uniform condition

(2.5) lead quite naturally to (2.4). This actually also helps to understand what (A3) means, i.e. the possibility of considering a change of parameter (transformation T) of the form $T(f_h) = f_{h-t\bar{\psi}_{t,n}}$, where $\bar{\psi}_{t,n}$ is of order $1/\sqrt{n}$, and such that the prior is hardly modified by this transformation. In parametric setups, continuity of the prior near the true value is enough to ensure that the prior would hardly be modified by such a transformation and this remains true in the semi-parametric setups where we can write the parameter as (θ, η) with θ the (finite dimensional) parameter of interest. Indeed as shown in [3], under certain conditions, the transformations on $f_{\theta,\eta}$ can be transferred to transformations on θ . Our setup is more complex since T applies on the infinite dimensional parameter f , so that a condition of the form $d\pi(T(f)) = d\pi(f)(1 + o(1))$ does not necessarily make sense.

Now, we can state the main result of this section.

THEOREM 2.2. *Let f_0 be a density on \mathcal{F} such that $\|\log(f_0)\|_\infty < \infty$. Assume that (A1), (A2) and (A3) are true. Then, we have for any z , in \mathbb{P}_0 -probability,*

$$(2.14) \quad \mathbb{P}^\pi \{ \sqrt{n}(\Psi(f) - \Psi(P_n)) \leq z | X^n \} - \Phi_{F_0(\bar{\psi}^2)}(z) \rightarrow 0.$$

The proof of Theorem 2.2 is given in Section 4.2. It is based on the asymptotic behavior of the Laplace transform of $\sqrt{n}(\Psi(f) - \Psi(P_n))\mathbb{1}_{A_n}$ calculated at the point t which is proved to be equivalent to $\exp(t^2 F_0(\bar{\psi}^2)/2)$ times the left hand side of (2.13) under (A1) and (A2), so that (A3) implies (2.14). We do not establish that (A3) is equivalent to (2.14) under (A1) and (A2) (the proof is based on the limit of the asymptotic behavior of the Laplace transform and not of the characteristic function), but we believe that it is close to being so.

3. Posterior rates for infinite dimensional exponential families.

Since one of the key conditions needed to obtain a Bernstein-von Mises theorem is a concentration rate of the posterior distribution of order ϵ_n , we now give two general results on concentration rates of posterior distributions based on the two different setups of orthonormal bases: the Fourier basis and the wavelet basis. These results have their own interest since we obtain in such contexts optimal adaptive rates of convergence. In a similar spirit [21] considers infinite dimensional exponential families and derives minimax and adaptive posterior concentration rates. Her work differs from the following theorem in two main aspects. Firstly she restricts her attention to the case of Sobolev spaces and Fourier basis, whereas we consider Besov spaces and secondly she obtains adaptivity by putting a prior on the smoothness of

the Sobolev class whereas we obtain adaptivity by constructing a prior on the size k of the parametric spaces, which to our opinion is a more natural approach. Moreover [21] merely considers Gaussian priors. Also related to this problem are the works of [13] and [10] who derive a general framework to obtain adaptive posterior concentration rates, the former applies her results to the Haar basis case. The limitation in her case, apart from the fact that she considers the Haar basis and no other wavelet basis is that she constraints the θ_λ 's in each k dimensional model to belong to a ball with fixed radius.

Note also that the family of priors defined in Section 1.2 has also been used in the infinitely many means model (equivalently in the white noise model) by [26] where minimax but non adaptive rates were obtained for the \mathbb{L}_2 -risk.

THEOREM 3.1. *We assume that $\| \log(f_0) \|_\infty < \infty$ and $\log(f_0) \in \mathcal{B}_{p,q}^\gamma$ with $p \geq 2$, $1 \leq q \leq \infty$ and $\gamma > 1/2$ is such that*

$$\beta < 1/2 + \gamma \quad \text{if } p_* \leq 2 \quad \text{and} \quad \beta < \gamma + 1/p_* \quad \text{if } p_* > 2.$$

Then,

$$(3.1) \quad \mathbb{P}^\pi \left\{ f_\theta \text{ s.t. } h(f_0, f_\theta) \leq \sqrt{\frac{\log n}{L(n)}} \epsilon_n | X^n \right\} = 1 + o_{\mathbb{P}_0}(1),$$

and

$$(3.2) \quad \mathbb{P}^\pi \left\{ f_\theta \text{ s.t. } \|\theta_0 - \theta\|_{\ell_2} \leq \log n \sqrt{\frac{\log n}{L(n)}} \epsilon_n | X^n \right\} = 1 + o_{\mathbb{P}_0}(1),$$

where in the case (PH),

$$\epsilon_n = \epsilon_0 \left(\frac{\log n}{n} \right)^{\frac{\gamma}{2\gamma+1}},$$

in the case (D), $L(n) = 1$,

$$\epsilon_n = \epsilon_0 (\log n) n^{-\frac{\beta}{2\beta+1}}, \quad \text{if } \gamma \geq \beta$$

$$\epsilon_n = \epsilon_0 n^{-\frac{\gamma}{2\beta+1}}, \quad \text{if } \gamma < \beta$$

and ϵ_0 is a constant large enough.

The proof of Theorem 3.1 is given in Section 4.1. If the density g only satisfies a tail condition of the form

$$g(x) \leq C_g |x|^{-p_*},$$

where C_g is a constant and $p_* > 1$, then, in the case (PH), if $\gamma > 1$, the rates defined by (3.1) and (3.2) remain valid. Note that in the case (PH) the posterior concentration rate is, up to a $\log n$ term, the minimax rate of convergence, whereas in the case (D) the minimax rate is achieved only when $\gamma = \beta$.

4. Proofs. In this section we prove results stated previously. We first prove Theorem 3.1. Then the proof of Theorem 2.2 is given. From these results, we finally deduce the proofs of Theorem 2.1 and of the related results. In the sequel, C denotes a generic positive constant whose value is of no importance and may change from line to line.

4.1. *Proof of Theorem 3.1.* We first give a preliminary lemma which will be extensively used in the sequel.

4.1.1. *Preliminary lemma.*

LEMMA 4.1. *Set $K_n = \{1, 2, \dots, k_n\}$ with $k_n \in \mathbb{N}^*$. Assume either of the following two cases:*

- $\gamma > 0$, $p = q = 2$ when Φ is the Fourier basis
- $0 < \gamma < r$, $2 \leq p \leq \infty$, $1 \leq q \leq \infty$ when Φ is the wavelet basis with r vanishing moments (see [12]).

Then the following results hold.

- There exists a constant $c_{1,\Phi}$ depending only on Φ such that for any $\theta = (\theta_\lambda)_\lambda \in \mathbb{R}^{k_n}$,

$$(4.1) \quad \left\| \sum_{\lambda \in K_n} \theta_\lambda \phi_\lambda \right\|_\infty \leq c_{1,\Phi} \sqrt{k_n} \|\theta\|_{\ell_2}.$$

- If $\log(f_0) \in \mathcal{B}_{p,q}^\gamma(R)$, then there exists $c_{2,\gamma}$ depending on γ only such that

$$(4.2) \quad \sum_{\lambda \notin K_n} \theta_{0\lambda}^2 \leq c_{2,\gamma} R^2 k_n^{-2\gamma}.$$

- If $\log(f_0) \in \mathcal{B}_{p,q}^\gamma(R)$ with $\gamma > \frac{1}{2}$, then there exists $c_{3,\Phi,\gamma}$ depending on Φ and γ only such that:

$$(4.3) \quad \left\| \sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda \right\|_\infty \leq c_{3,\Phi,\gamma} R k_n^{\frac{1}{2}-\gamma}.$$

PROOF. Let us first consider the Fourier basis. We have:

$$\begin{aligned} \left\| \sum_{\lambda \in K_n} \theta_\lambda \phi_\lambda \right\|_\infty &\leq \sum_{\lambda \in K_n} |\theta_\lambda| \times \|\phi_\lambda\|_\infty \\ &\leq \|\phi\|_\infty \sum_{\lambda \in K_n} |\theta_\lambda|, \end{aligned}$$

which proves (4.1). Inequality (4.2) follows from the definition of $\mathcal{B}_{2,2}^\gamma = W^\gamma$. To prove (4.3), we use the following inequality: for any x ,

$$\begin{aligned} \left| \sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda(x) \right| &\leq \|\phi\|_\infty \sum_{\lambda \notin K_n} |\theta_{0\lambda}| \\ &\leq \|\phi\|_\infty \left(\sum_{\lambda \notin K_n} |\lambda|^{2\gamma} \theta_{0\lambda}^2 \right)^{\frac{1}{2}} \left(\sum_{\lambda \notin K_n} |\lambda|^{-2\gamma} \right)^{\frac{1}{2}}. \end{aligned}$$

Now, we consider the wavelet basis. Without loss of generality, we assume that $\log_2(k_n + 1) \in \mathbb{N}^*$. We have for any x ,

$$\begin{aligned} \left| \sum_{\lambda \in K_n} \theta_\lambda \phi_\lambda(x) \right| &\leq \left(\sum_{\lambda \in K_n} \theta_\lambda^2 \right)^{\frac{1}{2}} \left(\sum_{\lambda \in K_n} \phi_\lambda^2(x) \right)^{\frac{1}{2}} \\ &\leq \|\theta\|_{\ell_2} \left(\sum_{-1 \leq j \leq \log_2(k_n)} \sum_{k < 2^j} \varphi_{jk}^2(x) \right)^{\frac{1}{2}}, \end{aligned}$$

with $\varphi_{-10} = \mathbb{1}_{[0,1]}$. Since $\varphi(x) = 0$ for $x \notin [-A, A]$, for $j \geq 0$,

$$\text{card} \left\{ k \in \{0, \dots, 2^j - 1\} \text{ s.t. } \varphi_{jk}(x) \neq 0 \right\} \leq 3(2A + 1).$$

(see [17], p. 282 or [18], p. 112). So, there exists c_φ depending only on φ such that

$$\left| \sum_{\lambda \in K_n} \theta_\lambda \phi_\lambda(x) \right| \leq \|\theta\|_{\ell_2} \left(\sum_{0 \leq j \leq \log_2(k_n)} 3(2A + 1) 2^j c_\varphi^2 \right)^{\frac{1}{2}},$$

which proves (4.1). For the second point, we just use the inclusion $\mathcal{B}_{p,q}^\gamma(R) \subset \mathcal{B}_{2,\infty}^\gamma(R)$ and

$$\sum_{\lambda \notin K_n} \theta_{0\lambda}^2 = \sum_{j > \log_2(k_n)} \sum_{k=0}^{2^j-1} \theta_{0jk}^2 \leq R^2 \sum_{j > \log_2(k_n)} 2^{-2j\gamma} \leq \frac{R^2}{1 - 2^{-2\gamma}} k_n^{-2\gamma}.$$

Finally, for the last point, we have for any x :

$$\begin{aligned} \left| \sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda(x) \right| &\leq \sum_{j > \log_2(k_n)} \left(\sum_{k=0}^{2^j-1} \theta_{0jk}^2 \right)^{\frac{1}{2}} \left(\sum_{k=0}^{2^j-1} \varphi_{jk}^2(x) \right)^{\frac{1}{2}} \\ &\leq C k_n^{\frac{1}{2}-\gamma}, \end{aligned}$$

where $C \leq R(3(2A+1))^{\frac{1}{2}} c_\varphi (1 - 2^{\frac{1}{2}-\gamma})^{-1}$. \square

4.1.2. *Proof of Theorem 3.1.* To prove Theorem 3.1, we use the following version of theorems on posterior convergence rates. Its proof is not given, since it is a slight modification of Theorem 2.4 of [9].

THEOREM 4.1. *Let f_0 be the true density and let π be a prior on \mathcal{F} satisfying the following conditions: There exist $(\epsilon_n)_n$ a positive sequence decreasing to zero with $n\epsilon_n^2 \rightarrow +\infty$ and a constant $c > 0$ such that for any n , there exists $\mathcal{F}_n^* \subset \mathcal{F}$ satisfying*

- (A)

$$\mathbb{P}^\pi \{ \mathcal{F}_n^{*c} \} = o(e^{-(c+2)n\epsilon_n^2}).$$

- (B) For any $j \in \mathbb{N}^*$, let

$$S_{n,j} = \{ f \in \mathcal{F}_n^* \text{ s.t. } j\epsilon_n < h(f_0, f) \leq (j+1)\epsilon_n \},$$

and $H_{n,j}$ the Hellinger metric entropy of $S_{n,j}$. There exists $J_{0,n}$ (that may depend on n) such that for all $j \geq J_{0,n}$,

$$H_{n,j} \leq (K-1)nj^2\epsilon_n^2,$$

where K is an absolute constant.

- (C) Let

$$B_n(\epsilon_n) = \{ f \in \mathcal{F} \text{ s.t. } K(f_0, f) \leq \epsilon_n^2, V(f_0, f) \leq \epsilon_n^2 \}.$$

Then,

$$\mathbb{P}^\pi \{ B_n(\epsilon_n) \} \geq e^{-cn\epsilon_n^2}.$$

We have:

$$\mathbb{P}^\pi \{ f \text{ s.t. } h(f_0, f) \leq J_{0,n}\epsilon_n | X^n \} = 1 + o_{\mathbb{P}_0}(1).$$

To prove Theorem 3.1 it is thus enough to verify conditions (A), (B) and (C). We consider $(\Lambda_n)_n$ the increasing sequence of subsets of \mathbb{N}^* defined by $\Lambda_n = \{1, 2, \dots, l_n\}$ with $l_n \in \mathbb{N}^*$. For any n , we set:

$$\mathcal{F}_n^* = \left\{ f_\theta \in \mathcal{F}_{l_n} \text{ s.t. } f_\theta = \exp \left(\sum_{\lambda \in \Lambda_n} \theta_\lambda \phi_\lambda - c(\theta) \right), \|\theta\|_{\ell_2} \leq w_n \right\},$$

with

$$w_n = \exp(w_0 n^\rho (\log n)^q), \quad \rho > 0, \quad q \in \mathbb{R}.$$

Recall that

- $\epsilon_n = \epsilon_0 n^{-\frac{\gamma}{2\gamma+1}} (\log n)^{\frac{\gamma}{2\gamma+1}}$ and $l_n = \frac{l_0 n \epsilon_n^2}{L(n)}$ in the case (PH),
- $\epsilon_n = \epsilon_0 n^{-\frac{\beta}{2\beta+1}}$ and $l_n = k_n^* = n^{\frac{1}{2\beta+1}}$ in the case (D).

Condition (A):. Since $\beta > 1/2$, $\sum_\lambda \tau_\lambda < \infty$ and for the sake of simplicity, without loss of generality, we assume that $\sum_\lambda \tau_\lambda \leq 1$. Using the tail assumption on g ,

$$\begin{aligned} \pi \{ \mathcal{F}_n^{*c} \} &\leq \sum_{\lambda > l_n} p(\lambda) + \mathbb{P}^\pi \left\{ \sum_{\lambda \leq l_n} \theta_\lambda^2 > w_n^2 \right\} \\ &\leq C \exp(-c_2 l_n L(l_n)) + \sum_{\lambda \leq l_n} \mathbb{P}^\pi \left\{ \frac{\theta_\lambda^2}{\tau_\lambda} > w_n^2 \right\} \\ &\leq C \exp(-c_2 l_0 n \epsilon_n^2) + \sum_{\lambda \leq l_n} \mathbb{P}^\pi \left\{ \exp\left(\frac{c_* |\theta_\lambda|^{p_*}}{2 \tau_\lambda^{p_*/2}}\right) > \exp\left(\frac{c_* w_n^{p_*}}{2}\right) \right\} \\ &\leq C \exp(-c_2 l_0 n \epsilon_n^2) + C l_n \exp\left(-\frac{c_* w_n^{p_*}}{2}\right) \\ &\leq C \exp(-c_2 l_0 n \epsilon_n^2) + C \exp(-n^H) \end{aligned}$$

for any positive $H > 0$. Hence,

$$\pi \{ \mathcal{F}_n^{*c} \} \leq C \exp(-c_2 (l_0 - 1) n \epsilon_n^2)$$

and Condition (A) is proved for l_0 large enough.

Condition (B):. We apply Lemma 4.1 with $K_n = \Lambda_n$ and $k_n = l_n$. For this purpose, we show that the Hellinger distance between two functions of \mathcal{F}_n^* is related to the ℓ_2 -distance of the associated coefficients. So, let us consider f_θ and $f_{\theta'}$ belonging to \mathcal{F}_n^* with

$$f_\theta = \exp\left(\sum_{\lambda \in \Lambda_n} \theta_\lambda \phi_\lambda - c(\theta)\right), \quad f_{\theta'} = \exp\left(\sum_{\lambda \in \Lambda_n} \theta'_\lambda \phi_\lambda - c(\theta')\right).$$

Let us assume that $\|\theta' - \theta\|_{\ell_1} \leq \tilde{c}_1 \epsilon_n l_n^{-1/2}$ with \tilde{c}_1 a positive constant. Then,

$$\left\| \sum_{\lambda \in \Lambda_n} (\theta'_\lambda - \theta_\lambda) \phi_\lambda \right\|_\infty \leq C \sqrt{l_n} \|\theta' - \theta\|_{\ell_2} \leq C \sqrt{l_n} \|\theta' - \theta\|_{\ell_1} \leq C \tilde{c}_1 \epsilon_n \rightarrow 0$$

and

$$\begin{aligned}
|c(\theta) - c(\theta')| &= \left| \log \left(\int_0^1 f_\theta(x) \exp \left(\sum_{\lambda \in \Lambda_n} (\theta'_\lambda - \theta_\lambda) \phi_\lambda(x) \right) dx \right) \right| \\
&\leq \left| \log \left(1 + C \left\| \sum_{\lambda \in \Lambda_n} (\theta'_\lambda - \theta_\lambda) \phi_\lambda \right\|_\infty \right) \right| \\
&\leq C \left\| \sum_{\lambda \in \Lambda_n} (\theta'_\lambda - \theta_\lambda) \phi_\lambda \right\|_\infty.
\end{aligned}$$

Then,

$$\begin{aligned}
h^2(f_\theta, f_{\theta'}) &= \int f_\theta(x) \left(\exp \left(\frac{1}{2} \sum_{\lambda \in \Lambda_n} (\theta'_\lambda - \theta_\lambda) \phi_\lambda(x) + \frac{1}{2} (c(\theta) - c(\theta')) \right) - 1 \right)^2 dx \\
&\leq \int_0^1 f_\theta(x) \left(\exp \left(C \left\| \sum_{\lambda \in \Lambda_n} (\theta'_\lambda - \theta_\lambda) \phi_\lambda \right\|_\infty \right) - 1 \right)^2 dx \\
&\leq C \sum_{\lambda \in \Lambda_n} (\theta_\lambda - \theta'_\lambda) \phi_\lambda \Big|_\infty^2 \\
(4.4) \leq Cl_n \|\theta - \theta'\|_{\ell_1}^2 &\leq Cl_n^2 \|\theta - \theta'\|_{\ell_2}^2.
\end{aligned}$$

The next lemma establishes a converse inequality.

LEMMA 4.2. *Let $c_0 = \inf_{x \in [0,1]} f_0(x) > 0$. There exists a constant $c \leq 1/2$ depending on γ, β, R and Φ such that if*

$$(j+1)^2 \epsilon_n^2 l_n \leq c \times \min \left(c_0, (1 - e^{-1})^2 \right)$$

then for $f_\theta \in S_{n,j}$,

$$\|\theta_0 - \theta\|_{\ell_2}^2 \leq \frac{1}{c_0 c} (\log n)^2 h^2(f_0, f_\theta).$$

PROOF. Using Theorem 5 of [25], with $M_1 = \left(\int_0^1 \frac{f_0^2(x)}{f_\theta(x)} dx \right)^{\frac{1}{2}}$, if

$$h^2(f_0, f_\theta) \leq \frac{1}{2} (1 - e^{-1})^2,$$

we have

$$(4.5) \quad V(f_0, f_\theta) \leq 5h^2(f_0, f_\theta) (|\log M_1| - \log(h(f_0, f_\theta)))^2.$$

But

$$\begin{aligned} M_1 &= \int_0^1 f_0(x) \exp \left(\sum_{\lambda \in \Lambda_n} (\theta_{0\lambda} - \theta_\lambda) \phi_\lambda(x) + \sum_{\lambda \notin \Lambda_n} \theta_{0\lambda} \phi_\lambda(x) - c(\theta_0) + c(\theta) \right) dx \\ &\leq \int_0^1 f_0(x) \exp \left(C[\sqrt{l_n} \|\theta_0 - \theta\|_{\ell_2} + R l_n^{\frac{1}{2}-\gamma}] - c(\theta_0) + c(\theta) \right) dx, \end{aligned}$$

by using (4.1) and (4.3). Furthermore,

$$(4.6) \quad |c(\theta_0) - c(\theta)| \leq C[\sqrt{l_n} \|\theta_0 - \theta\|_{\ell_2} + R l_n^{\frac{1}{2}-\gamma}].$$

So,

$$|\log M_1| \leq C[\sqrt{l_n} \|\theta_0 - \theta\|_{\ell_2} + R l_n^{\frac{1}{2}-\gamma}].$$

Finally, since $f_\theta \in S_{n,j}$ for $j \geq 1$,

$$\begin{aligned} V(f_0, f_\theta) &\leq 5h^2(f_0, f_\theta) \left(C[\sqrt{l_n} \|\theta_0 - \theta\|_{\ell_2} + R l_n^{\frac{1}{2}-\gamma}] - \log(\epsilon_n) \right)^2 \\ &\leq Ch^2(f_0, f_\theta) \left(l_n \|\theta_0 - \theta\|_{\ell_2}^2 + (\log n)^2 \right). \end{aligned}$$

Since $f_0(x) \geq c_0$ for any x and $\int_0^1 \phi_\lambda(x) dx = 0$ for any $\lambda \in \Lambda_n$, we have

$$(4.7) \quad V(f_0, f_\theta) \geq c_0 \|\theta_0 - \theta\|_{\ell_2}^2.$$

Combining (4.4) and (4.7), we conclude that

$$\|\theta_0 - \theta\|_{\ell_2}^2 \leq 2Cc_0^{-1}(\log n)^2 h^2(f_0, f_\theta),$$

if $h^2(f_0, f_\theta) l_n \leq (j+1)^2 \epsilon_n^2 l_n \leq c_0/(2C)$. Lemma 4.2 is proved by taking $c = (\max(C, 1))^{-1}/2$. \square

Now, under the assumptions of Lemma 4.2, using (4.4), we obtain

$$H_{n,j} \leq \log \left((Cl_n(j+1) \log n)^{l_n} \right) \leq l_n \log \left(C\epsilon_n^{-1} \sqrt{l_n} \log n \right).$$

Then, we have $H_{n,j} \leq (K-1)nj^2\epsilon_n^2$ as soon as $j \geq J_{0,n} = \sqrt{j_0 \log n L(n)^{-1}}$, where j_0 is a constant and condition (B) is satisfied for such j 's. Now, let j such that

$$(4.8) \quad c(j+1)^2 \epsilon_n^2 l_n > \min \left(\frac{c_0}{2}, \frac{1}{2}(1 - e^{-1})^2 \right).$$

In this case, since for $f_\theta \in \mathcal{F}_n^*$,

$$\|\theta\|_{\ell_1} \leq \sqrt{l_n} \|\theta\|_{\ell_2} \leq \sqrt{l_n} w_n,$$

for n large enough,

$$H_{n,j} \leq \log \left(\left(C l_n w_n \epsilon_n^{-1} \right)^{l_n} \right) \leq 2l_n \log(w_n) \leq 2w_0 l_n n^\rho (\log n)^q.$$

Choosing w_0 , q and ρ small enough such that $l_n^2 (\log n)^q \leq n^{1-\rho}$, together with (4.8), implies condition (B).

Condition (C). Let $k_n \in \mathbb{N}$ increasing to ∞ and $K_n = \{1, \dots, k_n\}$, define

$$A(u_n) = \left\{ \theta \text{ s.t. } \theta_\lambda = 0 \text{ for every } \lambda \notin K_n \text{ and } \sum_{\lambda \in K_n} (\theta_{0\lambda} - \theta_\lambda)^2 \leq u_n^2 \right\},$$

where u_n goes to 0 such that

$$(4.9) \quad \sqrt{k_n} u_n \rightarrow 0.$$

We define for any λ ,

$$\beta_\lambda(f_0) = \int_0^1 \phi_\lambda(x) f_0(x) dx.$$

Denote

$$f_{0K_n} = \exp \left(\sum_{\lambda \in K_n} \theta_{0\lambda} \phi_\lambda(x) - c(\theta_{0K_n}) \right), \quad f_{0\bar{K}_n} = \exp \left(\sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda(x) - c(\theta_{0\bar{K}_n}) \right).$$

We have

$$\begin{aligned} K(f_0, f_{0K_n}) &= \sum_{\lambda \notin K_n} \theta_{0\lambda} \beta_\lambda(f_0) + c(\theta_{0K_n}) - c(\theta_0) \\ &= \sum_{\lambda \notin K_n} \theta_{0\lambda} \beta_\lambda(f_0) + \log \left(\int_0^1 f_0(x) e^{-\sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda(x)} dx \right). \end{aligned}$$

Using inequality (4.3) of Lemma 4.1, we obtain

$$\begin{aligned} &\int_0^1 f_0(x) e^{-\sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda(x)} dx \\ &= 1 - \sum_{\lambda \notin K_n} \theta_{0\lambda} \beta_\lambda(f_0) + \frac{1}{2} \int_0^1 f_0(x) \left(\sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda(x) \right)^2 dx \times (1 + o(1)). \end{aligned}$$

We have

$$\left| \sum_{\lambda \notin K_n} \theta_{0\lambda} \beta_\lambda(f_0) \right| \leq \|f_0\|_2 \left(\sum_{\lambda \notin K_n} \theta_{0\lambda}^2 \right)^{\frac{1}{2}}$$

and

$$\int_0^1 f_0(x) \left(\sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda(x) \right)^2 dx \leq \|f_0\|_\infty \sum_{\lambda \notin K_n} \theta_{0\lambda}^2$$

So,

$$\begin{aligned} \log \left(\int_0^1 f_0(x) e^{-\sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda(x)} dx \right) &= - \sum_{\lambda \notin K_n} \theta_{0\lambda} \beta_\lambda(f_0) - \frac{1}{2} \left(\sum_{\lambda \notin K_n} \theta_{0\lambda} \beta_\lambda(f_0) \right)^2 \\ &\quad + \frac{1}{2} \int_0^1 f_0(x) \left(\sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda(x) \right)^2 dx + o \left(\sum_{\lambda \notin K_n} \theta_{0\lambda}^2 \right), \end{aligned}$$

and

$$K(f_0, f_{0K_n}) = \frac{1}{2} \int_0^1 f_0(x) \left(\sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda(x) \right)^2 dx - \frac{1}{2} \left(\sum_{\lambda \notin K_n} \theta_{0\lambda} \beta_\lambda(f_0) \right)^2 + o \left(\sum_{\lambda \notin K_n} \theta_{0\lambda}^2 \right).$$

This implies that for n large enough,

$$K(f_0, f_{0K_n}) \leq \|f_0\|_\infty \sum_{\lambda \notin K_n} \theta_{0\lambda}^2 \leq C k_n^{-2\gamma}.$$

Now, if $\theta \in A(u_n)$ we have

$$\begin{aligned} K(f_0, f_\theta) &= K(f_0, f_{0K_n}) + \sum_{\lambda \in K_n} (\theta_{0\lambda} - \theta_\lambda) \beta_\lambda(f_0) - c(\theta_{0K_n}) + c(\theta) \\ &\leq C k_n^{-2\gamma} + \sum_{\lambda \in K_n} (\theta_{0\lambda} - \theta_\lambda) \beta_\lambda(f_0) - c(\theta_{0K_n}) + c(\theta). \end{aligned}$$

We set for any x , $T(x) = \sum_{\lambda \in K_n} (\theta_\lambda - \theta_{0\lambda}) \phi_\lambda(x)$. Using (4.1), $\|T\|_\infty \leq C \sqrt{k_n} u_n \rightarrow 0$. So,

$$\int_0^1 f_{0K_n}(x) \exp(T(x)) dx = 1 + \int_0^1 f_{0K_n}(x) T(x) dx + \int_0^1 f_{0K_n}(x) T^2(x) v(n, x) dx,$$

where v is a bounded function. Since $\log(1 + u) \leq u$ for any $u > -1$, for $\theta \in A(u_n)$ and n large enough,

$$\begin{aligned} -c(\theta_{0K_n}) + c(\theta) &= \log \left(\int_0^1 f_{0K_n}(x) e^{T(x)} dx \right) \\ &\leq \int_0^1 f_{0K_n}(x) T(x) dx + \int_0^1 f_{0K_n}(x) T^2(x) v(n, x) dx \\ &\leq \sum_{\lambda \in K_n} (\theta_\lambda - \theta_{0\lambda}) \beta_\lambda(f_{0K_n}) + C k_n u_n^2. \end{aligned}$$

So,

$$\begin{aligned} K(f_0, f_\theta) &\leq C k_n^{-2\gamma} + \sum_{\lambda \in K_n} (\theta_{0\lambda} - \theta_\lambda) (\beta_\lambda(f_0) - \beta_\lambda(f_{0K_n})) \\ &\leq C k_n^{-2\gamma} + u_n \|f_0 - f_{0K_n}\|_2. \end{aligned}$$

Besides (4.3) implies

$$\|f_0 - f_{0K_n}\|_2^2 \leq \|f_0\|_\infty^2 \int_0^1 \left(1 - \exp \left(- \sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda(x) - c(\theta_{0K_n}) + c(\theta_0) \right) \right)^2 dx$$

and

$$|c(\theta_{0K_n}) - c(\theta_0)| \leq \left\| \sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda \right\|_\infty.$$

Finally,

$$\|f_0 - f_{0K_n}\|_2 \leq C \left\| \sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda \right\|_\infty \leq C k_n^{\frac{1}{2}-\gamma}$$

and

$$(4.10) \quad K(f_0, f_\theta) \leq C k_n^{-2\gamma} + C u_n k_n^{\frac{1}{2}-\gamma}.$$

We now bound $V(f_0, f_\theta)$. For this purpose, we refine the control of $|c(\theta_{0K_n}) - c(\theta_0)|$:

$$\begin{aligned} |c(\theta_{0K_n}) - c(\theta_0)| &= \left| \log \left(\int_0^1 f_0(x) \exp \left(- \sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda(x) \right) dx \right) \right| \\ &= \left| \log \int_0^1 f_0(x) \left(1 - \sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda(x) + w(n, x) \left(\sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda(x) \right)^2 \right) dx \right|, \end{aligned}$$

where w is a bounded function. So,

$$\begin{aligned} |c(\theta_{0K_n}) - c(\theta_0)| &\leq C \left(\sum_{\lambda \notin K_n} |\theta_{0\lambda} \beta_\lambda(f_0)| + \int_0^1 \left(\sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda(x) \right)^2 dx \right) \\ &\leq C \left(\sum_{\lambda \notin K_n} \theta_{0\lambda}^2 \right)^{\frac{1}{2}} \leq C k_n^{-\gamma}. \end{aligned}$$

In addition,

$$\begin{aligned} |c(\theta_{0K_n}) - c(\theta)| &\leq \sum_{\lambda \in K_n} |\theta_\lambda - \theta_{0\lambda}| |\beta_\lambda(f_{0K_n})| + C k_n u_n^2 \\ &\leq u_n (\|f_0 - f_{0K_n}\|_2 + \|f_0\|_2) + C k_n u_n^2 \\ &\leq C u_n + C k_n u_n^2. \end{aligned}$$

Finally,

$$(4.11) \quad V(f_0, f_\theta) \leq u_n^2 + C k_n^{-2\gamma} + C k_n^2 u_n^4.$$

Now, let us consider the case (PH). We take k_n and u_n such that

$$(4.12) \quad k_n = k_0 \epsilon_n^{-1/\gamma} \quad \text{and} \quad u_n = u_0 \epsilon_n k_n^{-\frac{1}{2}},$$

where k_0 and u_0 are constants depending on $\|f_0\|_\infty$, γ , R and Φ . Note that (4.9) is then satisfied. If ϵ_0 is large enough and u_0 is small enough, then, by using (4.10) and (4.11),

$$K(f_0, f_\theta) \leq \epsilon_n^2 \quad \text{and} \quad V(f_0, f_\theta) \leq \epsilon_n^2.$$

So, Condition (C) is satisfied if

$$\mathbb{P}^\pi \{A(u_n)\} \geq e^{-c n \epsilon_n^2},$$

where $A(u_n)$ is defined in (4.9). We have:

$$\mathbb{P}^\pi \{A(u_n)\} \geq \mathbb{P}^\pi \left\{ \theta \text{ s.t. } \sum_{\lambda \in K_n} (\theta_\lambda - \theta_{0\lambda})^2 \leq u_n^2 \middle| k_n \right\} \times \exp(-c_1 k_n L(k_n)).$$

The prior on θ implies that, with $G_\lambda = \lambda^\beta \theta_\lambda \tau_0^{-1/2}$,

$$\begin{aligned}
P_1 &= \mathbb{P}^\pi \left\{ \theta \text{ s.t. } \sum_{\lambda \in K_n} (\theta_\lambda - \theta_{0\lambda})^2 \leq u_n^2 \middle| k_n \right\} \\
&\geq \mathbb{P}^\pi \left\{ \theta \text{ s.t. } \sum_{\lambda \in K_n} \left| \sqrt{\tau_0} \lambda^{-\beta} G_\lambda - \theta_{0\lambda} \right| \leq u_n \middle| k_n \right\} \\
&= \mathbb{P}^\pi \left\{ \theta \text{ s.t. } \sum_{\lambda \in K_n} \lambda^{-\beta} \left| G_\lambda - \tau_0^{-\frac{1}{2}} \lambda^\beta \theta_{0\lambda} \right| \leq \tau_0^{-\frac{1}{2}} u_n \middle| k_n \right\} \\
&= \int \dots \int 1_{\left\{ \sum_{\lambda \in K_n} \lambda^{-\beta} \left| x_\lambda - \tau_0^{-\frac{1}{2}} \lambda^\beta \theta_{0\lambda} \right| \leq \tau_0^{-\frac{1}{2}} u_n \right\}} \prod_{\lambda \in K_n} g(x_\lambda) dx_\lambda \\
&\geq \int \dots \int 1_{\left\{ \sum_{\lambda \in K_n} \lambda^{-\beta} |y_\lambda| \leq \tau_0^{-\frac{1}{2}} u_n \right\}} \prod_{\lambda \in K_n} g\left(y_\lambda + \tau_0^{-\frac{1}{2}} \lambda^\beta \theta_{0\lambda}\right) dy_\lambda.
\end{aligned}$$

When $\gamma \geq \beta$, we have $\sup_{\lambda \in K_n} \left| \tau_0^{-\frac{1}{2}} \lambda^\beta \theta_{0\lambda} \right| < \infty$ and $\sup_n \left\{ \tau_0^{-\frac{1}{2}} k_n^\beta u_n \right\} < \infty$. Using assumptions on the prior, there exists a constant D such that

$$\begin{aligned}
P_1 &\geq D^{k_n} \int \dots \int 1_{\left\{ \sum_{\lambda \in K_n} \lambda^{-\beta} |y_\lambda| \leq \tau_0^{-\frac{1}{2}} u_n \right\}} \prod_{\lambda \in K_n} dy_\lambda \\
(4.13) \quad &\geq \exp(-Ck_n \log n).
\end{aligned}$$

When $\gamma < \beta$, there exist a and $b > 0$ such that $\forall |y| \leq M$ for some positive constant M

$$g(y + u) \geq a \exp(-b|u|^{p_*}).$$

Using the above calculations we obtain if $p_* \leq 2$

$$\begin{aligned}
P_1 &\geq D^{k_n} \exp\left\{-C \sum_{\lambda \in K_n} \lambda^{p_*\beta} |\theta_{0\lambda}|^{p_*}\right\} \int \dots \int 1_{\left\{ \sum_{\lambda \in K_n} \lambda^{-\beta} |y_\lambda| \leq \tau_0^{-\frac{1}{2}} u_n \right\}} \prod_{\lambda \in K_n} dy_\lambda \\
&\geq \exp\left[-Ck_n^{1-p_*/2+p_*(\beta-\gamma)}\right] \exp(-Ck_n \log n) \\
&\geq \exp(-Ck_n \log n) \quad \text{if } \beta \leq 1/2 + \gamma
\end{aligned}$$

and if $p_* > 2$, $\sum_{\lambda \in K_n} \lambda^{p_*\beta} |\theta_{0\lambda}|^{p_*} \leq k_n^{p_*\beta - p_*\gamma}$ so that

$$\begin{aligned}
P_1 &\geq D^{k_n} \exp\left\{-C \sum_{\lambda \in K_n} \lambda^{p_*\beta} |\theta_{0\lambda}|^{p_*}\right\} \exp(-Ck_n \log n) \\
&\geq \exp(-Ck_n \log n) \quad \text{if } \beta \leq \gamma + 1/p_*.
\end{aligned}$$

Condition (C) is established by choosing k_0 small enough. Similar computations lead to the result in the case (D).

4.2. *Proof of Theorem 2.2.* Let $Z_n = \sqrt{n}(\Psi(f) - \Psi(P_n))$. We have

$$(4.14) \quad \mathbb{P}^\pi \{A_n | X^n\} = 1 + o_{\mathbb{P}_0}(1).$$

So, it is enough to prove that conditionally on A_n and X^n , the distribution of Z_n converges to the distribution of a Gaussian variable whose variance is $F_0(\tilde{\psi}^2)$. This will be established if for any $t \in \mathbb{R}$,

$$(4.15) \quad \lim_{n \rightarrow +\infty} L_n(t) = \exp\left(\frac{t^2}{2} F_0[\tilde{\psi}^2]\right),$$

where $L_n(t)$ is the Laplace transform of Z_n conditionally on A_n and X^n :

$$\begin{aligned} L_n(t) &= \mathbb{E}^\pi [\exp(t\sqrt{n}(\Psi(f) - \Psi(P_n))) | A_n, X^n] \\ &= \frac{\mathbb{E}^\pi [\exp(t\sqrt{n}(\Psi(f) - \Psi(P_n))) \mathbb{1}_{A_n}(f) | X^n]}{\mathbb{P}^\pi \{A_n | X^n\}} \\ &= \frac{\int_{A_n} \exp(t\sqrt{n}(\Psi(f) - \Psi(P_n)) + \ell_n(f) - \ell_n(f_0)) d\pi(f)}{\int_{A_n} \exp(\ell_n(f) - \ell_n(f_0)) d\pi(f)}. \end{aligned}$$

We set for any x ,

$$(4.16) \quad B_{h,n}(x) = \int_0^1 (1-u) e^{uh(x)/\sqrt{n}} du.$$

So,

$$\exp\left(\frac{h(x)}{\sqrt{n}}\right) = 1 + \frac{h(x)}{\sqrt{n}} + \frac{h^2(x)}{n} B_{h,n}(x),$$

which implies that

$$f(x) - f_0(x) = f_0(x) \left(\frac{h(x)}{\sqrt{n}} + \frac{h^2(x)}{n} B_{h,n}(x) \right)$$

and

$$\begin{aligned} t\sqrt{n}(\Psi(f) - \Psi(P_n)) &= -tG_n(\tilde{\psi}) + t\sqrt{n} \left(\int \tilde{\psi}(x)(f(x) - f_0(x)) dx \right) \\ &= -tG_n(\tilde{\psi}) + tF_0(h\tilde{\psi}) + \frac{t}{\sqrt{n}} F_0(h^2 B_{h,n}\tilde{\psi}). \end{aligned}$$

Since

$$\ell_n(f) - \ell_n(f_0) = -\frac{F_0(h^2)}{2} + G_n(h) + R_n(h),$$

we have

$$\begin{aligned} L_n(t) &= \frac{\int_{A_n} \exp\left(G_n(h - t\tilde{\psi}) + tF_0(h\tilde{\psi}) + \frac{t}{\sqrt{n}}F_0(h^2B_{h,n}\tilde{\psi}) - \frac{F_0(h^2)}{2} + R_n(h)\right) d\pi(f)}{\int_{A_n} \exp\left(-\frac{F_0(h^2)}{2} + G_n(h) + R_n(h)\right) d\pi(f)} \\ &= \frac{\int_{A_n} \exp\left(-\frac{F_0((h-t\bar{\psi}_{t,n})^2)}{2} + G_n(h - t\bar{\psi}_{t,n}) + R_n(h - t\bar{\psi}_{t,n}) + U_{n,h}\right) d\pi(f)}{\int_{A_n} \exp\left(-\frac{F_0(h^2)}{2} + G_n(h) + R_n(h)\right) d\pi(f)}, \end{aligned}$$

where straightforward computations show that

$$\begin{aligned} U_{n,h} &= tF_0(h(\tilde{\psi} - \bar{\psi}_{t,n})) + \frac{t^2}{2}F_0(\bar{\psi}_{t,n}^2) + R_n(h) - R_n(h - t\bar{\psi}_{t,n}) + \frac{t}{\sqrt{n}}F_0(h^2B_{h,n}\tilde{\psi}) \\ &= tF_0(h\tilde{\psi}) + t\sqrt{n}F_0(\bar{\psi}_{t,n}) + \frac{t}{\sqrt{n}}F_0(h^2B_{h,n}\tilde{\psi}) \\ &= tF_0(h\tilde{\psi}) + n \log\left(F_0\left[\exp\left(\frac{h}{\sqrt{n}} - \frac{t\tilde{\psi}}{\sqrt{n}}\right)\right]\right) + \frac{t}{\sqrt{n}}F_0(h^2B_{h,n}\tilde{\psi}). \end{aligned}$$

Now, let us study each term of the last expression. We have

$$\begin{aligned} F_0\left[\exp\left(\frac{h}{\sqrt{n}} - \frac{t\tilde{\psi}}{\sqrt{n}}\right)\right] &= F_0\left[e^{\frac{h}{\sqrt{n}}}\left(1 - \frac{t\tilde{\psi}}{\sqrt{n}} + \frac{t^2}{2n}\tilde{\psi}^2\right)\right] + o(n^{-\frac{3}{2}}) \\ &= 1 - \frac{t}{\sqrt{n}}F_0\left[e^{\frac{h}{\sqrt{n}}}\tilde{\psi}\right] + \frac{t^2}{2n}F_0\left[e^{\frac{h}{\sqrt{n}}}\tilde{\psi}^2\right] + o(n^{-\frac{3}{2}}). \end{aligned}$$

So,

$$F_0\left[e^{\frac{h}{\sqrt{n}}}\tilde{\psi}\right] = \frac{F_0[h\tilde{\psi}]}{\sqrt{n}} + \frac{F_0[h^2B_{h,n}\tilde{\psi}]}{n}; \quad F_0\left[e^{\frac{h}{\sqrt{n}}}\tilde{\psi}^2\right] = F_0[\tilde{\psi}^2] + \frac{F_0[h\tilde{\psi}^2]}{\sqrt{n}} + \frac{F_0[h^2B_{h,n}\tilde{\psi}^2]}{n}.$$

Note that, on A_n , we have $F_0(h^2) = 0(nu_n^2)$ and $F_0(h^2B_{h,n}) = o(n)$. Therefore, uniformly on A_n ,

$$\begin{aligned} F_0\left[\exp\left(\frac{h}{\sqrt{n}} - \frac{t\tilde{\psi}}{\sqrt{n}}\right)\right] &= 1 - \frac{t}{\sqrt{n}}\left(\frac{F_0[h\tilde{\psi}]}{\sqrt{n}} + \frac{F_0[h^2B_{h,n}\tilde{\psi}]}{n}\right) \\ &\quad + \frac{t^2}{2n}\left(F_0[\tilde{\psi}^2] + \frac{F_0[h\tilde{\psi}^2]}{\sqrt{n}} + \frac{F_0[h^2B_{h,n}\tilde{\psi}^2]}{n}\right) + o(n^{-1}) \\ &= 1 - \frac{t}{n}\left[F_0[h\tilde{\psi}] + \frac{F_0[h^2B_{h,n}\tilde{\psi}]}{\sqrt{n}} - \frac{tF_0(\tilde{\psi}^2)}{2} + o(1)\right] \\ &= 1 + o(n^{-1/2}) \end{aligned}$$

and

$$n \log \left(F_0 \left[\exp \left(\frac{h}{\sqrt{n}} - \frac{t\tilde{\psi}}{\sqrt{n}} \right) \right] \right) = -t \left[F_0(h\tilde{\psi}) + \frac{F_0[h^2 B_{h,n}\tilde{\psi}]}{\sqrt{n}} - \frac{tF_0(\tilde{\psi}^2)}{2} \right] + o(1).$$

Finally,

$$U_{n,h} = \frac{t^2}{2} F_0 [\tilde{\psi}^2] + o(1)$$

and up to a multiplicative factor equal to $1 + o(1)$,

$$L_n(t) = \exp \left(\frac{t^2}{2} F_0 [\tilde{\psi}^2] \right) \frac{\int_{A_n} \exp \left(-\frac{F_0((h-t\bar{\psi}_{t,n})^2)}{2} + G_n(h-t\bar{\psi}_{t,n}) + R_n(h-t\bar{\psi}_{t,n}) \right) d\pi(f)}{\int_{A_n} \exp \left(-\frac{F_0(h^2)}{2} + G_n(h) + R_n(h) \right) d\pi(f)}.$$

Finally (A3) implies (4.15) and the theorem is proved.

4.3. *Proof of Theorem 2.1.* We apply Theorem 2.2 of Section 2.3, so we prove that conditions (A1), (A2) and (A3) are satisfied. Let ϵ_n be the posterior concentration rate obtained in Theorem 3.1. Let us consider $f = f_\theta \in \mathcal{F}_k$ for $1 \leq k \leq l_n$ where l_n is defined in Section 2.2. First, using (4.5), we have

$$(4.17) \quad V(f_0, f) \leq C(\log n)^3 \epsilon_n^2,$$

as soon as $h(f_0, f) \leq \sqrt{\log n} \epsilon_n$. Thus, using (3.1), we have

$$\mathbb{P}^\pi \left\{ A_{u_n}^1 | X^n \right\} = 1 + o_{\mathbb{P}_0}(1)$$

with $u_n = u_0(\log n)^3 \epsilon_n^2$, for a constant u_0 large enough. Note that we can restrict ourselves to $A_{u_n}^1 \cap (\cup_{k \leq l_n} \mathcal{F}_k)$, since $\mathbb{P}_\pi [(\cup_{k \leq l_n} \mathcal{F}_k)^c] \leq e^{-c n \epsilon_n^2}$ for any $c > 0$ by choosing l_0 large enough.

To establish (A2), we observe that

$$\begin{aligned} \|\log f_\theta - \log f_0\|_\infty &\leq \left\| \sum_{\lambda \in \mathbb{N}^*} (\theta_{0\lambda} - \theta_\lambda) \phi_\lambda \right\|_\infty + |c(\theta) - c(\theta_0)| \\ &\leq C \left(\sqrt{l_n} \|\theta - \theta_0\|_{\ell_2} + l_n^{\frac{1}{2}-\gamma} \right) = o(1), \end{aligned}$$

by using $\gamma > 1/2$, Lemma 4.1 and (4.6). So, (A2) is implied by (A1). Now, let us establish (A3). Without loss of generality, we can assume that A_n ,

the set defined in assumption (A2), is included into $(\cup_{k \leq l_n} \mathcal{F}_k)$. For any t , we study the term

$$\begin{aligned} I_n &= \frac{\int_{A_n} \exp\left(-\frac{F_0((h_f - t\bar{\psi}_{t,n})^2)}{2} + G_n(h_f - t\bar{\psi}_{t,n}) + R_n(h_f - t\bar{\psi}_{t,n})\right) d\pi(f)}{\int_{A_n} \exp\left(-\frac{F_0(h_f^2)}{2} + G_n(h_f) + R_n(h_f)\right) d\pi(f)} \\ &= \frac{\sum_{1 \leq k \leq l_n} p(k) \int_{A_n \cap \mathcal{F}_k} \exp\left(-\frac{F_0((h_f - t\bar{\psi}_{t,n})^2)}{2} + G_n(h_f - t\bar{\psi}_{t,n}) + R_n(h_f - t\bar{\psi}_{t,n})\right) d\pi_k(f)}{\sum_{1 \leq k \leq l_n} p(k) \int_{A_n \cap \mathcal{F}_k} \exp\left(-\frac{F_0(h_f^2)}{2} + G_n(h_f) + R_n(h_f)\right) d\pi_k(f)}. \end{aligned}$$

If we set

$$b_{n,k,t} = \frac{t\Pi_{f_0,k}\tilde{\psi} - t\tilde{\psi}_{\Pi,0}}{\sqrt{n}} = \frac{t}{\sqrt{n}} \sum_{\lambda=1}^k \tilde{\psi}_{\Pi,\lambda} \phi_\lambda,$$

we have, using (4.1) and since $k \leq l_n$,

$$\begin{aligned} \|b_{n,k,t}\|_\infty &\leq \frac{Ct\sqrt{k}}{\sqrt{n}} \|\Pi_{f_0,k}\tilde{\psi} - \tilde{\psi}_{\Pi,0}\|_{f_0} \\ &\leq \frac{Ct\sqrt{l_n}}{\sqrt{n}} \|\tilde{\psi}\|_\infty = O(\epsilon_n). \end{aligned}$$

Recall that for $f_\theta \in \mathcal{F}_k$,

$$h_\theta = \sqrt{n} \left(\sum_{\lambda \in \mathbb{N}^*} (\theta_\lambda - \theta_{0\lambda}) \phi_\lambda - c(\theta) + c(\theta_0) \right)$$

so, for $\theta' = \theta - t\frac{\tilde{\psi}_{\Pi}^{[k]}}{\sqrt{n}}$, with $H_n = (h_\theta - t\tilde{\psi})/\sqrt{n}$ and $\Delta_{\psi,k} = \tilde{\psi} - \Pi_{f_0,k}\tilde{\psi}$

$$\begin{aligned} h_{\theta'} &= h_\theta - \sqrt{n}b_{n,k,t} + \sqrt{n} \left(c(\theta) - c\left(\theta - t\frac{\tilde{\psi}_{\Pi}^{[k]}}{\sqrt{n}}\right) \right) \\ &= h_\theta - t\bar{\psi}_{t,n} + t(\tilde{\psi} - \Pi_{f_0,k}\tilde{\psi}) - \sqrt{n} \log \left[\frac{F_0(e^{H_n + t\Delta_{\psi,k}/\sqrt{n}})}{F_0(e^{H_n})} \right] \\ &= h_\theta - t\bar{\psi}_{t,n} + t\Delta_{\psi,k} - \Delta_n, \end{aligned}$$

with

$$\Delta_n = \sqrt{n} \log \left[\frac{F_0(e^{H_n + t\Delta_{\psi,k}/\sqrt{n}})}{F_0(e^{H_n})} \right].$$

Now, as previously, (4.6) implies $\|h_\theta\|_\infty/\sqrt{n} \leq \sqrt{k}\epsilon_n = o(1)$ and since $F(\Delta_{\psi,k}^2) = O(1)$, $\|\Delta_{\psi,k}\|_\infty = O(\sqrt{l_n}) = O(\sqrt{n}\epsilon_n)$,

$$\begin{aligned} F_0(e^{H_n+t\Delta_{\psi,k}/\sqrt{n}}) &= F_0\left(e^{H_n}\left(1 + \frac{t\Delta_{\psi,k}}{\sqrt{n}} + \frac{t^2\Delta_{\psi,k}^2}{2n}\right)\right) + o\left(F(\Delta_{\psi,k}^2)\frac{\|\Delta_{\psi,k}\|_\infty}{n^{3/2}}\right) \\ &= F_0\left(e^{H_n}\left(1 + \frac{t\Delta_{\psi,k}}{\sqrt{n}} + \frac{t^2\Delta_{\psi,k}^2}{2n}\right)\right) + o\left(\frac{\epsilon_n}{n}\right) \\ &= F_0(e^{H_n}) + \frac{t}{\sqrt{n}}F_0(e^{H_n}\Delta_{\psi,k}) + \frac{t^2}{2n}F_0(e^{H_n}\Delta_{\psi,k}^2) + o\left(\frac{1}{n}\right). \end{aligned}$$

Furthermore, for any function v satisfying $F_0(|v|) < \infty$,

$$(4.18) \quad F_0(e^{H_n}v) = F_0\left(v e^{h_\theta/\sqrt{n}}\right) - \frac{t}{\sqrt{n}}F_0\left(v e^{h_\theta/\sqrt{n}}\tilde{\psi}\right) + o\left(\frac{1}{n}\right).$$

Note that in the case $v = 1$ since $F_0(e^{h_\theta/\sqrt{n}}) = 1$ we can be more precise and we obtain

$$\begin{aligned} F_0(e^{H_n}) &= 1 - \frac{t}{\sqrt{n}}F_0\left(e^{h_\theta/\sqrt{n}}\tilde{\psi}\right) + O(1/n) \\ (4.19) \quad &= 1 - \frac{tF_0(h_\theta\tilde{\psi})}{n} + O\left(\frac{\epsilon_n^2}{\sqrt{n}} + \frac{1}{n}\right) = 1 + o\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Moreover,

$$(4.20) \quad F_0\left(v e^{h_\theta/\sqrt{n}}\right) = F_0(v) + o(F_0(|v|)).$$

Therefore, using (4.18) with $v = \Delta_{\psi,k}^2$ leads to

$$\frac{F_0(e^{H_n+t\Delta_{\psi,k}/\sqrt{n}})}{F_0(e^{H_n})} = 1 + \frac{t}{\sqrt{n}}\frac{F_0(e^{H_n}\Delta_{\psi,k})}{F_0(e^{H_n})} + \frac{t^2}{2n}F_0(\Delta_{\psi,k}^2) + o\left(\frac{1}{n}\right),$$

and using (4.18) with $v = \Delta_{\psi,k}$ together with (4.19) and using (4.20)

$$\frac{t}{\sqrt{n}}F_0(e^{H_n}\Delta_{\psi,k}) = \frac{t}{\sqrt{n}}F_0\left(\Delta_{\psi,k}e^{h_\theta/\sqrt{n}}\right) - \frac{t^2}{n}F_0\left(\Delta_{\psi,k}\tilde{\psi}\right) + o\left(\frac{1}{n}\right).$$

Moreover,

$$F_0\left(\Delta_{\psi,k}e^{h_\theta/\sqrt{n}}\right) = \frac{1}{\sqrt{n}}\left[F_0(h_\theta\Delta_{\psi,k}) + \frac{1}{\sqrt{n}}F_0\left(h_\theta^2B_{h_\theta,n}\Delta_{\psi,k}\right)\right],$$

where $B_{h,n}$ is defined by (4.16). Since $F_0(e^{h_\theta/\sqrt{n}}\tilde{\psi}) = F_0(\tilde{\psi}) + o(1) = o(1)$, $F_0(e^{H_n}) = 1 + o\left(\frac{1}{\sqrt{n}}\right)$ and $F_0(\tilde{\psi}\Delta_{\psi,k}) = F_0(\Delta_{\psi,k}^2)$ we obtain

$$\frac{F_0(e^{H_n+t\Delta_{\psi,k}/\sqrt{n}})}{F_0(e^{H_n})} = 1 + \frac{t}{\sqrt{n}}F_0\left(e^{h_\theta/\sqrt{n}}\Delta_{\psi,k}\right) - \frac{t^2}{2n}F_0\left(\Delta_{\psi,k}^2\right) + o\left(\frac{1}{n}\right)$$

and finally,

$$\begin{aligned} \Delta_n &= \sqrt{n} \log \left[\frac{F_0(e^{H_n+t\Delta_{\psi,k}/\sqrt{n}})}{F_0(e^{H_n})} \right] \\ &= tF_0\left(e^{h_\theta/\sqrt{n}}\Delta_{\psi,k}\right) - \frac{t^2}{2\sqrt{n}}F_0\left(\Delta_{\psi,k}^2\right) + o\left(\frac{1}{\sqrt{n}}\right) \\ (4.21) &= \frac{t}{\sqrt{n}} \left[F_0(h_\theta\Delta_{\psi,k}) + \frac{F_0(h_\theta^2 B_{h_\theta,n}\Delta_{\psi,k})}{\sqrt{n}} - \frac{t}{2}F_0(\Delta_{\psi,k}^2) \right] + o\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Moreover

$$F_0\left(h_\theta^2 B_{h_\theta,n}\Delta_{\psi,k}\right) = \frac{1}{2}F_0\left(h_\theta^2\Delta_{\psi,k}\right) + o\left(F_0\left(h_\theta^2|\Delta_{\psi,k}|\right)\right)$$

and by using (4.17),

$$\begin{aligned} \frac{F_0\left(h_\theta^2|\Delta_{\psi,k}|\right)}{\sqrt{n}} &\leq \|\Delta_{\psi,k}\|_\infty \frac{F_0\left(h_\theta^2\right)}{\sqrt{n}} \\ &\leq C\|\Delta_{\psi,k}\|_\infty \sqrt{n} (\log n)^3 \epsilon_n^2. \end{aligned}$$

To bound $\|\Delta_{\psi,k}\|_\infty$, we set $\psi_{+k} = \sum_{\lambda>k} \tilde{\psi}_\lambda \phi_\lambda$, so

$$\Delta_{\psi,k} = \psi_{+k} - \Pi_{f_0,k}(\psi_{+k}).$$

Then by using (4.1),

$$\begin{aligned} \|\Delta_{\psi,k}\|_\infty &\leq \|\psi_{+k}\|_\infty + \|\Pi_{f_0,k}\psi_{+k}\|_\infty \\ &\leq \|\psi_{+k}\|_\infty + C\sqrt{k}\|\Pi_{f_0,k}\psi_{+k}\|_{f_0} \\ &\leq \|\psi_{+k}\|_\infty + C\sqrt{k}\|\psi_{+k}\|_{f_0} \\ &\leq \|\psi_{+k}\|_\infty + C\sqrt{k}\|\psi_{+k}\|_2. \end{aligned}$$

Under (2.8), we obtain

$$\Delta_n = \frac{t}{\sqrt{n}} \left[F_0(h_\theta\Delta_{\psi,k}) - \frac{t}{2}F_0(\Delta_{\psi,k}^2) \right] + o(n^{-1/2}).$$

Note that $\Delta_n = o(1)$. Finally,

$$\begin{aligned} R_n(h_{\theta'}) &= \sqrt{n}F_0(h_{\theta'}) + \frac{F_0(h_{\theta'}^2)}{2} \\ &= R_n(h_\theta - t\bar{\psi}_{t,n}) - \sqrt{n}\Delta_n - \frac{t^2}{2}F_0(\Delta_{\psi,k}^2) + tF_0(h_\theta\Delta_{\psi,k}) - \Delta_n F_0(h_\theta) + o(1) \\ &= R_n(h_\theta - t\bar{\psi}_{t,n}) - \Delta_n F_0(h_\theta) + o(1). \end{aligned}$$

Recall that $h_{\theta'} = h_\theta - t\bar{\psi}_{t,n} + t\Delta_{\psi,k} - \Delta_n$, $\Delta_n = o(1)$ and $F_0(\Delta_{\psi,k}) = 0$. Note also that

$$\bar{\psi}_{t,n}(x) = \tilde{\psi}(x) + \frac{\sqrt{n}}{t} \log \left(F_0(e^{H_n}) \right) = \tilde{\psi}(x) + o(1)$$

so that $F_0(\Delta_{\psi,k}\bar{\psi}_{t,n}) = F_0(\Delta_{\psi,k}^2) + o(1)$ and

$$\begin{aligned} -\frac{F_0(h_{\theta'}^2)}{2} &= -\frac{F_0((h_\theta - t\bar{\psi}_{t,n})^2)}{2} - \frac{F_0((t\Delta_{\psi,k} - \Delta_n)^2)}{2} - F_0((h_\theta - t\bar{\psi}_{t,n})(t\Delta_{\psi,k} - \Delta_n)) \\ &= -\frac{F_0((h_\theta - t\bar{\psi}_{t,n})^2)}{2} + \frac{t^2 F_0(\Delta_{\psi,k}^2)}{2} - tF_0(h_\theta\Delta_{\psi,k}) + \Delta_n F_0(h_\theta) + o(1). \end{aligned}$$

Furthermore,

$$G_n(h_{\theta'}) = G_n(h_\theta - t\bar{\psi}_{t,n}) + tG_n(\Delta_{\psi,k}).$$

We set

$$\mu_{n,k} = -F_0(h_\theta\Delta_{\psi,k}) + G_n(\Delta_{\psi,k})$$

and we finally obtain,

$$\begin{aligned} &-\frac{F_0((h_{\theta'})^2)}{2} + G_n(h_{\theta'}) + R_n(h_{\theta'}) \\ &= -\frac{F_0((h_\theta - t\bar{\psi}_{t,n})^2)}{2} + R_n(h_\theta - t\bar{\psi}_{t,n}) + G_n(h_\theta - t\bar{\psi}_{t,n}) + t\mu_{n,k} \\ &\quad + \frac{t^2 F_0(\Delta_{\psi,k}^2)}{2} + o(1). \end{aligned}$$

Note that $F_0(h_\theta\Delta_{\psi,k}) = -\sqrt{n}F_0[(\tilde{\psi} - \Pi_{f_0,k}\tilde{\psi}) \sum_{\lambda \geq k+1} \theta_{0\lambda}\phi_\lambda]$ so that $\mu_{n,k}$ does not depend on θ and setting $T_k\theta = \theta - t\frac{\tilde{\psi}^{[k]}}{\sqrt{n}}$ for all θ , we can write

$$\begin{aligned} J_k &:= \frac{\int_{A_n \cap \mathcal{F}_k} \exp\left(-\frac{F_0((h_f - t\bar{\psi}_{t,n})^2)}{2} + G_n(h_f - t\bar{\psi}_{t,n}) + R_n(h_f - t\bar{\psi}_{t,n})\right) d\pi_k(f)}{\int_{A_n \cap \mathcal{F}_k} \exp\left(-\frac{F_0(h_f^2)}{2} + G_n(h_f) + R_n(h_f)\right) d\pi_k(f)} \\ &= e^{-\frac{t^2 F_0(\Delta_{\psi,k}^2)}{2}} e^{-t\mu_{n,k}} \frac{\int_{\Theta_k \cap A_n} e^{-\frac{F_0(h_{T_k\theta}^2)}{2} + G_n(h_{T_k\theta}) + R_n(h_{T_k\theta})} d\pi_k(\theta)}{\int_{\Theta_k \cap A_n} e^{-\frac{F_0(h_\theta^2)}{2} + G_n(h_\theta) + R_n(h_\theta)} d\pi_k(\theta)} (1 + o(1)), \end{aligned}$$

where $A'_n = \{\theta \text{ s.t. } f_\theta \in A_n\}$. Moreover, for $k \leq l_n$, $\|\tilde{\psi}_\Pi^{[k]}\|_{\ell_2} \leq C$. So, if we set

$$T_k(A'_n) = \left\{ \theta \in \Theta_k \cap A'_n \text{ s.t. } \theta + t \frac{\tilde{\psi}_\Pi^{[k]}}{\sqrt{n}} \in A'_n \right\}$$

for all $\theta \in T_k(A'_n)$,

$$\|\theta - \theta_0\|_{\ell_2}^2 \leq 2(\log n)^3 \epsilon_n^2 + \frac{2t^2 C^2}{n} \leq 2\epsilon_n^2 (\log n)^3 (1 + o(1))$$

since $n\epsilon_n^2 \rightarrow +\infty$. For all $\theta \in \Theta_k \cap A'_n$ such that $\|\theta - \theta_0\|_{\ell_2} \leq \frac{(\log n)^{3/2} \epsilon_n}{2}$

$$\theta + t \frac{\tilde{\psi}_\Pi^{[k]}}{\sqrt{n}} \in A'_n \cap \Theta_k$$

for n large enough and we can write

$$A'_{n,1} = \left\{ \theta \in A'_n : \|\theta - \theta_0\|_{\ell_2} \leq \frac{(\log n)^{3/2} \epsilon_n}{2} \right\}, \quad A'_{n,2} = \left\{ \theta \in A'_n : \|\theta - \theta_0\|_2 \leq 3(\log n)^{3/2} \epsilon_n \right\}$$

then

$$(4.22) \quad \Theta_k \cap A'_{n,1} \subset T_k(A'_n) \subset \Theta_k \cap A'_{n,2}$$

and under assumption (2.7),

$$\begin{aligned} J_k &\leq e^{-t^2 \frac{F_0(\Delta_{\psi,k}^2)}{2}} e^{-t\mu_{n,k}} \frac{\int_{\Theta_k \cap A'_{n,2}} e^{-\frac{F_0(h_\theta^2)}{2} + G_n(h_\theta) + R_n(h_\theta)} d\pi_k(\theta)}{\int_{\Theta_k \cap A'_n} e^{-\frac{F_0(h_\theta^2)}{2} + G_n(h_\theta) + R_n(h_\theta)} d\pi_k(\theta)} (1 + o(1)), \\ J_k &\geq e^{-t^2 \frac{F_0(\Delta_{\psi,k}^2)}{2}} e^{-t\mu_{n,k}} \frac{\int_{\Theta_k \cap A'_{n,1}} e^{-\frac{F_0(h_\theta^2)}{2} + G_n(h_\theta) + R_n(h_\theta)} d\pi_k(\theta)}{\int_{\Theta_k \cap A'_n} e^{-\frac{F_0(h_\theta^2)}{2} + G_n(h_\theta) + R_n(h_\theta)} d\pi_k(\theta)} (1 + o(1)). \end{aligned}$$

Therefore,

$$\begin{aligned} \zeta_n(t) &:= \mathbb{E}^\pi[\exp(t\sqrt{n}(\Psi(f) - \Psi(\mathbb{P}_n))) \mathbb{1}_{A_n}(f) | X^n] \\ &= e^{\frac{t^2 F_0(\tilde{\psi}^2)}{2}} \left[\sum_{k=1}^{l_n} p(k | X^n) J_k \right] (1 + o(1)) \\ &\leq \left[\sum_{k=1}^{l_n} p(k | X^n) \mathbb{1}_{\Theta_k \cap A'_n \neq \emptyset} e^{-t\mu_{n,k}} e^{t^2 \frac{F_0(\tilde{\psi}^2) - F_0(\Delta_{\psi,k}^2)}{2}} \right] (1 + o(1)) \end{aligned}$$

and

$$\zeta_n(t) \geq e^{t^2 \frac{F_0(\tilde{\psi}^2)}{2}} \sum_{k=1}^{l_n} p(k|X^n) e^{-t\mu_{n,k}} e^{-t^2 \frac{F_0(\Delta_{\psi,k}^2)}{2}} \pi \left[A'_{n,1} | X^n, k \right].$$

Besides under the above conditions on the prior, with probability converging to 1,

$$\pi \left[(A'_{n,1})^c | X^n \right] \leq e^{-nc\epsilon_n^2},$$

for some positive constant $c > 0$. Then uniformly over k such that $\Theta_k \cap A'_{n,1} \neq \emptyset$

$$\pi \left[(A'_{n,1})^c | X^n, k \right] e^{-t\mu_{n,k}} = o(1)$$

and

$$\zeta_n(t) \geq e^{t^2 \frac{F_0(\tilde{\psi}^2)}{2}} \sum_{k=1}^{l_n} p(k|X^n) \mathbb{1}_{\Theta_k \cap A_n \neq \emptyset} e^{-t\mu_{n,k}} e^{-t^2 \frac{F_0(\Delta_{\psi,k}^2)}{2}} (1 + o(1)).$$

This proves that the posterior distribution of $\sqrt{n}(\Psi(f) - \Psi(P_n))$ is asymptotically equal to a mixture of Gaussian distributions with variance $V_{0k} = F_0(\tilde{\psi}^2) - F_0(\Delta_{\psi,k}^2)$, means $-\mu_{n,k}$ and weights $p(k|X^n)$. Now if $\|\Delta_{\psi,k}\| = o(1)$ ($k \rightarrow +\infty$) $G_n(\Delta_{\psi,k}) = o_{\mathbb{P}_0}(1)$ and with probability converging to 1,

$$|\mu_{n,k}| \leq C\sqrt{n} \left(\sum_{\lambda=k+1}^{+\infty} \tilde{\psi}_\lambda^2 \right)^{1/2} \left(\sum_{\lambda=k+1}^{+\infty} \theta_{0\lambda}^2 \right)^{1/2} + o(1).$$

Thus, under (2.10), Equality (2.11) is proved.

4.4. *Proof of Corollary 2.1.* Let $k \leq l_n$ ($k = k_n^*$ in the case (D)) and $\lambda \leq k$. If $\theta_\lambda \sim \mathcal{N}(0, \tau_0^2 \lambda^{-2\beta})$, we have:

$$\begin{aligned} \frac{\sum_{\lambda=1}^k \tilde{\psi}_{\Pi,\lambda}^2 \lambda^{2\beta}}{n} &\leq \frac{Ck^{2\beta}}{n} \\ \frac{\sum_{\lambda=1}^k \theta_\lambda \tilde{\psi}_{\Pi,\lambda} \lambda^{2\beta}}{\sqrt{n}} &= O\left(\frac{1}{\sqrt{n}} \left[\|\theta - \theta_0\|_{\ell_2} k^{2\beta} + (k^{2\beta-\gamma} + 1) \right]\right). \end{aligned}$$

Similar computations hold when g is the Student density since

$$\begin{aligned} &\left| \sum_{\lambda=1}^k \log \left(1 + C\lambda^{2\beta}\theta_\lambda^2 \right) - \log \left(1 + C\lambda^{2\beta}(\theta_\lambda - t\tilde{\psi}_{\Pi,\lambda}/\sqrt{n})^2 \right) \right| \\ &= O\left(\sum_{\lambda=1}^k \lambda^{2\beta} \left| (\theta_\lambda - t\tilde{\psi}_{\Pi,\lambda}/\sqrt{n})^2 - \theta_\lambda^2 \right| \right). \end{aligned}$$

Under conditions of Corollary 2.1, above terms are negligible when n goes to 0 if $\sum_{\lambda=1}^k (\theta_\lambda - \theta_{0\lambda})^2 \leq (\log n)^4 \epsilon_n^2 / L(n)^2$. If g is the Laplace density,

$$\left| \log \left(g \left(\frac{\theta_\lambda - t \tilde{\psi}_{\Pi, \lambda} / \sqrt{n}}{\sqrt{\tau_\lambda}} \right) \right) - \log \left(g \left(\frac{\theta_\lambda}{\sqrt{\tau_\lambda}} \right) \right) \right| \leq C \frac{|\tilde{\psi}_{\Pi, \lambda}| \lambda^\beta}{\sqrt{n}},$$

so that

$$\left| \log \left(\frac{\pi_k(\theta - t \frac{\tilde{\psi}_{\Pi}^{[k]}}{\sqrt{n}})}{\pi_k(\theta)} \right) \right| \leq C \frac{\sum_{\lambda=1}^k \lambda^\beta |\tilde{\psi}_{\Pi, \lambda}|}{\sqrt{n}} \leq o \left(\frac{k^{\beta+1/2}}{\sqrt{n}} \right) = o(1).$$

4.5. *Proof of Proposition 2.1.* We set

$$(4.23) \quad k_n = n^{1/(2\gamma+1)} (\log n)^{-2/(2\gamma+1)} (\log \log n)^{-2/(2\gamma+1)}.$$

Let $J_1 > 3$. We have

$$\begin{aligned} \sum_{\lambda \geq J_1} \theta_{0\lambda}^2 \lambda^{2\gamma} &\leq \sum_{\lambda \geq J_1} \frac{1}{\lambda \log \lambda (\log \log \lambda)^2} \\ &\leq \int_{J_1}^{\infty} \frac{1}{x \log x (\log \log x)^2} dx = \frac{1}{\log \log J_1}, \end{aligned}$$

and similarly

$$\begin{aligned} \sum_{\lambda \geq J_1} \theta_{0\lambda}^2 &\leq \sum_{\lambda \geq J_1} \frac{1}{\lambda^{2\gamma+1} \log \lambda (\log \log \lambda)^2} \leq \int_{J_1}^{\infty} \frac{1}{x^{2\gamma+1} \log x (\log \log x)^2} dx \\ &= \left[-\frac{1}{2\gamma x^{2\gamma} \log x (\log \log x)^2} \right]_{J_1}^{\infty} (1 + o(1)) \\ (4.24) \quad &= \frac{1}{2\gamma J_1^{2\gamma} \log J_1 (\log \log J_1)^2} (1 + o(1)) \end{aligned}$$

when $J_1 \rightarrow \infty$. Thus, for k_1 large enough,

$$\mathbb{P}^\pi [k \leq k_1 k_n | X^n] = 1 + o(1).$$

We now study the terms $\mu_{n,k}$ and we show that there are some k 's for which neither $\mu_{n,k}$ nor $p(k|X^n)$ can be neglected. First note that when $k \rightarrow \infty$ $G_n(\Delta_{\psi,k}) = o(1)$ and

$$\begin{aligned} \mu_{n,k} &= \sqrt{n} F_0 \left(\Delta_{\psi,k} \sum_{\lambda \geq k+1} \theta_{0\lambda} \phi_\lambda \right) + o(1) \\ &= \sqrt{n} \int \Delta_{\psi,k} \sum_{\lambda \geq k+1} \theta_{0\lambda} \phi_\lambda - \sqrt{n} \int (1 - f_0) \Delta_{\psi,k} \sum_{\lambda \geq k+1} \theta_{0\lambda} \phi_\lambda + o(1) \\ &:= \mu_{n,k,1} + \mu_{n,k,2} + o(1). \end{aligned}$$

We first consider $\mu_{n,k,1}$:

$$\begin{aligned}\mu_{n,k,1} &= \sqrt{n} \int \tilde{\psi} \sum_{\lambda \geq k+1} \theta_{0\lambda} \phi_\lambda = \sqrt{n} \sum_{\lambda \geq k+1} \theta_{0\lambda} \int 1_{u \leq x_0} \phi_\lambda(u) du \\ &= \sqrt{2n} \sum_{l \geq (k+1)/2} \theta_{0,2l} \frac{\sin(2\pi l x_0)}{2\pi l} = \frac{\sqrt{n}}{\sqrt{2\pi}} \sum_{l \geq (k+1)/2} \frac{\sin^2(2\pi l x_0)}{l^{\gamma+3/2} \sqrt{\log l} \log \log l}.\end{aligned}$$

With $x_0 = 1/4$, we finally obtain:

$$\mu_{n,k,1} = \frac{\sqrt{n}}{\sqrt{2\pi}} \sum_{m \geq (k+1)/4 - 1/2} \frac{1}{(2m+1)^{\gamma+3/2} \log^{1/2}(2m+1) \log \log(2m+1)},$$

so that there exist two constants c_1 and c_2 such that for all $k \leq k_n$,

$$|\mu_{n,k,1}| \geq c_1 \sqrt{n} k_n^{-\gamma-1/2} (\log k_n)^{-1/2} (\log \log k_n)^{-1} \geq c_2 \sqrt{\log n}.$$

Now, let us deal with $\mu_{n,k,2}$. We have

$$\Delta_{\psi,k} = \sum_{\lambda \geq k+1} \tilde{\psi}_\lambda \phi_\lambda - \Pi_{f_0,k} \left(\sum_{\lambda \geq k+1} \tilde{\psi}_\lambda \phi_\lambda \right)$$

and

$$\left\| \Pi_{f_0,k} \left(\sum_{\lambda \geq k+1} \tilde{\psi}_\lambda \phi_\lambda \right) \right\|_2^2 \leq C \sum_{\lambda \geq k+1} \tilde{\psi}_\lambda^2.$$

So,

$$\begin{aligned}\mu_{n,k,2} &\leq C \sqrt{n} \|f_0 - 1\|_\infty \left(\sum_{\lambda \geq k+1} \tilde{\psi}_\lambda^2 \right)^{1/2} \left(\sum_{\lambda \geq k+1} \theta_{0\lambda}^2 \right)^{1/2} \\ &\leq C \sqrt{n} \|f_0 - 1\|_\infty \frac{k^{-\gamma-1/2}}{\sqrt{\log k} \log \log k}.\end{aligned}$$

By choosing k_0 large enough $\|f_0 - 1\|_\infty$ can be made as small as needed, so that we finally obtain that there exists $c > 0$ such that for all $k \leq k_n$

$$|\mu_{n,k}| \geq c \sqrt{\log n}.$$

REFERENCES

- [1] BERGER, J. (1985) *Statistical decision theory and Bayesian analysis* (2nd edition) Springer - Verlag, New-York.
- [2] BOUCHERON, S. and GASSIAT, E. (2009) A Bernstein-von Mises theorem for discrete probability distributions. *Electron. J. Stat.* **3**, 114–148
- [3] CASTILLO, I. (2008) A semi-Parametric Bernstein-von Mises theorem. *Preprint*
- [4] CHAMBAZ, A. and ROUSSEAU, J. (2008) Bounds for Bayesian order identifications. *Ann. Statist.* **36**, 938-962.
- [5] COHEN, A., DAUBECHIES, I. and VIAL, P. (1993) Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* **1**, 54–81.
- [6] DIACONIS, P. and FREEDMAN, D. (1986) On the consistency of Bayes estimates. *Ann. Statist.* **14**, 1-26.
- [7] FREEDMAN, D. (1999) On the Bernstein-von Mises theorem with infinite- dimensional parameters. *Ann. Statist.* **27** 1119–1140.
- [8] GHOSAL S. (2000) Asymptotic normality of the posterior distributions for exponential families with many parameters. *Journ. Mult. Anal.*, **74** 49-69.
- [9] GHOSAL, S., GHOSH, J.K. and VAN DER VAART, A. (2000) Convergence rates of posterior distributions. *Ann. Statist* **28** 500–531.
- [10] GHOSAL, S., LEMBER, J. and VAN DER VAART, A. (2008) Nonparametric Bayesian model selection and averaging. *Electr. Journ. Statist.* **2**, 63-89.
- [11] GHOSH, J.K. and RAMAMOORTHY, R.V. (2003) *Bayesian nonparametrics* Springer, New York.
- [12] HRDLE, W., KERKYACHARIAN, G., PICARD, D. and TSYBAKOV, A. (1998) *Wavelets, approximation, and statistical applications*. Lecture Notes in Statistics, 129. Springer-Verlag, New York
- [13] HUANG, T.Z. (2004) Convergence rates for posterior distributions and adaptive estimation. *Ann. Statist.* **32**, 1556-1593.
- [14] KIM, Y. and LEE, J. (2004) A Bernstein-von Mises theorem in the nonparametric right-censoring model. *Ann. Statist.* **32**, 1492-1512.
- [15] KIM, Y. (2006) The Bernstein-von Mises theorem for the proportional hazard model. *Ann. Statist.* **34**, 1678-1700.
- [16] LeCAM L. (1953) On some asymptotic properties properties of maximum likelihood estimates and related Bayes estimates/ *Univ. California Publ. Statist.* **1** 277-330.
- [17] MALLAT, S. (1998) *A wavelet tour of signal processing*, Academic Press, San Diego.
- [18] MEYER, Y. (1992) *Wavelets and operators*, Cambridge University Press, Cambridge.
- [19] ROBERT, C.P. (2007) *Bayesian Choice* (2nd edition) Springer-Verlag.
- [20] SHEN, X. (2002) Asymptotic normality of semiparametric and nonparametric posterior distributions. *J. Amer. Statist. Assoc.* **97**, 222-235.
- [21] SCRICCILOLO, C. (2006) Convergence rates for Bayesian density estimation of infinite-dimensional exponential families. *Ann. Statist.* **34**, 2897-2920.
- [22] SHEN, X. and WASSERMAN, L. (2001) Rates of convergence of posterior distributions. *Ann. Statist.* **29**, 687-714.
- [23] VAN DER VAAT, A. W. (2002). Semiparametric statistics. *Ecole d'Eté de St Flour 1999*, 331-457, (eds: P. Bernard). Springer Verlag, New York.
- [24] VERDINELLI, I. and WASSERMAN, L. (1998) Bayesian goodness-of-fit testing using infinite- dimensional exponential families. *Ann. Statist.* **26**, 1215-1241.

- [25] WONG, W.H. and SHEN, X. (1995) Probability inequalities for likelihood ratios and convergence rates of sieve mles *Ann. Statist.* **23**, 339-362.
- [26] ZHAO, L.H. (2000) Bayesian aspects of some nonparametric problems. *Ann. Statist.* **28**, 532-552.

LABORATOIRE DE MATHÉMATIQUES
UNIVERSITÉ PARIS SUD
15 RUE GEORGES CLEMENCEAU
91405 ORSAY
FRANCE

CEREMADE
UNIVERSITÉ PARIS DAUPHINE,
PLACE DU MARÉCHAL DE LATTRE DE TASSIGNY
75016 PARIS, FRANCE.
E-MAIL: rousseau@ceremade.dauphine.fr