Submitted to the Annals of Statistics

BERNSTEIN-VON MISES THEOREM FOR LINEAR FUNCTIONALS OF THE DENSITY

BY VINCENT RIVOIRARD*,

Université Paris Dauphine
E-MAIL: Vincent.Rivoirard@ceremade.dauphine.fr

BY JUDITH ROUSSEAU*,
Université Paris Dauphine and CREST-ENSAE
E-MAIL: Judith.Rousseau@ceremade.dauphine.fr

In this paper, we study the asymptotic posterior distribution of linear functionals of the density by deriving general conditions to obtain a semi-parametric version of the Bernstein-von Mises theorem. The special case of the cumulative distributive function evaluated at a specific point is widely considered. In particular, we show that for infinite dimensional exponential families, under quite general assumptions, the asymptotic posterior distribution of the functional can be either Gaussian or a mixture of Gaussian distributions with different centering points. This illustrates the positive but also the negative phenomena that can occur for the study of Bernstein-von Mises results.

1. Introduction.

The Bernstein-von Mises property, in Bayesian analysis, concerns the asymptotic form of the posterior distribution of a quantity of interest θ, and more specifically it corresponds to the asymptotic normality of the posterior distribution of θ with mean $\hat{\theta}$ and asymptotic variance σ^2 and where, if θ is the true parameter, $\hat{\theta}$ is asymptotically distributed as a Gaussian random variable with mean θ and variance σ^2. Such results are well known in regular 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 also have good frequentist properties. These results are given in many Bayesian textbooks see for instance [20] or [1].

*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

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 feasible. But of course, the more complex the model the harder it is to derive Bernstein-von Mises theorems.

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 $\mathbb{P}^\pi\left[U_n|X^n\right]$ where $\mathbb{P}^\pi[.|X^n]$ 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 [5] consistency is an important condition, even from a subjectivist view point. 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 from above $\mathbb{P}^\pi\left[U_n|X^n\right]$, as in the studies of consistency and concentration rates of the posterior distribution but also 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 hardly any 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 parameters. In a discrete setting, [2] derive Bernstein-von Mises results. 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 an infinite dimensional nuisance parameter, [3] obtains interesting conditions leading to a Bernstein-von Mises theorem on the parameter of interest, clarifying an earlier work of [21]. More precisely, when the parameter of interest is handled in the case of no loss of information, then some classical parametric tools can be used (such as the continuity around the true parameter). The framework considered in the sequel does not make possible such a separation. Other differences with our paper have to be pointed out: The centering considered by [21] is based on the sieve maximum likelihood estimate, whereas priors considered by [3] are merely Gaussian in the information loss case. In Section 2.1 we describe more precisely results by [3, 21] and compare them with ours.

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 and 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 [24] 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 functionals 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, $X_1, ..., X_n$ are independent and identically distributed according to $\mathbb{P}$. We set $X^n = (X_1, ..., X_n)$ 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

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

We denote $\ell_n(f)$ the log-likelihood associated with the density f. For any integrable function g, we set $F(g) = \int_0^1 f(u)g(u)du$. We denote by $< .,. >_f$ the inner product and by $\|.\|_f$ 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 $< .,. >_2$, and $\|.\|_2$, the associated norm. The Kullback-Leibler divergence and the Hellinger distance between two densities f_1 and f_2 will be respectively denoted $K(f_1, f_2)$ and $h(f_1, f_2)$. We recall that

$$K(f_1, f_2) = F_1 \left(\log(f_1/f_2) \right), \quad h(f_1, f_2) = \left[\int \left(\sqrt{f_1(x)} - \sqrt{f_2(x)} \right)^2 dx \right]^{1/2}.$$

In the sequel, we shall also use

$$V(f_1, f_2) = F_1 \left((\log(f_1/f_2))^2 \right).$$

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 for any measurable function g such that $F_0(|g|) < \infty$,

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

and F_n is the empirical distribution function.

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

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

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 the cumulative distribution function at a fixed point x_0:

$$\Psi_{x_0}(f) = F(x_0) = \int_0^1 1\!\!1_{x \le x_0} f(x)dx, \quad x_0 \in [0,1].$$

Let π be a prior on $\mathcal{F}$. The aim of this paper is to study the posterior distribution of $\Psi(f)$ and to derive conditions under which, for all $z \in \mathbb{R}$

$$\mathbb{P}^\pi\left[\sqrt{n}(\Psi(f) - \Psi(P_n)) \le z | X^n\right] \to \Phi_{V_0}(Z) \quad \text{ in } \mathbb{P}_0\text{-probability}, \tag{1.2}$$

where V_0 is the variance of $\sqrt{n}\Psi(P_n)$ under $\mathbb{P}_0$ and Φ_{V_0} 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, credible regions for $\Psi(f)$ (such as highest posterior density regions, equal tail or one-sided intervals) have also the correct asymptotic frequentist coverage. In Section 2.3, 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:

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

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 D.1), wavelet bases constitute unconditional bases of Besov spaces $\mathcal{B}^{\gamma}_{p,q}$ 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: $W^{\gamma} = \mathcal{B}^{\gamma}_{2,2}$. 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 $k \in \{0, \dots, 2^j - 1\}$,

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

Here and in the sequel, $\mathbb{N}$ denotes the set of non negative integers, and $\mathbb{N}^*$ the set of positive integers. 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 $\|.\|_\gamma$ and $\|.\|_{\gamma,p,q}$ the norms associated with W^γ and $\mathcal{B}^{\gamma}_{p,q}$ 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

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

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 [22]. The special case of log-spline priors has been studied by [8] and [13], whereas the prior considered by [26] is based on Legendre polynomials. The wavelet basis is treated in [13] in the special case of the Haar basis.

We now define the class of priors π considered for these models, which we call *the class of sieve priors.*

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 and $r \in \{0,1\}$ such that for any $k \in \mathbb{N}^*$,*

$$\exp(-c_1 kL(k)) \leq p(k) \leq \exp(-c_2 kL(k)), \tag{1.4}$$

where $L(x) = (\log x)^r$.

[Case (D)] *Let $k_n^* = \lfloor k_0 n^{1/(2\beta+1)} \rfloor$, i.e. the largest integer smaller than $k_0 n^{1/(2\beta+1)}$, where k_0 is some fixed positive real number, then k is deterministic and we set $k := k_n^*$ (p is then the Dirac mass at the point k_n^*).*

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

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

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 with 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 geometric priors.

1.3. *Organization of the paper.* We first give very general conditions under which we obtain a Bernstein-von Mises Theorem (see Theorem 2.1 in Section 2.1). Section 2.2 gives a first illustration of Theorem 2.1, based on random histograms. In Section 2.3, we focus on infinite dimensional exponential families. Theorem 2.3 gives the asymptotic posterior distribution of $\Psi(f)$ which can be either Gaussian or a mixture of Gaussian distributions. Corollary 2.2 illustrates positive results with respect to our purpose, but Proposition 2.1 shows that some bad phenomenons may happen. Finally, our message is summarized in Section 2.4. Proofs of the results are given in Section 3. Other technical aspects are given in Appendix.

2. Bernstein-von Mises theorems.

2.1. *The general case.* 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

$$\tilde{\psi}(x) = \psi(x) - F_0(\psi). \tag{2.1}$$

Note that this notation is coherent with the definition of the influence function associated with the tangent set $\{s \in \mathbb{L}_2(F_0) \ \text{ s.t. } \ F_0(s) = 0\}$, defined for instance in Chapter 25 of [24] or used by [23].

For each density function $f \in \mathcal{F}$, 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, as defined in Chapter 25 of [24] 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 $A^1_{u_n} = \left\{f \in \mathcal{F} \ \text{ s.t. } \ V(f_0, f) \leq u_n^2\right\}$ 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) There exists $\tilde{u}_n = o(1)$ such that if A_n is the subset of functions $f \in A^1_{u_n}$ such that

$$\int \left|\log\left(\frac{f(x)}{f_0(x)}\right)\right| f(x)dx \leq \tilde{u}_n \tag{2.2}$$

then

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

(A3) Let

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

and for any x,

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

We have for any t,

$$\frac{\int_{A_n} \exp\left(-\frac{F_0((h_f - t\bar{\psi}_{h_f,n})^2)}{2} + G_n(h_f - t\bar{\psi}_{h_f,n}) + R_n(h_f - t\bar{\psi}_{h_f,n})\right) d\pi(f)}{\int_{A_n} \exp\left(-\frac{F_0(h_f^2)}{2} + G_n(h_f) + R_n(h_f)\right) d\pi(f)}$$
$$= 1 + o_{\mathbb{P}_0}(1). \tag{2.3}$$

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

THEOREM 2.1. *Let f_0 be a density on $\mathcal{F}$ such that $\|\log(f_0)\|_\infty < \infty$. Assume that (A1), (A2) and (A3) are true. Then, we have for any $z \in \mathbb{R}$, as n goes to infinity,*

$$\mathbb{P}^\pi\left\{\sqrt{n}(\Psi(f) - \Psi(P_n)) \le z | X^n\right\} - \Phi_{F_0(\tilde{\psi}^2)}(z) \to 0, \tag{2.4}$$

in $\mathbb{P}_0$-probability.

The proof of Theorem 2.1 is given in Section 3.1. 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 $\exp(t^2 F_0(\tilde{\psi}^2)/2)$ times the left hand side of (2.3) under (A1) and (A2), so that (A3) implies (2.4).

Now, we discuss assumptions. Condition (A1) concerns concentration rates of the posterior distribution and there exists now a large literature on such results, see for instance [23] or [8] for general results. The difficulty here comes from the use of V instead of the Hellinger or the $\mathbb{L}_1$ distances. However note that u_n does not need to be optimal. In our examples, obtaining a posterior concentration rate in terms of V leads us to modify the prior in the case of random histograms and thus to suboptimal posterior concentration rates but has no impact in the case of exponential families. It is also interesting to note that the loss function V is similar to the $\|.\|_L$-norm considered in [3] (i.e. the norm induced by the LAN expansion associated to linear paths on $\log f$) and to the Fisher norm considered in [21]. Indeed, the proof of Theorem 2.1 gives:

$$\ell_n(f) - \ell_n(f_0) = -\frac{nV(f_0, f)}{2} + G_n(h_f) + R_n(h_f) \tag{2.5}$$

with $R_n(h_f) = o_{\mathbb{P}_0}(1)$ pointwise (i.e. for a fixed function h_f). Condition (A1) is thus to be related to Condition **C** in [3] and to Condition (9) in [21]. However, the formulation of Condition (9) in [21] is not quite as general as Condition **C** in [3] or as our conditions since [21] also requires (stated in our framework):

$$\sup_{f:\, V(f_0,f) > \epsilon_n^2} \{\ell_n(f) - \ell_n(f_0)\} \le -cn\epsilon_n^2.$$

Indeed, a concern of [21] is to obtain a Bernstein-von Mises theorem with a centering point which is the maximum likelihood (or a sieve maximum likelihood estimator), for which such a condition is quite natural. It is known now, see for instance [9], that weaker conditions can be obtained to derive the posterior concentration rate.

Condition (A2) could be viewed as a symmetrization of (A1) since if on $A^1_{u_n}$, we also have $V(f, f_0) \leq u_n^2$ then (A2) is true. Actually, (A2) is a weaker condition since it is only based on the first moment of $\log(f/f_0)$ with respect to the density f.

The main difficulty comes from condition (A3). Roughly speaking, (A3) means that a change of parameter induced by a transformation T of the form $T(f_h) = f_{h-t\bar{\psi}_{h,n}}$, or close enough to it, can be considered 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. A similar condition can be found in [21] (see Condition (14)). We emphasize two major differences between Shen's condition ([21]) and ours: first Shen's condition is based on the sieve MLE of $\log f$, which we do not consider since we re-center on the empirical $\Psi(P_n)$. Secondly and more importantly, Condition (14) in [21] is expressed in terms of the conditional prior distribution of f given $\theta = \Psi(f)$ which is very difficult to control in most non-parametric models, whereas in our case the expectation is taken with respect to the prior on f.

However, (A3) still remains a demanding condition (the most demanding one) to verify in general models, and it is often the condition which is not verified when the Bernstein-von Mises theorem is not satisfied, as illustrated in our examples below. Interestingly, this condition can also be found in [3], but in a less explicit way. Indeed, in [3], the parameter is split into (θ, g) say, where g is a function (so it is infinite dimensional) and θ is the parameter of interest and is finite dimensional. Two cases are then considered, namely the case without loss of information and the case with loss. In the former, the computations simplify greatly and the change of parameter is only made on the parametric part θ, which usually is easy to verify. In the latter, the non-parametric part is more influential and this case is handled merely in the setup of Gaussian priors for which an interesting discussion on how this change of parameter is influenced by the respective smoothness of the prior (see page 14 of [3]) and of the true parameter is lead. In our context, the smoothness of the functional Ψ, of the true density f_0 and of the prior are certainly influential, as will be illustrated in the examples below. However, for non-Gaussian priors, the notion of smoothness of the prior is not so clearly defined. In particular priors leading to adaptive posterior concentration rates cannot be said to have a smoothness of their own. We rather view this condition as a *no bias condition*, which also applies to the Gaussian case. Indeed, choosing a less regular Gaussian prior allows for correct approximation of rougher curves and thus avoids biases in the estimation of rough functionals. To make this statement more precise we consider now the framework of sieve models.

Consider $(\mathcal{F}_k)_k$, a sequence of subsets, such that $\cup_k \mathcal{F}_k \subset \mathcal{F}$ and $\mathcal{F}_k = \{f_\theta \text{ s.t. } \theta \in \Theta_k\}$ with $\Theta_k \subset \mathbb{R}^{r_k}$ and $(r_k)_k$ is an increasing sequence going to infinity. A prior on $\mathcal{F}$ is then defined as a probability on k, say $p(.)$ and given k a probability on θ, say π_k. This setup is quite general and it includes in particular the two types of examples considered in the paper, namely the random histograms (see Section 2.2) and the exponential families (see Section 2.3). For notational ease, we write h_θ instead of h_{f_θ} and $\bar{\psi}_{h,n}$ instead of $\bar{\psi}_{h_{f_\theta},n}$. Assumption (A3) then corresponds to a change of *parameter* from h_θ to $h_\theta - t\bar{\psi}_{h,n}$. So, a first difficulty comes from expressing this change in terms of θ. This change of parameter has just to be done locally and more precisely has to be defined on a set $\tilde{A}_n$ closely related to the set A_n introduced in (A2). In other words, we construct $\tilde{A}_n$ and for each k, a map $T_k : \tilde{A}_n \cap \Theta_k \to \Theta_k$ and we define $\psi_{k,\theta}$ such that

$$h_{T_k\theta} = h_\theta - t\psi_{k,\theta}$$

(equivalently $f_{T_k\theta} = f_\theta e^{-t\psi_{k,\theta}/\sqrt{n}}$). The aim of this construction is to build T_k such that $\psi_{k,\theta} \approx \bar{\psi}_{h,n}$. Mathematically, this approximation is expressed via log-likelihoods and we set

$$\rho_n(\theta) := \ell_n(f_{T_k\theta}) - \ell_n(f_\theta e^{-t\bar{\psi}_{h,n}/\sqrt{n}}).$$

By using (2.5) or (3.6), note that

$$\begin{aligned}\rho_n(\theta) \quad &= \quad -\frac{F_0(h^2_{T_k\theta})}{2} + G_n(h_{T_k\theta}) + R_n(h_{T_k\theta}) \\ &\quad - \left(-\frac{F_0((h_\theta - t\bar{\psi}_{h,n})^2)}{2} + G_n(h_\theta - t\bar{\psi}_{h,n}) + R_n(h_\theta - t\bar{\psi}_{h,n})\right).\end{aligned}$$

Then, Relation (2.3) of assumption (A3) can be reduced to the following one: for all k such that $\Theta_k \cap \tilde{A}_n \neq \emptyset$

$$\frac{\int_{\tilde{A}_n \cap \Theta_k} \exp\left(-\frac{F_0(h^2_{T_k\theta})}{2} + G_n(h_{T_k\theta}) + R_n(h_{T_k\theta})\right) e^{-\rho_n(\theta)} \pi_k(\theta) d\theta}{\int_{\tilde{A}_n \cap \Theta_k} \exp\left(-\frac{F_0(h^2_\theta)}{2} + G_n(h_\theta) + R_n(h_\theta)\right) \pi_k(\theta) d\theta}$$

$$= 1 + o_{\mathbb{P}_0}(1). \tag{2.6}$$

Proposition A.1 in Appendix gives the rigorous construction of $\tilde{A}_n$ and of the map T_k. Proposition A.1 also states that under mild conditions

$$\begin{aligned}\rho_n(\theta) \quad &= \quad -tF_0[\Delta_{k,\theta} h_\theta] + tG_n(\Delta_{k,\theta}) \\ &\quad -\frac{t^2}{2} F_0((\bar{\psi}_{h,n} - \psi_{k,\theta})^2) + t^2 F_0\left[(\bar{\psi}_{h,n} - \psi_{k,\theta})\bar{\psi}_{h,n}\right] + o_{\mathbb{P}_0}(1),\end{aligned}$$

uniformly in $\theta \in \tilde{A}_n \cap \Theta_k$, where $\Delta_{k,\theta}$ is the difference $\bar{\psi}_{h,n} - \psi_{k,\theta}$ up to an additive constant, i.e. there exists a constant $b_{k,\theta} \in \mathbb{R}$ such that

$$\Delta_{k,\theta}(x) = \bar{\psi}_{h,n}(x) - \psi_{k,\theta}(x) + b_{k,\theta}, \quad x \in [0;1].$$

Note that the function $\psi_{k,\theta}$ is related to the approximation γ_n of the least favorable direction considered in [3].

As will be illustrated in subsequent examples, under many priors, we can obtain $\pi_k(T_k\theta) = \pi_k(\theta)(1 + o(1))$ uniformly over $\tilde{A}_n \cap \Theta_k$ and integration over $\tilde{A}_n \cap \Theta_k$ is hardly modified by T_k. Therefore, the key condition to verify (A3) is $\rho_n(\theta) = o_{\mathbb{P}_0}(1)$ uniformly in $\theta \in \tilde{A}_n$, which is implied by

$$F_0(h_\theta \Delta_{k,\theta}) = o(1) \quad \& \quad F_0((\bar{\psi}_{h,n} - \psi_{k,\theta})^2) = o(1) \tag{2.7}$$

uniformly over $\tilde{A}_n$ (see Proposition A.1). Condition (2.7) expresses that the difference $\bar{\psi}_{h,n} - \psi_{k,\theta}$ has to be small enough, illustrating in this context what we mean by *a no bias condition.*

Before studying in details infinite dimensional exponential families, we illustrate this general result in the setup of priors based on random histograms with random partitions. Such priors have often been considered in the Bayesian non-parametric literature, see for instance [11], since they are both simple to implement and flexible.

2.2. *Bernstein-von Mises for random histograms with random partitions.* Let $f_{\eta,z}$ be the density on $[0,1]$ defined for any x, by

$$f_{\eta,z}(x) = \sum_{j=1}^{k} \frac{\eta_j}{\Delta z_j} 1\!\!1_{I_j}(x), \quad \eta_j \geq 0, \quad \sum_{j=1}^{k} \eta_j = 1, \quad k \in \mathbb{N}^*,$$

where $\Delta z_j = z_j - z_{j-1}$, $I_j = (z_{j-1}, z_j]$, $0 = z_0 < z_1 < ... < z_k = 1$. To define $\pi_{z|k}$ the conditional density of z given k, we consider the parametrization $(\Delta z_1, \ldots, \Delta z_k)$, which lies in the k-dimensional simplex:

$$\mathcal{S}_k = \left\{ (x_1, ..., x_k) \in [0,1]^k \ \text{ s.t. } \ \sum_{i=1}^{k} x_i = 1 \right\}.$$

Given $c_1 > 0$ and $\alpha > 0$, we consider the prior on $\mathcal{S}_k$ whose density with respect to the Lebesgue measure is

$$\pi_{z|k}(\Delta z_1, ..., \Delta z_k) = \frac{e^{-c_1 \sum_{i=1}^{k} \Delta z_i^{-\alpha}}}{b_k(c_1)} 1\!\!1_{\mathcal{S}_k}(\Delta z_1, ..., \Delta z_k)$$

with for $c > 0$,

$$b_k(c) = \int_{\mathcal{S}_k} e^{-c\sum_{i=1}^k x_i^{-\alpha}} dx_1 \dots dx_k.$$

Conditionally on k and z, η has a density with respect to the Lebesgue measure on $\mathcal{S}_k$ satisfying

$$\pi_{\eta|k}(\eta_1, ..., \eta_k) = \frac{e^{-c_2\sum_{i=1}^k \eta_i^{-\alpha}}}{b_k(c_2)} \mathbb{1}_{\mathcal{S}_k}(\eta_1, ..., \eta_k)$$

where $c_2 > 0$. We finally consider a prior on k which satisfies the following property: there exist $C_0, C_0', c_0, c_0' > 0$ such that when k is large,

$$C_0 e^{-c_0 k^{\alpha+1}} \le p(k) \le C_0' e^{-c_0' k^{\alpha+1}}.$$

We have the following result.

THEOREM 2.2. *We consider $\psi(x) = x$. Under the above prior, if f_0 is γ-Hölderian with $(\alpha+1)/2 < \gamma \le 1$, strictly positive on $[0,1]$ and almost surely differentiable on $(0,1)$ with derivative $f_0'(x) > c_0$ for almost all $x \in (0,1)$ then assumptions (A1), (A2) and (A3) are satisfied and the conclusion of Theorem 2.1 holds. Moreover, under the above prior, the rate of concentration of the posterior distribution around f_0 is of order $O((n/\log n)^{-2\gamma/(2\gamma+\alpha+1)} \log n)$.*

The proof is given in Appendix C. The prior is assumed to follow quite stringent tail conditions, inducing a loss in the concentration rate. This is due to the need of controlling the posterior concentration rate in terms of the divergence V. We do not claim that this concentration rate is sharp, and it is quite possible that it can be improved. However this would not shed more lights on the Bernstein-von Mises property per se.

As explained before, a key condition for the Bernstein-von Mises theorem to be satisfied is that for any z and any η such that

$$V(f_0, f_{\eta,z}) = O((n/\log n)^{-2\gamma/(2\gamma+\alpha+1)} \log n)$$

we have:

$$\sqrt{n}\left(\sum_{j=1}^k \int_{I_j} f_0(x)\left(\log(f_0(x)) - \log(\eta_j/\Delta z_j)\right)(\tilde{\psi}(x) - \tilde{\psi}_j)dx\right) = o(1), \tag{2.8}$$

with $\tilde{\psi}_j = (\int_{I_j} \tilde{\psi}(x) f_0(x)dx)/\Delta z_j$, which is a transcription of the first part of relation (2.7). This is satisfied under the condition $f_0'(x) > c_0$. The proof of

Theorem 2.2 shows that the result remains valid if ψ is any Lipschitz function. Equation (2.8) shows that we have to approximate conveniently both f_0 and $\tilde{\psi}$ by piecewise constant functions. Since the asymptotic posterior distribution is driven by the smoothness of f_0, it is sometimes necessary to force the prior to remove the bias due to the shift by $\tilde{\psi}$ in the approximation of $fe^{-t\bar{\psi}_{h,n}/\sqrt{n}}$ in the prior model to validate (2.7). This is illustrated in the following section.

2.3. *Bernstein-von Mises for 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 $\log(f_0) \in \mathbb{L}_2[0,1]$. 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 $(\phi_\lambda)_{0\le\lambda\le k}$ for the scalar product $< .,. >_{f_0}$ and

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

where $\tilde{\psi}$ is defined in (2.1). 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 $(\tilde{\psi}_{\Pi,\lambda})_{\lambda\le k}$ denote the sequences of coefficients of the expansions of the functions $\tilde{\psi}$ and $\Pi_{f_0,k}\tilde{\psi}$ respectively. We finally note:

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

Let $(\epsilon_n)_n$ be the sequence decreasing to zero defined in Theorem B.1 (see Appendix B). The sequence $L(n)$ is based on the function L defined in the case (PH) of Definition 1.1 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 $\mathbb{P}_p\left(k > \frac{l_0 n\epsilon_n^2}{L(n)}\right) \le e^{-an\epsilon_n^2}$. Following for instance [9] p. 221, it implies that there exists $c > 0$ and l_0 large enough such that

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

Define $l_n = l_0 n\epsilon_n^2/L(n)$ in the case (PH). In the case (D) we set $l_n = k_n^*$, where k_n^* is defined in Definition 1.1 . We have the following result.

THEOREM 2.3. *We consider the prior defined in Definition 1.1. We assume that* $\|\log(f_0)\|_\infty < \infty$ *and* $\log(f_0) \in \mathcal{B}^\gamma_{p,q}$ *with* $p \geq 2$*,* $1 \leq q \leq \infty$ *and* $\gamma > 1/2$ *is such that*

$$\beta < 1/2 + \gamma \quad \textit{if} \quad p_* \leq 2 \qquad \textit{and} \qquad \beta < \gamma + 1/p_* \quad \textit{if} \quad p_* > 2,$$

where p_* *is defined in Definition 1.1. For any* $k \in \mathbb{N}^*$*, set*

$$B_k = \left\{ \theta \in \mathbb{R}^k \;\; s.t. \;\; \sum_{\lambda=1}^{k} (\theta_\lambda - \theta_{0\lambda})^2 \leq \frac{4\,(\log n)^3}{L^2(n)} \epsilon_n^2 \right\},$$

and assume that for any $t \in \mathbb{R}$*,*

$$\lim_{n\to+\infty} \max_{k \leq l_n} \sup_{\theta \in B_k} \frac{\pi_k(\theta)}{\pi_k\left(\theta - \frac{t\tilde{\psi}_\Pi^{[k]}}{\sqrt{n}}\right)} = 1 \tag{2.9}$$

and

$$\sup_{k \leq l_n} \left\{ \| \sum_{\lambda > k} \tilde{\psi}_\lambda \phi_\lambda \|_\infty + \sqrt{k} \| \sum_{\lambda > k} \tilde{\psi}_\lambda \phi_\lambda \|_2 \right\} = o\left(\frac{(\log n)^{-3}}{\sqrt{n}\epsilon_n^2} \right) \tag{2.10}$$

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

$$\mathbb{P}^\pi \left[\sqrt{n}(\Psi(f) - \Psi(P_n)) \leq z | X^n \right] = \sum_k p(k|X^n) \Phi_{V_{0k}}\left(z + \mu_{n,k}\right) + o_{\mathbb{P}_0}(1), \tag{2.11}$$

where

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

In the case (D), if

$$\sum_{\lambda > k_n^*} \tilde{\psi}_\lambda^2 = o\left(n^{\frac{2\gamma}{2\beta+1} - 1} \right) \tag{2.12}$$

then, for any $z \in \mathbb{R}$*,*

$$\mathbb{P}^\pi \left[\sqrt{n}(\Psi(f) - \Psi(P_n)) \leq z | X^n \right] \;\; = \;\; \Phi_{V_0}\left(z\right) + o_{\mathbb{P}_0}(1), \tag{2.13}$$

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

The proof of Theorem 2.3 is given in Section 3.2. This result is a consequence of Theorem 2.1. Conditions (A1) and (A2) are verified using Theorem B.1. Conditions (2.9) and (2.10) are needed to study the asymptotic behavior of the ratio defined in equation (2.3) which must go to 1 for condition (A3) to be satisfied. As explained in Section 2.1, to control the ratio defined in (2.3) we need to express the change of parameter h to $h - t\bar{\psi}_{h,n}$ in terms of a change of parameter from $\theta \in B_k$ to $\theta - t\tilde{\psi}_{\Pi}^{[k]}/\sqrt{n}$. Condition (2.9) ensures that the prior is not dramatically modified by this change of parameter. The following three examples of priors illustrate this condition. For the sake of simplicity, we only consider the case $p = q = 2$.

COROLLARY 2.1. *Assume that* $\log(f_0) \in W^\gamma$. *We still assume that* $\beta > 1/2$ *and* $\gamma > 1/2$. *Condition (2.9) 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 $\gamma > \beta - 1/2$ for the case (PH) (no further condition 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 $\psi(x) = 1\!\!1_{x \le x_0}$, conditions on γ and β can be relaxed. In particular, in the case (PH), if g is the Laplace density, (2.9) is satisfied as soon as $\gamma > \beta - 1/2$. By choosing $1/2 < \beta \le 1$, this is satisfied for any $\gamma > 1/2$ as imposed by Theorem 2.3. Note that in the case (PH), Theorem B.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.

Condition (2.10) is needed to obtain $\|\Delta_k\|_\infty = o(\sqrt{n}u_n^2)^{-1})$ for all $k \le l_n$ as required by Proposition A.1, which is the first step used to establish Theorem 2.3. Indeed, (3.13) gives $\sqrt{n}u_n^2 = \sqrt{n}\epsilon_n^2(\log n)^3$ which goes to 0 with n, so that Condition (2.10) is quite mild. It requires some minimal smoothness on ψ through the decay to zero of its coefficients. Note that we require $\epsilon_n = o(n^{-1/4})$, which is a consequence of the conditions imposed on β, γ and p_*, but which is necessary in various parts of the proof. The threshold $n^{-1/4}$ is often encountered in semi-parametric analysis as the *no bias* condition (see for instance [25], Section 25.8) and is also required in [3] in the Cox model example (i.e. with information loss).

Conditions (2.9) and (2.10) are rather mild, so that quite generally, the posterior distribution of $\sqrt{n}(\Psi(f) - \Psi(P_n))$ is asymptotically a mixture of Gaussian distributions with variances $V_{0k} - F_0(\Delta_k^2)$ 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 $\mu_{n,k}$ and $F_0(\Delta_k^2)$ to be small whenever $p(k|X^n)$ is not. The situation where $F_0(\Delta_k^2) \neq o(1)$ under the posterior distribution corresponds to the case where there exists k_0 such that $f_0 \in \mathcal{F}_{k_0}$. 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 $\Psi(f_{\hat{\theta}_{k_0}})$. The posterior distribution therefore satisfies a Bernstein-von Mises theorem, but it is a parametric result and not a non-parametric Bernstein-von Mises Theorem. However, even if $F_0(\Delta_k^2) = o(1)$, in the setup of exponential families, it may happen that $\sqrt{n}F_0\left(\Delta_k \sum_{\lambda \geq k+1} \theta_{0\lambda}\phi_\lambda\right) \neq o(1)$ so $\mu_{n,k} \neq o_{\mathbb{P}_0}(1)$ and the posterior distribution would not satisfy the non-parametric Bernstein-von Mises property. The term $\mu_{n,k}$ is a bias term in the posterior distribution. It is related to the term $\gamma_n - \gamma$ in [3] in the case of information loss, since $\Pi_{f_0,k}\tilde{\psi}$ plays the same role as γ_n. In the case of Gaussian priors the control of $\gamma_n - \gamma$ is induced by a smoothness assumption on the prior. Here the notion of smoothness is not so clearly defined and the control of $\mu_{n,k}$ strongly depends on a lower bound on the set of k's such that $\sum_{\lambda \geq k+1} \theta_{0\lambda}^2 \leq \epsilon_n^2$, which can be interpreted as a no bias condition. Indeed $|\mu_{n,k}| \leq C\sqrt{n}\left(\sum_{\lambda > k} \tilde{\psi}_\lambda^2\right)^{1/2} \left(\sum_{\lambda > k} \theta_{0\lambda}^2\right)^{1/2}$. Therefore for the Bernstein-von Mises property to be satisfied over a class of functions f_0, the posterior on k needs to be almost 0 for k's such that $\left(\sum_{\lambda > k} \tilde{\psi}_\lambda^2\right)^{1/2}$ is larger than $[\sqrt{n}\left(\sum_{\lambda > k} \theta_{0\lambda}^2\right)^{1/2}]^{-1}$. In general we cannot assess a lower bound on k for which $\sum_{\lambda > k} \theta_{0\lambda}^2 \leq \epsilon_n^2$ unless we assume some extra conditions on the behavior of the $\theta_{0\lambda}$'s. Thus in the case (PH), the Bernstein-von Mises theorem will often not be satisfied, even for regular functional $\tilde{\psi}$ unless strong assumptions are put on the behavior of the coefficients $(\theta_{0\lambda})_\lambda$. This remark is illustrated in Proposition 2.1, where we prove the non validity of the Bernstein-von Mises theorem for a given family of functions f_0 (with various smoothness parameters).

The Bernstein-von Mises theorem is however satisfied in the case of a prior of type (D), under condition (2.12). The latter is verified if either $\gamma > \beta + 1/2$ or if $\gamma > \beta$ and ψ is a smooth function like a continuously differentiable function in the case of the Fourier basis or a piecewise constant function (as in the case of the cumulative distribution function). Therefore to obtain a BVM theorem, the true density f_0 and the functional $\tilde{\psi}$ are required to have a minimal smoothness ($\gamma > 1/2$ for f_0 and condition (2.10) on $\tilde{\psi}$). Conditions (2.12), $k = k_n^*$ and the constraints on β, force the prior to approximate correctly functions that are potentially less regular than f_0.

We illustrate this issue in the special case of the cumulative distribution function calculated at a given point x_0: $\psi(x) = 1\!\!1_{x\leq x_0}$. 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. *Let $x_0 \in [0;1]$. Assume that ψ is a piecewise constant function. Consider the class of sieve priors defined in Definition 1.1 in the case (D) with g is either 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 a Student density and if $\gamma \geq \beta > 1$, the same result holds.*

We now illustrate the fact that for the case (PH), the Bernstein-von Mises property may be not valid.

PROPOSITION 2.1. *Let us consider the Fourier basis and 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, $f_0 \in W^\gamma$ and 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 (3.23).

2.4. *A conclusion.* As a conclusion on the existence of Bernstein von Mises theorem for linear functionals of the density, we see that apart from the usual concentration results of the posterior distribution, the key condition is to be able to define a change of parameter from f to $fe^{-t\bar{\psi}_{h,n}/\sqrt{n}}$ which does not significantly modify the prior. Such a construction differs, depending on the family of priors considered. In this paper we have called this a *no bias* condition since it means that not only f_0 needs to be well approximated with such a prior but also $fe^{-t\bar{\psi}_{h,n}/\sqrt{n}}$, for all f in a neighborhood of f_0. Remember that $\bar{\psi}_{h,n}$ is equal to the functional ψ up to a constant, so the influence of ψ is of course non-negligible. This *no bias* condition can be problematic since the posterior (being driven by the likelihood) is targeted to approximate correctly f_0 and in the case of adaptive posterior distributions such as (PH), it is thus adapted to the smoothness of f_0, which might not be the same as the smoothness of $fe^{-t\bar{\psi}_{h,n}/\sqrt{n}}$. In the case of Gaussian priors, as considered in [3], this implies that the prior is not too smooth so that $fe^{-t\bar{\psi}_{h,n}/\sqrt{n}}$ can be correctly approximated by sequences in the associated RKHS. In the family of sieve priors it means that the posterior distribution concentrates on k's that are large enough.

3. Proofs.

This section contains the proofs of all the results except Theorem 2.2 for which we need to establish rates of posterior distributions. So, its proof is deferred in Appendix which contains all the aspects about concentration rates.

In the sequel, C denotes a generic positive constant whose value is of no importance and may change from line to line. To simplify some expressions, we omit at some places the integer part $\lfloor \cdot \rfloor$.

3.1. *Proof of Theorem 2.1.* Let $Z_n = \sqrt{n}(\Psi(f) - \Psi(P_n))$. We have

$$\mathbb{P}^\pi \{A_n | X^n\} = 1 + o_{\mathbb{P}_0}(1). \tag{3.1}$$

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 $t \in \mathbb{R}$,

$$\lim_{n \to +\infty} L_n(t) = \exp\left(\frac{t^2}{2} F_0\left[\tilde{\psi}^2\right]\right), \tag{3.2}$$

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

$$\begin{aligned} L_n(t) &= \mathbb{E}^\pi\left[\exp(t\sqrt{n}(\Psi(f) - \Psi(P_n)))|A_n, X^n\right] \\ &= \frac{\mathbb{E}^\pi\left[\exp(t\sqrt{n}(\Psi(f) - \Psi(P_n)))\mathbb{1}_{A_n}(f)|X^n\right]}{\mathbb{P}^\pi\{A_n|X^n\}} \\ &= \frac{\int_{A_n} \exp\left(t\sqrt{n}(\Psi(f) - \Psi(P_n)) + \ell_n(f) - \ell_n(f_0)\right) d\pi(f)}{\int_{A_n} \exp\left(\ell_n(f) - \ell_n(f_0)\right) d\pi(f)}. \end{aligned} \tag{3.3}$$

We first deal with $t\sqrt{n}(\Psi(f) - \Psi(P_n))$. For this purpose, we introduce for any x,

$$B_{h,n}(x) = \int_0^1 (1-u)e^{uh(x)/\sqrt{n}}du. \tag{3.4}$$

Note that, with $h = h_f = \sqrt{n}(\log f - \log f_0)$,

$$\begin{aligned} B_{h,n}(x) &\leq 0.5 \times 1_{\{f(x)\leq f_0(x)\}} + 1_{\{f(x)>f_0(x)\}} \int_0^1 e^{u(\log f(x) - \log f_0(x))} du \\ &\leq 1_{\{f(x)\leq f_0(x)\}} + 1_{\{f(x)>f_0(x)\}} (\log f(x) - \log f_0(x))^{-1} \frac{f(x)}{f_0(x)}. \qquad (3.5) \end{aligned}$$

So, using (3.4), a Taylor expansion gives

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

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 we have

$$\ell_n(f) - \ell_n(f_0) = -\frac{F_0(h^2)}{2} + G_n(h) + R_n(h), \tag{3.6}$$

we can write $L_n(t)$ as

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

where straightforward computations show that

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

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

$$\|\tilde{\psi}\|_\infty \leq 2\|\psi\|_\infty < \infty, \tag{3.7}$$

the Taylor expansion of $\exp\left(-t\tilde{\psi}/\sqrt{n}\right)$ and the formula

$$f(x) = f_0(x) \exp\left(\frac{h(x)}{\sqrt{n}} \right),$$

we obtain:

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

Also,

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

Note that, on A_n, we have $F_0(h^2) = O(nu_n^2)$ and, by using (3.5),

$$\begin{aligned} F_0\left(h^2 B_{h,n} \right) &\leq nF_0\left[\left(\log\left(\frac{f}{f_0} \right) \right)^2 \right] + nF\left[\left| \log\left(\frac{f}{f_0} \right) \right| \right] \\ &\leq nu_n^2 + n\tilde{u}_n \end{aligned}$$

so, $F_0\left(h^2 B_{h,n}\right) = o(n)$. So, uniformly on A_n, since $\tilde{\psi}$ is bounded (see (3.7))

$$\begin{aligned} F_0\left[\exp\left(\frac{h}{\sqrt{n}} - \frac{t\tilde{\psi}}{\sqrt{n}}\right)\right] &= 1 - \frac{t}{\sqrt{n}}\left(\frac{F_0[h\tilde{\psi}]}{\sqrt{n}} + \frac{F_0[h^2 B_{h,n}\tilde{\psi}]}{n}\right) \\ &\quad + \frac{t^2}{2n}\left(F_0\left[\tilde{\psi}^2\right] + \frac{F_0[h\tilde{\psi}^2]}{\sqrt{n}} + \frac{F_0[h^2 B_{h,n}\tilde{\psi}^2]}{n}\right) + o\left(n^{-1}\right) \\ &= 1 - \frac{t}{n}\left[F_0[h\tilde{\psi}] + \frac{F_0[h^2 B_{h,n}\tilde{\psi}]}{\sqrt{n}} - \frac{tF_0(\tilde{\psi}^2)}{2} + o(1)\right] \\ &= 1 + o\left(n^{-1/2}\right) \qquad (3.8) \end{aligned}$$

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,

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

and up to the multiplicative factor $1 + o(1)$, $L_n(t)$ is equal to

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

Finally (A3) implies (3.2) and the theorem is proved.

3.2. *Proof of Theorem 2.3.* We use the same approach as in Theorem 2.1. We first prove that conditions (A1) and (A2) are satisfied. Let ϵ_n be the posterior concentration rate obtained in Theorem B.1. 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),
- $\epsilon_n = \epsilon_0(\log n)^{1_{\{\gamma\geq\beta\}}}n^{-\frac{\beta\wedge\gamma}{2\beta+1}}$ and $l_n = k_n^* = k_0 n^{\frac{1}{2\beta+1}}$ in the case (D).

Note that for any $a \geq 0$, since $\gamma > 1/2$ and $\beta > 1/2$, we have:

$$(\log n)^a l_n\epsilon_n^2 = o(1). \qquad (3.9)$$

Note also that in the sequel we can restrict ourselves to $\cup_{k\leq l_n}\mathcal{F}_k$. Indeed, in the case of the prior (PH)

$$\begin{aligned} \mathbb{P}_\pi\left[(\cup_{k\leq l_n}\mathcal{F}_k)^c\right] &= \sum_{\lambda>l_n} p(\lambda) \\ &\leq C\exp\left(-c_2 l_n L(l_n)\right) = o(e^{-cn\epsilon_n^2}), \qquad (3.10) \end{aligned}$$

for some positive constant c and in the case of the prior (D) $\mathbb{P}_\pi\left[(\cup_{k\le l_n}\mathcal{F}_k)^c\right]=0$ by definition.

In the sequel, for any $k\le l_n$ and any $\theta\in\mathbb{R}^k$, we still denote θ the sequence whose λ-th component is equal to θ_λ for $\lambda\le k$ and whose λ-th component is equal to 0 for $\lambda>k$. Then we can define

$$\tilde{A}_n=\left\{\theta\in\cup_{k\le l_n}\mathbb{R}^k \ \text{ s.t. } \ \|\theta-\theta_0\|_{\ell_2}\le\frac{2(\log n)^{3/2}\epsilon_n}{L(n)^{1/2}}\right\}.$$

In the same spirit as for Proposition A.1, with a slight abuse of notation, we also denote

$$\tilde{A}_n=\{f_\theta \ \text{ s.t. } \ \theta\in\tilde{A}_n\}.$$

Note that from Theorem B.1,

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

To prove (A1) and (A2) we control $V(f_0,f_\theta)$ and $F_\theta[|\log(f_\theta/f_0)|]$ for $f_\theta\in\tilde{A}_n$. For any $\theta\in\tilde{A}_n$, we have:

$$V(f_0,f_\theta) \quad\le\quad 2\|f_0\|_\infty\|\theta-\theta_0\|_{\ell_2}^2+2(c(\theta)-c(\theta_0))^2.$$

We show that the second term of the right hand side is smaller than the first one up to a constant. For this purpose, note that for $\theta\in\tilde{A}_n$, by using (D.1), (D.3) and (3.9),

$$\|\sum_{\lambda=1}^{+\infty}(\theta_{0\lambda}-\theta_\lambda)\phi_\lambda\|_\infty\le C\sqrt{l_n}\|\theta-\theta_0\|_{\ell_2}+Cl_n^{\frac{1}{2}-\gamma}=o(1). \tag{3.11}$$

Therefore for $\theta\in\tilde{A}_n$,

$$\begin{aligned}
c(\theta)-c(\theta_0) &= \log\left(\int_0^1 f_0(x)e^{-\sum_{\lambda=1}^{+\infty}(\theta_{0\lambda}-\theta_\lambda)\phi_\lambda(x)}dx\right)\\
&= \log\left\{1-\sum_{\lambda=1}^{+\infty}(\theta_{0\lambda}-\theta_\lambda)F_0(\phi_\lambda)+\frac{1}{2}F_0\left[\left(\sum_{\lambda=1}^{+\infty}(\theta_{0\lambda}-\theta_\lambda)\phi_\lambda\right)^2\right](1+o(1))\right\}\\
&= -\sum_{\lambda=1}^{+\infty}(\theta_{0\lambda}-\theta_\lambda)F_0(\phi_\lambda)(1+o(1))+O(\|\theta-\theta_0\|_{\ell_2}^2)
\end{aligned}$$

and for n large enough,

$$|c(\theta)-c(\theta_0)|\le 2\|f_0\|_2\|\theta-\theta_0\|_{\ell_2}. \tag{3.12}$$

Using Theorem B.1, this implies that on $\tilde{A}_n$,

$$V(f_0, f_\theta) \quad = \quad O(\epsilon_n^2 (\log n)^3 / L(n)). \tag{3.13}$$

Thus (A1) is verified with $u_n^2 = u_0^2 \epsilon_n^2 (\log n)^3 / L(n)$ and u_0 large enough. To establish (A2), we observe that we have:

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

by using (3.11). So, on $\tilde{A}_n$,

$$V(f_\theta, f_0) \leq C V(f_0, f_\theta)$$

and (A2) is implied by (A1). Conditions 1 and 2 of Proposition A.1 are also true with $w_n = 1$ for any n. Now, let us study the validity of (A3): For any t, we study the term

$$I_n \quad = \quad \frac{\int_{\tilde{A}_n} \exp\left(-\frac{F_0((h_f - t\bar{\psi}_{h,n})^2)}{2} + G_n(h_f - t\bar{\psi}_{h,n}) + R_n(h_f - t\bar{\psi}_{h,n}) \right) d\pi(f)}{\int_{\tilde{A}_n} \exp\left(-\frac{F_0(h_f^2)}{2} + G_n(h_f) + R_n(h_f) \right) d\pi(f)}.$$

We introduce

$$J_{k,n} \quad := \quad \frac{\int_{\tilde{A}_n \cap \mathcal{F}_k} \exp\left(-\frac{F_0((h_f - t\bar{\psi}_{h,n})^2)}{2} + G_n(h_f - t\bar{\psi}_{h,n}) + R_n(h_f - t\bar{\psi}_{h,n}) \right) d\pi_k(f)}{\int_{\tilde{A}_n \cap \mathcal{F}_k} \exp\left(-\frac{F_0(h_f^2)}{2} + G_n(h_f) + R_n(h_f) \right) d\pi_k(f)},$$

so that

$$I_n = \frac{\sum_k J_{k,n} \mathbb{P}^\pi[\tilde{A}_n \cap \mathcal{F}_k | X^n]}{\sum_k \mathbb{P}^\pi[\tilde{A}_n \cap \mathcal{F}_k | X^n]}. \tag{3.14}$$

We now study $J_{k,n}$, using the approach described in Section 2.1 and Proposition A.1. At this stage, we have only to focus on Condition 3. of Proposition A.1. To define $\psi_{k,\theta}$ and $T_k \theta$ for all $\theta \in \tilde{A}_n \cap \mathbb{R}^k$, first define

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

We have, using (D.1) and since $k \le l_n$,

$$\begin{aligned}\|D_{n,k,t}\|_\infty &\le \frac{t\sqrt{k}}{\sqrt{n}}\|\tilde\psi_\Pi^{[k]}\|_{\ell_2} \le \frac{t\sqrt{k}}{\sqrt{n}}\|\Pi_{f_0,k}\tilde\psi\|_2 \le \frac{t\sqrt{k}}{\sqrt{c_0}\sqrt{n}}\|\Pi_{f_0,k}\tilde\psi\|_{f_0}\\ &\le \frac{t\sqrt{k}}{\sqrt{c_0}\sqrt{n}}\|\tilde\psi\|_{f_0} \le \frac{t\sqrt{l_n}}{\sqrt{c_0}\sqrt{n}}\|\tilde\psi\|_\infty = O(\epsilon_n),\end{aligned}$$

where c_0 is a lower bound for f_0. Now, we can set

$$T_k\theta = \theta - t\frac{\tilde\psi_\Pi^{[k]}}{\sqrt{n}}$$

and

$$\psi_{k,\theta} = \frac{\sqrt{n}D_{n,k,t}}{t} - \frac{\sqrt{n}}{t}\left(c(\theta) - c\left(\theta - t\frac{\tilde\psi_\Pi^{[k]}}{\sqrt{n}}\right)\right).$$

So, $f_{T_k\theta} = f_\theta e^{-t\psi_{k,\theta}/\sqrt{n}}$ and $\bar\psi_{h,n} - \psi_{k,\theta} = \Delta_{k,\theta} - b_{k,\theta}$ with $\Delta_{k,\theta} = \tilde\psi - \Pi_{f_0,k}\tilde\psi$ and straightforward computations show that:

$$\begin{aligned}b_{k,\theta} &= -\tilde\psi_{\Pi,0} - \frac{\sqrt{n}}{t}\left(c(\theta) - c\left(\theta - t\frac{\tilde\psi_\Pi^{[k]}}{\sqrt{n}}\right)\right) - \frac{\sqrt{n}}{t}\log\left(F_0\left[\exp\left(\frac{h_\theta}{\sqrt{n}} - \frac{t\tilde\psi}{\sqrt{n}}\right)\right]\right)\\ &= \frac{\sqrt{n}}{t}\log\left[\frac{F_0(e^{H_n + t\Delta_{k,\theta}/\sqrt{n}})}{F_0(e^{H_n})}\right],\end{aligned}$$

with $H_n = (h_\theta - t\tilde\psi)/\sqrt{n}$.

We first bound $\|\Delta_{k,\theta}\|_\infty$. To emphasize the fact that $\Delta_{k,\theta}$ does not depend on θ, we write hereafter $\Delta_k := \Delta_{k,\theta}$. Since $\|\tilde\psi\|_\infty = O(1)$ and since on $\tilde A_n$,

$$\begin{aligned}n^{-1/2}\|h_\theta\|_\infty &\le \|\sum_{\lambda=1}^{+\infty}(\theta_{0\lambda} - \theta_\lambda)\phi_\lambda(x) - c(\theta_0) + c(\theta)\|_\infty\\ &\le C\sqrt{l_n}\|\theta_0 - \theta\|_{\ell_2} + C\|\theta_0 - \theta\|_{\ell_2} = o(1),\end{aligned}$$

$\|H_n\|_\infty = o(1)$. To bound $\|\Delta_k\|_\infty$, we set $\psi_{+k} = \sum_{\lambda>k}\tilde\psi_\lambda\phi_\lambda$, so $\Delta_k = \psi_{+k} - \Pi_{f_0,k}(\psi_{+k})$. Then by using (D.1),

$$\begin{aligned}\|\Delta_k\|_\infty &\le \|\psi_{+k}\|_\infty + \|\Pi_{f_0,k}\psi_{+k}\|_\infty\\ &\le \|\psi_{+k}\|_\infty + C\sqrt{k}\|\Pi_{f_0,k}\psi_{+k}\|_{f_0}\\ &\le \|\psi_{+k}\|_\infty + C\sqrt{k}\|\psi_{+k}\|_{f_0}\\ &\le \|\psi_{+k}\|_\infty + C\sqrt{k}\|\psi_{+k}\|_2 = o\left(\frac{1}{\sqrt{n}u_n^2}\right),\end{aligned}$$

where the last inequality comes from condition (2.10).

We now bound $b_{k,\theta}$. Since $F_0(\Delta_k^2) = O(1)$ and $\|\Delta_k\|_\infty = o(\sqrt{n})$,

$$\begin{aligned} F_0(e^{H_n + t\Delta_k/\sqrt{n}}) &= F_0\left(e^{H_n}\left(1 + \frac{t\Delta_k}{\sqrt{n}}\right)\right) + O\left(\frac{F_0(\Delta_k^2)}{n}\right) \\ &= F_0\left(e^{H_n}\right) + \frac{t}{\sqrt{n}} F_0(e^{H_n}\Delta_k) + o(1/\sqrt{n}). \end{aligned}$$

Note also that from (3.8), $F_0(e^{H_n}) = 1 + o(1/\sqrt{n})$. Furthermore, since $F_0(|\Delta_k|) < \infty$, $\|\tilde{\psi}\|_\infty < \infty$ and since $n^{-1/2}\|h_\theta\|_\infty = o(1)$,

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

We thus obtain that

$$b_{k,\theta} = \frac{\sqrt{n}}{t} \log\left[\frac{F_0(e^{H_n + t\Delta_k/\sqrt{n}})}{F_0(e^{H_n})}\right] = o(1). \tag{3.15}$$

Note that $F_0((\bar{\psi}_{h,n} - \psi_{k,\theta})^2) = F_0((\Delta_k - b_{k,\theta})^2) = F_0(\Delta_k^2) + o(1) = O(1)$ uniformly over $\tilde{A}_n$ and condition 3 of Proposition A.1 is satisfied with $w_n = 1$ for any n, which implies that

$$\begin{aligned} \rho_n(\theta) &= -tF_0[\Delta_k h_\theta] + tG_n(\Delta_k) \\ &\quad -\frac{t^2}{2} F_0((\bar{\psi}_{h,n} - \psi_{k,\theta})^2) + t^2 F_0\left[(\bar{\psi}_{h,n} - \psi_{k,\theta})\bar{\psi}_{h,n}\right] + o(1). \end{aligned}$$

We can simplify this last expression. Indeed, since Δ_k is orthogonal to any $\phi_\lambda, \lambda \leq k$ including $\phi_0 = 1$, we obtain, using the expression of h_θ in exponential families,

$$F_0(h_\theta \Delta_k) = -\sqrt{n} F_0((\sum_{\lambda > k} \theta_{0\lambda}\phi_\lambda)\Delta_k),$$

which is independent of θ. Also,

$$F_0((\bar{\psi}_{h,n} - \psi_{k,\theta})\bar{\psi}_{h,n}) = F_0(\Delta_k^2) + F_0(\psi_{k,\theta}\Delta_k) + o(1) = F_0(\Delta_k^2) + o(1),$$

where the last equality comes from the orthogonality between Δ_k and $\psi_{k,\theta}$. Therefore, uniformly in θ,

$$\rho_n(\theta) = t\mu_{n,k} + \frac{t^2}{2} F_0(\Delta_k^2) + o(1)$$

and for all k such that $\mathcal{F}_k \cap \tilde{A}_n \neq \emptyset$, by using the definition of $\rho_n(\theta)$ in Section 2.1, we obtain finally:

$$J_{k,n} = e^{\frac{-t^2 F_0(\Delta_k^2)}{2}} e^{-t\mu_{n,k}} \frac{\int_{\mathbb{R}^k \cap \tilde{A}_n} e^{-\frac{F_0\left(h^2_{T_k\theta}\right)}{2} + G_n\left(h_{T_k\theta}\right) + R_n\left(h_{T_k\theta}\right)} d\pi_k(\theta)}{\int_{\mathbb{R}^k \cap \tilde{A}_n} e^{-\frac{F_0(h_\theta^2)}{2} + G_n(h_\theta) + R_n(h_\theta)} d\pi_k(\theta)} (1 + o(1)).$$

We now prove that the prior π_k is not affected by the change of parameter $\theta \to T_k\theta$. For $k \leq l_n$, $\|\tilde{\psi}_{\Pi}^{[k]}\|_{\ell_2} \leq C$, where C does not depend on k and n. So, if we set

$$T_k(\tilde{A}_n) = \left\{ \theta + t\frac{\tilde{\psi}_{\Pi}^{[k]}}{\sqrt{n}} \ \text{ s.t. } \ \theta \in \mathbb{R}^k \cap \tilde{A}_n \right\},$$

for all $\theta \in T_k(\tilde{A}_n)$, using Theorem B.1 and $(a+b)^2 \leq 2a^2 + 2b^2$,

$$\|\theta - \theta_0\|_{\ell_2}^2 \leq \frac{4\epsilon_n^2 \log^3 n}{L(n)} + \frac{2t^2 C^2}{n} \leq \frac{4\epsilon_n^2 \log^3 n}{L(n)} (1 + o(1))$$

since $n\epsilon_n^2 \to +\infty$. Conversely, for all $\theta \in \mathbb{R}^k \cap \tilde{A}_n$ such that $\|\theta - \theta_0\|_{\ell_2} \leq 1.5(\log n)^{3/2} L(n)^{-1/2} \epsilon_n$

$$\theta - t\frac{\tilde{\psi}_{\Pi}^{[k]}}{\sqrt{n}} \in \tilde{A}_n \cap \mathbb{R}^k$$

for n large enough and we can write (still for n large enough)

$$\mathbb{R}^k \cap \tilde{A}_{n,1} \subset T_k(\tilde{A}_n) \subset \mathbb{R}^k \cap \tilde{A}_{n,2} \tag{3.16}$$

with

$$\tilde{A}_{n,1} = \left\{ \theta \in \tilde{A}_n \ \text{ s.t. } \ \|\theta - \theta_0\|_{\ell_2} \leq 1.5(\log n)^{3/2} L(n)^{-1/2} \epsilon_n \right\},$$

$$\tilde{A}_{n,2} = \left\{ \theta \ \text{ s.t. } \ \|\theta - \theta_0\|_{\ell_2} \leq \sqrt{5}(\log n)^{3/2} L(n)^{-1/2} \epsilon_n \right\}.$$

Therefore,

$$J_{k,n} \leq e^{-t^2 \frac{F_0(\Delta_k^2)}{2}} e^{-t\mu_{n,k}} \frac{\int_{\mathbb{R}^k \cap \tilde{A}_{n,2}} e^{-\frac{F_0(h_\theta^2)}{2} + G_n(h_\theta) + R_n(h_\theta)} d\pi_k(\theta)}{\int_{\mathbb{R}^k \cap \tilde{A}_n} e^{-\frac{F_0(h_\theta^2)}{2} + G_n(h_\theta) + R_n(h_\theta)} d\pi_k(\theta)} (1 + o(1)),$$

$$J_{k,n} \geq e^{-t^2 \frac{F_0(\Delta_k^2)}{2}} e^{-t\mu_{n,k}} \frac{\int_{\mathbb{R}^k \cap \tilde{A}_{n,1}} e^{-\frac{F_0(h_\theta^2)}{2} + G_n(h_\theta) + R_n(h_\theta)} d\pi_k(\theta)}{\int_{\mathbb{R}^k \cap \tilde{A}_n} e^{-\frac{F_0(h_\theta^2)}{2} + G_n(h_\theta) + R_n(h_\theta)} d\pi_k(\theta)} (1 + o(1)).$$

Therefore, using (3.3) and the last equality of the proof of Theorem 2.1 and combining the above inequalities with (3.14) we obtain

$$\begin{aligned}\zeta_n(t) &:= \mathbb{E}^{\pi}[\exp(t\sqrt{n}(\Psi(f)-\Psi(\mathbb{P}_n)))\mathbb{1}_{\tilde{A}_n}(f)|X^n] \\ &= L_n(t)\times\mathbb{P}^{\pi}\left\{\tilde{A}_n|X^n\right\} = L_n(t)(1+o(1)) \\ &= e^{\frac{t^2F_0(\tilde{\psi}^2)}{2}}I_n(1+o(1)).\end{aligned}$$

Therefore,

$$\begin{aligned}\zeta_n(t) &\leq e^{\frac{t^2F_0(\tilde{\psi}^2)}{2}}\frac{\sum_{k=1}^{l_n}e^{-t^2\frac{F_0(\Delta_k^2)}{2}}e^{-t\mu_{n,k}}\mathbb{P}^{\pi}\left[\tilde{A}_{n,2}\cap\mathbb{R}^k|X^n\right]}{\sum_{k=1}^{l_n}\mathbb{P}^{\pi}\left[\tilde{A}_{n,1}\cap\mathbb{R}^k|X^n\right]}(1+o(1)) \\ &\leq e^{\frac{t^2F_0(\tilde{\psi}^2)}{2}}\sum_{k=1}^{l_n}e^{-t^2\frac{F_0(\Delta_k^2)}{2}}e^{-t\mu_{n,k}}\mathbb{P}^{\pi}[k|X^n](1+o(1)). \qquad (3.17)\end{aligned}$$

We now have to provide a lower bound of $\zeta_n(t)$. Using the exponential rate pointed out in Theorem B.1, We first observe that with $\mathbb{P}_0$-probability tending to 1, $\mathbb{P}^{\pi}\left[\tilde{A}_{n,i}|X^n\right]\geq 1-e^{-nc\epsilon_n^2}$, for some positive $c>0$ and $i=1,2$. We also have that for $0<a<1/(2\max(\beta;\gamma)+1)$,

$$\begin{aligned}\mathbb{P}_0\left[\max_{k\leq l_n}|G_n(\Delta_k)|>n^{-a}n\epsilon_n^2\right] &\leq \sum_{k=1}^{l_n}\frac{F_0[\Delta_k^2]}{n^{-2a}(n\epsilon_n^2)^2} \\ &\leq \frac{F_0[\tilde{\psi^2}]l_n}{n^{-2a}(n\epsilon_n^2)^2}=o(1).\end{aligned}$$

Define the event

$$\Omega_n=\left\{\max_{k\leq l_n}|G_n(\Delta_k)|\leq n^{-a}n\epsilon_n^2,\quad \mathbb{P}^{\pi}\left[\tilde{A}_{n,i}|X^n\right]\geq 1-e^{-nc\epsilon_n^2}, i=1,2\right\},$$

so that $\mathbb{P}_0[\Omega_n^c]=o(1)$ and on Ω_n we have

$$\begin{aligned}\zeta_n(t) &\geq e^{\frac{t^2F_0(\tilde{\psi}^2)}{2}}\frac{\sum_{k=1}^{l_n}e^{-t^2\frac{F_0(\Delta_k^2)}{2}}e^{-t\mu_{n,k}}\mathbb{P}^{\pi}\left[\tilde{A}_{n,1}\cap\mathbb{R}^k|X^n\right]}{\sum_{k=1}^{l_n}\mathbb{P}^{\pi}\left[\tilde{A}_{n,2}\cap\mathbb{R}^k|X^n\right]}(1+o(1)) \\ &\geq e^{\frac{t^2F_0(\tilde{\psi}^2)}{2}}\sum_{k=1}^{l_n}e^{-t^2\frac{F_0(\Delta_k^2)}{2}}e^{-t\mu_{n,k}}\left[\mathbb{P}^{\pi}[k|X^n]-\mathbb{P}^{\pi}\left[(\tilde{A}_{n,1})^c\cap\mathbb{R}^k|X^n\right]\right].\end{aligned}$$

Now, we introduce

$$I_0 = \left\{ k \le l_n \ \text{ s.t. } \ \mathbb{P}^\pi[k|X^n] \ge r_n^{-1} \mathbb{P}^\pi \left[(\tilde{A}_{n,1})^c \cap \mathbb{R}^k | X^n \right] \right\}$$
$$I_1 = \left\{ k \le l_n \ \text{ s.t. } \ \mathbb{P}^\pi[k|X^n] < r_n^{-1} \mathbb{P}^\pi \left[(\tilde{A}_{n,1})^c \cap \mathbb{R}^k | X^n \right] \right\},$$

with $r_n = e^{-\frac{nc\epsilon_n^2}{2}}$. We have

$$\zeta_n(t) \ \ge \ (1 - r_n) e^{t^2 \frac{F_0(\tilde{\psi}^2)}{2}} \sum_{k \in I_0} e^{-t\mu_{n,k}} e^{-t^2 \frac{F_0(\Delta_k^2)}{2}} \mathbb{P}^\pi[k|X^n]$$

and

$$\sum_{k \in I_1} \mathbb{P}^\pi[k|X^n] \le r_n^{-1} \mathbb{P}^\pi \left[(\tilde{A}_{n,1})^c | X^n \right] \le e^{-\frac{nc\epsilon_n^2}{2}}. \tag{3.18}$$

Moreover, on Ω_n,

$$\begin{aligned} |\mu_{n,k}| \ &\le \ C \left[\sqrt{n} \left(\sum_{\lambda=k+1}^{+\infty} \tilde{\psi}_\lambda^2 \right)^{1/2} \left(\sum_{\lambda=k+1}^{+\infty} \theta_{0\lambda}^2 \right)^{1/2} + n^{-a} n \epsilon_n^2 \right] \\ &\le \ 2n^{-a} n \epsilon_n^2, \end{aligned}$$

for n large enough. This yields

$$\sum_{k \in I_1} e^{-t\mu_{n,k}} e^{-t^2 \frac{F_0(\Delta_k^2)}{2}} \mathbb{P}^\pi[k|X^n] \ \le \ C e^{2tn^{-a} n \epsilon_n^2 - \frac{nc\epsilon_n^2}{2}}.$$

Using (3.10) and (3.18), for n large enough,

$$\begin{aligned} \sum_{k \in I_0} e^{-t\mu_{n,k}} e^{-t^2 \frac{F_0(\Delta_k^2)}{2}} \mathbb{P}^\pi[k|X^n] \ &\ge \ e^{-3tn^{-a} n \epsilon_n^2} \\ &\ge \ e^{\frac{nc\epsilon_n^2}{4}} \sum_{k \in I_1} e^{-t\mu_{n,k}} e^{-t^2 \frac{F_0(\Delta_k^2)}{2}} \mathbb{P}^\pi[k|X^n]. \end{aligned}$$

This yields

$$\zeta_n(t) \ge e^{t^2 \frac{F_0(\tilde{\psi}^2)}{2}} \sum_{k=1}^{l_n} e^{-t\mu_{n,k}} e^{-t^2 \frac{F_0(\Delta_k^2)}{2}} \mathbb{P}^\pi[k|X^n](1 + o(1)). \tag{3.19}$$

Inequalities (3.17) and (3.19) prove that the posterior distribution of $\sqrt{n}(\Psi(f) - \Psi(P_n))$ is asymptotically equal to a mixture of Gaussian distributions with variance $V_{0k} = F_0(\tilde{\psi}^2) - F_0(\Delta_k^2)$, means $-\mu_{n,k}$ and weights $\mathbb{P}^\pi(k|X^n)$. Straightforward computations prove the last part of the theorem.

3.3. *Proof of Corollary 2.1.* Condition (2.9) is satisfied if

$$\max_{1\le k\le l_n}\sup_{\theta\in B_k}\sum_{\lambda=1}^{k}\left|\log g\left(\frac{\theta_\lambda}{\sqrt{\tau_\lambda}}\right)-\log g\left(\frac{\theta_\lambda}{\sqrt{\tau_\lambda}}-\frac{t}{\sqrt{n}}\frac{\tilde\psi_{\Pi,\lambda}}{\sqrt{\tau_\lambda}}\right)\right|=o(1). \tag{3.20}$$

So, if we introduce the conditions

$$\sum_{\lambda=1}^{l_n}\tilde\psi_{\Pi,\lambda}^2\lambda^{2\beta}=o(n) \tag{3.21}$$

and

$$\max_{1\le k\le l_n}\sum_{\lambda=1}^{k}\theta_\lambda\tilde\psi_{\Pi,\lambda}\lambda^{2\beta}=o(\sqrt{n}), \tag{3.22}$$

then:

- if g is the Laplace density, (3.20) is satisfied if (3.21) is satisfied,
- if g is the Gaussian or the Student density, (3.20) is satisfied if (3.21) and (3.22) are satisfied.

Now, since $\sum_{\lambda=1}^{l_n}\tilde\psi_{\Pi,\lambda}^2\le C$ (see the arguments in the proof of Theorem 2.3),

$$\sum_{\lambda=1}^{l_n}\tilde\psi_{\Pi,\lambda}^2\lambda^{2\beta}\le Cl_n^{2\beta}=o(n)$$

under the assumptions of Corollary 2.1. Furthermore, for $1\le k\le l_n$,

$$\begin{aligned}\sum_{\lambda=1}^{k}\theta_\lambda\tilde\psi_{\Pi,\lambda}\lambda^{2\beta} &\le l_n^{2\beta}\sum_{\lambda=1}^{k}|\theta_\lambda-\theta_{0\lambda}||\tilde\psi_{\Pi,\lambda}|+\sum_{\lambda=1}^{l_n}\lambda^{2\beta-\gamma}\lambda^\gamma|\theta_{0\lambda}||\tilde\psi_{\Pi,\lambda}|\\ &\le C\left(l_n^{2\beta}\epsilon_n\frac{(\log n)^{3/2}}{L(n)}+l_n^{2\beta-\gamma}+1\right).\end{aligned}$$

The expressions of l_n and ϵ_n in the cases (PH) and (D) and straightforward computations allow to finish the proof.

3.4. *Proof of Proposition 2.1.* We first prove that $f_0\in W^\gamma$ for any k_0. Indeed, for any $J_1>3$,

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

In the same spirit, observe that for $J_1 > 3$,

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

and if $x_0 = 1/4$

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

when $J_1 \to \infty$. Set $\epsilon_n^2 = n^{-2\gamma/(2\gamma+1)} (\log n)^{(2\gamma-1)/(2\gamma+1)} (\log\log n)^{-2/(2\gamma+1)}$ and

$$k_n = n^{1/(2\gamma+1)} (\log n)^{-2/(2\gamma+1)} (\log\log n)^{-2/(2\gamma+1)}. \tag{3.23}$$

Previous computations lead to the existence of $0 < c_1 < c_2 < +\infty$ such that

$$c_1 \epsilon_n^2 \le \sum_{\lambda \ge k_n} \theta_{0\lambda}^2 \le c_2 \epsilon_n^2. \tag{3.24}$$

Furthermore, $n\epsilon_n^2$ is of the same order as $k_n \log n$ and it is straightforward to prove that there exists $c > 0$ such that

$$\mathbb{P}^{\pi} \left\{ K(f_0, f_\theta) \le \epsilon_n^2, \quad V(f_0, f_\theta) \le \epsilon_n^2 \right\} \ge e^{-cn\epsilon_n^2},$$

which implies, using Lemma 8.1 of [8], that

$$\int_{\mathcal{F}} e^{\ell_n(f) - \ell_n(f_0)} d\pi(f) \ge e^{-cn\epsilon_n^2}$$

for some constant $c > 0$ with probability going to 1. If k_1 is large enough, then $\mathbb{P}^\pi[k > k_1 k_n] \leq e^{-2cn\epsilon_n^2}$ and the computation of [8] p. 525 implies

$$\mathbb{P}^\pi[k \leq k_1 k_n | X^n] = 1 + o_{\mathbb{P}_0}(1).$$

This also implies from Theorem B.1 that the posterior concentration rate (for the ℓ_2-loss) is less than $M_0 \epsilon_n \log n$ for some positive M_0.

Moreover, inequality (3.23) implies that there exists k_2 such that for all $k \leq k_2 k_n (\log n)^{-1/(2\gamma)}$ and all $\theta \in \mathbb{R}^k$, $\|\theta - \theta_0\|_{\ell_2} > M_0 \epsilon_n \log n$, so we can restrict ourselves to $k \geq k_2 k_n (\log n)^{-1/(2\gamma)}$. So, there exist $k_1 > 0$ and $k_2 > 0$ such that

$$\mathbb{P}^\pi \left\{ k_2 k_n (\log n)^{-1/(2\gamma)} \leq k \leq k_1 k_n | X^n \right\} = 1 + o_{\mathbb{P}_0}(1).$$

We now show that

$$\mathbb{P}_0 \left[\min_{k_2 k_n (\log n)^{-1/(2\gamma)} \leq k \leq k_1 k_n} \mu_{n,k} \geq c\sqrt{\log n} \right] \to 1. \tag{3.25}$$

First, by using the same arguments as in the proof of Theorem 2.3, we note that when $k \in [k_2 k_n (\log n)^{-1/(2\gamma)}, k_1 k_n]$ $G_n(\Delta_k) = o_{\mathbb{P}_0}(1)$ and also that

$$\max_{k_2 k_n (\log n)^{-1/(2\gamma)} \leq k \leq k_1 k_n} |G_n(\Delta_k)| = O_{\mathbb{P}_0}(1).$$

Now, we have:

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

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 $k \leq k_1 k_n$,

$$\mu_{n,k,1} \geq c'_1 \sqrt{n} k^{-\gamma-1/2} (\log k)^{-1/2} (\log\log k)^{-1} \geq c'_2 \sqrt{\log n}.$$

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

$$\Delta_k = \sum_{\lambda \geq k+1} \tilde{\psi}_\lambda \phi_\lambda - \Pi_{f_{0,k}} \left(\sum_{\lambda \geq k+1} \tilde{\psi}_\lambda \phi_\lambda \right)$$

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, for any $k \in [k_2 k_n (\log n)^{-1/(2\gamma)}, k_1 k_n]$,

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

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 (3.25) is true.

APPENDIX A: PRELIMINARY RESULT TO PROVE (A3)

We state the following technical result that constitutes the first step to prove the condition (A3) which expresses the change of parameter. We use notations of Section 2.1.

PROPOSITION A.1. *For a sequence $(u_n)_n$ such that $\sqrt{n} u_n \to +\infty$, we assume that the following three conditions are satisfied.*

1. *Assumption (A1) is satisfied with $(u_n)_n$ and there exists a sequence $(l_n)_n$ of integers such that $\mathbb{P}^\pi [k > l_n | X^n] = o_{\mathbb{P}_0}(1)$.*
2. *There exists a sequence $(w_n)_n$ lower bounded by a positive constant such that $w_n \sqrt{n} u_n^2 = o(1)$ and*

$$\tilde{A}_n \subset A^1_{u_n} \cap (\cup_{k \leq l_n} \mathcal{F}_k) \cap \{f_\theta \ \ s.t. \ \ V(f_\theta, f_0) \leq w_n u_n^2\},$$

satisfies

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

To simplify notations, in the sequel, $\tilde{A}_n$ will also denote the set of sequences θ such that $f_\theta \in \tilde{A}_n$.

3. *For each k such that $\mathcal{F}_k \cap \tilde{A}_n \neq \emptyset$, there exists a map : $T_k : \tilde{A}_n \cap \Theta_k \to \Theta_k$ and a function $\psi_{k,\theta}$ such that for all $\theta \in \tilde{A}_n \cap \Theta_k$*
   (a) $f_{T_k\theta} = f_\theta e^{-t\psi_{k,\theta}/\sqrt{n}}$
   (b)
   $$\max_{k\leq l_n} \sup_{\theta\in\Theta_k\cap\tilde{A}_n} F_0\left[(\bar{\psi}_{h,n} - \psi_{k,\theta})^2\right] = O(1)$$
   (c) *for all $\theta \in \Theta_k$ such that $f_\theta \in \tilde{A}_n$, $\bar{\psi}_{h,n}(x) - \psi_{k,\theta}(x)$ can be decomposed as*
   $$\bar{\psi}_{h,n}(x) - \psi_{k,\theta}(x) = \Delta_{k,\theta}(x) - b_{k,\theta},$$
   where $b_{k,\theta}$ is a constant such that
   $$\max_{k\leq l_n} \sup_{\theta\in\Theta_k\cap\tilde{A}_n} |b_{k,\theta}| = o(u_n^{-1} w_n^{-1/2})$$
   and $\Delta_{k,\theta}(x)$ is a function satisfying
   $$\max_{k\leq l_n} \sup_{f\in\mathcal{F}_k\cap\tilde{A}_n} \|\Delta_{k,\theta}\|_\infty = o(w_n^{-1} n^{-1/2} u_n^{-2} \wedge n^{1/2}).$$

Then, we have uniformly over $\cup_{k\leq l_n}\mathcal{F}_k \cap \tilde{A}_n$

$$\begin{aligned}\rho_n(\theta) &= -tF_0[\Delta_{k,\theta}h_\theta] + tG_n(\Delta_{k,\theta}) \\ &\quad -\frac{t^2}{2}F_0((\bar{\psi}_{h,n} - \psi_{k,\theta})^2) + t^2 F_0\left[(\bar{\psi}_{h,n} - \psi_{k,\theta})\bar{\psi}_{h,n}\right] + o(1).\end{aligned}$$

The conditions considered in Proposition A.1 are mild and, apart from condition 3, are slightly stronger versions of assumptions (A1) and (A2). In the two types of examples considered in this paper $w_n = 1$ and in many cases w_n increases to infinity at most as a power of $\log n$. The constraints on $b_{k,\theta}$ and $\Delta_{k,\theta}$ are mild since the right hand terms go to infinity.

Proof of Proposition A.1: We consider the change of parameter $\theta \mapsto T_k\theta$ for all θ such that $f_\theta \in \tilde{A}_n \cap \mathcal{F}_k$ and we study

$$\begin{aligned}\rho_n(\theta) &= -\frac{F_0(h_{T_k\theta}^2)}{2} + G_n(h_{T_k\theta}) + R_n(h_{T_k\theta}) \\ &\quad -\left(-\frac{F_0((h_\theta - t\bar{\psi}_{h,n})^2)}{2} + G_n(h_\theta - t\bar{\psi}_{h,n}) + R_n(h_\theta - t\bar{\psi}_{h,n})\right),\end{aligned}$$

with $h_{T_k\theta} = \sqrt{n}\log(f_{T_k\theta}/f_0)$. Recall that

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

and $\|\tilde{\psi}\|_\infty < \infty$. From (3.8)

$$\frac{\sqrt{n}}{t}\log\left(F_0\left[\exp\left(\frac{h_\theta}{\sqrt{n}} - \frac{t\tilde{\psi}}{\sqrt{n}}\right)\right]\right) = o(1),$$

so that $\|\bar{\psi}_{h,n}\|_\infty < +\infty$. Writing $h_{T_k\theta} = h_\theta - t\bar{\psi}_{h,n} + t(\bar{\psi}_{h,n} - \psi_{k,\theta})$ and combining the above upper bound with Condition 3 of Proposition A.1, we obtain

$$\begin{aligned} F_0(h_{T_k\theta}^2) &= F_0((h_\theta - t\bar{\psi}_{h,n})^2) + t^2F_0((\bar{\psi}_{h,n} - \psi_{k,\theta})^2) + 2tF_0(h_\theta(\bar{\psi}_{h,n} - \psi_{k,\theta})) \\ &\quad -2t^2F_0(\bar{\psi}_{h,n}(\bar{\psi}_{h,n} - \psi_{k,\theta})), \end{aligned}$$

$$G_n(h_{T_k\theta}) = G_n(h_\theta - t\bar{\psi}_{h,n}) + tG_n(\bar{\psi}_{h,n} - \psi_{k,\theta}) = G_n(h_\theta - t\bar{\psi}_{h,n}) + tG_n(\Delta_{k,\theta})$$

and

$$\begin{aligned} R_n(h_{T_k\theta}) &= R_n(h_\theta - t\bar{\psi}_{h,n}) + t\sqrt{n}F_0(\bar{\psi}_{h,n} - \psi_{k,\theta}) + \frac{t^2}{2}F_0((\bar{\psi}_{h,n} - \psi_{k,\theta})^2) \\ &\quad +tF_0((\bar{\psi}_{h,n} - \psi_{k,\theta})(h_\theta - t\bar{\psi}_{h,n})) \end{aligned}$$

so that

$$\rho_n(\theta) = t\sqrt{n}F_0(\bar{\psi}_{h,n} - \psi_{k,\theta}) + tG_n(\Delta_{k,\theta}).$$

To give the final expression of $\rho_n(\theta)$, we use following computations. Recall that $\bar{\psi}_{h,n} - \psi_{k,\theta} = \Delta_{k,\theta} - b_{k,\theta}$ where $b_{k,\theta}$ is a constant with respect to x and note that by definition of $\psi_{k,\theta}$, $F_0(e^{(h_\theta - t\psi_{k,\theta})/\sqrt{n}}) = 1$ so that since

$$\|\bar{\psi}_{h,n} - \psi_{k,\theta}\|_\infty = \|\Delta_{k,\theta} - b_{k,\theta}\|_\infty = o(\sqrt{n}),$$

we have:

$$\begin{aligned} 1 &= F_0(e^{(h_\theta - t\bar{\psi}_{h,n})/\sqrt{n} + t(\bar{\psi}_{h,n} - \psi_{k,\theta})/\sqrt{n}}) \\ &= F_0(e^{(h_\theta - t\bar{\psi}_{h,n})/\sqrt{n}}) + \frac{t}{\sqrt{n}}F_0(e^{(h_\theta - t\bar{\psi}_{h,n})/\sqrt{n}}(\bar{\psi}_{h,n} - \psi_{k,\theta})) \\ &\quad +\frac{t^2}{n}F_0(e^{(h_\theta - t\bar{\psi}_{h,n})/\sqrt{n}}(\bar{\psi}_{h,n} - \psi_{k,\theta})^2 B_{(\bar{\psi}_{h,n} - \psi_{k,\theta}),n}), \end{aligned}$$

where $B_{h,n}$ is defined in (3.4). Note that $F_0(e^{(h_\theta - t\bar{\psi}_{h,n})/\sqrt{n}}) = 1$ and multiplying the previous expression by n, we obtain: $\sum_{i=1}^6 S_i = 0$, with

$$S_1 = t\sqrt{n}F_0(\bar{\psi}_{h,n} - \psi_{k,\theta}), \quad S_2 = tF_0[(\bar{\psi}_{h,n} - \psi_{k,\theta})(h_\theta - t\bar{\psi}_{h,n})],$$

$$S_3 = \frac{t}{\sqrt{n}}F_0[(\bar{\psi}_{h,n} - \psi_{k,\theta})(h_\theta - t\bar{\psi}_{h,n})^2 B_{h_\theta - t\bar{\psi}_{h,n},n}], \quad S_4 = t^2 F_0[(\bar{\psi}_{h,n} - \psi_{k,\theta})^2 B_{(\bar{\psi}_{h,n} - \psi_{k,\theta}),n}],$$

$$S_5 = \frac{t^2}{\sqrt{n}}F_0[(\bar{\psi}_{h,n} - \psi_{k,\theta})^2(h_\theta - t\bar{\psi}_{h,n})B_{(\bar{\psi}_{h,n} - \psi_{k,\theta}),n}]$$

and

$$S_6 = \frac{t^2}{n}F_0[(\bar{\psi}_{h,n} - \psi_{k,\theta})^2(h_\theta - t\bar{\psi}_{h,n})^2 B_{(\bar{\psi}_{h,n} - \psi_{k,\theta}),n}B_{h_\theta - t\bar{\psi}_{h,n},n}].$$

We successively study each term except the first one.

$$S_2 = tF_0(\Delta_{k,\theta}h_\theta) - tb_{k,\theta}F_0(h_\theta) - t^2F_0[(\bar{\psi}_{h,n} - \psi_{k,\theta})\bar{\psi}_{h,n}].$$

Since $\|\bar{\psi}_{h,n}\|_\infty \le C$, $B_{h_\theta - t\bar{\psi}_{h,n},n} = B_{h_\theta,n}(1 + O(1/\sqrt{n}))$ uniformly over $\tilde{A}_n$. Then,

$$\begin{aligned} S_3 \quad &= \quad -\frac{tb_{k,\theta}}{\sqrt{n}}F_0[h_\theta^2 B_{h_\theta - t\bar{\psi}_{h,n},n}] + o(u_n^{-1}w_n^{-1/2}n^{-1/2} + w_n^{-1}n^{-1}u_n^{-2}) \\ &\quad +O(n^{-1/2}\|\Delta_{k,\theta}\|_\infty F_0\left(h_\theta^2 B_{h_\theta,n}\right))) + O(n^{-1/2}(|b_{k,\theta}| + \|\Delta_{k,\theta}\|_\infty)F_0(|h_\theta|B_{h_\theta,n})) \end{aligned}$$

since from condition (3c) of Proposition A.1, we have $|b_{k,\theta}| = o(u_n^{-1}w_n^{-1/2})$ and $\|\Delta_{k,\theta}\|_\infty = o(w_n^{-1}n^{-1/2}u_n^{-2})$. We have also used that

$$2f_0(x)B_{h_\theta,n}(x) = 2\int_0^1 (1-u)f_0^{1-u}(x)f_\theta^u(x)du \le f_0(x) + f_\theta(x).$$

This inequality implies

$$2F_0\left(h_\theta^2 B_{h_\theta,n}\right) \le F_0(h_\theta^2) + F_\theta(h_\theta^2) \le 2nu_n^2 w_n, \quad 2F_0\left(|h_\theta|B_{h_\theta,n}\right) \le 2u_n\sqrt{nw_n}$$

therefore,

$$S_3 = -\frac{tb_{k,\theta}}{\sqrt{n}}F_0[h_\theta^2 B_{h_\theta,n}] + o(1).$$

Using $\|\bar{\psi}_{h,n} - \psi_{k,\theta}\|_\infty = o(\sqrt{n})$ (see previously), we have:

$$\|B_{(\bar{\psi}_{h,n} - \psi_{k,\theta}),n} - 0.5\|_\infty = \|\int_0^1 (1-u)e^{u(\bar{\psi}_{h,n} - \psi_{k,\theta})/\sqrt{n}}du - 0.5\|_\infty = o(1),$$

and

$$S_4 = \frac{t^2 F_0((\bar\psi_{h,n} - \psi_{k,\theta})^2)}{2}(1 + o(1)).$$

The fifth term is controlled as follows:

$$\begin{aligned} |S_5| &\leq \frac{t^2\|\bar\psi_{h,n} - \psi_{k,\theta}\|_\infty}{\sqrt{n}} \left(F_0(h_\theta^2)\right)^{1/2} \left(F_0((\bar\psi_{h,n} - \psi_{k,\theta})^2\right)^{1/2} + o(1) \\ &= o\left(\frac{\sqrt{n}u_n}{\sqrt{n}u_n\sqrt{w_n}} + \frac{\sqrt{n}u_n}{nu_n^2 w_n}\right) + o(1) = o(1). \end{aligned}$$

Finally,

$$|S_6| \leq \frac{t^2\|\bar\psi_{h,n} - \psi_{k,\theta}\|_\infty^2 F_0(h_\theta^2 B_{h_\theta,n})}{n} + o(1) = o(1).$$

So, we finally obtain

$$\begin{aligned} 0 = t\sqrt{n}F_0(\bar\psi_{h,n} - \psi_{k,\theta}) + tF_0(\Delta_{k,\theta}h_\theta) - tb_{k,\theta}F_0(h_\theta) - \frac{tb_{k,\theta}}{\sqrt{n}}F_0[h_\theta^2 B_{h_\theta,n}] \\ + \frac{t^2 F_0((\bar\psi_{h,n} - \psi_{k,\theta})^2)}{2} - t^2 F_0[(\bar\psi_{h,n} - \psi_{k,\theta})\bar\psi_{h,n}] + o(1). \end{aligned}$$

Using the relation

$$F_0(h_\theta) + \frac{F_0(h_\theta^2 B_{h_\theta,n})}{\sqrt{n}} = 0,$$

which comes from a Taylor expansion of $1 = F_0(e^{h_\theta/\sqrt{n}})$, we obtain

$$0 = t\sqrt{n}F_0(\bar\psi_{h,n} - \psi_{k,\theta}) + tF_0(\Delta_{k,\theta}h_\theta) + \frac{t^2 F_0((\bar\psi_{h,n} - \psi_{k,\theta})^2)}{2} - t^2 F_0[(\bar\psi_{h,n} - \psi_{k,\theta})\bar\psi_{h,n}] + o(1).$$

We finally obtain that uniformly on $\tilde A_n$,

$$\rho_n(\theta) = -tF_0[\Delta_{k,\theta}h_\theta] + tG_n(\Delta_{k,\theta}) - \frac{t^2}{2}F_0((\bar\psi_{h,n} - \psi_{k,\theta})^2) + t^2 F_0[(\bar\psi_{h,n} - \psi_{k,\theta})\bar\psi_{h,n}] + o(1)$$

and Proposition A.1 is proved.

APPENDIX B: 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, we now state the following result established in [19]. We use notations of Section 1.2.

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

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

Then, there exists $c > 0$ *such that if*

$$\Omega_n = \left\{ \theta \ \text{ s.t. } \ h(f_0, f_\theta) \leq \sqrt{\frac{\log n}{L(n)}} \epsilon_n \ \text{ and } \|\theta_0 - \theta\|_{\ell_2} \leq \sqrt{\frac{(\log n)^3}{L(n)}} \epsilon_n \right\}$$

$$\lim_{n \to +\infty} \mathbb{P}_0 \left\{ \mathbb{P}^\pi \left\{ \Omega_n | X^n \right\} \geq 1 - \exp(-cn\epsilon_n^2) \right\} = 1,$$

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 given constant. We also have that there exists* $a > 0$ *such that*

$$\mathbb{P}^\pi \left\{ f_\theta \ \text{ s.t. } \ K(f_0, f_\theta) \leq \epsilon_n^2; V(f_0, f_\theta) \leq \epsilon_n^2 \right\} \geq e^{-an\epsilon_n^2}.$$

APPENDIX C: PROOF OF THEOREM 2.2

To prove Theorem 2.2, we need to prove that assumptions (A1), (A2) and (A3) are satisfied. Obtaining a posterior concentration rate in terms of the Hellinger distance, for such a prior is quite straightforward following [8], however some technical details are required to obtain a posterior concentration rate in terms of the divergence V and to check condition (A2). We first give the main arguments to check assumptions (A1) and (A2). Recall that f_0 is Hölderian with smoothness $\gamma \leq 1$:

- Let $z_j^* = j/k, j = 0, ..., k$, $z^* = (z_0^*, ..., z_k^*)$ and $\eta_j^* = \int_{z_{j-1}^*}^{z_j^*} f_0(x)dx$, $j \geq 1$ then using the fact that for some positive constants $a < b$, we have $a \leq f_0(x) \leq b$ for all x, so that $f_{\eta^*,z^*}(x) \in [a, b]$ for all x and

$$\begin{aligned} \|f_{\eta^*,z^*} - f_0\|_2^2 &= \sum_{j=1}^k \int_{z_{j-1}^*}^{z_j^*} (f_0(x) - k\eta_j^*)^2 \leq C \sum_{j=1}^k (z_j^* - z_{j-1}^*)^{2\gamma+1} \\ &\leq Ck^{-2\gamma}, \end{aligned}$$

we deduce that $\| \log f_0 - \log f_{\eta^*,z^*}\|_{f_0}^2 \leq Ca^{-1}k^{-2\gamma}$, $|f_0/f_{\eta^*,z^*}(x)| \leq b/a$ and $h(f_0, f_{\eta^*,z^*}) \leq Ca^{-1/2}k^{-\gamma/2}$. Then, using Lemma 8.2 of [8] and the above bounds, we have:

$$K(f_0, f_{\eta^*,z^*}) \leq Ck^{-2\gamma} \quad V(f_0, f_{\eta^*,z^*}) \leq Ck^{-2\gamma}.$$

Consider $z = (z_j)_{j=0,\dots,k}$, $\eta = (\eta_j)_{j=1,\dots,k}$ with

$$|z_j - z_j^*| \leq k^{-H_1}, \quad |\eta_j - \eta_j^*| \leq k^{-H_1}, \quad j \in \{1, ..., k-1\}$$

with $H_1 > 1$ large enough, then it is easy to prove that $V(f_0, f_{\eta,z}) \leq 2Ck^{-2\gamma}$ and $K(f_0, f_{\eta,z}) \leq 2Ck^{-2\gamma}$. Set $\epsilon_n \geq \epsilon_0 n^{-\gamma/(2\gamma+\alpha+1)}$ and $k = \lfloor k_0 n^{1/(2\gamma+\alpha+1)} \rfloor$, for some positive constants k_0 and ϵ_0. Remember that both η and $(\Delta z_1, ..., \Delta z_k)$ have a density with respect to the Lebesgue measure on $\mathcal{S}_k$ that has the following form:

$$\pi_k(x_1, ..., x_k) = \frac{e^{-c\sum_{i=1}^k x_i^{-\alpha}}}{b_k(c)},$$

where $b_k(c)$ is the normalization constant. We have for all $\tau < 1$

$$\begin{aligned} b_k(c) &= \int_{\mathcal{S}_k} e^{-c\sum_{i=1}^k x_i^{-\alpha}} dx \leq \frac{1}{(k-1)!} \\ b_k(c) &\geq e^{-c\tau^{-\alpha}k^{\alpha+1}} \int \prod_{i=1}^{k-1} \mathbb{1}_{[\tau/k,1/k]}(\eta_i) d\eta_i = e^{-c\tau^{-\alpha}k^{1+\alpha}}(1-\tau)^{k-1}k^{-(k-1)}. \end{aligned}$$

Therefore,

$$\begin{aligned} &\mathbb{P}^\pi(\{f \ \text{ s.t. } \ K(f_0, f) \leq \epsilon_n^2, \ V(f_0, f) \leq \epsilon_n^2\}) \\ \geq &\ \frac{\int_{\mathcal{S}_k \cap \{|x_i - \eta_i^*| \leq k^{-H_1}, \forall i\}} e^{-c_2\sum_{i=1}^k x_i^{-\alpha}} dx}{b_k(c_2)} \frac{\int_{\mathcal{S}_k \cap \{|x_i - \Delta z_i^*| \leq k^{-H_1}, \forall i\}} e^{-c_1\sum_{i=1}^k x_i^{-\alpha}} dx}{b_k(c_1)} \end{aligned}$$

and since $\eta_i^* \geq k^{-1}a$ and $\Delta z_i^* := z_i^* - z_{i-1}^* = 1/k$, $|\eta_i^* - x_i| \leq n^{-H_1}$ implies that $x_i > a/(2k)$ for n large enough and $|x_i - \Delta z_i^*| \leq k^{-H_1}$ implies that $x_i > 1/(2k)$. This leads to

$$\begin{aligned} \mathbb{P}^\pi(\{f \ \text{ s.t. } \ K(f_0, f) \leq \epsilon_n^2, V(f_0, f) \leq \epsilon_n^2\}) &\geq C((k-1)!)^2 e^{-(c_0 + 2^\alpha(c_1 + a^{-\alpha}c_2))k^{\alpha+1}} \\ &\geq e^{-n\epsilon_n^2}, \end{aligned}$$

if ϵ_0 is large enough.

- For k_1 a constant, let $k_n = k_1 n^{1/(2\gamma+\alpha+1)}$, $A, H > 0$ and

$$\mathcal{F}_n = \{f_{\eta,z} \text{ s.t. } k \leq k_n, n^{-H} \leq \Delta z_j, \forall j = 1, .., k-1, \sum_{j=1}^{k} \eta_j^{-\alpha} \leq A n\epsilon_n^2\},$$

  $H > (\alpha+1)/[\alpha(2\gamma+\alpha+1)]$. Note that for all k if there exists j such that $\Delta z_j \leq n^{-H}$ then $\sum_j \Delta z_j^{-\alpha} \geq n^{\alpha H} \geq A n\epsilon_n^2$ for any $A > 0$ if n is large enough, and vice-versa if $\sum_{j=1}^{k} \eta_j^{-\alpha} \leq A n\epsilon_n^2$ then $\eta_j \geq (A n\epsilon_n^2)^{1/\alpha}$ for all $j = 1, ..., k$. Therefore

$$\begin{aligned} \mathbb{P}^\pi(\mathcal{F}_n^c) &\leq p(k > k_n) + e^{-Ac_2 n\epsilon_n^2} \sum_{k \leq k_n} \frac{1}{b_k(c_2)} + e^{-Ac_1 n\epsilon_n^2} \sum_{k \leq k_n} \frac{1}{b_k(c_1)} \\ &\leq e^{-Cn\epsilon_n^2}, \end{aligned}$$

  for any $C > 0$, by choosing k_1 and A large enough.
- For all $k \leq k_n$ and all $(f_{\eta,z}, f_{\eta',z'}) \in \mathcal{F}_n{}^2$, we have: $h^2(f_{\eta',z'}, f_{\eta,z}) \leq \epsilon_n^2$ as soon as $|\Delta z_j - \Delta z'_j| \leq n^{-H_1-H}$ for all $j = 1, ..., k-1$ and $|\eta_j - \eta'_j| \leq n^{-H_1}$ for some positive value H_1 large enough, where $\Delta z'_j = z'_j - z'_{j-1}$. Indeed, note that under the above conditions $|z_j - z'_j| \leq \sum_{i=1}^{j} |\Delta z_i - \Delta z'_i| \leq k n^{-H_1-H}$ for all j and also that $I_j \cap (I'_j)^c = (I_j \cap I'_{j-1}) \cup (I_j \cap I'_{j+1})$ for all j, with $I'_j = (z'_{j-1}, z'_j]$, and vice-versa, so that

$$\begin{aligned} h(f_{\eta,z}, f_{\eta',z'})^2 &\leq \sum_{j=1}^{k} (\Delta z_j \wedge \Delta z'_j) \left(\frac{\eta_j^{1/2}}{\Delta z_j^{1/2}} - \frac{(\eta'_j)^{1/2}}{(\Delta z'_j)^{1/2}} \right)^2 \\ &\quad + 4 \int_{z_{j-1} \wedge z'_{j-1}}^{z_{j-1} \vee z'_{j-1}} \left(\frac{\eta_j}{\Delta z_j} + \frac{\eta'_j}{\Delta z'_j} \right) \\ &\leq 2k n^{-2H_1+H} + 3n^{-2H_1} \leq \epsilon_n^2, \end{aligned}$$

  for H_1 large enough. We conclude that the covering number of $\mathcal{F}_n$ is bounded by $\exp(Ck \log n) = o(\exp(n\epsilon_n^2))$.

This implies that the posterior distribution satisfies : if M is large enough,

$$\mathbb{P}^\pi\left[\{f \text{ s.t. } h(f_0, f) \leq M\epsilon_n\} | X^n\right] = o_{\mathbb{P}_0}(1).$$

To obtain the posterior concentration rate in terms of V, note that, if $f_{\eta,z} \in \mathcal{F}_n$, with $h^2(f_0, f_{\eta,z}) \le \epsilon_n^2$,

$$\begin{aligned} M_1 &= \left[\int_0^1 \frac{f_0^2(x)}{f_{\eta,z}(x)}dx\right]^{\frac{1}{2}} \le \|f_0\|_\infty \left[\int_0^1 \sum_{j=1}^k \frac{\Delta z_j}{\eta_j} 1_{I_j}(x)dx\right]^{\frac{1}{2}} \\ &\le \|f_0\|_\infty \left(\max_j \frac{\Delta z_j}{\eta_j}\right)^{1/2} \le \|f_0\|_\infty n^{H/2} \end{aligned}$$

and using Theorem 5 of [27],

$$V(f_0, f_{\eta,z}) \le 5h^2(f_0, f_{\eta,z}) \left(|\log M_1| + |\log h(f_0, f_{\eta,z})|\right)^2 \le C\epsilon_n^2 (\log n)^2$$

and $K(f_0, f_{\eta,z}) \le C\epsilon_n^2 \log n$, which achieves the proof of condition (A1). Assumption(A2) is proved along the same lines. Indeed whenever $f_{\eta,z} \in \mathcal{F}_n$,

$$\left[\int_0^1 \frac{f_{\eta,z}^2(x)}{f_0(x)}dx\right]^{\frac{1}{2}} \le \|f_0^{-1}\|_\infty^{1/2} n^{H/2}$$

so that using Theorem 5 of [27], when $h^2(f_{\eta,z}, f_0) \le \epsilon_n^2$

$$V(f_{\eta,z}, f_0) \le C\epsilon_n^2 (\log n)^2,$$

which together with the inequality

$$\int f_{\eta,z}(x) |\log f_{\eta,z} - \log f_0|(x)dx \le V(f_{\eta,z}, f_0)^{1/2}$$

proves Assumption (A2) with $u_n \asymp \tilde{u}_n = O(\epsilon_n \log n)$. We now validate Assumption (A3) using Proposition A.1. Set $\tilde{A}_n = \mathcal{F}_n \cap \{f_{\eta,z} \text{ s.t. } V(f_0, f_{\eta,z}) \le u_n^2, V(f_{\eta,z}, f_0) \le u_n^2\}$ with $u_n^2 = C\epsilon_n^2 (\log n)^2$, the above results imply

$$\mathbb{P}^\pi\left[\tilde{A}_n | X^n\right] = 1 + o_{\mathbb{P}_0}(1),$$

which proves that conditions 1 and 2 of Proposition A.1 are satisfied with $w_n = 1$. Note that $u_n^2 = o(n^{-1/4})$ and $nu_n^2 \to +\infty$. We prove that condition 3 of Proposition A.1 is satisfied. Let $A_n = \{(k, \eta, z) \text{ s.t. } f_{\eta,z} \in \tilde{A}_n\}$, $\theta = (\eta, z)$ and for any x,

$$\tilde{f}(x) = \sum_{j=1}^k \tilde{\psi}_j 1\!\!1_{I_j}(x), \quad \tilde{\psi}_j = \frac{F_0(1\!\!1_{I_j}\tilde{\psi})}{F_0(1\!\!1_{I_j})}.$$

Denote also

$$\psi_{k,\theta} = \sum_j \tilde{\psi}_j 1\!\!1_{I_j} + \frac{\sqrt{n}}{t} \log \left(\sum_j \eta_j e^{-t\tilde{\psi}_j/\sqrt{n}} \right),$$

and note that $\bar{\psi}_{h,n} - \psi_{k,\theta} = \Delta_{k,\theta} - b_{k,\theta}$ with

$$\Delta_{k,\theta} = \tilde{\psi} - \tilde{f}, \quad b_{k,\theta} = \frac{\sqrt{n}}{t} \log \left[\frac{F_0\left(e^{h_\theta/\sqrt{n}} e^{-t\tilde{f}/\sqrt{n}}\right)}{F_0\left(e^{h_\theta/\sqrt{n}} e^{-t\tilde{\psi}/\sqrt{n}}\right)} \right].$$

For every $\theta = (\eta, z)$ we set $T_k\theta = (\eta', z')$ with

$$\eta'_j = \frac{\eta_j e^{-\frac{t\tilde{\psi}_j}{\sqrt{n}}}}{\sum_i \eta_i e^{-\frac{t\tilde{\psi}_i}{\sqrt{n}}}}, \quad j = 1, ..., k \quad z' = z, \tag{C.1}$$

then by construction $h_{T_k\theta} = f_\theta e^{-t\psi_{k,\theta}/\sqrt{n}}$. For all partition z of $[0, 1]$, if

$$u^* = (u^*_1, ..., u^*_k), \quad u^*_j = F_0(1\!\!1_j \log f_0)/F_0(I_j),$$

then u^* minimizes in $u \in \mathbb{R}^k$

$$\sum_{j=1}^{k} \int_{I_j} f_0(x) \left(u_j - \log f_0(x)\right)^2 dx.$$

Note also that, since $f'_0(x) \geq c_0 > 0$ (or equivalently $f'_0(x) < -c_0$) for all $x \in (0, 1)$, there exists $c'_0 > 0$ such that $(\log f_0(x))' > c'_0$ for all $x \in (0, 1)$ (or $< -c'_0$) and

$$\sum_{j=1}^{k} \int_{I_j} f_0(x) \left(u^*_j - \log f_0(x)\right)^2 dx \geq {c'_0}^2 \sum_j \Delta z_j^3,$$

so that for all $\theta = (\eta, z)$,

$$V(f_0, f_{\eta,z}) \geq {c'_0}^2 \sum_j \Delta z_j^3. \tag{C.2}$$

Therefore,

$$\sum_{j=1}^{k} \Delta z_j^3 \leq C_0(\epsilon_n \log n)^2,$$

for some positive constant C_0 on A_n. Since $\psi(x) = x$,

$$\|\tilde{f} - \tilde{\psi}\|_2^2 = O(\sum_j \Delta z_j^3) = O((\epsilon_n \log n)^2)$$

and since for any $x \in I_j$, $|e^{-t(\tilde{f}(x) - \tilde{\psi}(x))/\sqrt{n}} - 1| \le \frac{2t\Delta z_j}{\sqrt{n}}$,

$$\begin{aligned} b_{k,\theta} &= \frac{\sqrt{n}}{t} \log \left[\frac{F_0\left(e^{h_\theta/\sqrt{n}} e^{-t\tilde{\psi}/\sqrt{n}} e^{-t(\tilde{f}-\tilde{\psi})/\sqrt{n}}\right)}{F_0\left(e^{h_\theta/\sqrt{n}} e^{-t\tilde{\psi}/\sqrt{n}}\right)} \right] \\ &\le \frac{\sqrt{n}}{t} \log \left[1 + C\frac{t}{\sqrt{n}} \sum \Delta z_j^2 \eta_j \right] \\ &\le C(\sum_j \Delta z_j^3)^{1/2}. \end{aligned}$$

We deduce:

$$\begin{aligned} \sup_{(k,\theta)\in A_n} F_0((\psi_{k,\theta} - \bar{\psi}_{h,n})^2) &= O(\sum_j \Delta z_j^3) = O((\epsilon_n \log n)^2), \\ \sup_{(k,\theta)\in A_n} [\|\Delta_{k,\theta}\|_\infty + |b_{k,\theta}|] &= O((\epsilon_n \log n)^{2/3}). \end{aligned} \tag{C.3}$$

So Condition 3 of Proposition A.1 is verified and

$$\rho_n(\theta) = -tF_0(h_\theta(\tilde{\psi} - \tilde{f})) + tG_n(\tilde{\psi} - \tilde{f}) + o_{\mathbb{P}_0}(1).$$

We have $|F_0(h_\theta(\tilde{\psi} - \tilde{f}))| \le \sqrt{n} u_n \|\tilde{\psi} - \tilde{f}\|_{f_0}$ which from (C.3) is of order $O(\sqrt{n} u_n^2) = o(1)$ uniformly over A_n. We now prove that $G_n(\bar{\psi}_{h,n} - \psi_{k,\theta}) = o_{\mathbb{P}_0}(1)$ uniformly. Note that $\psi_{k,\theta}$ is monotone, so that $[\psi_{k,\theta} - \tilde{\psi}]\epsilon_n^{-2/3}$ belongs to the class $\mathcal{G}$ of functions that are of variation bounded by 1, with values in $[-1, 1]$. Hence from [25] p 273, since $\sup_{(k,\theta)\in A_n} \|\psi_{k,\theta}\|_\infty = O(1)$, the bracketing entropy of $\mathcal{G}$ is bounded as follows

$$\log N_{[]}(\epsilon, \mathcal{G}, \mathbb{L}_2(\mathbb{P}_0)) \le K\epsilon^{-1} \quad \forall \epsilon > 0$$

for some positive constant K and Theorem 19.5 of [25] leads to

$$\sup_{g\in\mathcal{G}} |G_n(g) - G(g)| = o_p(1)$$

where G is a tight centered Gaussian process in $l^\infty(\mathcal{G})$, the set of bounded functions on $\mathcal{G}$, with covariance matrix given by $\mathbb{E}_{f_0}(G(f_1)G(f_2)) = F_0(f_1 f_2) -$

$F_0(f_1)F_0(f_2)$. Thus $\mathbb{P}[\sup_{g\in\mathcal{G}}|G(g)|<+\infty]=1$. Moreover note that for all n and all $(k,\theta)\in A_n$, $\epsilon_n^{-2/3}(\psi_{k,\theta}-\tilde{\psi})\in\mathcal{G}$, therefore

$$\lim_{a\to+\infty}\text{limsup}_n\mathbb{P}[\sup_{(k,\theta)\in A_n}|G(\epsilon_n^{-2/3}(\psi_{k,\theta}-\tilde{\psi}))|>a]=0,$$

and $\sup_{(k,\theta)\in A_n}|G_n(\psi_{k,\theta}-\tilde{\psi})|=o_{\mathbb{P}_0}(1)$. We finally obtain that

$$\begin{aligned} I_n &= \frac{\int_{\tilde{A}_n}\exp\left(-\frac{F_0((h_f-t\bar{\psi}_{h_f,n})^2)}{2}+G_n(h_f-t\bar{\psi}_{h_f,n})+R_n(h_f-t\bar{\psi}_{h_f,n})\right)d\pi(f)}{\int_{\tilde{A}_n}\exp\left(-\frac{F_0(h_f^2)}{2}+G_n(h_f)+R_n(h_f)\right)d\pi(f)} \\ &= \sum_{k\leq l_n}p(k|X^n)\frac{\int_{\tilde{A}_n\cap\mathcal{S}_k^2}\exp\left(-\frac{F_0(h_{T_k\theta}^2)}{2}+G_n(h_{T_k\theta})+R_n(h_{T_k\theta})\right)d\pi_k(\theta)}{\int_{\tilde{A}_n\cap\mathcal{S}_k^2}\exp\left(-\frac{F_0(h_\theta^2)}{2}+G_n(h_\theta)+R_n(h_\theta)\right)d\pi_k(\theta)}+o(1). \end{aligned}$$

Using (C.3) we obtain that for all $k\leq l_n$, $T_kA_n\subset\{\theta\in\mathcal{S}_k^2;V(f_0,f_\theta)\leq 2u_n^2,V(f_\theta,f_0)\leq 2u_n^2\}:=A_{n,2}\cap\mathcal{S}_k^2$ and vice-versa $A_{n,1}\cap\mathcal{S}_k^2:=\{\theta\in\mathcal{S}_k^2;V(f_0,f_\theta)\leq u_n^2/2,V(f_\theta,f_0)\leq u_n^2/2\}\subset T_kA_n$ and also under the prior on $\theta=(\eta,z)$, using $T_k\theta=(\eta',z)$

$$\begin{aligned} \left|\log\left(\frac{\pi_{\eta|k}(\eta')}{\pi_{\eta|k}(\eta)}\right)\right| &\leq \frac{C}{\sqrt{n}}\sum_j\eta_j^{-\alpha} \\ &\leq C\sqrt{n}\epsilon_n^2=o(1) \end{aligned}$$

as soon as $\alpha+1<2\gamma$ on $\mathcal{F}_n$. Hence using (3.3) and the last equality of the proof of Theorem 2.1 we finally obtain

$$\begin{aligned} \zeta_n(t) &:= \mathbb{E}^\pi[\exp(t\sqrt{n}(\Psi(f)-\Psi(\mathbb{P}_n)))1\!\!1_{\tilde{A}_n}(f)|X^n] \\ &= L_n(t)\times\mathbb{P}^\pi\left\{\tilde{A}_n|X^n\right\}=L_n(t)(1+o(1)) \\ &= e^{\frac{t^2F_0(\tilde{\psi}^2)}{2}}I_n(1+o(1)). \end{aligned}$$

Therefore,

$$\begin{aligned} \zeta_n(t) &\leq e^{\frac{t^2F_0(\tilde{\psi}^2)}{2}}\frac{\sum_{k=1}^{l_n}\mathbb{P}^\pi\left[A_{n,2}\cap\mathcal{S}_k^2|X^n\right]}{\sum_{k=1}^{l_n}\mathbb{P}^\pi\left[A_{n,1}\cap\mathcal{S}_k^2|X^n\right]}(1+o(1)) \\ &\leq e^{\frac{t^2F_0(\tilde{\psi}^2)}{2}}(1+o(1)) \\ \zeta_n(t) &\geq e^{\frac{t^2F_0(\tilde{\psi}^2)}{2}}\frac{\sum_{k=1}^{l_n}\mathbb{P}^\pi\left[A_{n,1}\cap\mathcal{S}_k^2|X^n\right]}{\sum_{k=1}^{l_n}\mathbb{P}^\pi\left[A_{n,2}\cap\mathcal{S}_k^2|X^n\right]}(1+o(1)) \\ &\geq e^{\frac{t^2F_0(\tilde{\psi}^2)}{2}}(1+o(1)), \end{aligned}$$

which achieves the proof of Theorem 2.2.

APPENDIX D: TECHNICAL LEMMA

In Section 3, we use at many places results of the following lemma. We use notations of Section 1.2.

LEMMA D.1. *Set $K_n = \{1, 2, \ldots, k_n\}$ with $k_n \in \mathbb{N}^*$. Assume one 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 $\theta = (\theta_\lambda)_\lambda \in \mathbb{R}^{k_n}$,*

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

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

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

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

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

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

$$\begin{aligned} \left\| \sum_{\lambda \in K_n} \theta_\lambda \phi_\lambda \right\|_\infty &\leq \sum_{\lambda \in K_n} |\theta_\lambda| \times \|\phi_\lambda\|_\infty \\ &\leq \sqrt{2} \sum_{\lambda \in K_n} |\theta_\lambda|, \end{aligned}$$

which proves (D.1). Inequality (D.2) follows from the definition of $\mathcal{B}^{\gamma}_{2,2} = W^{\gamma}$. To prove (D.3), we use the following inequality: for any x,

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

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,

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

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

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

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

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

which proves (D.1). For the second point, we just use the inclusion $\mathcal{B}^{\gamma}_{p,q}(R)\subset\mathcal{B}^{\gamma}_{2,\infty}(R)$ and

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

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

$$\begin{aligned}\left|\sum_{\lambda\notin K_n}\theta_{0\lambda}\phi_\lambda(x)\right| &\leq \sum_{j>\log_2(k_n)}\left(\sum_{k=0}^{2^j-1}\theta_{0jk}^2\right)^{\frac12}\left(\sum_{k=0}^{2^j-1}\varphi_{jk}^2(x)\right)^{\frac12}\\ &\leq Ck_n^{\frac12-\gamma},\end{aligned}$$

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

Acknowledgment. The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR), under grant BANHDITS. The authors would like to thank the anonymous Associate Editor and Referees for valuable comments and suggestions.

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 for Gaussian process priors. *Preprint*

[4] Chambaz, A. and Rousseau, J. (2008) Bounds for Bayesian order identifications. *Ann. Statist.* **36**, 938-962.

[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] Ghosal, S. and van der Vaart, A. (2007) Convergence rates of posterior distributions for non-i.i.d. observations. *Ann. Statist.* **35** 192223. MR2332274

[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 Ramamoorthi, R.V. (2003) *Bayesian nonparametrics* Springer, New York.

[12] Härdle, 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 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] Rivoirard, V. and Rousseau, J. (2010) Posterior concentration rates for infinite dimensional exponential families. *Submitted* Available on http://www.ceremade.dauphine.fr/∼rivoirar/ratesBA-rev.pdf

[20] Robert, C.P. (2007) *Bayesian Choice* (2nd edition) Springer-Verlag.

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

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

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

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

[25] Van der Vaat, A. W. (1998). *Asymptotic statistics.* Cambridge Series in Statistical and Probabilistic Mathematics, vol 3. Cambridge University Press, Cambridge.

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

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

CEREMADE
Université Paris Dauphine
Place du Maréchal de Lattre de Tassigny
75775 Paris Cedex 16
France

CEREMADE
Université Paris Dauphine,
Place du Maréchal de Lattre de Tassigny
75775 Paris Cedex 16
E-mail: Judith.Rousseau@ceremade.dauphine.fr