Bernstein Von Mises Theorem for linear functionals of the density

Vincent Rivoirard and Judith Rousseau

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 variance being equal to the asymptotic frequentist variance of the centering point. Such results are well know in parametric frameworks, see for instance [13] 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 behaviour 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 [16] or [1].

In a frequentist perspective the Bernstein-von Mises property enables the construction of confidence regions since under this property a Bayesian credible 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 feasable. However the more complex the model the harder it is to derive Bernstein - von Mises theorems. In infinite dimensional setups, the mecanisms are even more complex.

Semi-parametric and non parametric models are widely popular both from a theoretical and practical persepective 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 nonparametric models. Most existing works on the asymptotic posterior distributions deal with consistency of the posterior or rates of concentration of the posterior. In other words it consists in controling objects in the form

$$\mathbb{P}^{\pi}\left[U_n|X^n\right]$$

where $\mathbb{P}^{\pi}[.|X^n]$ denotes the posterior distribution given a n vector of observationsX^n and U_n denotes either a fixed neighbourhood (consistency) or a sequence of shrinking neighbourhoods (rates of concentration). As remarked by [5] consistency is an important condition since it is not possible to construct subjective prior in a nonparametric framework. Obtaining concentration rates of the posterior helps in understanding 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 $\mathbb{P}^{\pi}\left[U_n|X^n\right]$ but to determine an equivalent of $\mathbb{P}^{\pi}\left[U_n|X^n\right]$ for some specific types of sets U_n. This difficulty explains that there is up to now very little work on Bernstein Von Mises theorems in infinite dimensional models. The most well known results are negative results and are given in [6]. Some positive results are provided by [7] on the asymptotic normality of the posterior distribution of the parameter in an exponential family with increasing number of paramters. Nice positive results are obtained in [11] and [12], however they rely heavily on a conjuguacy type of property of the family of priors they consider 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 parametric part, which is the parameter of interest and a non parametric part, which is the nuisance parameter, [2] obtains interesting conditions leading to a Bernstein - von Mises theorem on the parametric part, clarifying an earlier work of [17].

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 nuisance parameter, which is 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 [20] for an excellent presentation of a general theory on such problems.

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

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 $F_n(x)$, see for instance [9].This result is obtained using the conjuguate nature of the Dirichlet prior, leading to an explicit posterior disitribution. Other frequentist estimators, based on frequentist estimates of the density have also been studied in the frequentist literature, in particular estimates based on kernel estimators. 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 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 the cumulative distribution function $F(x)$, with centering its empirical counterpart, for instance $F_n(x)$ the empirical cumulative distribution function, when the prior puts positive mass on absolutely continuous densities with respect to Lebesgue measures. 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 Notations and aim

In this paper we assume that given a distribution $\mathbb{P}$ with density f with respect to Lebesgue measure, $X_1, ..., X_n$ are independent and indentically distributed by $\mathbb{P}$ and denote $X^n = (X_1, ..., X_n)$ and without loss of generality we assume that the $X_i \in [0, 1]$. Let $\mathcal{F} = \{f : [0, 1] \to \mathbb{R}^+, \int_0^1 f(x)dx = 1\}$. We now define some notations that will be used throughout the paper. Denote $l_n(f)$ the log-likelihood associated with the density f and $l_n(\theta) = l_n(f_\theta)$. Let f_0 be the true density of the observations X^n and denote by F_0 its cumulative distributon function and let x be a fixed real number.

We denote by $< .,. >$ the inner product in $\mathbb{L}_2(F_0) = \left\{g : \quad \int g^2(x) f_0(x)dx\right\}$. and by $\|.\|$ the corresponding norm. For an integrable function g we sometimes denote $F_0(g) = \int_0^1 f_0(u)g(u)du =< 1, g >$. We also consider the inner product in $L^2([0, 1])$ noted $< .,. >_2$ with its corresponding norm whichdenoted $\|.\|_2$. Let $K(f, f') = F[\log(f/f')]$ be the Kullback-Leibler divergence between two densities f and f'. Let $h(f, f') = \left[\int(\sqrt{f}(y) - \sqrt{f'}(y))^2 dy\right]^{1/2}$ be the Hellinger distance between f and f'. Denote also $V(f, f') = F[(\log(f/f'))^2]$.

We consider the usual notations on the empirical process, namely

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

Let F_n be the empirical distribution function : $F_n(x) = P_n(\mathbb{1}_{.\le x})$.

Consider a prior Π on the set $\mathcal{F}$ of densities on $[0, 1]$ with respect to Lebesgue measure. The aim of this paper is to study the posterior distribution of $\Psi(f)$ a continuous linear form on $L^2[0, 1]$. A typical example is $\Psi(f) = F(x) = \mathbb{P}[X \le x]$ for a given x and to derive conditions under which

$$\mathbb{P}^\pi\left[\sqrt{n}(\Psi(f) - \Psi(P_n)) \le z|X^n\right] \to \Phi\left[\frac{z}{V_0}\right] \quad \text{in } \mathbb{P}_0 \text{ Probability},$$

where V_0 is the variance of $\sqrt{n}\Psi(P_n)$. We also denote by $\Phi_V(z)$ the cumulative distribution function of a Gaussian random variable centered at 0 with variance V, by $\mathbb{E}^\pi[.|X^n]$ posterior expectations and BVM for Bernstein Von Mises.

1.2 Organization of the paper

In Section 2 we present the general Bernstein Von Mises theorem, which is given in the formal way in the case where linear submodels are adapted to the prior. We then apply this general theorem to the case where the prior is based on infinite dimensional exponential families, in Section 3. In this Section we first give a general results giving the asymptotic posterior distribution of the $\Psi(f)$ which can be either Gaussian or a mixture of Gaussian distributions. We also provide a theorem describing the posterior rate of concentration under such priors, see Section 3.1. Finally, in Section 3.3, using an example we explain how things can go wrong. The proofs are postponed in Section 4.

2 Bernstein Von Mises theorems

2.1 Some heuristiscs for proving Bernstein Von Mises theorems

We first define some notions that are useful in the study of asymptotic properties of semi-parametric models. These notions can be found for instance in [20].

As in Chapter 25 of [20], to study the asymptotic behaviour of semi-parametric models we consider 1-dimensional differentiable paths locally around the true parameter f_0, that is submodels of the form : $t \to f_t$ for $0 < t < t_0$, for some $t_0 > 0$ such that for each path there exists a measurable function g called the score function for the submodel $\{f_t, , 0 < t < t_0\}$ at $t = 0$ satisfying

$$\lim_{t \to 0} \int_{\mathbb{R}} \left(\frac{f_t^{1/2}(u) - f_0^{1/2}(u)}{t} - \frac{1}{2} g(u) f_0^{1/2}(u) \right)^2 du = 0. \tag{1}$$

We denote by $\mathcal{F}_{f_0}$ the tangent set, i.e. the collection of score functions g associated with these differentiable paths. Using (1), $\mathcal{F}_{f_0}$ can be identified with a subset of $\{v \in \mathbb{L}_2(F_0) : F_0(v) = 0\}$. For instance, when considering all probability laws, the most usual collection of differentiable paths is given by

$$f_t(x) = c(t) f_0(x) e^{t g(x)} \tag{2}$$

with $\|g\|_\infty < \infty$ and c such that $c(0) = 1$ and $c'(0) = 0$. In this case, g is the score function. Actually, as emphasized by [20], we can prove, by considering the collection of differentiable paths of the form $f_t(x) = 2c(t) f_0(x)(1 + \exp(-2tg(x)))^{-1}$ (with previous conditions on c), that the tangent space is the subspace of $\mathbb{L}_2(F_0)$ of functions g verifying $F_0(g) = 0$.

Now, consider a continuous linear form Ψ on $\mathbb{L}_2$. We can identify such a functional by a function $\in \mathbb{L}_2$ such that for all $f \in \mathbb{L}_2$

$$\Psi(f) = \int f(u) \psi(u) du. \tag{3}$$

Then for any differentiable path $t \to f_t$ with score function g, if the function ψ is bounded on $\mathbb{R}$ (or on the support of f_t for all $0 \le t < t_0$),

$$
\begin{aligned}
\frac{\Psi(f_t) - \Psi(f_0)}{t} &= \int \psi(u) g(u) f_0(u) du + \int \frac{\left(f_t^{1/2}(u) - f_0^{1/2}(u)\right)^2}{t} \psi(u) du \\
&\quad + 2 \int \psi(u) \left(\frac{f_t^{1/2}(u) - f_0^{1/2}(u)}{t} - \frac{1}{2} g(u) f_0^{1/2}(u) \right) f_0^{1/2}(u) du \\
&= < \psi, g > + o(1).
\end{aligned}
$$

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 $g \in \mathcal{F}_{f_0}$,

$$
\int \tilde{\psi}(u) g(u) f_0(u) du = \int \psi(u) g(u) f_0(u) du.
$$

This implies :

$$
\lim_{t \to 0} \frac{\Psi(f_t) - \Psi(f_0)}{t} = < \tilde{\psi}, g > . \tag{4}
$$

The efficient influence function will play an important role for our purpose. The efficient influence function is also the key notion to characterize asymptotically efficient estimators (see Section 25.3 of [20]).

Now, let us provide some examples by specifying different types of continuous linear forms that can be considered.

Example 1. *An important example is provided by the cumulative distribution function. If $x_0 \in \mathbb{R}$ is fixed, consider for any density function $f \in \mathbb{L}_2$ whose cdf is F,*

$$
\Psi(f) = \int \mathbb{1}_{u \le x_0} f(u) du = F(x_0)
$$

so that in this case, $\psi(u) = \mathbb{1}_{u \le x_0}$, which is a bounded function and if $\mathcal{F}_{f_0}$ is the subspace of $\mathbb{L}^2(F_0)$ of functions g satisfying $F_0(g) = 0$ then $\tilde{\psi}(u) = \mathbb{1}_{u \le x_0} - F_0(x_0)$.

Example 2. *More generally, for any measurable set A then consider $\psi(u) = \mathbb{1}_{u \in A}$ and for any density function $f \in \mathbb{L}_2$*

$$
\Psi(f) = \int \mathbb{1}_{u \in A} f(u) du.
$$

Example 3. *If f_0 has bounded support, say on $[0, 1]$ then the functional*

$$
\Psi(f) = \mathbb{E}_f[X] = \int_0^1 u f(u) du
$$

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

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

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

Now, let us set $f_{g,n} = f_t$ if $t = n^{-\frac{1}{2}}$. We have :

$$\begin{aligned} \sqrt{n}\left(\Psi(f_{g,n}) - \Psi(P_n)\right) &= \sqrt{n} \int \psi(x)(f_{g,n}(x) - f_0(x))dx - G_n(\tilde{\psi}) \\ &= \Delta_n(g) + < \tilde{\psi}, g > - G_n(\tilde{\psi}), \end{aligned}$$

with

$$\Delta_n(g) = \sqrt{n}\left(\Psi(f_{g,n}) - \Psi(f_0)\right) - < \tilde{\psi}, g > .$$

Furthermore,

$$\begin{aligned} l_n(f_{g,n}) - l_n(f_0) &= \sum_{i=1}^{n} \log\left(\frac{f_{g,n}(X_i)}{f_0(X_i)}\right) \\ &= R_n(g) + G_n(g) - \frac{F_0(g^2)}{2}, \end{aligned}$$

with

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

So,

$$\begin{aligned} & t\sqrt{n}\left(\Psi(f_{g,n}) - \Psi(P_n)\right) + l_n(f_{g,n}) - l_n(f_0) \\ &\quad = R_n(g) - \frac{F_0(g^2)}{2} + G_n(g - t\tilde{\psi}) + t\Delta_n(g) + t < \tilde{\psi}, g > \\ &\quad = R_n(g - t\tilde{\psi}) + G_n(g - t\tilde{\psi}) - \frac{F_0((g - t\tilde{\psi})^2)}{2} + \frac{t^2 F_0(\tilde{\psi}^2)}{2} + U_n, \end{aligned}$$

with

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

Lemma 25.14 of [20] shows that under (1), $R_n(g) = o(1)$ and (4) yields $\Delta_n(g) = o(1)$ for a fixed g. 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 $U_n = o(1)$,

$$\sqrt{n}\left(\Psi(f_{g,n}) - \Psi(f)\right) + l_n(f_{g,n}) - l_n(f) = o(1)$$

and the equality

$$\frac{\int e^{R_n(g-t\tilde{\psi}) + G_n(g-t\tilde{\psi}) - \frac{F_0((g-t\tilde{\psi})^2)}{2}} d\pi(f)}{\int e^{R_n(g) + G_n(g) - \frac{F_0(g^2)}{2}} d\pi(f)} = \frac{\int \exp\left(l_n(f) - l_n(f_0)\right) d\pi(f_{g+t\tilde{\psi}})}{\int \exp\left(l_n(f) - l_n(f_0)\right) d\pi(f)} = 1 + o(1), \quad (5)$$

then

$$L_n(t) = \exp\left(\frac{t^2 F_0(\tilde{\psi}^2)}{2}\right)(1+o(1)).$$

In this case, our goal is reached. However, it is not obvious that a given prior π satisfies all these properties. In particular in a nonparametric framework $R_n(g) \neq o(1)$ uniformly over a set whose posterior probability goes to 1. An alternative is to give up the framework of tangent sets and score functions and to consider local linear submodels.

2.2 Bernstein Von Mises under linear submodels

In this section we study the case where linear local models are adapted to the prior. More precisely, we assume that $\|\log(f_0)\|_\infty < \infty$ so, 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 underline the relationship between f and h. Note that in this context h is not the score function since $F_0(h) \neq 0$. It would be equivalent to consider local models of the form $f = f_0(1 + h/\sqrt{n})$, except that we would have to impose constraints on h for f to be positive. We consider a continuous linear form Ψ on $\mathbb{L}_2$ such that for any $f \in \mathbb{L}_2$, we consider ψ such that (3) is satisfied and we set for any x,

$$\psi_c(x) = \psi(x) - F_0(\psi). \tag{6}$$

Note that ψ_c coincides with the influence function $\tilde{\psi}$ associated with the tangent set $\{g \in \mathbb{L}_2(F_0); F_0(g) = 0\}$. 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 $A^1_{u_n} = \left\{f \in \mathcal{F} : \quad V(f_0, f) \leq u_n^2\right\}$ the posterior distribution of $A^1_{u_n}$ satisfies

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

(A2) The posterior distribution of the subset $A_n \subset A^1_n$ of densities such that

$$\int \left|\log\left(\frac{f(x)}{f_0(x)}\right)\right|^3 (f_0(x) + f(x))\,dx = o(1) \tag{7}$$

satisfies

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

(A3) Let

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

and for any x,

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

We have

$$\frac{\int_{A_n} \exp\left(-\frac{F_0((h_f - t\bar{\psi}_{t,n})^2)}{2} + \mathbb{G}_n(h_f - t\bar{\psi}_{t,n}) + R_n(h_f - t\bar{\psi}_{t,n})\right) d\pi(f)}{\int_{A_n} \exp\left(-\frac{F_0(h_f^2)}{2} + G_n(h_f) + R_n(h_f)\right) d\pi(f)} = 1 + o_{\mathbb{P}_0}(1). \quad (8)$$

Before stating our main 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 [19] or [8] for general results. The difficulty here comes from the use of V instead of the Hellinger or the $\mathbb{L}_1$-distance. 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 below.

The main difficulty comes from condition (A3). To prove it, we need to be able to construct a tranformation T such that $Tf_h = f_{h-t\bar{\psi}_{t,n}}$ exists 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 transform and this remains true in semi-parametric setups where we can write down the parameter as (θ, η) where θ is the parameter of interest and is finite dimensional. Indeed as shown in [2] under certain conditions the transformations can be transfered to transformations on θ which is finite dimensional. Here this aspect is more complex since T is a transformation on f which is infinite dimensional so that a condition in the form $d\pi(Tf) = d\pi(f)(1 + o(1))$ does not necessarily make sense. We study this aspect in more details in Section 3.

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

Theorem 2.1. *Let f_0 be a density on $\mathcal{F}$ such that $\|\log(f_0)\|_\infty < \infty$ and $\|\psi\|_\infty < \infty$. Assume that (A1), (A2) and (A3) are true. Then, if*

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

we have for any z, in probability with respect to $\mathbb{P}_0$,

$$\mathbb{P}^\pi\left\{\sqrt{n}(\Psi(f) - \Psi(P_n)) \le z | X^n\right\} - \Phi_{F_0(\psi_c^2)}(z) \to 0.$$

Sieve priors lead to interesting behaviours of the posterior distribution as is illustrated in the following section. Indeed they have a behaviour which is half way between parametric and non parametric and depending on where the true density lies their asymptotic behaviour can vary significantly. We illustrate these features in the following two sections.

3 Bernstein Von Mises theorem under infinite dimensional exponential families

In this section, we study a specific class of priors based on infinite dimensional exponential families on the following class of densities supported by $[0,1]$:

$$\mathcal{F} = \left\{ f \geq 0 \quad f \text{ is 1-periodic}, \int_0^1 f(x)dx = 1,\ \log(f) \in \mathbb{L}^2([0,1]) \right\}.$$

We assume that $f_0 \in \mathcal{F}$ and we consider two types of orthonormal bases defined in the following section.

3.0.1 Orthonormal bases

We now describe the two types of orthonormal bases considered, namely the Fourier and wavelet bases. Fourier bases constitute unconditional bases of periodized Sobolev spaces W^γ where γ is the smoothness parameter. Our results are also valid for a wide range of Besov spaces. In this case, we consider wavelet bases which allow for the following expansions :

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

where $\theta_{-10} = \int_0^1 f(t)dt$ and $\theta_{jk} = \int_0^1 f(t)\psi_{jk}(t)dt$. We recall that the functions Υ_{jk} are obtained by dilations and translations of a periodized mother wavelet ψ that can be assumed to be supported by the compact set $[-A, A]$:

$$\Upsilon_{jk}(t) = 2^{\frac{j}{2}}\sum_{l=-\infty}^{+\infty}\Upsilon(2^j t - k + 2^j l), \quad t \in [0,1].$$

If Υ belongs to the Hölder space C^r and has r vanishing moments then the wavelet basis constitutes an unconditional basis of the Besov space $\mathcal{B}^\gamma_{p,q}$ for $1 \leq p, q \leq +\infty$ and $\max\left(0, \frac{1}{p} - \frac{1}{2}\right) < \gamma < r$. In this case, $\mathcal{B}^\gamma_{p,q}$ is the set of functions f of $\mathbb{L}_2[0,1]$ such that $\|f\|_{\gamma,p,q} < \infty$ where

$$\|f\|_{\gamma,p,q} = \begin{cases} |\theta_{-10}| + \left(\sum_{j=0}^{+\infty} 2^{jq(\gamma+\frac{1}{2}-\frac{1}{p})}\left(\sum_{k=0}^{2^j-1}|\theta_{jk}|^p\right)^{\frac{q}{p}}\right)^{\frac{1}{q}} & \text{if } q < \infty \\ |\theta_{-10}| + \sup_{j\geq 0}\left\{2^{j(\gamma+\frac{1}{2}-\frac{1}{p})}\left(\sum_{k=0}^{2^j-1}|\theta_{jk}|^p\right)^{\frac{1}{p}}\right\} & \text{if } q = \infty. \end{cases}$$

We refer the reader to [15] for a good review of wavelets and Besov spaces. We just mention that Besov spaces include in particular Sobolev spaces ($W^\gamma = \mathcal{B}^\gamma_{2,2}$) and, when γ is not an integer, Hölder spaces ($C^\gamma = \mathcal{B}^\gamma_{\infty,\infty}$). To shorten notations, the orthonormal basis will be denoted $(\phi_\lambda)_{\lambda\in\mathbb{N}}$, where $\phi_0 = \mathbb{1}_{[0,1]}$ and

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

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

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

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

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

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

where $\theta_\lambda = \int_0^1 f(t)\phi_\lambda(t)dt$. Recall that when the Fourier basis is used, f lies in W^γ for $\gamma > 0$ if and only if $\|f\|_\gamma < \infty$, where

$$\|f\|_\gamma = \left(\theta_0^2 + \sum_{\lambda\in\mathbb{N}^*} |\lambda|^{2\gamma}\theta_\lambda^2\right)^{\frac{1}{2}}.$$

We respectively use $\|.\|_\gamma$ and $\|.\|_{\gamma,p,q}$ to define the radius of the balls of W^γ and $\mathcal{B}^\gamma_{p,q}$ respectively. We now present the general result on posterior concentration rates associated with such prior models.

3.1 Posterior rates

Assume that $f_0 \in \mathcal{F}$ and let Φ be one of the orthonormal basis introduced in Section 3.0.1, then

$$\log(f_0) - \int_0^1 \log(f_0(x))dx = \sum_{\lambda\in\mathbb{N}^*} \theta_{0\lambda}\phi_\lambda.$$

Set $\theta_0 = (\theta_{0\lambda})_{\lambda\in\mathbb{N}^*}$ and define

$$c(\theta_0) = -\int_0^1 \log(f_0(x))dx,$$

we have

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

We consider the following family of models : for any $k \in \mathbb{N}^*$, we set

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

where

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

So, we define a prior π on the set $\mathcal{F} = \cup_k \mathcal{F}_k$ 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 [18]. The special case of log-spline priors has been studied by [8] and [10], whereas the prior considered by [21] is based on Legendre polynomials. For the wavelet case, [10] considered the special case of the Haar basis.

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 first 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 [18] 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 Scricciolo merely considers Gaussian priors. Also related to this problem is the work of [10] who derives a general framework to obtain adaptive posterior concentration rates and apply 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_j's in each k dimensional model to belong to a ball with fixed radius.

Now, we specify the conditions on the prior π :

Definition 3.1. *Let $\beta > 1/2$ be fixed and let g be a continuous and positive density on $\mathbb{R}$ bounded (up to a contant) by the function $M_p(x) = \exp(-|x|^p)$ for a positive constant p and assume that for all $M > 0$ there exists a, b such that*

$$g(y+u) \geq a \exp\{-b(|y|^p + |u|^p)\}, \quad \forall |y| \leq M, \quad \forall u \in \mathbb{R}$$

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 $k \in \mathbb{N}^*$,

$$\exp(-c_1 k L(k)) \leq p(k) \leq \exp(-c_2 k L(k)), \tag{9}$$

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

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

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

Conditionally on k we define the prior on $\mathcal{F}_k$ by assuming that the prior distribution π_k on $\theta = (\theta_\lambda)_{1 \leq \lambda \leq k}$ is given by

$$\frac{\theta_\lambda}{\sqrt{\tau_\lambda}} \sim g, \quad \tau_\lambda = \tau_0 \lambda^{-2\beta} \quad i.i.d.$$

where $\beta < 1/2 + p/2$ if $p \leq 2$ and $\beta < 1/2 + 1/p$ if $p > 2$.

Observe that we do not necessarily consider Gaussian prior since we allow for densities g to have heavier tails. The prior on k can be non random, which corresponds to the Dirac case (D). For the case (PH), $L(x) = \log(x)$ corresponds to a Poisson prior on k and the case $L(x) = 1$ corresponds to hypergeometric priors. Now, we have the the following result.

Theorem 3.1. *Assume that* $\|\log(f_0)\|_\infty < \infty$ *and that there exists* $\gamma > 1/2$ *such that* $\log(f_0) \in \mathcal{B}^{\gamma}_{p,\infty}$, *with* $p \geq 2$. *Then,*

$$\mathbb{P}^{\pi}\left\{f_\theta : \quad h(f_0, f_\theta) \leq \frac{\log n}{L(n)}\epsilon_n | X^n\right\} = 1 + o_P(1), \tag{10}$$

and

$$\mathbb{P}^{\pi}\left\{f_\theta : \quad \|\theta_0 - \theta\|_2 \leq \frac{(\log n)^2}{L(n)}\epsilon_n | X^n\right] = 1 + o_P(1), \tag{11}$$

where in case (PH),

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

in case (D),

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

and ϵ_0 *is a constant large enough.*

Remark 1. *If the density* g *only satisfies a tail condition of the form*

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

with $p > 1$, *then, in case (PH), if* $\gamma > 1$ *the rates defined by (10) and (11) remain valid.*

3.2 Bernstein Von Mises under these models

In this section, we apply Theorem 2.1 of Section 2.2 to establish the following Bernstein Von Mises-type result. For this purpose, let us expand the function ψ_c defined in (6) on the basis $(\phi_\lambda)_{\lambda\in\mathbb{N}}$:

$$\psi_c = \sum_{\lambda\in\mathbb{N}} \psi_{c,\lambda}\phi_\lambda.$$

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 $< f, g >= F_0(fg)$ and $\Delta_\psi = \psi_c - \Pi_{f_0,k}\Delta_\psi$. So we can write for any $x \in [0, 1]$,

$$\Pi_{f_0,k}\psi_c(x) = \psi_{\Pi,c,0} + \sum_{\lambda=1}^{k} \psi_{\Pi,c,\lambda}\phi_\lambda(x),$$

since $\phi_0(x) = 1$. We denote $B_{n,k}$ the renormalized sequence of coefficients that appear in the above sum :

$$B_{n,k} = \frac{\psi_{\Pi,c,[k]}}{\sqrt{n}}, \quad \psi_{\Pi,c,[k]} = (\psi_{\Pi,c,\lambda})_{1\leq\lambda\leq k}.$$

Such quantites will play a key role in the sequel. Let $l_0 > 0$ be large enough so that

$$\mathbb{P}^\pi \left[k > l_0 n\epsilon_n^2 | X^n\right] \leq e^{-cn\epsilon_n^2},$$

for some positive $c > 0$, where ϵ_n is the posterior concentration rate defined in Theorem 3.1 and define $l_n = l_0 n\epsilon_n^2$.

We have the following result.

Theorem 3.2. *Let us assume that the prior is defined as in Definition 3.1 and for all* $t \in \mathbb{R}$, $1 \leq k \leq l_n$ *and* θ

$$\frac{\pi_k(\theta)}{\pi_k(\theta - tB_{n,k})} = 1 + o(1) \tag{12}$$

uniformly. Under assumptions of Theorem 3.1

– *for all* $z \in \mathbb{R}$

$$\mathbb{P}^\pi \left[\sqrt{n}(\Psi(f) - \Psi(P_n)) \leq z | X^n\right] = \sum_k p(k|X^n)\Phi\left(\frac{z + \mu_{n,k}}{\sqrt{V_0}}\right) + o_{\mathbb{P}_0}(1), \tag{13}$$

where

– $V_0 = F_0(\tilde{\psi}^2)$, *which corresponds to the variance of* $\sqrt{n}\psi(\mathbb{P}_n)$ *under* $\mathbb{P}_0$.

– $\mu_{n,k} = -\sqrt{n}F_0[(\tilde{\psi} - M_{f_0,k}\tilde{\psi})\sum_{j\geq k+1}\theta_{0j}\phi_j] + \mathbb{G}_n(\Delta_\psi)$

– *In the case (D), if*

$$\sum_{\lambda=\ell_n+1}^{\infty} \psi_{c,\lambda}^2 = o\left(n^{\frac{2\gamma-2\beta-1}{2\beta+1}}(\log n)^{-2\gamma}\right) \tag{14}$$

then,

$$\mathbb{P}^\pi \left[\sqrt{n}(\Psi(f) - \Psi(P_n)) \leq z | X^n\right] = \Phi\left(\frac{z}{V_0}\right) + o_{\mathbb{P}_0}(1). \tag{15}$$

The first part of Theorem 3.2 shows that the posterior distribution of $\sqrt{n}(\Psi(f) - \Psi(P_n))$ is asymptotically a mixture of Gaussian distributions having the same variance V_0 but with different mean values $\mu_{n,k}$ with weight $p(k|X^n)$. To obtain an asymptotic Gaussian distribution with mean zero it is necessary for $\mu_{n,k}$ to be small whenever $p(k|X^n)$ is not. The conditions given in the second part of Theorem 3.2 ensure that this is the case, however they are not necessary conditions.

We now discuss condition (12) in three different examples.

– Gaussian : If g is Gausian then for all $k \leq l_n$ (or $k_{n,\beta}$ in the case of a type (D) prior) and all $j \leq k$, $\theta_j \sim \mathcal{N}(0, \tau_0^2 j^{-2\beta})$ and for all $\theta \in A_n \cap \mathcal{F}_k$

$$\begin{aligned} \frac{\sum_{j=1}^k \tilde{\psi}_j^2 j^{2\beta}}{n} &\leq \frac{k^{2\beta}}{n} \leq O(n^{2\beta-1}\epsilon_n^{4\beta}) = o(1) \\ \frac{\sum_{j=1}^k \theta_j \tilde{\psi}_j j^{2\beta}}{\sqrt{n}} &= \frac{\sum_{j=1}^k (\theta_j - \theta_{0j})\tilde{\psi}_j j^{2\beta} + \sum_{j=1}^k \theta_{0j}\tilde{\psi}_j j^{2\beta}}{\sqrt{n}} \leq \frac{C}{\sqrt{n}}\left[\|\theta - \theta_0\| k^{2\beta} + (k^{2\beta-\gamma} + 1)\right] \\ &= o(1). \end{aligned}$$

This implies that uniformly over A_n

$$\pi_k(\theta - B_{n,k}) = \pi_k(\theta)(1 + o(1))$$

– Laplace : If g is Laplace, $g(x) \propto e^{-|x|}$,

$$|\log(g(\frac{\theta_j - t\tilde{\psi}_j/\sqrt{n}}{\sqrt{\tau_j}})) - \log(g(\frac{\theta_j}{\sqrt{\tau_j}}))| \leq C\frac{|\tilde{\psi}_j|}{n}$$

So that

$$\begin{aligned} \left|\log\left[\frac{\pi_k(\theta - B_n)}{\pi_k(\theta)}\right]\right| &\leq C\frac{\sum_{j=1}^k j^\beta |\tilde{\psi}_j|}{\sqrt{n}} \\ &\leq C\frac{k^{\beta+1/2}}{\sqrt{n}} = o(1), \quad \text{if } k \leq k_n^* = n^{1/(2\beta+1)}(\log n)^{-1} \end{aligned}$$

Hence if $\gamma \geq \beta$ condition (12) is satisfied.

– Student : In the Student case for g we can use the calculations made in the Gaussian case since

$$\begin{aligned} &\sum_{j=1}^k \log\left(1 + Cj^{2\beta}\theta_j^2\right) - \log\left(1 + Cj^{2\beta}(\theta_j - t\tilde{\psi}_j/\sqrt{n})^2\right) \\ =\ & 0\left(\sum_{j=1}^k j^{2\beta}[(\theta_j - t\tilde{\psi}_j/\sqrt{n})^2 - \theta_j^2]\right) \\ =\ & o(1) \end{aligned}$$

Therefore in all these cases condition (12) is satisfied.

Interestingly Theorem 3.2 shows that parametric sieve models (increasing sequence of models) have a behaviour which is a mix between parametric and nonparametric 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 on the empirical (nonparametric MLE) estimator with the correct variance whereas if it allows for k's that are not large enough (corresponding to $\sum_{j=1}^k j^{2\beta}[(\theta_j - t\tilde{\psi}_j/\sqrt{n})^2 - \theta_j^2]$ or Δ_ψ not small enough) then the posterior distribution is not asymptotically Gaussian with the right centering. We illustrate in the following section this issue in the special case of the cumulative distribution function.

3.3 An example : the cumulative distribution function

As a special case consider the functional on f to be the cumulative distribution function calculated at a given point x. As seen in Section 2, $\tilde{\psi}(u) = 1\!\!1_{u\leq x} - F_0(x)$. Denote $F_n(x) = \mathbb{P}_n(\psi)$ and recall that the covariance of $\mathbb{G}_n(\psi)$ under P_0^n is equal to $V_0 = F_0(x)(1 - F_0(x))$.

As an illustration, consider the case of a non localised basis and more specifically the Fourier basis. The case of wavelet bases is dealt with in the same way. In other words set $\phi_{2j+1}(u) = \sqrt{2}\sin(2\pi(j+1)u)$ and $\phi_{2j}(u) = \sqrt{2}\cos(2\pi ju)$, $j \geq 1$ and $\phi_0(u) \equiv 1$.

Corollary 3.1. *If the prior density g on the coefficients is Gaussian or Laplace then if $f_0 \in \mathcal{S}_\gamma$, with $\gamma \geq \beta > 1/2$ and if the prior on k is the dirac mass on $k_n = k_{n,\beta}$ then the posterior distribution of $\sqrt{n}(F(x) - F_n(x))$ is asymptotically Gaussian with mean 0 and variance V_0.*

If the prior density g is Student and if $\gamma \geq \beta \geq 1$ then the same result remains valid.

Proof. To prove this Corollary we need only verify conditions (14) since condition (12) is proved in the previous section.

To prove condition (14) note that

$$\psi_{2j} = -\frac{\sin(2\pi jx)}{2\pi j}, \psi_{2j-1} = -\frac{1 - cos(2\pi jx)}{2\pi j}$$

Therefore

$$\sum_{j\geq k_{n,\beta}} \psi_j^2 = 0(k_{n,\beta}^{-1}) = o(n^{\frac{2(\gamma-\beta)-1}{2\beta+1}})$$

as soon as $\gamma \geq \beta$, which achieves the proof.

■

Remark : In this remark we illustrate the fact that in the case of a random k, which leads to an adaptive minimax rate of convergence for the posterior distribution we might not have a Bernstein - von Mises theorem. Consider a density f_0 in the form

$$f_0 = \exp\left(\sum_{j\geq k_0} \theta_{0j}\phi_j(u)du - c(\theta_0)\right)$$

where k_0 is fixed but can be large and $\theta_{0,2j} = 0$ and $\theta_{0,2j-1} = \sin(2\pi jx)/[j^{\gamma+1/2}\sqrt{\log j}\log\log j]$. Then for $J_1 > 3$

$$\begin{aligned}\sum_{j\geq J_1} \theta_{0j}^2 j^{2\gamma} &\leq \sum_{j\geq J_1} \frac{1}{j\log j \log\log j^2} \\ &\leq \int_{j_1}^{\infty} \frac{1}{x\log x(\log\log x)^2}dx \\ &= \frac{1}{\log\log J_1},\end{aligned}$$

and similarly

$$\begin{aligned}
\sum_{j\geq J_1}\theta_{0j}^2 &\leq \sum_{j\geq J_1}\frac{1}{j^{2\gamma+1}\log j\log\log j^2}\\
&\leq \int_{j_1}^{\infty}\frac{1}{x^{2\gamma+1}\log x(\log\log x)^2}dx\\
&= \left[-\frac{1}{2\gamma x^{2\gamma}\log x(\log\log x)^2}\right]_{J_1}^{\infty}(1+o(1))\\
&= \frac{1}{2\gamma J_1^{2\gamma}\log J_1(\log\log J_1)^2}(1+o(1)) \qquad (16)
\end{aligned}$$

when $J_1 \to \infty$.

Consider a Poisson distribution with parameter $\nu > 0$ fixed on k is then for such f_0, if $k_n = n^{1/(2\gamma+1)}(\log n)^{-2/(2\gamma+1)}(\log\log n)^{-2/(2\gamma+1)}$, if k_1 is large enough

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

We now study the mean terms $\mu_{n,k}$ and we show that there are some k's for which neither $\mu_{n,k}$ nor $\pi(k|X^n)$ can be neglected.

First note that when $k \to \infty$ $G_n(\Delta_\psi) = o(1)$

$$\begin{aligned}
\mu_{n,k} &= \sqrt{n}F_0\left[(\tilde{\psi} - M_{f_0,k}\tilde{\psi})(l_0 - M_{f_0,k}l_0)\right] \qquad (17)\\
&= \sqrt{n}F_0\left[(\sum_{j=k+1}^{\infty}\psi_j\phi_j)(l_0 - M_{f_0,k}l_0)\right] \qquad (18)\\
&= \sqrt{n}\int\left[(\tilde{\psi} - M_{f_0,k}\tilde{\psi})(l_0 - M_{f_0,k}l_0)\right] + \sqrt{n}\int(f_0 - 1)\left[(\tilde{\psi} - M_{f_0,k}\tilde{\psi})(l_0 - M_{f_0,k}l_0)\right] \qquad (19)
\end{aligned}$$

We first consider the first term of the right hand side of (17).

$$\begin{aligned}
\mu_{n,k,1} &= \sqrt{n}\int\left[(\sum_{j=k+1}^{\infty}\psi_j\phi_j)(l_0 - M_{f_0,k}l_0)\right]\\
&= \sqrt{n}\int\left[\sum_{j=k+1}^{\infty}\psi_j\sum_{j=k+1}^{\infty}\theta_{0j}\phi_j\right]\\
&= \sqrt{n}\sum_{j=k+1}^{\infty}\psi_j\theta_{0j}\\
&= \sqrt{n}\sum_{l\geq k/2}\frac{\sin^2(2\pi xl)}{(2l+1)^{\gamma+3/2}\log(2l+1)^{1/2}\log\log(2l+1)}
\end{aligned}$$

and if $x = 1/4$ we have

$$\begin{aligned}
\mu_{n,k,1} &= \sqrt{n} \sum_{j \geq k/4-1/2} \frac{1}{(4j+3)^{\gamma+3/2}(\log 4j+3)^{1/2}\log\log(4j+3)} \\
\mu_{n,k,1} &\leq \sqrt{n} \sum_{j \geq k/4+1/2} \frac{1}{j^{\gamma+3/2}(\log j)^{1/2}\log\log j} \\
&\leq 2^{2\gamma+1}\sqrt{n}\left[\frac{x^{-\gamma-1/2}}{2\gamma\sqrt{\log x}\log\log x}\right]_k^\infty \\
&\leq C\sqrt{n}\frac{k^{-\gamma-1/2}}{\sqrt{\log k}\log\log k} \\
\mu_{n,k,1} &\geq 2^{-2\gamma-1}\sqrt{n} \sum_{j \geq k/4+1/2} \frac{1}{j^{\gamma+3/2}(\log j)^{1/2}\log\log j} \\
&\geq C'\sqrt{n}\frac{k^{-\gamma-1/2}}{\sqrt{\log k}\log\log k}.
\end{aligned}$$

Note that there exists $c > 0$ such that for all $k \leq k_n$

$$\mu_{n,k,1} \geq c\sqrt{\log n}.$$

We now consider the second term of (17). Let $M_{1,k}$ denote the projection on $(\phi_0, ..., \phi_k)$ with respect to the scalar product $< f, g >_2 = \int fg(u)du$ and note that

$$M_{f_0,k}l_0 = M_{1,k}l_0 + M_{f_0,k}[\sum_{j=k+1}^{\infty} \theta_{0j}\phi_j]$$

$$\begin{aligned}
|\mu_{n,k,2}| &= \left|\sqrt{n}\int (f_0-1)\left[(\sum_{j=k+1}^{\infty}\psi_j\phi_j)(l_0 - M_{f_0,k}l_0)\right]\right| \\
&= \left|\sqrt{n}\int (f_0-1)\left[(\sum_{j=k+1}^{\infty}\psi_j\phi_j)(l_0 - M_{1,k}l_0)\right]\right| \\
&\quad + \left|\sqrt{n}\int (f_0-1)\left[(\sum_{j=k+1}^{\infty}\psi_j\phi_j)(M_{1,k}l_0 - M_{f_0,k}l_0)\right]\right| \\
&\leq 2|f_0-1|_\infty \left(\sum_{j=k+1}^{\infty}\psi_j^2\right)^{1/2}\left(\sum_{j=k+1}^{\infty}\theta_{0,j}^2\right)^{1/2} \\
&\leq C\sqrt{n}|f_0-1|_\infty \frac{k^{-\gamma-1/2}}{\sqrt{\log k}\log\log k}
\end{aligned}$$

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

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

Thus in this case the posterior distribution is not asymptotically Gaussian with mean $F_n(x)$ and variance $F_0(x)(1 - F_0(x))/n$. Wether it is asymptotically equivalent to a mixtures of Gaussian is not clear. It would be a consequence of the way the posterior distribution of k concentrates as n goes to infinity.

4 Proofs

In the sequel, C denotes a generic positive constant whose value is of no importance.

4.1 Preliminary lemma

Let us first state the following lemma.

Lemma 1. *Set $K_n = \{1, 2, \ldots, k_n\}$ 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.*

Then the following results hold.

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

$$\left\| \sum_{\lambda \in K_n} \theta_\lambda \phi_\lambda \right\|_\infty \leq c_{1,\Phi}\sqrt{k_n}\|\theta\|_{\ell_2}. \tag{20}$$

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

$$\sum_{\lambda \notin K_n} \theta^2_{0\lambda} \leq c_{2,\gamma}\, R^2 k_n^{-2\gamma}. \tag{21}$$

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

$$\left\| \sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda \right\|_\infty \leq c_{3,\Phi,\gamma}\, R\, k_n^{\frac{1}{2}-\gamma}. \tag{22}$$

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 (20). Inequality (21) follows from the definition of $\mathcal{B}^{\gamma}_{2,2} = W^{\gamma}$. To prove (22), we use the following inequality : for any x,

$$\begin{aligned}\left|\sum_{\lambda\notin K_n}\theta_{0\lambda}\phi_\lambda(x)\right| &\leq \|\phi\|_\infty \sum_{\lambda\notin K_n}|\theta_{0\lambda}|\\ &\leq \|\phi\|_\infty\left(\sum_{\lambda\notin K_n}|\lambda|^{2\gamma}\theta^2_{0\lambda}\right)^{\frac12}\left(\sum_{\lambda\notin K_n}|\lambda|^{-2\gamma}\right)^{\frac12}.\end{aligned}$$

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,

$$\begin{aligned}\left|\sum_{\lambda\in K_n}\theta_{\lambda}\phi_\lambda(x)\right| &\leq \left(\sum_{\lambda\in K_n}\theta^2_{\lambda}\right)^{\frac12}\left(\sum_{\lambda\in K_n}\phi^2_\lambda(x)\right)^{\frac12}\\ &\leq \|\theta\|_{\ell_2}\left(\sum_{0\leq j\leq\log_2(k_n)}\sum_{k=0}^{2^j-1}\Upsilon^2_{jk}(x)\right)^{\frac12}.\end{aligned}$$

Since $\Upsilon(x)=0$ for $x\notin[-A,A]$,

$$\mathrm{card}\left\{k\in\{0,\dots,2^j-1\}:\quad \Upsilon_{jk}(x)\neq 0\right\}\leq 3(2A+1).$$

(see [14], p. 282 or [15], p. 112). So, there exists c_Υ depending only on Υ such that

$$\left|\sum_{\lambda\in K_n}\theta_{\lambda}\phi_\lambda(x)\right| \leq \|\theta\|_{\ell_2}\left(\sum_{0\leq j\leq\log_2(k_n)}3(2A+1)2^jc^2_\Upsilon\right)^{\frac12},$$

which proves (20). For the second point, we just use the inclusion $\mathcal{B}^{\gamma}_{p,q}(R)\subset\mathcal{B}^{\gamma}_{2,\infty}(R)$ and

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

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

$$\begin{aligned}\left|\sum_{\lambda\notin K_n}\theta_{0\lambda}\phi_\lambda(x)\right| &\leq \sum_{j>\log_2(k_n)}\left(\sum_{k=0}^{2^j-1}\theta^2_{0jk}\right)^{\frac12}\left(\sum_{k=0}^{2^j-1}\Upsilon^2_{jk}(x)\right)^{\frac12}\\ &\leq \sum_{j>\log_2(k_n)}\left(R^22^{-2j\gamma}\right)^{\frac12}\left(3(2A+1)2^jc^2_\Upsilon\right)^{\frac12}\\ &\leq R(3(2A+1))^{\frac12}c_\Upsilon\sum_{j>\log_2(k_n)}2^{j\left(\frac12-\gamma\right)}\\ &\leq \frac{R(3(2A+1))^{\frac12}c_\Upsilon}{1-2^{\frac12-\gamma}}k_n^{\frac12-\gamma}.\end{aligned}$$

■

4.2 Proof of Theorem 2.1

Let $Z_n = \sqrt{n}(\Psi(f) - \Psi(P_n))$. Using (A1) and (A2), we have

$$\mathbb{P}^\pi \left\{A_n | X^n\right\} = 1 + o_{\mathbb{P}_0}(1). \tag{23}$$

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

$$\lim_{n\to+\infty} L_n(t) = \exp\left(\frac{t^2}{2} F_0\left[\psi_c^2\right]\right), \tag{24}$$

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

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

Recall that we have set for any x,

$$B_{h,n}(x) = \int_0^1 (1-u) \exp\left(\frac{uh(x)}{\sqrt{n}}\right) du,$$

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

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

Since

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

we have

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

where straightforward computations show that

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

Now, let us study each term of the last expression. For this purpose, let us state the following lemma.

Lemma 2. *On A_n, we have*
- $F_0(h^2) = 0(nu_n^2)$,
- $F_0\left(h^2 \left|B_{h,n} - \frac{1}{2}\right|\right) = o(n)$

Proof. For the first point, we observe :

$$
F_0(h^2) = n \int \left(\log\left(\frac{f(x)}{f_0(x)}\right)\right)^2 f_0(x)dx = nV(f_0, f) \le nu_n^2.
$$

For the second point, we use following inequalities.

$$
\begin{aligned}
F_0\left(h^2 \left|B_{h,n} - \frac{1}{2}\right|\right) &= F_0\left(h^2 \left|B_{h,n} - \frac{1}{2}\right| \left(\mathbb{1}_{|h|\le\sqrt{n}} + \mathbb{1}_{|h|>\sqrt{n}}\right)\right) \\
&\le \int h^2(x) f_0(x) \left(\frac{e^1}{6}\frac{|h(x)|}{\sqrt{n}} \mathbb{1}_{|h(x)|\le\sqrt{n}} + \left(\frac{e^{\frac{h(x)}{\sqrt{n}}} + 1}{2}\right) \mathbb{1}_{|h(x)|>\sqrt{n}}\right) dx \\
&\le \left(\frac{1}{2} + \frac{e^1}{6}\right) \frac{F_0\left(|h|^3\right)}{\sqrt{n}} + \frac{1}{2} F(h^2 \mathbb{1}_{|h|>\sqrt{n}}) \\
&\le n \int \left|\log\left(\frac{f(x)}{f_0(x)}\right)\right|^3 \left(\left(\frac{1}{2} + \frac{e^1}{6}\right) f_0(x) + \frac{1}{2} f(x)\right) dx \\
&= o(n).
\end{aligned}
$$

■

Now, we have

$$\begin{aligned} F_0\left[\exp\left(\frac{h}{\sqrt{n}}-\frac{t\psi_c}{\sqrt{n}}\right)\right] &= F_0\left[e^{\frac{h}{\sqrt{n}}}\left(1-\frac{t\psi_c}{\sqrt{n}}+\frac{t^2}{2n}\psi_c^2\right)\right]+0(n^{-\frac{3}{2}}) \\ &= 1-\frac{t}{\sqrt{n}}F_0\left[e^{\frac{h}{\sqrt{n}}}\psi_c\right]+\frac{t^2}{2n}F_0\left[e^{\frac{h}{\sqrt{n}}}\psi_c^2\right]+0(n^{-\frac{3}{2}}). \end{aligned}$$

So,

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

Note that $F_0(|h|)=0(\sqrt{n}u_n)$. Therefore, uniformly on A_n,

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

and

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

Finally,

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

and

$$L_n(t)=\exp\left(\frac{t^2}{2}F_0\left[\psi_c^2\right]\right)\frac{\int_{A_n}\exp\left(-\frac{F_0((h-t\bar{\psi}_{t,n})^2)}{2}+G_n(h-t\bar{\psi}_{t,n})+R_n(h-t\bar{\psi}_{t,n})\right)d\pi(f)}{\int_{A_n}\exp\left(-\frac{F_0(h^2)}{2}+G_n(h)+R_n(h)\right)d\pi(f)}(1+o(1)).$$

Finally (A3) implies (24) and the theorem is proved.

4.3 Proof of Theorem 3.1

Denote for any n,

$$B_n(\epsilon_n)=\{f\in\mathcal{F}:\quad K(f_0,f)\le\epsilon_n^2,\ V(f_0,f)\le\epsilon_n^2\},$$

where $K(f_0,f)$ is the Kullback-Leibler divergence between f_0 and f :

$$K(f_0,f)=\int_0^1\log\left(\frac{f_0(x)}{f(x)}\right)f_0(x)dx$$

and

$$V(f_0, f) = \int_0^1 \left(\log\left(\frac{f_0(x)}{f(x)}\right)\right)^2 f_0(x)dx.$$

To prove Theorem 3.1, we use the following version of the theorem on posterior convergence rates. Its proof is not given, but it is a slight modification of Theorem 2.4 of [8].

Theorem 4.1. *Let f_0 be the true density. We assume that there exists a constant c such that for any n, there exists $\mathcal{F}_n^* \subset \mathcal{F}$ and a prior π on $\mathcal{F}$ satisfying the following conditions :*

- *(A)*

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

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

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

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*

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

Then,

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

We have :

$$\mathbb{P}^\pi\left\{f : \quad h(f_0, f) \leq J_{0,n}\epsilon_n | X^n\right\} = 1 + o_P(1)$$

To prove Theorem 3.1 it is thus enough to prove that conditions (A), (B) and (C) of the previous result are satisfied. We consider $(\Lambda_n)_n$ the increasing sequence of subets of $\mathbb{N}^*$ defined by

$$\Lambda_n = \{1, 2, \ldots, l_n\}$$

with $l_n \in \mathbb{N}^*$. For any n, we set :

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

with

$$w_n = \exp(w_0 n^\rho (\log n)^q), \quad \rho > 0$$

Recall that

- $\epsilon_n = \epsilon_0 n^{-\frac{\gamma}{2\gamma+1}} (\log n)^{\frac{\gamma}{2\gamma+1}}$ in case (PH)

- $\epsilon_n = \epsilon_0 n^{-\frac{\beta}{2\beta+1}}$ in case (D).
Define l_n by

$$l_n = \frac{l_0 n\epsilon_n^2}{L(n)}, \tag{25}$$

where l_0 is some positive constant. When $\gamma, \beta > \frac{1}{2}$, we have

$$l_n\epsilon_n^2 \to 0. \tag{26}$$

Proof of condition (A) : We have, since $\sum_k \tau_k < \infty$

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

for any positive $H > 0$. Hence,

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

and Condition (A) is proved.

Proof of condition (B) : We apply Lemma 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 $\|\theta' - \theta\|_{\ell_1} \leq c_1\epsilon_n l_n^{-1/2}$ with c_1 a positive constant, then using (20) and (26),

$$\left\|\sum_{\lambda\in\Lambda_n} (\theta'_\lambda - \theta_\lambda)\phi_\lambda\right\|_\infty \leq C\sqrt{l_n}\|\theta'-\theta\|_{\ell_2} \leq C\sqrt{l_n}\|\theta'-\theta\|_{\ell_1} \leq Cc_1\epsilon_n \to 0$$

and

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

Then,

$$\begin{aligned}
h^2(f_\theta,f_{\theta'}) &= \int f_\theta(x)\left(\exp\left(\frac{1}{2}\sum_{\lambda\in\Lambda_n}(\theta'_\lambda-\theta_\lambda)\phi_\lambda(x)+\frac{1}{2}\left(c(\theta)-c(\theta')\right)\right)-1\right)^2 dx \\
&\le \int_0^1 f_\theta(x)\left(\exp\left(C\|\sum_{\lambda\in\Lambda_n}(\theta'_\lambda-\theta_\lambda)\phi_\lambda\|_\infty\right)-1\right)^2 dx \\
&\le C\|\sum_{\lambda\in\Lambda_n}(\theta_\lambda-\theta'_\lambda)\phi_\lambda\|_\infty^2 \\
&\le Cl_n\|\theta-\theta'\|_{\ell_1}^2 \le Cl_n^2\|\theta-\theta'\|_{\ell_2}^2 \qquad (27)
\end{aligned}$$

The next lemma establishes a converse inequality.

Lemma 3. *There exists a constant $c\le 1/2$ depending on γ, β,R and Φ such that if*

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

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

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

Note that in cases where $\epsilon^2 l_n$ is not bounded (which only occures in some cases where $\gamma<\beta$) then there is no such j.

Proof. Using Theorem 5 of [22], with $M_1=\left(\int_0^1\frac{f_0^2(x)}{f_\theta(x)}dx\right)^{\frac{1}{2}}$, if

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

we have

$$V(f_0,f_\theta) \le 5h^2(f_0,f_\theta)\left(|\log M_1|-\log(h(f_0,f_\theta)\right)^2. \qquad (28)$$

But

$$\begin{aligned} M_1 &= \int_0^1 f_0(x) \exp\left(\sum_{\lambda \in \Lambda_n} (\theta_{0\lambda} - \theta_\lambda)\phi_\lambda(x) + \sum_{\lambda \notin \Lambda_n} \theta_{0\lambda}\phi_\lambda(x) - c(\theta_0) + c(\theta) \right) dx \\ &\leq \int_0^1 f_0(x) \exp\left(c_{1,\Phi}\sqrt{l_n}\|\theta_0 - \theta\|_{\ell_2} + c_{3,\Phi,\gamma} R \, l_n^{\frac{1}{2}-\gamma} - c(\theta_0) + c(\theta) \right) dx, \end{aligned}$$

by using (20) and (22). Furthermore,

$$\begin{aligned} |c(\theta_0) - c(\theta)| &= \left| \log\left(\int_0^1 f_0(x) \exp\left(-\sum_{\lambda \in \Lambda_n} (\theta_{0\lambda} - \theta_\lambda)\phi_\lambda(x) - \sum_{\lambda \notin \Lambda_n} \theta_{0\lambda}\phi_\lambda(x) \right) dx \right) \right| \\ &\leq \left| \log\left(\int_0^1 f_0(x) \exp\left(c_{1,\Phi}\sqrt{l_n}\|\theta_0 - \theta\|_{\ell_2} + c_{3,\Phi,\gamma} R \, l_n^{\frac{1}{2}-\gamma} \right) dx \right) \right| \\ &= c_{1,\Phi}\sqrt{l_n}\|\theta_0 - \theta\|_{\ell_2} + c_{3,\Phi,\gamma} R \, l_n^{\frac{1}{2}-\gamma}. \end{aligned} \tag{29}$$

So,

$$|\log M_1| \leq 2c_{1,\Phi}\sqrt{l_n}\|\theta_0 - \theta\|_{\ell_2} + 2c_{3,\Phi,\gamma} R \, l_n^{\frac{1}{2}-\gamma}.$$

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

$$\begin{aligned} V(f_0, f_\theta) &\leq 5h^2(f_0, f_\theta) \left(2c_{1,\Phi}\sqrt{l_n}\|\theta_0 - \theta\|_{\ell_2} + 2c_{3,\Phi,\gamma} R \, l_n^{\frac{1}{2}-\gamma} - \log(\epsilon_n) \right)^2 \\ &\leq c_{4,\gamma,\beta,\Phi,R} h^2(f_0, f_\theta) \left(\sqrt{l_n}\|\theta_0 - \theta\|_{\ell_2} + \log n \right)^2 \\ &\leq 2c_{4,\gamma,\beta,\Phi,R} h^2(f_0, f_\theta) \left(l_n\|\theta_0 - \theta\|_{\ell_2}^2 + (\log n)^2 \right), \end{aligned}$$

where $c_{4,\gamma,\beta,R,\Phi}$ is a constant that only depends on γ, β,R and Φ. Since $f_0(x) \geq c_0$ for any x and $\int_0^1 \phi_\lambda(x)dx = 0$ for any $\lambda \in \Lambda$, we have

$$\begin{aligned} V(f_0, f_\theta) &\geq c_0 \int_0^1 \left(\sum_{\lambda \in \Lambda_n} (\theta_{0\lambda} - \theta_\lambda)\phi_\lambda(x) + \sum_{\lambda \notin \Lambda_n} \theta_{0\lambda}\phi_\lambda(x) + c(\theta) - c(\theta_0) \right)^2 dx \\ &\geq c_0 \left[\sum_{\lambda \in \Lambda_n} (\theta_{0\lambda} - \theta_\lambda)^2 + \sum_{\lambda \notin \Lambda_n} \theta_{0\lambda}^2 \right] \\ &\geq c_0 \|\theta_0 - \theta\|_{\ell_2}^2. \end{aligned} \tag{30}$$

Combining (27) and (30), we conclude that

$$\begin{aligned} \|\theta_0 - \theta\|_{\ell_2}^2 &\leq \frac{2c_{4,\gamma,\beta,\Phi,R}}{c_0} h^2(f_0, f_\theta) \left(l_n\|\theta_0 - \theta\|_{\ell_2}^2 + (\log n)^2 \right) \\ &\leq \frac{4c_{4,\gamma,\beta,\Phi,R}}{c_0} (\log n)^2 h^2(f_0, f_\theta), \end{aligned}$$

if
$$\frac{2c_{4,\gamma,\beta,\Phi,R}}{c_0} h^2(f_0,f_\theta) l_n \leq \frac{2c_{4,\gamma,\beta,\Phi,R}}{c_0}(j+1)^2\epsilon_n^2 l_n \leq \frac{1}{2}.$$
Lemma 3 is proved by taking $c = (\max(2c_{4,\gamma,\beta,\Phi,R},1))^{-1}/2$. ■

Now, under assumptions of Lemma 3, using (27), 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, since $l_n L(n) = l_0 n\epsilon_n^2$, we have
$$H_{n,j} \leq (K-1)nj^2\epsilon_n^2$$
as soon as
$$J_{0,n}^2 \geq \frac{j_0 \log n}{L(n)},$$
where j_0 is a constant and condition (B) is satisfied for such j's. Now, let j be such that
$$c(j+1)^2\epsilon_n^2 l_n > \min\left(\frac{c_0}{2}, \frac{1}{2}(1-e^{-1})^2\right). \tag{31}$$
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,
$$H_{n,j} \leq \log\left(\left(Cl_n w_n \epsilon_n^{-1}\right)^{l_n}\right) \leq 2l_n \log(w_n) \leq 2w_0 l_n n^\rho (\log n)^q.$$
Then, using (31), condition (B) is satisfied if w_0 and q are small enough and if
$$l_n^2(\log n)^q \leq n^{1-\rho},$$
which is true for n large enough, since $\gamma,\beta > \frac{1}{2}$, for ρ small enough.

Proof of condition (C) : In the sequel, D is a constant that only depends on $\|f_0\|_\infty$, γ, R and Φ and that may change from line to line. Now, we use the set K_n defined in Lemma 1, such that $k_n = \mathrm{card}(K_n)$, that will be specified later, goes to ∞. We assume that θ belongs to $A(u_n)$ where
$$A(u_n) = \left\{\theta : \quad \theta_\lambda = 0 \text{ for every } \lambda \notin K_n \text{ and } \sum_{\lambda\in K_n}(\theta_{0\lambda}-\theta_\lambda)^2 \leq u_n^2\right\}, \tag{32}$$
where u_n goes to 0 such that
$$\sqrt{k_n}u_n \to 0. \tag{33}$$

We define for any $\lambda \in \Lambda$,

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

Let us introduce the following notations :

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

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 (22) of Lemma 1 and a Taylor expansion of the function e^x we obtain

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

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,

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

So, finally,

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

This implies that for n large enough,

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

Now, if $f_\theta \in \mathcal{F}_{k_n}$ with

$$f_\theta = \exp\left(\sum_{\lambda\in K_n}\theta_\lambda\phi_\lambda - c(\theta)\right),$$

we have

$$\begin{aligned} K(f_0, f_\theta) &= \sum_{\lambda\notin K_n}\theta_{0\lambda}\beta_\lambda(f_0) + \sum_{\lambda\in K_n}(\theta_{0\lambda}-\theta_\lambda)\beta_\lambda(f_0) - c(\theta_0) + c(\theta_{0K_n}) - c(\theta_{0K_n}) + c(\theta) \\ &= K(f_0, f_{0K_n}) + \sum_{\lambda\in K_n}(\theta_{0\lambda}-\theta_\lambda)\beta_\lambda(f_0) - c(\theta_{0K_n}) + c(\theta) \\ &\le Dk_n^{-2\gamma} + \sum_{\lambda\in K_n}(\theta_{0\lambda}-\theta_\lambda)\beta_\lambda(f_0) - c(\theta_{0K_n}) + c(\theta). \end{aligned}$$

We set for any x,

$$T(x) = \sum_{\lambda\in K_n}(\theta_\lambda - \theta_{0\lambda})\phi_\lambda(x).$$

Using (20),

$$\|T\|_\infty \le c_{1,\Phi}\sqrt{k_n}u_n \to 0.$$

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 $\log(1+u) \le u$ for any $u > -1$ and for n large enough,

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

Then, for $\theta \in A(u_n)$ and n large enough,

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

So,

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

Using (22), we have

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

and

$$\begin{aligned}
|c(\theta_{0K_n}) - c(\theta_0)| &= \left|\log\left(\int_0^1 f_0(x) \exp\left(-\sum_{\lambda\notin K_n} \theta_{0\lambda}\phi_\lambda(x)\right) dx\right)\right| \\
&\le \|\sum_{\lambda\notin K_n} \theta_{0\lambda}\phi_\lambda\|_\infty.
\end{aligned}$$

Finally,

$$\|f_0 - f_{0K_n}\|_2 \le D\|\sum_{\lambda\notin K_n} \theta_{0\lambda}\phi_\lambda\|_\infty \le D k_n^{\frac{1}{2}-\gamma}.$$

and

$$K(f_0, f_\theta) \le D k_n^{-2\gamma} + D u_n k_n^{\frac{1}{2}-\gamma}. \tag{34}$$

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,

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

In addition,

$$\begin{aligned}
|c(\theta_{0K_n}) - c(\theta)| &\leq \sum_{\lambda\in K_n} |\theta_\lambda - \theta_{0\lambda}|\,|\beta_\lambda(f_{0K_n})| + Dk_n u_n^2 \\
&\leq u_n\left(\|f_0 - f_{0K_n}\|_2 + \|f_0\|_2\right) + Dk_n u_n^2 \\
&\leq Du_n + Dk_n u_n^2
\end{aligned}$$

Finally,

$$\begin{aligned}
V(f_0, f_\theta) &= \int_0^1 f_0(x)\left(\log\left(\frac{f_0(x)}{f_\theta(x)}\right)\right)^2 dx \\
&= \int_0^1 \left(\sum_{\lambda\in\Lambda}(\theta_{0\lambda} - \theta_\lambda)\phi_\lambda(x) - c(\theta_0) + c(\theta)\right)^2 f_0(x)dx \\
&\leq \|f_0\|_\infty \int_0^1 \left(\sum_{\lambda\in\Lambda}(\theta_{0\lambda} - \theta_\lambda)\phi_\lambda(x) + c(\theta_{0K_n}) - c(\theta_0) + c(\theta) - c(\theta_{0K_n})\right)^2 dx \\
&\leq \|f_0\|_\infty \left(\int_0^1 \left(\sum_{\lambda\in\Lambda}(\theta_{0\lambda} - \theta_\lambda)\phi_\lambda(x)\right)^2 dx + 2(c(\theta_{0K_n}) - c(\theta_0))^2 + 2(c(\theta) - c(\theta_{0K_n}))^2\right) \\
&\leq u_n^2 + Dk_n^{-2\gamma} + Dk_n u_n^2. \qquad (35)
\end{aligned}$$

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

$$k_n^{-2\gamma} \leq k_0\epsilon_n^2 \quad \text{and } u_n = u_0\epsilon_n k_n^{-\frac{1}{2}}, \qquad (36)$$

where k_0 and u_0 are constants depending on $\|f_0\|_\infty$, γ, R and Φ. If k_0 and u_0 are small enough, then, by using (34) and (35),

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

So, Condition (C) is satisfied if

$$\mathbb{P}^\pi\left\{A(u_n)\right\} \geq e^{-cn\epsilon_n^2},$$

where, $A(u_n)$ is defined in (32). We have :

$$\mathbb{P}^\pi\left\{A(u_n)\right\} \geq \mathbb{P}^\pi\left\{\theta : \sum_{\lambda\in K_n}(\theta_\lambda-\theta_{0\lambda})^2 \leq u_n^2\right\}\times\exp\left(-c_1k_nL(k_n)\right\}$$

The prior on θ implies that

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

Using (21), 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 since

$$\sup_n\left\{\tau_0^{-\frac{1}{2}}k_n^\beta u_n\right\}<\infty \tag{37}$$

using assumptions on the prior, there exists a constant D_3 such that

$$\begin{aligned}
P_1 &\geq D_3^{k_n}\int\ldots\int 1_{\left\{\sum_{\lambda\in K_n}\lambda^{-\beta}|y_\lambda|\leq\tau_0^{-\frac{1}{2}}u_n\right\}}\prod_{\lambda\in K_n}dy_\lambda\\
&\geq D_3^{k_n}\int\ldots\int 1_{\left\{\sum_{\lambda\in K_n}|y_\lambda|\leq\tau_0^{-\frac{1}{2}}u_n\right\}}\prod_{\lambda\in K_n}dy_\lambda\\
&\geq \exp\left(-D_4k_n\log n\right),
\end{aligned}\tag{38}$$

where D_4 is a constant. When $\gamma<\beta$, since there exists $a,b>0$ such that $\forall|y|\leq M$ for some positive M

$$g(y+u)\geq a\exp(-b|u|^p)$$

using the above calculations we obtain if $p\leq 2$

$$\begin{aligned}
P_1 &\geq D_3^{k_n}\exp\{-C\sum_{\lambda\in K_n}\lambda^{p\beta}|\theta_{0\lambda}|^p\}\sum\int\ldots\int 1_{\left\{\sum_{\lambda\in K_n}\lambda^{-\beta}|y_\lambda|\leq\tau_0^{-\frac{1}{2}}u_n\right\}}\prod_{\lambda\in K_n}dy_\lambda\\
&\geq \exp\left[-Ck_n^{1-p/2+\beta-\gamma}\right]\exp\left(-D_4k_n\log n\right)\\
&\geq \exp\left(-(D_4+1)k_n\log n\right) \quad \text{if } \beta\leq 1/2+p/2
\end{aligned}$$

and if $t > 2$

$$\begin{aligned} P_1 &\geq D_3^{k_n} \exp\{-C \sum_{\lambda \in K_n} \lambda^{p\beta} |\theta_{0\lambda}|^p\} \exp(-D_4 k_n \log n) \\ &\geq \exp(-(D_4+1) k_n \log n) \quad \text{ if } \beta \leq 1/2 + 1/p \end{aligned}$$

So, Condition (C) is established as soon as $D_4 k_n \log n \leq c n \epsilon_n^2$. Using (36), this can be satisfied if and only if we take k_n such that

$$k_0^{-\frac{1}{2\gamma}} \epsilon_n^{-\frac{1}{\gamma}} \leq k_n \leq \frac{c n \epsilon_n^2}{D_4 \log n}, \tag{39}$$

which is possible if and only if ϵ_0 is large enough. In particular, this implies that

$$\sup_n \left\{ \epsilon_n \left(\frac{\log n}{n} \right)^{-\frac{\gamma}{2\gamma+1}} \right\} < \infty.$$

Note that when k_n satisfies (39), Conditions (33) and (37) are satisfied as well.

Similar computations show the result for the case (D) with

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

by taking $k_n = k_n^*$ and $u_n = u_0 \epsilon_n k_n^{-\frac{1}{2}}$. The rate (11) is proved by using Lemma 3 in the case of $\gamma \geq \beta$. If $\gamma < \beta$, the constraints become

$$\epsilon_n^2 \geq (k_n^*) {-2\gamma} = \epsilon_0^2 n^{-2\gamma/(2\beta+1)} (\log n)^{2\gamma}, \quad \text{and} \quad n\epsilon_n^2 \geq k_n^* \log n$$

therefore the posterior concentration rate is bounded by $O(n^{-2\gamma/(2\beta+1)} (\log n)^{2\gamma})$.

4.4 Proof of Theorem 3.2

Our goal is to prove conditions (A1), (A2) and (A3) of Section 2.2. Let ϵ_n be the posterior concentration rate as obtained in Theorem 3.1.

Let us consider $f = f_\theta \in \mathcal{F}_k$ for $1 \leq k \leq l_n$, where $l_n = l_0 n \epsilon_n^2$ in the case of type (PH) priors and $l_n = k_n^*$ in the case of type (D) priors. First, we have, using the same upper bound as in the proof of Lemma 3 we have

$$\begin{aligned} V(f_0, f) &= \int_0^1 \left(\log \left(\frac{f_0(x)}{f(x)} \right) \right)^2 f_0(x) dx \\ &\leq C h^2(f_0, f) [l_n \|\theta - \theta_0\|_{l_2}^2 + (\log n)^2] \\ &\leq 2C (\log n)^2 \epsilon_n^2, \end{aligned}$$

as soon as $\theta \in \{f; h(f_0, f) \leq \epsilon_n\}$. Thus, using (11), we have

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

with

$$u_n = u_0(\log n)^2 \epsilon_n^2 = o(n^{-1/2})$$

for a constant u_0 large enough. Note that we can restrict ourselves to $A^1_{u_n} \cap (\cup_{k\leq l_n} \mathcal{F}_k)$, since $\mathbb{P}^\pi\left[(\cup_{k\leq l_n}\mathcal{F}_k)^c\right] \leq e^{-cn\epsilon_n^2}$ for any $c > 0$ by choosing l_0 large enough, see the proof of Theorem 3.1.

To establish (A2), we observe that

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

by using Lemma 1 and (29). So, (A2) is implied by (A1). Now, let us establish (A3). Denote A_n the set defined in assumption (A2) and restricted to $(\cup_{k\leq l_n}\mathcal{F}_k)$. For any t, we study the term

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

and we show that $I_n = ... + o_{\mathbb{P}_0}(1)$. If we set

$$b_{n,k,t} = \frac{t\Pi_{f_0,k}\psi_c - t\psi_{\Pi,c,0}}{\sqrt{n}} = \frac{t}{\sqrt{n}}\sum_{\lambda=1}^{k}\psi_{\Pi,c,\lambda}\phi_\lambda,$$

we have using (20) and since $k \leq l_n$:

$$\begin{aligned}
\|b_{n,k,t}\|_\infty &\leq \frac{t\sqrt{k}}{\sqrt{n}}\|\psi_{\Pi,c,[k]}\|_{\ell_2} \\
&\leq \frac{2t\sqrt{k}}{\sqrt{c_0}\sqrt{n}}\|\psi_c\|_{f_0,2} = O\left(\sqrt{\frac{l_n}{n}}\right) = O(\epsilon_n).
\end{aligned}$$

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

$$h_\theta = \sqrt{n}\left(\sum_{\lambda\in\mathbb{N}^*}(\theta_\lambda - \theta_{0\lambda})\phi_\lambda - c(\theta) + c(\theta_0)\right) \quad \text{and} \quad B_{n,k} = \frac{\psi_{\Pi,c,[k]}}{\sqrt{n}}$$

so, for $\theta' = \theta - B_{n,k}$, with $H_n = (h_\theta - t\psi_c)/\sqrt{n}$ and $\Delta_\psi = \psi_c - \Pi_{f_0,k}\psi_c$

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

with

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

Now, we use

$$\begin{aligned} F_0(e^{H_n}) &= F_0\left(e^{h_\theta/\sqrt{n}}\left(1-\frac{t}{\sqrt{n}}\psi_c+O\left(\frac{1}{n}\right)\right)\right) \\ &= 1-\frac{t}{\sqrt{n}}F_0\left(e^{h_\theta/\sqrt{n}}\psi_c\right)+O\left(\frac{1}{n}\right) \\ &= 1+O\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Therefore,

$$\begin{aligned} F_0(e^{H_n+t\Delta_\psi/\sqrt{n}}) &= F_0\left(e^{H_n}\left(1+\frac{t\Delta_\psi}{\sqrt{n}}+\frac{t^2\Delta_\psi^2}{2n}\right)\right)+0\left(F(\Delta_\psi^2)\frac{\|\Delta_\psi\|_\infty}{n^{3/2}}\right) \\ &= F_0\left(e^{H_n}\left(1+\frac{t\Delta_\psi}{\sqrt{n}}+\frac{t^2\Delta_\psi^2}{2n}\right)\right)+0\left(\frac{\epsilon_n}{n}\right) \\ &= F_0\left(e^{H_n}\right)+\frac{t}{\sqrt{n}}F_0(e^{H_n}\Delta_\psi)+\frac{t^2}{2n}F_0(e^{H_n}\Delta_\psi^2)+o\left(\frac{1}{n}\right), \end{aligned}$$

so,

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

Using

$$\begin{aligned} \frac{t}{\sqrt{n}}F_0(e^{H_n}\Delta_\psi) &= \frac{t}{\sqrt{n}}F_0\left(e^{h_\theta/\sqrt{n}}\Delta_\psi\left(1-t\psi_c/\sqrt{n}+O(n^{-1})\right)\right) \\ &= \frac{t}{\sqrt{n}}F_0\left(e^{h_\theta/\sqrt{n}}\Delta_\psi\right)-\frac{t^2}{n}F_0\left(e^{h_\theta/\sqrt{n}}\Delta_\psi\psi_c\right)+o\left(\frac{1}{n}\right) \end{aligned}$$

Since $F_0(e^{H_n}\Delta_\psi)=F(\Delta_\psi)=0(1)$, $F_0(e^{H_n}\Delta_\psi^2)=F(\Delta_\psi^2)=0(1)$ and $F_0(e^{h_\theta/\sqrt{n}}\psi_c)=F_0(\psi_c)+o(1)=o(1)$, we obtain

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

and finally,

$$\begin{aligned} \Delta_n &= \sqrt{n}\log\left[\frac{F_0(e^{H_n+t\Delta_\psi/\sqrt{n}})}{F_0(e^{H_n})}\right] \\ &= tF_0\left(e^{h_\theta/\sqrt{n}}\Delta_\psi\right)-\frac{t^2}{\sqrt{n}}F_0\left(\Delta_\psi\psi_c\right)+\frac{t^2}{2\sqrt{n}}F_0(\Delta_\psi^2)+o\left(\frac{1}{\sqrt{n}}\right) \\ &= \frac{t}{\sqrt{n}}\left[F_0\left(h_\theta\Delta_\psi\right)+\frac{F_0\left(h_\theta^2B_{h_\theta,n}\Delta_\psi\right)}{\sqrt{n}}-tF_0\left(\Delta_\psi\psi_c\right)+\frac{t}{2}F_0(\Delta_\psi^2)\right]+o(n^{-1}) \qquad (40) \end{aligned}$$

where

$$B_{h,n} = \int_0^1 (1-u)e^{uh/\sqrt{n}}du. \tag{41}$$

This implies that

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

We finally obtain, using the fact that $\Delta_n = 0(\epsilon_n) = o(n^{-1/6})$,

$$\begin{aligned} &-\frac{F_0((h_{\theta'})^2)}{2} + \mathbb{G}_n(h_{\theta'}) + R_n(h_{\theta'}) \\ &= -\frac{F_0((h_\theta - t\bar{\psi})^2)}{2} + \mathbb{G}_n(h_\theta - t\bar{\psi}) + R_n(h_\theta - t\bar{\psi}) - tF_0(h_\theta\Delta_\psi) + t\mathbb{G}_n(\Delta_\psi) \\ &\quad -\frac{t^2}{2}[F_0(\Delta_\psi^2) - F_0(\psi_c\Delta_\psi)] + o(1) \end{aligned}$$

Note that $F_0(\Delta_\psi^2) - F_0(\psi_c\Delta_\psi) = 0$ thus to the order $o(1)$, setting $\mu_{n,k} = -F_0(h_\theta\Delta_\psi) + \mathbb{G}_n(\Delta_\psi)$

$$-\frac{F_0((h_{\theta'})^2)}{2} + \mathbb{G}_n(h_{\theta'}) + R_n(h_{\theta'}) = -\frac{F_0((h_\theta - t\bar{\psi})^2)}{2} + \mathbb{G}_n(h_\theta - t\bar{\psi}) + R_n(h_\theta - t\bar{\psi}) + t\mu_{n,k}.$$

Note that by orthogonality $F_0(h_\theta\Delta_\psi) = \sqrt{n}F_0[(\tilde{\psi} - M_{f_0,k}\tilde{\psi})\sum_{j\geq k+1}\theta_{0j}\phi_j]$ so that $\mu_{n,k}$ does not depend on θ and setting $T_k\theta = \theta - B_{n,k}$ for all $\theta \in \Theta_k \cap B'_{\epsilon_n}$, we can write

$$I_k = e^{t^2\frac{F_0(\bar{\psi}^2)}{2}} e^{t\mu_{n,k}} \frac{\int_{\Theta_k \cap B'_{\epsilon_n}} e^{-\frac{F_0(h_{T_k\theta}^2)}{2} + \mathbb{G}_n(h_{T_k\theta}) + R_n(h_{T_k\theta})} d\pi(f_\theta(h))}{\int_{\Theta_k \cap B'_{\epsilon_n}} e^{-\frac{F_0(h_\theta^2)}{2} + \mathbb{G}_n(h_\theta) + R_n(h_\theta)} d\pi(f_\theta(h))}.$$

Note that

$$T_k(B'_{\epsilon_n}) = \{\theta \in \Theta_k; f_{\theta + B_{n,k}} \in B_{\epsilon_n}\}, k \leq k_n$$

and $\|B_{n,k}\|_2 \leq t/\sqrt{n}$, so that for all $\theta \in T_k(B'_{\epsilon_n})$ $\|\theta - \theta_0\|_2^2 \leq \epsilon_n^2 + t^2/n \leq 2\epsilon_n^2$ since $n\epsilon_n^2 \to +\infty$ and for all $\theta \in B_{\epsilon_n(1-t^2/n\epsilon_n^2)} \cap \Theta_k$, $\theta + B_{n,k} \in B_{\epsilon_n} \cap \Theta_k$ so that

$$B_{\epsilon_n(1-t^2/(n\epsilon_n^2))} \cap \Theta_k \subset T_k B_{\epsilon_n} \subset B_{\epsilon_n(1+t^2/(n\epsilon_n^2))}. \tag{42}$$

Write $\epsilon_{n,1} = \epsilon_n(1-t^2/(n\epsilon_n^2))$ and $\epsilon_{n,2} = \epsilon_n(1+t^2/(n\epsilon_n^2))$ Therefore, under the assumption that

$$\frac{\pi_k(\theta)}{\pi_k(\theta - B_{n,k})} = 1 + o(1),$$

uniformly over B'_{ϵ_n},

$$
\begin{aligned}
I_k &\leq e^{t^2 \frac{F_0(\tilde{\psi}^2)}{2}} e^{t\mu_{n,k}} \frac{\int_{\Theta_k \cap B_{\epsilon_{n,2}}} e^{-\frac{F_0(h_\theta^2)}{2} + \mathbb{G}_n(h_\theta) + R_n(h_\theta)} d\pi(f_\theta(h))}{\int_{\Theta_k \cap B_{\epsilon_n}} e^{-\frac{F_0(h_\theta^2)}{2} + \mathbb{G}_n(h_\theta) + R_n(h_\theta)} d\pi(f_\theta(h))} (1 + o(1)) \\
&\geq e^{t^2 \frac{F_0(\tilde{\psi}^2)}{2}} e^{t\mu_{n,k}} \frac{\int_{\Theta_k \cap B_{\epsilon_{n,1}}} e^{-\frac{F_0(h_\theta^2)}{2} + \mathbb{G}_n(h_\theta) + R_n(h_\theta)} d\pi(f_\theta(h))}{\int_{\Theta_k \cap B_{\epsilon_n}} e^{-\frac{F_0(h_\theta^2)}{2} + \mathbb{G}_n(h_\theta) + R_n(h_\theta)} d\pi(f_\theta(h))} (1 + o(1))
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\zeta_n(x,t) &= \mathbb{E}\left[\exp(t\sqrt{n}(\psi(f) - \psi(\mathbb{P}_n))) \mathbb{1}_{B'_{\epsilon_n}}(f) | X^n\right] \\
&= \left[\sum_{k=1}^{k_n} p(k|X^n) I_k\right] (1 + o(1)) \\
&\leq e^{t^2 \frac{F_0(\tilde{\psi}^2)}{2}} \left[\sum_{k=1}^{k_n} p(k|X^n) \mathbb{1}_{\Theta_k \cap B_{\epsilon_n} \neq \emptyset} e^{t\mu_{n,k}}\right] (1 + o(1))
\end{aligned}
$$

and

$$
\zeta_n(x,t) \geq e^{t^2 \frac{F_0(\tilde{\psi}^2)}{2}} \sum_{k=1}^{k_n} p(k|X^n) e^{t\mu_{n,k}} \pi\left[B_{\epsilon_{n,1}} | X^n, k\right].
$$

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

$$
\pi\left[B^c_{\epsilon_{n,1}} | X^n\right] \leq e^{-nc\epsilon_n^2},
$$

for some positive constant $c > 0$. Since for all k such that $\Theta_k \cap B_{\epsilon_n/2} \neq \emptyset$

$$
|\mu_{n,k}| = 0_P(\sqrt{n}\epsilon_n)
$$

then uniformly over k such that $\Theta_k \cap B_{\epsilon_n/2} \neq \emptyset$

$$
\pi\left[B^c_{\epsilon_{n,1}} | X^n, k\right] e^{\mu_{n,k}} = o(1)
$$

and

$$
\zeta_n(x,t) \geq e^{t^2 \frac{F_0(\tilde{\psi}^2)}{2}} \sum_{k=1}^{k_n} p(k|X^n) \mathbb{1}_{\Theta_k \cap B_{\epsilon_n} \neq \emptyset} e^{t\sqrt{n}\mu_{n,k}} + o(1).
$$

This proves that the posterior distribution of $\sqrt{n}(\psi(f) - \psi(\mathbb{P}_n))$ is asymptotically equal to a mixture of Gaussian distributions with variance $V_0 = E_0^n[\mathbb{G}_n(\psi)]$, means $-\mu_{n,k}$ and weights $p(k|X^n)$.

Now if $\|\Delta_\psi\| = o(1)$ $(k \to +\infty)$ $\mathbb{G}_n(\Delta_\psi) = o_P(1)$ and with probability converging to 1, $\forall \epsilon > 0$

$$|\mu_{n,k}| \leq \|f_0\|_\infty \sqrt{n} \left(\sum_{k+1}^{\infty} \psi_j^2 \right)^{1/2} \left(\sum_{k+1}^{\infty} \theta_{0j}^2 \right)^{1/2} + \epsilon$$

Thus if $k = k_{n,\beta}$ and

$$\sum_{j=k_{n,\beta}}^{\infty} \psi_j^2 = o(n^{\frac{2\gamma-2\beta-1}{2\beta+1}} (\log n)^{-2\gamma}),$$

$\mu_{n,k} = o_P(1)$ and Theorem (15) is proved.

Références

[1] Berger, J.

[2] Castillo, I. (2008) A semi-Parametric Bernstein - Von Mises theorem. *Preprint*

[3] Chambaz, A. and Rousseau, J. (2008) Bounds for Bayesian order identifications. *Ann. Statist.*

[4] Cohen, A., Daubechies, I. and Vial, P. (1993) Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* **1**, 54–81.

[5] Diaconis, P. and Freedman, D. (1986) On the consistency of Bayes estimates. *Ann. Statist.* **14** 1-26.

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

[7] Ghosal S. (2000) Asymptotic normality of the posterior distributions for exponential families with many parameters. *Journ. Mult. Anal.*, **74** 49-69.

[8] Ghosal, S., Ghosh, J.K. and van der Vaart, A. (2000) Convergence rates of posterior distributions. *Ann. Statist* **28** 500–531.

[9] Ghosh, J.K. and Ramamoorthi, R.V. (2003) *Bayesian nonparametrics* Springer, New York.

[10] Huang, T.Z. (2004) Convergence rates for posterior distributions and adaptive estimation. *Ann. Statist.* **32**, 1556-1593.

[11] Kim, Y. and Lee, J. (2004) A Bernstein - von Mises theorem in the nonparametric right-censoring model. *Ann. Statist.* **32**, 1492-1512.

[12] Kim, Y. (2006) The Bernstein-von Mises theorem for the proportional hazard model. *Ann. Statist.* **34**, 1678-1700.

[13] LeCam L. (1953) On some asymptotic properties properties of maximum likelihood estimates and related Bayes estimates/ *Univ. California Publ. Statist.* **1** 277-330.

[14] Mallat, S. (1998) *A wavelet tour of signal processing*, Academic Press, San Diego.

[15] Meyer, Y. (1992) *Wavelets and operators*, Cambridge University Press, Cambridge.

[16] Robert, C.P. *Bayesian Choice*

[17] Shen, X. (2002) Asymptotic normality of semiparametric and nonparametric posterior distributions. *J. Amer. Statist. Assoc.* **97**, 222-235.

[18] Scricciolo, C. (2006) Convergence rates for Bayesian density estimation of infinite-dimensional exponential families. *Ann. Statist.* **34**, 2897-2920.

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

[20] Van der Vaat, A. W. (2002). Semiparametric statistics. *Ecole d'Eté de St Flour 1999*, 331-457, (eds : P. Bernard). Springer Verlag, New York.

[21] Verdinelli, I. and Wasserman, L. (1998) Bayesian goodness-of-fit testing using infinite- dimensional exponential families. *Ann. Statist.* **26**, 1215-1241.

[22] Wong, W.H. and Shen, X. (1995) Probability inequalities for likelihood ratios and convergence rates of sieve mles *Ann. Statist.* **23**, 339-362.