The Annals of Statistics
2012, Vol. 40, No. 3, 1489–1523
DOI: 10.1214/12-AOS1004

BERNSTEIN–VON MISES THEOREM FOR LINEAR FUNCTIONALS OF THE DENSITY[1]

BY VINCENT RIVOIRARD AND JUDITH ROUSSEAU

Université Paris Dauphine, and Université Paris Dauphine and CREST-ENSAE

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 in 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, [19] 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, [1] or [24].

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,

Received March 2010; revised March 2012.

[1]Supported in part by the ANR-SP Bayes grant.

MSC2010 subject classifications. 62G07, 62G20.

Key words and phrases. Bayesian nonparametric, rates of convergence, Bernstein–von Mises, adaptive estimation.

the more complex the model, the harder it is to derive Bernstein–von Mises theorems.

Semi-parametric and nonparametric 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 nonparametric or semi-parametric approaches is more recent and has been made possible mainly by the development of algorithms such as Markov chain Monte Carlo algorithms, but has grown rapidly over the past decade.

However, there is still little work on asymptotic properties of Bayesian procedures in semi-parametric models or even in nonparametric models. Most of existing works on the asymptotic posterior distributions deal with consistency or rates of concentration of the posterior. In other words, it consists of controlling objects of the form $\mathbb{P}^\pi[U_n|X^n]$, where $\mathbb{P}^\pi[\cdot|X^n]$ denotes the posterior distribution given a n-vector of observations X^n, and U_n denotes either a fixed neighborhood (consistency) or a sequence of shrinking neighborhoods (rates of concentration). However, to obtain a Bernstein–von Mises theorem, it is necessary not only to bound from above $\mathbb{P}^\pi[U_n|X^n]$, as in the studies of consistency and concentration rates of the posterior distribution, but also to determine an equivalent of $\mathbb{P}^\pi[U_n|X^n]$ for some specific types of sets U_n. This difficulty explains that there is, up to now, hardly any work on Bernstein–von Mises theorems in infinite-dimensional models. The most well-known results are negative results and are given in [7] and [8]. Recently [4] has proposed another interesting counter-example where the Bernstein–von Mises property does not hold, due to a subtle behavior of the prior which introduces a bias term in the posterior distribution. This phenomenon will be also encountered in our framework and discussed below. Some positive results are provided by [9] on the asymptotic normality of the posterior distribution of the parameter in an exponential family with increasing number of parameters. In a discrete setting, [3] derive Bernstein–von Mises results. Nice positive results are obtained in [16] 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, [5] obtains interesting conditions leading to a Bernstein–von Mises theorem on the parameter of interest, clarifying an earlier work of [26]. 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). In this paper, such a separation cannot be considered. Other differences with our paper have to be pointed out: The centering considered by [26] is based on the sieve maximum likelihood estimate, whereas priors considered by [5] are merely Gaussian in the information loss case. In Section 2.1 we describe more precisely results by [5, 26] and establish connections with ours. An alternative set of assumptions has been recently proposed in [2] in the same context as in [5]. Interesting results have also been obtained in the Gaussian white noise model, by [18] in the case of direct estimation and by [17] in the context of linear inverse problems. These are based on explicit expressions of the posterior distribution.

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, [28] 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 $F(x_0)(1 - F(x_0))/n$.

A 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, [12]. 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, $X_1, \dots, X_n$ are independent and identically distributed according to $\mathbb{P}$. We set $X^n = (X_1, \dots, X_n)$ and denote F the cumulative distribution function associated with f. Without loss of generality we assume that for any i, $X_i \in [0, 1]$, and we set

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

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

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

We also consider the classical inner product in $\mathbb{L}_2[0, 1]$, denoted $\langle \cdot, \cdot \rangle_2$, and $\|\cdot\|_2$, the associated norm. The Kullback–Leibler divergence and the Hellinger distance between two densities f_1 and f_2 will be, respectively, denoted $K(f_1, f_2)$ and $h(f_1, f_2)$. We recall that

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

In the sequel, we shall also use

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

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

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

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

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

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,

$$\Psi_{x_0}(f) = F(x_0) = \int_0^1 \mathbb{1}_{x \le x_0} f(x)\,dx, \qquad x_0 \in \mathbb{R}.$$

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 $z \in \mathbb{R}$,

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

where V_0 is the variance of $\sqrt{n}\Psi(P_n)$ under $\mathbb{P}_0$, and $\Phi_{V_0}(z)$ is the cumulative distribution function of a centered Gaussian random variable with variance V_0. Note that under this duality between the Bayesian and the frequentist behaviors, 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 2.2, we study in detail the special case of infinite-dimensional exponential families as described in the following section.

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

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

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 $1 \le p, q \le +\infty$ and $\gamma > \max(0, \frac{1}{p} - \frac{1}{2})$. We refer the reader to [13] for a good review of wavelets and Besov spaces. We just mention that the scale of Besov spaces includes Sobolev spaces, that is, $W^\gamma = \mathcal{B}^\gamma_{2,2}$. 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 $\lambda \ge 1$,

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

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

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

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

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

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

We use such expansions to build nonparametric 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 $\mathcal{F}_\infty = \bigcup_k \mathcal{F}_k \subset \mathcal{F}$ 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 nonparametric literature. See, for instance, [25]. The special case of log-spline priors has been studied by [10] and [14], whereas the prior considered by [30] is based on Legendre polynomials. The wavelet basis is treated in [14] 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 $k \in \mathbb{N}^*$,

$$\exp(-c_1 k L(k)) \le p(k) \le \exp(-c_2 k L(k)), \tag{1.4}$$

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

[*Case* (*D*)] Let $k_n^* = \lfloor k_0 n^{1/(2\beta+1)} \rfloor$, that is, 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

$$\frac{\theta_\lambda}{\sqrt{\tau_\lambda}} \overset{\text{i.i.d.}}{\sim} g, \qquad \tau_\lambda = \tau_0 \lambda^{-2\beta}, \qquad 1 \le \lambda \le k,$$

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

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

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 nonrandom. For the case (PH), Poisson prior satisfies the condition with $L(x) = \log x$, and geometric prior satisfies the condition with $L(x) = 1$.

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. In Section 2.2, we focus on infinite-dimensional exponential families. Theorem 2.2 gives the asymptotic posterior distribution of $\Psi(f)$ which can be either Gaussian or a mixture of Gaussian distributions. Corollary 2.1 illustrates positive results with respect to our purpose, but Proposition 2.1 shows that some bad phenomena may happen. Proofs of the results are given in Section 3, except for the proofs of Proposition 2.1 and Lemma 2.1 which are given in the supplementary material [23].

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 [28] or used by [26].

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 [28] since $F_0(h) \neq 0$ (in particular h depends on n and f). 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 $A^1_{u_n} = \{f \in \mathcal{F} \text{ s.t. } V(f_0, f) \le u_n^2\}$, the posterior distribution of $A^1_{u_n}$ satisfies

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

(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

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

then

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

(A3) Let

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

and for any x,

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

We have for any t,

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

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*, *as n goes to infinity*,

$$\sup_{z\in\mathbb{R}} \bigl|\mathbb{P}^\pi\bigl\{\sqrt{n}\bigl(\Psi(f) - \Psi(P_n)\bigr) \le z | X^n\bigr\} - \Phi_{F_0(\tilde{\psi}^2)}(z)\bigr| \to 0 \tag{2.4}$$

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 the assumptions. Condition (A1) concerns concentration rates of the posterior distribution, and there exists now a large literature on such results; see, for instance, [27] or [10] 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 the case of exponential families, obtaining a posterior concentration rate in terms of V has no impact on posterior concentration rates; see Section 2.2. It is also interesting to note that the loss function V is similar to the $\|\cdot\|_L$-norm considered in [5] (i.e., the norm induced by the LAN expansion associated to linear paths on $\log f$) and to the Fisher norm considered in [26]. Indeed, the proof of Theorem 2.1 gives

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

with $R_n(h_f) = o_{\mathbb{P}_0}(1)$ pointwise (i.e., for a fixed function f). This condition is thus to be related to Condition **C** in [5] and to Condition (9) in [26]. However, the formulation of Condition (9) in [26] is not quite as general as Condition **C** in [5], or as our conditions, since [26] also requires (stated in our framework)

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

Indeed, a concern of [26] 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, e.g., [11]) 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 $V(f, f_0) \le u_n^2$, 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 [26] [see Condition (14)]. We emphasize two major differences between Shen's condition [26] 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 $\psi(P_n)$. Second and more importantly, Condition (14) in [26] is expressed in terms of the conditional prior distribution of f given $\theta = \Psi(f)$, which is very difficult to control in most nonparametric 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 example below. Interestingly, this condition can also be found in [5], but in a less explicit way. Indeed, in [5], 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 nonparametric 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 [5]) 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 relevant. 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 a sequence of subsets $\mathcal{F}_k$ such that $\mathcal{F} = \bigcup_k \mathcal{F}_k$ and $\mathcal{F}_k = \{f_\theta \text{ s.t. } \theta \in \Theta_k\}$ with $\Theta_k \subset \mathbb{R}^{r_k}$, and (r_k) is an increasing sequence going to infinity. A prior on $\mathcal{F}$ is then defined as a probability on k, say $p(\cdot)$, and given k a probability on θ, say π_k. This set up is quite general, and it includes in particular random histograms, free-knot splines, mixtures models with random numbers of components and our example about exponential families; see Section 2.2. 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. In other words, for each k such that $A_n \cap \Theta_k$ is nonempty, construct a map $T_k : A_n \cap \Theta_k \to \Theta_k$ and 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 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

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

Note that

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

Relation (2.3) of assumption (A3) is then verified if uniformly over k such that $\mathcal{F}_k \cap \tilde{A}_n \neq \varnothing$

$$\begin{aligned}&\frac{\int_{\tilde{A}_n\cap\Theta_k} \exp(-F_0(h^2_{T_k\theta})/2 + G_n(h_{T_k\theta}) + R_n(h_{T_k\theta}))e^{-\rho_{n,k}(\theta)}\pi_k(\theta)\,d\theta}{\int_{\tilde{A}_n\cap\Theta_k} \exp(-F_0(h^2_\theta)/2 + G_n(h_\theta) + R_n(h_\theta))\pi_k(\theta)\,d\theta} \\ &\quad = 1 + o_{\mathbb{P}_0}(1).\end{aligned} \tag{2.5}$$

Proposition A.1 in the Appendix states that under mild conditions uniformly over $\bigcup_{k\le l_n} \Theta_k \cap \tilde{A}_n$,

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

where $\Delta_{k,\theta}$ is the difference $\bar{\psi}_{h,n} - \psi_{k,\theta}$ up to an additive constant, that is, there exists a constant $b_{k,\theta} \in \mathbb{R}$ such that

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

The function $\psi_{k,\theta}$ is related to the approximation γ_n of the least favorable direction considered in [5].

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 $T_k(\tilde{A}_n \cap \Theta_k)$ so that the key (sufficient) condition to verify (A3) is $\rho_{n,k} = o_{\mathbb{P}_0}(1)$, which is implied by

$$F_0(h_\theta\Delta_{k,\theta}) = o(1) \quad \text{and} \quad F_0\big((\bar{\psi}_{h,n} - \psi_{k,\theta})^2\big) = o(1) \tag{2.6}$$

uniformly over $\tilde{A}_n$; see Section 3. Condition (2.6) 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*.

The following family of examples shows the importance of this *no bias condition*.

2.2. *Bernstein–von Mises for exponential families.* In this section, we consider the nonparametric models (priors) defined in Section 1.2. Assume that f_0 is 1-periodic and $f_0 \in \mathbb{L}^2[0, 1]$. 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

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

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 $\langle\cdot,\cdot\rangle_{f_0}$ 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.

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

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 that

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

Let $(\varepsilon_n)_n$ 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 $\mathbb{P}_p(k > \frac{l_0 n\varepsilon_n^2}{L(n)}) \le e^{-an\varepsilon_n^2}$. Following, for instance, [11], page 221, it implies that there exists $c > 0$ and l_0 large enough such that

$$\mathbb{P}_0\biggl[\mathbb{P}^\pi\biggl(k > \frac{l_0 n\varepsilon_n^2}{L(n)}\Big| X^n\biggr) \le e^{-cn\varepsilon_n^2}\biggr] = 1 + o(1).$$

Now, setting $l_n = l_0 n\varepsilon_n^2/L(n)$, we have the following result.

THEOREM 2.2. *We consider the prior (PH) defined in Definition 1.1. We assume that* $\|\log(f_0)\|_\infty < \infty$ *and* $\log(f_0) \in \mathcal{B}_{p,q}^{\gamma}$ *with* $p \ge 2$, $1 \le q \le \infty$ *and* $\gamma > 1/2$ *is such that*

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

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

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

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

$$\lim_{n\to+\infty} \max_{k \le l_n} \sup_{\theta \in B_k} \frac{\pi_k(\theta)}{\pi_k(\theta - t\tilde{\psi}_\Pi^{[k]}/\sqrt{n})} = 1 \tag{2.7}$$

and

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

Then, for all $z \in \mathbb{R}$,

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

where:

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

REMARK 1. A stronger result holds for the case (D). Indeed, by replacing $k \le l_n$ with $k = k_n^*$ in (2.7) and (2.8), where k_n^* is defined in Definition 1.1, and by assuming that

$$\sum_{\lambda > k_n^*} \tilde{\psi}_\lambda^2 = o(n^{2\gamma/(2\beta+1)-1}), \tag{2.10}$$

then we have

$$\sup_{z\in\mathbb{R}} \left|\mathbb{P}^\pi[\sqrt{n}(\Psi(f) - \Psi(P_n)) \le z|X^n] - \Phi_{V_0}(z)\right| = o_{\mathbb{P}_0}(1), \tag{2.11}$$

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

The proof of Theorem 2.2 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 (2.7) and (2.8) 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 (Proposition A.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 (2.7) 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 $p = q = 2$.

LEMMA 2.1. *Assume that* $\log(f_0) \in W^\gamma$. *We still assume that* $\beta > 1/2$ *and* $\gamma > 1/2$. *Condition* (2.7) *is satisfied in the following cases*:

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

The proof is given is Section 2 of the supplementary material [23]. Lemma 2.1 holds for any bounded function ψ. For the special case $\psi(x) = \mathbb{1}_{x \le x_0}$, conditions on γ and β can be relaxed. In particular, in the case (PH), if g is the Laplace density, (2.7) 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 2.2. 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 (2.8) 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. Indeed, (3.12) gives $\sqrt{n}u_n^2 = \sqrt{n}\varepsilon_n^2(\log n)^3$ which goes to 0 with n, so that condition (2.8) is quite mild. It requires some minimal smoothness on ψ through the decay to zero of its coefficients. Note that we require $\varepsilon_n = o(n^{-1/4})$, 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, e.g., [29], Section 25.8) and is also required in [5] in the Cox model example (i.e., with information loss).

Conditions (2.7) and (2.8) 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 $F_0(\Delta_k^2) \ne o(1)$ 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 [6]), 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 instead of a nonparametric Bernstein–von Mises theorem. However, even if $F_0(\Delta_k^2) = o(1)$, the posterior distribution might not satisfy the nonparametric Bernstein–von Mises property since then $\mu_{n,k} = \sqrt{n}F_0(\Delta_k \sum_{\lambda \ge k+1} \theta_{0\lambda}\phi_\lambda) + o_{\mathbb{P}_0}(1)$, which might not be a $o_{\mathbb{P}_0}(1)$ [note that

$\sqrt{n}F_0(\Delta_k \sum_{\lambda\ge k+1}\theta_{0\lambda}\phi_\lambda)$ corresponds to $F_0(h_\theta \Delta_{k,\theta})$ in (2.6), so that the first part of (2.6) leads to $\mu_{n,k} = o_{\mathbb{P}_0}(1)$ in the setup of exponential families]. The term $\mu_{n,k}$ is a bias term in the posterior distribution. It is related to the term $\gamma_n - \gamma$ in [5] 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 $\sum_{\lambda\ge k+1}\theta_{0\lambda}^2 \le \varepsilon_n^2$, which can be interpreted as a no bias condition. Indeed $|\mu_{n,k}| \le C\sqrt{n}(\sum_{\lambda>k}\tilde{\psi}_\lambda^2)^{1/2}(\sum_{\lambda>k}\theta_{0\lambda}^2)^{1/2}$. 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 $(\sum_{\lambda>k}\tilde{\psi}_\lambda^2)^{1/2}$ is larger than $[\sqrt{n}(\sum_{\lambda>k}\theta_{0\lambda}^2)^{1/2}]^{-1}$. In general we cannot assess a lower bound on k for which $\sum_{\lambda>k}\theta_{0\lambda}^2 \le \varepsilon_n^2$ unless we assume some extra conditions on the behavior on 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 nonvalidity 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 (2.10). 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 (2.8) on $\tilde{\psi}$]. Conditions (2.10), $k = k_n^*$ and the constraints on β on the contrary, force the prior to approximate correctly functions that are potentially less smooth.

We illustrate this issue in the special case of the cumulative distribution function calculated at a given point $x_0 : \psi(x) = \mathbb{1}_{x\le x_0}$. 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 2.1. *Assume that ψ is a piecewise constant function. Consider the class of sieve priors defined in Definition 1.1 in the case (D) where g is either the Gaussian or the Laplace density. Then if $f_0 \in W^\gamma$, with $\gamma \ge \beta > 1/2$, 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 $\gamma \ge \beta > 1$, 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, 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 $\sqrt{n}(F(x_0) - F_n(x_0))$ is a mixture of Gaussian distributions with means bounded from below by $\tilde{c}\sqrt{\log n}$ for some positive constant $\tilde{c}$ and variance $F_0(x_0)(1 - F_0(x_0))$ 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 by

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

see the proof of Proposition 2.1. It does not exclude the fact that the posterior distribution of $\sqrt{n}(F(x_0) - F_n(x_0))$ could be asymptotically Gaussian, but even if it were, it would not have mean equal to zero, and therefore, the Bernstein–von Mises property is not valid.

2.3. *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 $fe^{-t\bar{\psi}_f/\sqrt{n}}$, which does not modify much 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 $fe^{-t\bar{\psi}_f/\sqrt{n}}$, for all f in a neighborhood of f_0. This 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 such as (PH), it is thus adapted to the smoothness of f_0, which might not be the same as the smoothness of $fe^{-t\bar{\psi}_f/\sqrt{n}}$ or even $f_0e^{-t\bar{\psi}_{f_0}/\sqrt{n}}$. In the case of Gaussian priors, as considered in [5], this implies that the prior is not too smooth so that $fe^{-t\bar{\psi}_f/\sqrt{n}}$ 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. 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 $t \in \mathbb{R}$,

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

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

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

We set for any x,

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

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

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

So, using (3.4),

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

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

$$t\sqrt{n}\big(\Psi(f) - \Psi(P_n)\big) = -tG_n(\tilde{\psi}) + t\sqrt{n}\bigg(\int \tilde{\psi}(x)\big(f(x) - f_0(x)\big)\,dx\bigg)$$
$$= -tG_n(\tilde{\psi}) + tF_0(h\tilde{\psi}) + \frac{t}{\sqrt{n}}F_0\big(h^2 B_{h,n}\tilde{\psi}\big).$$

Since

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

we have

$$\begin{aligned}
L_n(t) = \int_{A_n} \exp\bigg(&G_n(h - t\tilde{\psi}) + tF_0(h\tilde{\psi}) \\
&+ \frac{t}{\sqrt{n}}F_0\big(h^2 B_{h,n}\tilde{\psi}\big) - \frac{F_0(h^2)}{2} + R_n(h)\bigg)\,d\pi(f) \\
&\times \bigg(\int_{A_n} \exp\bigg(-\frac{F_0(h^2)}{2} + G_n(h) + R_n(h)\bigg)\,d\pi(f)\bigg)^{-1} \\
= \int_{A_n} \exp\bigg(&-\frac{F_0((h - t\bar{\psi}_{h,n})^2)}{2} \\
&+ G_n(h - t\bar{\psi}_{h,n}) + R_n(h - t\bar{\psi}_{h,n}) + U_{h,n}\bigg)\,d\pi(f) \\
&\times \bigg(\int_{A_n} \exp\bigg(-\frac{F_0(h^2)}{2} + G_n(h) + R_n(h)\bigg)\,d\pi(f)\bigg)^{-1},
\end{aligned}$$

where straightforward computations show that

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

Now let us expand the second term of the last expression. The first and third terms of the expression will then cancel out with this expansion. Using

$$\|\tilde{\psi}\|_\infty \le 2\|\psi\|_\infty < \infty, \tag{3.6}$$

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

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

we obtain

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

Also,

$$\begin{aligned} F_0[e^{h/\sqrt{n}}\tilde{\psi}] &= \frac{F_0[h\tilde{\psi}]}{\sqrt{n}} + \frac{F_0[h^2 B_{h,n}\tilde{\psi}]}{n}; \\ F_0[e^{h/\sqrt{n}}\tilde{\psi}^2] &= F_0[\tilde{\psi}^2] + \frac{F_0[h\tilde{\psi}^2]}{\sqrt{n}} + \frac{F_0[h^2 B_{h,n}\tilde{\psi}^2]}{n}. \end{aligned}$$

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

$$\begin{aligned} F_0\bigl(h^2 B_{h,n}\bigr) &\le nF_0\biggl[\biggl(\log\biggl(\frac{f}{f_0}\biggr)\biggr)^2\biggr] + nF\biggl[\biggl|\log\biggl(\frac{f}{f_0}\biggr)\biggr|\biggr] \\ &\le nu_n^2 + n\tilde{u}_n \end{aligned}$$

so, $F_0(h^2 B_{h,n}) = o(n)$. So, uniformly on A_n, since $\tilde{\psi}$ is bounded [see (3.6)],

$$\begin{aligned} &F_0\biggl[\exp\biggl(\frac{h}{\sqrt{n}} - \frac{t\tilde{\psi}}{\sqrt{n}}\biggr)\biggr] \\ &\quad = 1 - \frac{t}{\sqrt{n}}\biggl(\frac{F_0[h\tilde{\psi}]}{\sqrt{n}} + \frac{F_0[h^2 B_{h,n}\tilde{\psi}]}{n}\biggr) \\ &\qquad + \frac{t^2}{2n}\biggl(F_0[\tilde{\psi}^2] + \frac{F_0[h\tilde{\psi}^2]}{\sqrt{n}} + \frac{F_0[h^2 B_{h,n}\tilde{\psi}^2]}{n}\biggr) + o\bigl(n^{-1}\bigr) \\ &\quad = 1 - \frac{t}{n}\biggl[F_0[h\tilde{\psi}] + \frac{F_0[h^2 B_{h,n}\tilde{\psi}]}{\sqrt{n}} - \frac{tF_0(\tilde{\psi}^2)}{2} + o(1)\biggr] \\ &\quad = 1 + o\bigl(n^{-1/2}\bigr) \end{aligned} \tag{3.7}$$

and

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

Finally,

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

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

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

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

3.2. *Proof of Theorem* 2.2. 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:

- $\varepsilon_n = \varepsilon_0 n^{-\gamma/(2\gamma+1)} (\log n)^{\gamma/(2\gamma+1)}$ and $l_n = \frac{l_0 n \varepsilon_n^2}{L(n)}$ in the case (PH);
- $\varepsilon_n = \varepsilon_0 (\log n)^{1_{\{\gamma \geq \beta\}}} n^{-(\beta \wedge \gamma)/(2\beta+1)}$ and $l_n = k_n^* = k_0 n^{1/(2\beta+1)}$ in the case (D).

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

$$(\log n)^a l_n \varepsilon_n^2 = o(1). \tag{3.8}$$

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

$$\begin{aligned} \mathbb{P}_\pi\left[\left(\bigcup_{k \leq l_n} \mathcal{F}_k\right)^c\right] &= \sum_{\lambda > l_n} p(\lambda) \\ &\leq C \exp(-c_2 l_n L(l_n)) = o(e^{-cn\varepsilon_n^2}) \end{aligned} \tag{3.9}$$

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

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

$$A_n = \left\{\theta \in \bigcup_{k \leq l_n} \mathbb{R}^k \text{ s.t. } \|\theta - \theta_0\|_{\ell_2} \leq \frac{2(\log n)^{3/2} \varepsilon_n}{L(n)^{1/2}}\right\}, \qquad \tilde{A}_n = \{f_\theta \text{ s.t. } \theta \in A_n\}$$

and note that from Theorem B.1

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

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 $\theta \in A_n$, we have

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

Note that for $\theta \in A_n$, by using (D.1), (D.3) and (3.8),

$$\left\|\sum_{\lambda=1}^{+\infty}(\theta_{0\lambda} - \theta_\lambda)\phi_\lambda\right\|_\infty \le C\sqrt{l_n}\|\theta - \theta_0\|_{\ell_2} + C l_n^{1/2-\gamma} = o(1). \tag{3.10}$$

Therefore for $\theta \in A_n$,

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

and for n large enough,

$$\bigl|c(\theta) - c(\theta_0)\bigr| \le 2\|f_0\|_2\|\theta - \theta_0\|_{\ell_2}. \tag{3.11}$$

This implies that on A_n,

$$V(f_0, f_\theta) = O\bigl(\varepsilon_n^2(\log n)^3/L(n)\bigr). \tag{3.12}$$

Thus (A1) is verified with $u_n^2 = u_0^2\varepsilon_n^2(\log n)^3/L(n)$ and u_0 large enough. To establish (A2), we observe that we have on $\tilde{A}_n$,

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

by using (3.10). So, on $\tilde{A}_n$,

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

and (A2) is implied by (A1). Conditions (1) and (2) of Proposition A.1 are also true.

Now, let us study the validity of (A3): For any t, we study the term

$$I_n = \int_{\tilde{A}_n} \exp\left(-\frac{F_0((h_f - t\bar{\psi}_{h,n})^2)}{2} + G_n(h_f - t\bar{\psi}_{h,n}) + R_n(h_f - t\bar{\psi}_{h,n})\right) d\pi(f) \times \left(\int_{\tilde{A}_n} \exp\left(-\frac{F_0(h_f^2)}{2} + G_n(h_f) + R_n(h_f)\right) d\pi(f)\right)^{-1}.$$

We introduce

$$J_{k,n} := \int_{\tilde{A}_n \cap \mathcal{F}_k} \exp\left(-\frac{F_0((h_f - t\bar{\psi}_{h,n})^2)}{2} + G_n(h_f - t\bar{\psi}_{h,n}) + R_n(h_f - t\bar{\psi}_{h,n})\right) d\pi_k(f) \times \left(\int_{\tilde{A}_n \cap \mathcal{F}_k} \exp\left(-\frac{F_0(h_f^2)}{2} + G_n(h_f) + R_n(h_f)\right) d\pi_k(f)\right)^{-1},$$

so that

$$I_n = \frac{\sum_k J_{k,n} \mathbb{P}^\pi[\tilde{A}_n \cap \mathcal{F}_k | X^n]}{\sum_k \mathbb{P}^\pi[\tilde{A}_n \cap \mathcal{F}_k | X^n]}. \tag{3.13}$$

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

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

then define

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

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

so that $f_{T_k\theta} = f_\theta e^{-t\psi_{k,\theta}/\sqrt{n}}$. The function $\psi_{k,\theta}$ can then be understood as the projection of $\bar{\psi}_{h,n}$ on the first k components of ϕ_λ with the constraint that $f_\theta e^{-t\psi_{k,\theta}/\sqrt{n}}$ is indeed a probability density. Then we define $\Delta_{k,\theta} = \tilde{\psi} - \Pi_{f_0,k}\tilde{\psi}$ and $b_{k,\theta}$ satisfying $\bar{\psi}_{h,n} - \psi_{k,\theta} = \Delta_{k,\theta} - b_{k,\theta}$. Straightforward computations show that

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

To emphasize the fact that $\Delta_{k,\theta}$ does not depend on θ, we write hereafter $\Delta_k := \Delta_{k,\theta}$. The following lemma controls the terms H_n, $b_{k,\theta}$ and Δ_k:

LEMMA 3.1. *We have*

$$
\|H_n\|_\infty = o(1), \qquad \|\Delta_k\|_\infty = o\left(\frac{1}{\sqrt{n}u_n^2}\right), \qquad |b_{k,\theta}| = o(1)
$$

uniformly on k such that $\mathcal{F}_k \cap \tilde{A}_n \neq \varnothing$.

The proof is given in Appendix C.1.

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 A_n, and condition (3) of Proposition A.1 is satisfied with $w_n = 1$ for any n, which implies that

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

Since Δ_k is orthogonal to any $\phi_\lambda, \lambda \le k$ including $\phi_0 = 1$, we obtain, using the expression of h_θ in exponential families,

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

which is independent of θ. Also,

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

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

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

and for all k such that $\mathcal{F}_k \cap \tilde{A}_n \neq \varnothing$,

$$J_{k,n} = e^{-t^2 F_0(\Delta_k^2)/2} e^{-t\mu_{n,k}} \times \left(\int_{\mathbb{R}^k \cap A_n} e^{-F_0(h_{T_k\theta}^2)/2 + G_n(h_{T_k\theta}) + R_n(h_{T_k\theta})} \, d\pi_k(\theta) \times \left(\int_{\mathbb{R}^k \cap A_n} e^{-F_0(h_\theta^2)/2 + G_n(h_\theta) + R_n(h_\theta)} \, d\pi_k(\theta) \right)^{-1} \right) \big(1 + o(1)\big).$$

In the following lemma, using the above approximation, we obtain an equivalent of I_n.

LEMMA 3.2.

$$I_n = \sum_{k=1}^{l_n} e^{-t^2 F_0(\Delta_k^2)/2} e^{-t\mu_{n,k}} \mathbb{P}^\pi\big[k|X^n\big]\big(1 + o_{\mathbb{P}_0}(1)\big).$$

The proof of Lemma 3.2 is given in Appendix C.2. Using (3.3) and the last equality of the proof of Theorem 2.1 and combining the above inequalities with (3.13), we obtain

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

which, combined with Lemma 3.2, implies 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.

APPENDIX A: A TECHNICAL RESULT

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

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*

$$\tilde{A}_n \subset A_{u_n}^1 \cap \left(\bigcup_{k \le l_n} \mathcal{F}_k \right) \cap \{ f_\theta \ s.t. \ V(f_\theta, f_0) \le w_n u_n^2 \},$$

satisfies

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

(3) *For each* k *such that* $\mathcal{F}_k \cap \tilde{A}_n \neq \varnothing$, *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)

$$\max_{k \le l_n} \sup_{\theta \in \Theta_k \cap \tilde{A}_n} F_0\big[(\bar{\psi}_{h,n} - \psi_{k,\theta})^2\big] = O(1);$$

(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*

$$\max_{k \le l_n} \sup_{\theta \in \Theta_k \cap \tilde{A}_n} |b_{k,\theta}| = o\big(u_n^{-1} w_n^{-1/2}\big)$$

and $\Delta_{k,\theta}(x)$ *is a function satisfying*

$$\max_{k \le l_n} \sup_{f \in \mathcal{F}_k \cap \tilde{A}_n} \|\Delta_{k,\theta}\|_\infty = o\big(w_n^{-1} n^{-1/2} u_n^{-2}\big).$$

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

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

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 example 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

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

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 (3.7) of the main text,

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

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

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

and

$$\begin{aligned}
R_n(h_{T_k\theta}) &= R_n(h_\theta - t\bar{\psi}_{h,n}) \\
&\quad + t\sqrt{n}F_0(\bar{\psi}_{h,n} - \psi_{k,\theta}) + \frac{t^2}{2}F_0\big((\bar{\psi}_{h,n} - \psi_{k,\theta})^2\big) \\
&\quad + tF_0\big((\bar{\psi}_{h,n} - \psi_{k,\theta})(h_\theta - t\bar{\psi}_{h,n})\big)
\end{aligned} \tag{A.1}$$

so that

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

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

$$\|\bar{\psi}_{h,n} - \psi_{k,\theta}\|_\infty = \|\Delta_{k,\theta} - b_{k,\theta}\|_\infty = o(\sqrt{n}), \tag{A.2}$$

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

where $B_{h,n}$ is defined in (3.4) in the main text. 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

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

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

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

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

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.

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

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,

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

since from condition (3)(c) 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

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

therefore,

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

Using (A.2),

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

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:

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

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, combining the bounds on S_2–S_6, we obtain

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

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

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

We finally obtain that, uniformly on $\tilde{A}_n$,

$$\begin{aligned}\rho_{n,k} &= -t F_0[\Delta_{k,\theta} h_\theta] + t G_n(\Delta_{k,\theta}) - \frac{t^2}{2} F_0\big((\bar{\psi}_{h,n} - \psi_{k,\theta})^2\big) \\ &\quad + t^2 F_0\big[(\bar{\psi}_{h,n} - \psi_{k,\theta})\bar{\psi}_{h,n}\big] + o(1),\end{aligned}$$

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 [22].

THEOREM B.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*

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

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

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

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

where in the case (*PH*),

$$\varepsilon_n = \varepsilon_0\left(\frac{\log n}{n}\right)^{\gamma/(2\gamma+1)},$$

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

$$\varepsilon_n = \varepsilon_0(\log n)n^{-\beta/(2\beta+1)} \qquad \text{if } \gamma \geq \beta,$$
$$\varepsilon_n = \varepsilon_0 n^{-\gamma/(2\beta+1)} \qquad \text{if } \gamma < \beta,$$

and ε_0 *is a given constant. We also have that there exists* $a > 0$ *such that*

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

APPENDIX C: PROOFS OF LEMMAS 3.1 AND 3.2

C.1. Proof of Lemma 3.1. Let c_0 be a positive lower bound of f_0. We then have using (D.1),

$$\|D_{n,k,t}\|_\infty \leq \frac{t\sqrt{k}}{\sqrt{n}}\|\tilde{\psi}_\Pi^{[k]}\|_{\ell_2} \leq \frac{t\sqrt{k}}{\sqrt{n}}\|\Pi_{f_0,k}\tilde{\psi}\|_2 \leq \frac{t\sqrt{k}}{\sqrt{c_0}\sqrt{n}}\|\Pi_{f_0,k}\tilde{\psi}\|_{f_0}$$
$$\leq \frac{t\sqrt{k}}{\sqrt{c_0}\sqrt{n}}\|\tilde{\psi}\|_{f_0} \leq \frac{t\sqrt{l_n}}{\sqrt{c_0}\sqrt{n}}\|\tilde{\psi}\|_\infty = O(\varepsilon_n).$$

Since $\|\tilde{\psi}\|_\infty = O(1)$ and since on A_n,

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

$\|H_n\|_\infty = o(1)$.

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 (2.8).

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

$$\begin{aligned}
F_0\left(e^{H_n + t\Delta_k/\sqrt{n}}\right) &= F_0\left(e^{H_n}\left(1 + \frac{t\Delta_k}{\sqrt{n}}\right)\right) + 0\left(\frac{F_0(\Delta_k^2)}{n}\right) \\
&= F_0\left(e^{H_n}\right) + \frac{t}{\sqrt{n}}F_0\left(e^{H_n}\Delta_k\right) + o(1/\sqrt{n}).
\end{aligned}$$

Note also that from (3.7), $F_0(e^{H_n}) = 1 + o(1/\sqrt{n})$. Furthermore, since $F_0(|\Delta_k|) < \infty$, $\|\tilde{\psi}\|_\infty < \infty$ and since $n^{-1/2}\|h_\theta\|_\infty = o(1)$,

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

We thus obtain that

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

which completes the proof of Lemma 3.1.

C.2. Proof of Lemma 3.2. To prove Lemma 3.2 we first show that the prior π_k is not affected by the change of parameter $\theta \to T_k\theta$. For $k \le l_n$, $\|\tilde{\psi}_\Pi^{[k]}\|_{\ell_2} \le C$, where C does not depend on k and n. So, if we set

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

for all $\theta \in T_k(A_n)$, using Theorem B.1 and $(a+b)^2 \le 2a^2 + 2b^2$,

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

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

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

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

$$\mathbb{R}^k \cap A_{n,1} \subset T_k(A_n) \subset \mathbb{R}^k \cap A_{n,2} \tag{C.2}$$

with

$$A_{n,1} = \{\theta \in A_n \text{ s.t. } \|\theta - \theta_0\|_{\ell_2} \leq 1.5(\log n)^{3/2} L(n)^{-1/2}\varepsilon_n\},$$
$$A_{n,2} = \{\theta \text{ s.t. } \|\theta - \theta_0\|_{\ell_2} \leq \sqrt{5}(\log n)^{3/2} L(n)^{-1/2}\varepsilon_n\}.$$

Therefore,

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

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

Therefore,

$$\begin{aligned} I_n &\leq \frac{\sum_{k=1}^{l_n} e^{-t^2 F_0(\Delta_k^2)/2} e^{-t\mu_{n,k}} \mathbb{P}^\pi[A_{n,2} \cap \mathbb{R}^k | X^n]}{\sum_{k=1}^{l_n} \mathbb{P}^\pi[A_{n,1} \cap \mathbb{R}^k | X^n]} (1 + o(1)) \\ &\leq \sum_{k=1}^{l_n} e^{-t^2 F_0(\Delta_k^2)/2} e^{-t\mu_{n,k}} \mathbb{P}^\pi[k | X^n](1 + o(1)). \end{aligned} \tag{C.3}$$

Actually, using the exponential rate pointed out in Theorem B.1, we have with $\mathbb{P}_0$-probability tending to 1, $\mathbb{P}^\pi[A_{n,i}|X^n] \geq 1 - e^{-nc\varepsilon_n^2}$, for some positive $c > 0$ and $i = 1, 2$. We also have that for $0 < a < 1/(2\max(\beta; \gamma) + 1)$,

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

Define the event

$$\Omega_n = \Big\{\max_{k \leq l_n} |G_n(\Delta_k)| \leq n^{-a} n\varepsilon_n^2, \mathbb{P}^\pi[A_{n,i}|X^n] \geq 1 - e^{-nc\varepsilon_n^2}, i = 1, 2\Big\},$$

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

$$I_n \geq \frac{\sum_{k=1}^{l_n} e^{-t^2 F_0(\Delta_k^2)/2} e^{-t\mu_{n,k}} \mathbb{P}^\pi[A_{n,1} \cap \mathbb{R}^k | X^n]}{\sum_{k=1}^{l_n} \mathbb{P}^\pi[A_{n,2} \cap \mathbb{R}^k | X^n]} \big(1 + o(1)\big)$$

$$\geq \sum_{k=1}^{l_n} e^{-t^2 F_0(\Delta_k^2)/2} e^{-t\mu_{n,k}} \big[\mathbb{P}^\pi[k|X^n] - \mathbb{P}^\pi[(A_{n,1})^c \cap \mathbb{R}^k | X^n]\big].$$

Now, we introduce

$$I_0 = \{k \leq l_n \text{ s.t. } \mathbb{P}^\pi[k|X^n] \geq r_n^{-1} \mathbb{P}^\pi[(A_{n,1})^c \cap \mathbb{R}^k | X^n]\},$$
$$I_1 = \{k \leq l_n \text{ s.t. } \mathbb{P}^\pi[k|X^n] < r_n^{-1} \mathbb{P}^\pi[(A_{n,1})^c \cap \mathbb{R}^k | X^n]\}$$

with $r_n = e^{-nc\varepsilon_n^2/2}$. We have

$$I_n \geq (1 - r_n) \sum_{k \in I_0} e^{-t\mu_{n,k}} e^{-t^2 F_0(\Delta_k^2)/2} \mathbb{P}^\pi[k|X^n]$$

and

$$\sum_{k \in I_1} \mathbb{P}^\pi[k|X^n] \leq r_n^{-1} \mathbb{P}^\pi[(A'_{n,1})^c | X^n] \leq e^{-nc\varepsilon_n^2/2}. \tag{C.4}$$

Moreover, on Ω_n, since $\sqrt{n}\varepsilon_n = o(n\varepsilon_n^2 n^{-a})$ for some $a > 0$ small enough,

$$|\mu_{n,k}| \leq C\left[\sqrt{n}\left(\sum_{\lambda=k+1}^{+\infty} \tilde{\psi}_\lambda^2\right)^{1/2} \left(\sum_{\lambda=k+1}^{+\infty} \theta_{0\lambda}^2\right)^{1/2} + n^{-a} n\varepsilon_n^2\right]$$
$$\leq O\big(\sqrt{n}\varepsilon_n + n^{-a} n\varepsilon_n^2\big)$$
$$\leq O\big(n^{-a} n\varepsilon_n^2\big).$$

This yields

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

Using (3.9) and (C.4), for n large enough,

$$\sum_{k \in I_0} e^{-t\mu_{n,k}} e^{-t^2 F_0(\Delta_k^2)/2} \mathbb{P}^\pi[k|X^n]$$
$$\geq e^{-3tn^{-a} n\varepsilon_n^2} \geq e^{nc\varepsilon_n^2/4} \sum_{k \in I_1} e^{-t\mu_{n,k}} e^{-t^2 F_0(\Delta_k^2)/2} \mathbb{P}^\pi[k|X^n].$$

This yields

$$I_n \geq \sum_{k=1}^{l_n} e^{-t\mu_{n,k}} e^{-t^2 F_0(\Delta_{\psi,k}^2)/2} \mathbb{P}^\pi[k|X^n]\big(1 + o(1)\big). \tag{C.5}$$

APPENDIX D: TECHNICAL LEMMA

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

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

- $\gamma > 0$, $p = q = 2$ *when* Φ *is the Fourier basis*;
- $0 < \gamma < r$, $2 \le p \le \infty$, $1 \le q \le \infty$ *when* Φ *is the wavelet basis with* r *vanishing moments*; *see* [13].

Then the following results hold:

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

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

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

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

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

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

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

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

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,

$$\begin{aligned} \Big| \sum_{\lambda \notin K_n} \theta_{0\lambda} \phi_\lambda(x) \Big| &\le \sqrt{2} \sum_{\lambda \notin K_n} |\theta_{0\lambda}| \\ &\le \sqrt{2} \Big(\sum_{\lambda \notin K_n} |\lambda|^{2\gamma} \theta_{0\lambda}^2 \Big)^{1/2} \Big(\sum_{\lambda \notin K_n} |\lambda|^{-2\gamma} \Big)^{1/2}. \end{aligned}$$

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

$$\left|\sum_{\lambda\in K_n} \theta_\lambda \phi_\lambda(x)\right| \le \left(\sum_{\lambda\in K_n} \theta_\lambda^2\right)^{1/2} \left(\sum_{\lambda\in K_n} \phi_\lambda^2(x)\right)^{1/2}$$
$$\le \|\theta\|_{\ell_2} \left(\sum_{-1\le j\le \log_2(k_n)} \sum_{k<2^j} \varphi_{jk}^2(x)\right)^{1/2}$$

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

$$\operatorname{card}\{k \in \{0, \dots, 2^j - 1\} \text{ s.t. } \varphi_{jk}(x) \ne 0\} \le 3(2A+1);$$

see [20], page 282, or [21], page 112. So, there exists c_φ depending only on φ such that

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

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

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

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

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

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

Acknowledgments. 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.

SUPPLEMENTARY MATERIAL

Bernstein–von Mises theorem for linear functionals of the density: Supplementary material (DOI: 10.1214/12-AOS1004SUPP; .pdf). The supplementary material gives the proofs of Proposition 2.1 and of Lemma 2.1.

REFERENCES

[1] BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York. MR0804611

[2] BICKEL, P. and KLEIJN, B. (2011). The semiparametric Bernstein–von Mises theorem. Preprint. Available at arXiv:1007.0179v2.

[3] BOUCHERON, S. and GASSIAT, E. (2009). A Bernstein–von Mises theorem for discrete probability distributions. *Electron. J. Stat.* **3** 114–148. MR2471588

[4] CASTILLO, I. (2012). Semiparametric Bernstein–von Mises theorem and bias, illustrated with Gaussian process priors. *Sankhyā*. To appear.

[5] CASTILLO, I. (2012). A semi-parametric Bernstein–von Mises theorem for Gaussian process priors. *Probab. Theory Related Fields* **152** 53–99.

[6] CHAMBAZ, A. and ROUSSEAU, J. (2008). Bounds for Bayesian order identification with application to mixtures. *Ann. Statist.* **36** 938–962. MR2396820

[7] COX, D. D. (1993). An analysis of Bayesian inference for nonparametric regression. *Ann. Statist.* **21** 903–923. MR1232525

[8] FREEDMAN, D. (1999). On the Bernstein–von Mises theorem with infinite-dimensional parameters. *Ann. Statist.* **27** 1119–1140. MR1740119

[9] GHOSAL, S. (2000). Asymptotic normality of posterior distributions for exponential families when the number of parameters tends to infinity. *J. Multivariate Anal.* **74** 49–68. MR1790613

[10] GHOSAL, S., GHOSH, J. K. and VAN DER VAART, A. W. (2000). Convergence rates of posterior distributions. *Ann. Statist.* **28** 500–531. MR1790007

[11] GHOSAL, S. and VAN DER VAART, A. (2007). Convergence rates of posterior distributions for non-i.i.d. observations. *Ann. Statist.* **35** 192–223. MR2332274

[12] GHOSH, J. K. and RAMAMOORTHI, R. V. (2003). *Bayesian Nonparametrics*. Springer, New York. MR1992245

[13] HÄRDLE, W., KERKYACHARIAN, G., PICARD, D. and TSYBAKOV, A. (1998). *Wavelets, Approximation, and Statistical Applications. Lecture Notes in Statistics* **129**. Springer, New York. MR1618204

[14] HUANG, T.-M. (2004). Convergence rates for posterior distributions and adaptive estimation. *Ann. Statist.* **32** 1556–1593. MR2089134

[15] KIM, Y. (2006). The Bernstein–von Mises theorem for the proportional hazard model. *Ann. Statist.* **34** 1678–1700. MR2283713

[16] KIM, Y. and LEE, J. (2004). A Bernstein–von Mises theorem in the nonparametric right-censoring model. *Ann. Statist.* **32** 1492–1512. MR2089131

[17] KNAPIK, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2011). Bayesian inverse problems. Preprint. Available at http://arxiv.org/abs/1103.2692v1.

[18] LEAHU, H. (2011). On the Bernstein–von Mises phenomenon in the Gaussian white noise model. *Electron. J. Stat.* **5** 373–404. MR2802048

[19] LECAM, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *Univ. California Publ. Statist.* **1** 277–329. MR0054913

[20] MALLAT, S. (1998). *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, CA. MR1614527

[21] MEYER, Y. (1992). *Wavelets and Operators. Cambridge Studies in Advanced Mathematics* **37**. Cambridge Univ. Press, Cambridge. MR1228209

[22] RIVOIRARD, V. and ROUSSEAU, J. (2012). Posterior concentration rates for infinite dimensional exponential families. *Bayesian Anal.* **7** 311–334.

[23] RIVOIRARD, V. and ROUSSEAU, J. (2012). Supplement to "Bernstein–von Mises theorem for linear functionals of the density." DOI:10.1214/12-AOS1004SUPP.

[24] ROBERT, C. P. (2007). *The Bayesian Choice*. *From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed. Springer, New York. MR2723361

[25] SCRICCIOLO, C. (2006). Convergence rates for Bayesian density estimation of infinite-dimensional exponential families. *Ann. Statist.* **34** 2897–2920. MR2329472

[26] SHEN, X. (2002). Asymptotic normality of semiparametric and nonparametric posterior distributions. *J. Amer. Statist. Assoc.* **97** 222–235. MR1947282

[27] SHEN, X. and WASSERMAN, L. (2001). Rates of convergence of posterior distributions. *Ann. Statist.* **29** 687–714. MR1865337

[28] VAN DER VAART, A. (2002). Semiparametric statistics. In *Lectures on Probability Theory and Statistics* (*Saint-Flour*, 1999). *Lecture Notes in Math.* **1781** 331–457. Springer, Berlin. MR1915446

[29] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

[30] VERDINELLI, I. and WASSERMAN, L. (1998). Bayesian goodness-of-fit testing using infinite-dimensional exponential families. *Ann. Statist.* **26** 1215–1241. MR1647645

CEREMADE
UNIVERSITÉ PARIS DAUPHINE
PLACE DU MARÉCHAL DE LATTRE DE TASSIGNY
75775 PARIS CEDEX 16
FRANCE
E-MAIL: Vincent.Rivoirard@ceremade.dauphine.fr
Judith.Rousseau@ceremade.dauphine.fr