The Annals of Statistics
2015, Vol. 43, No. 6, 2353–2383
DOI: 10.1214/15-AOS1336

A BERNSTEIN–VON MISES THEOREM FOR SMOOTH FUNCTIONALS IN SEMIPARAMETRIC MODELS[1]

BY ISMAËL CASTILLO[*,†] AND JUDITH ROUSSEAU[‡,§]

CNRS–LPMA[*], Universités Paris VI & VII[†], CREST–ENSAE[‡] and Université Paris Dauphine[§]

A Bernstein–von Mises theorem is derived for general semiparametric functionals. The result is applied to a variety of semiparametric problems in i.i.d. and non-i.i.d. situations. In particular, new tools are developed to handle semiparametric bias, in particular for nonlinear functionals and in cases where regularity is possibly low. Examples include the squared L^2-norm in Gaussian white noise, nonlinear functionals in density estimation, as well as functionals in autoregressive models. For density estimation, a systematic study of BvM results for two important classes of priors is provided, namely random histograms and Gaussian process priors.

1. Introduction. Bayesian approaches are often considered to be close asymptotically to frequentist likelihood-based approaches so that the impact of the prior disappears as the information brought by the data—typically the number of observations—increases. This common knowledge is verified in most parametric models, with a precise expression of it through the so-called Bernstein–von Mises theorem or property (hereafter, BvM). This property says that, as the number of observations increases the posterior distribution can be approached by a Gaussian distribution centered at an efficient estimator of the parameter of interest and with variance the inverse of the Fisher information matrix of the whole sample; see, for instance, van der Vaart [32], Berger [2] or Ghosh and Ramamoorthi [23]. The situation becomes, however, more complicated in non- and semiparametric models. Semiparametric versions of the BvM property consider the behaviour of the marginal posterior in a parameter of interest, in models potentially containing an infinite-dimensional nuisance parameter. There some care is typically needed in the choice of the nonparametric prior and a variety of questions linked to prior choice and techniques of proofs arise. Results on semiparametric BvM applicable to general models and/or general priors include Shen [31], Castillo [10], Rivoirard and Rousseau [30] and Bickel and Kleijn [3]. The variety of possible interactions between prior and model and the subtleties of prior choice are illustrated in the previous general papers and in recent results in specific models such as Kim

Received May 2013; revised April 2015.

[1]Supported in part by ANR Grant projects "Banhdits" ANR-2010-BLAN-0113-03 and "Calibration" ANR 2011 BS01 010 01.

MSC2010 subject classifications. Primary 62G20; secondary 62M15.

Key words and phrases. Bayesian nonparametrics, Bernstein–von Mises theorem, posterior concentration, semiparametric inference.

[24], De Blasi and Hjort [17], Leahu [29], Knapik et al. [26], Castillo [11] and Kruijer and Rousseau [27]. In between semi- and nonparametric results, BvM for parameters with growing dimension have been obtained in, for example, Ghosal [21], Boucheron and Gassiat [7] and Bontemps [6]. Finally, although there is no immediate analogue of the BvM property for infinite dimensional parameters, as pointed out by Cox [16] and Freedman [19], some recent contributions have introduced possible notions of nonparametric BvM; see Castillo and Nickl [13] and also Leahu [29]. In fact, the results of the present paper are relevant for these, as discussed below.

For semiparametric BvM, it is of particular interest to obtain generic sufficient conditions that do not depend on the specific form of the considered model. In this paper, we provide a general result, Theorem 2.1 in Section 2, on the existence of the BvM property for generic models and functionals of the parameter. Let us briefly discuss the scope of our results; see Section 2 for precise definitions. Consider a model parameterised by η varying in a (subset of a) metric space S equipped with a σ-field $\mathcal{S}$. Let $\psi : S \to \mathbb{R}^d$, $d \geq 1$, be a measurable functional of interest and let Π be a probability distribution on S. Given observations Y^n from the model, we study the asymptotic posterior distribution of $\psi(\eta)$, denoted $\Pi[\psi(\eta)|Y^n]$. Let $\mathcal{N}(0, V)$ denote the centered normal law with covariance matrix V. We give general conditions under which a BvM-type property is valid,

$$\Pi\big[\sqrt{n}\big(\psi(\eta) - \hat{\psi}\big)|Y^n\big] \rightsquigarrow \mathcal{N}(0, V), \tag{1.1}$$

as $n \to \infty$ in probability, where $\hat{\psi}$ is a (random) centering point, and V a covariance matrix, both to be specified, and where $\rightsquigarrow$ stands for weak convergence. An interesting and well-known consequence of BvM is that posterior credible sets, such as equal-tail credible intervals, highest posterior density regions or one-sided credible intervals are also confidence regions with the same asymptotic coverage.

The contributions of the present paper can be regrouped around the following aims:

1. Provide general conditions on the model and on the functional ψ to guarantee (1.1) to hold, in a variety of frameworks both i.i.d. and non-i.i.d. This includes investigating how the choice of the prior influences bias $\hat{\psi}$ and variance V. This also includes studying the case of nonlinear functionals, which involves specific techniques for the bias. This is done via a Taylor-type expansion of the functional involving a linear term as well as, possibly, an additional quadratic term.

2. In frameworks with low regularity, second-order properties in the functional expansion may become relevant. We study this as an application of the main theorem in the important case of estimation of the squared L^2-norm of an unknown regression function in the case where the convergence rate for the functional is still parametric but where the "plug-in" property in the sense of Bickel and Ritov [5] is not necessarily satisfied.

3. Provide simple and ready-to-use sufficient conditions for BvM in the important example of density estimation on the unit interval. We present extensions and refinements in particular of results of Castillo [10] and Rivoirard and Rousseau [30] regarding, respectively, the use of Gaussian process priors in the context of density estimation, and, the possibility to consider nonlinear functionals. The class of random density histogram priors is also studied in details systematically for the first time in the context of Bayesian semiparametrics.

4. Provide simple sufficient conditions on the prior for BvM to hold in a more complex example involving dependent data, namely the nonlinear autoregressive model. To our knowledge, this is the first result of this type in such a model.

The techniques and results of the paper, as it turned out, have also been useful for different purposes in a recent series of works developing a multiscale approach for posteriors, in particular: (a) to prove functional limiting results, such as Bayesian versions of Donsker's theorem, or more generally BvM results as in Castillo and Nickl [14], a first step consists in proving the result for finite dimensional projections: this is exactly asking for a semiparametric BvM to hold, and results from Section 4 can be directly applied; (b) related to this is the study of many functionals simultaneously: this is used in the study of posterior contraction rates in the supremum norm in Castillo [12]. Finally, along the way, we shall also derive posterior rate results for Gaussian processes which are of independent interest; see Proposition 2 in the supplemental article (Castillo and Rousseau [15]).

Our results show that the most important condition is a *no-bias* condition, which will be seen to be essentially necessary. This condition is written in a nonexplicit way in the general Theorem 2.1, since the study of such a condition depends heavily on the family of priors that are considered together with the statistical model. Extensive discussions on the implication of this no-bias condition are provided in the context of the white noise model and density models for two families of priors. In the examples, we have considered the main tool used to verify this condition consists in constructing a change of parameterisation in the form $\eta \to \eta + \Gamma/\sqrt{n}$ for some given Γ depending on the functional of interest, which leaves the prior approximately unchanged. Roughly speaking, for the no-bias condition to be valid, it is necessary that both η_0 and Γ are well approximated under the prior. If this condition is not verified, then BvM may not hold: an example of this phenomenon is provided in Section 4.3.

Theorem 2.1 does not rely on a specific type of model, nor on a specific family of functionals. In Section 3, it is applied to the study of a nonlinear functional in the white noise model, namely the squared-norm of the signal. Applications to density estimation with three different types of functionals and to an autoregressive model can be found respectively in Section 4 and Section 5. Section 6 is devoted to proofs, together with the supplemental article (Castillo and Rousseau [15]).

Model, prior and notation. Let $(\mathcal{Y}^n, \mathcal{G}^n, P_\eta^n, \eta \in S)$ be a statistical experiment, with observations Y^n sitting on a space $\mathcal{Y}^n$ equipped with a σ-field $\mathcal{G}^n$, and where n is an integer quantifying the available amount of information. We typically consider the asymptotic framework $n \to \infty$. We assume that S is equipped with a σ-field $\mathcal{S}$, that S is a subset of a linear space and that for all $\eta \in S$, the measures P_η^n are absolutely continuous with respect to a dominating measure μ_n. Denote by p_η^n the associated density and by $\ell_n(\eta)$ the log-likelihood. Let η_0 denote the true value of the parameter and $P_{\eta_0}^n$ the frequentist distribution of the observations Y^n under η_0. Throughout the paper, we set $P_0^n := P_{\eta_0}^n$ and $P_0 := P_0^1$. Similarly, $E_0^n[\cdot]$ and $E_0[\cdot]$ denote the expectation under P_0^n and P_0, respectively, and E_η^n and E_η are the corresponding expectations under P_η^n and P_η. Given any prior probability Π on S, we denote by $\Pi[\cdot|Y^n]$ the associated posterior distribution on S, given by Bayes formula: $\Pi[B|Y^n] = \int_B p_\eta^n(Y^n)\, d\Pi(\eta) / \int p_\eta^n(Y^n)\, d\Pi(\eta)$. Throughout the paper, we use the notation o_p in the place of $o_{P_0^n}$ for simplicity.

The quantity of interest in this paper is a functional $\psi : S \to \mathbb{R}^d, d \geq 1$. We restrict in this paper to the case of real-valued functionals $d = 1$, noting that the presented tools do have natural multivariate counterparts not detailed here for notational simplicity.

For η_1, η_2 in S, the Kullback–Leibler divergence between $P_{\eta_1}^n$ and $P_{\eta_2}^n$ is

$$KL(P_{\eta_1}^n, P_{\eta_2}^n) := \int_{\mathcal{Y}^n} \log\left(\frac{dP_{\eta_1}^n}{dP_{\eta_2}^n}(y^n)\right) dP_{\eta_1}^n(y^n),$$

and the corresponding variance of the likelihood ratio is denoted by

$$V_n(P_{\eta_1}^n, P_{\eta_2}^n) := \int_{\mathcal{Y}^n} \log^2\left(\frac{dP_{\eta_1}^n}{dP_{\eta_2}^n}(y^n)\right) dP_{\eta_1}^n(y^n) - KL(P_{\eta_1}^n, P_{\eta_2}^n)^2.$$

Let $\|\cdot\|_2$ and $\langle\cdot,\cdot\rangle_2$ denote respectively the L_2 norm and the associated inner product on $[0, 1]$. We use also $\|\cdot\|_1$ to denote the L_1 norm on $[0, 1]$. For all $\beta \geq 0$, $\mathcal{C}^\beta$ denotes the class of β-Hölder functions on $[0, 1]$ where $\beta = 0$ corresponds to the case of continuous functions. Let $h(f_1, f_2) = (\int_0^1 (\sqrt{f_1} - \sqrt{f_2})^2\, d\mu)^{1/2}$ stand for the Hellinger distance between two densities f_1 and f_2 relative to a measure μ. For g integrable on $[0, 1]$ with respect to Lebesgue measure, we often write $\int_0^1 g$ or $\int g$ instead of $\int_0^1 g(x)\, dx$. For two real-valued functions A, B (defined on $\mathbb{R}$ or on $\mathbb{N}$), we write $A \lesssim B$ if A/B is bounded and $A \asymp B$ if $|A/B|$ is bounded away from 0 and ∞.

2. Main result. In this section, we give the general theorem which provides sufficient conditions on the model, the functional and the prior for BvM to be valid.

We say that the posterior distribution for the functional $\psi(\eta)$ is *asymptotically normal* with centering ψ_n and variance V if, for β the bounded Lipschitz metric (also known as the Lévy–Prohorov metric) for weak convergence (see Section 1 in

the supplemental article Castillo and Rousseau [15], and τ_n the mapping $\tau_n : \eta \to \sqrt{n}(\psi(\eta) - \psi_n))$, it holds, as $n \to \infty$, that

$$\beta\big(\Pi\big[\cdot|Y^n\big] \circ \tau_n^{-1}, \mathcal{N}(0, V)\big) \to 0, \tag{2.1}$$

in P_0^n-probability, which we also denote $\Pi[\cdot|Y^n] \circ \tau_n^{-1} \rightsquigarrow \mathcal{N}(0, V)$.

In models where an efficiency theory at rate $\sqrt{n}$ is available, we say that the posterior distribution for the functional $\psi(\eta)$ at $\eta = \eta_0$ *satisfies the BvM theorem* if (2.1) holds with $\psi_n = \hat{\psi}_n + o_p(1/\sqrt{n})$, for $\hat{\psi}_n$ a linear efficient estimator of $\psi(\eta)$ and V the efficiency bound for estimating $\psi(\eta)$. For instance, for i.i.d. models and a differentiable functional ψ with efficient influence function $\tilde{\psi}_{\eta_0}$ (see, e.g., [32] Chapter 25), the efficiency bound is attained if $V = P_0^n[\tilde{\psi}_{\eta_0}^2]$. Let us now state the assumptions which will be required.

Let A_n be a sequence of measurable sets such that, as $n \to \infty$,

$$\Pi\big[A_n|Y^n\big] = 1 + o_p(1). \tag{2.2}$$

We assume that there exists a Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle_L)$ with associated norm denoted $\|\cdot\|_L$, and for which the inclusion $A_n - \eta_0 \subset \mathcal{H}$ is satisfied for n large enough. Note that we do not necessarily assume that $S \subset \mathcal{H}$, as $\mathcal{H}$ gives a local description of the parameter space near η_0 only. Note also that $\mathcal{H}$ may depend on n. The norm $\|\cdot\|_L$ typically corresponds to the LAN (locally asymptotically normal) norm as described in (2.3) below.

Let us first introduce some notation which corresponds to expanding both the log-likelihood $\ell_n(\eta) := \ell_n(\eta, Y^n)$ in the model and the functional of interest $\psi(\eta)$. Both expansions have remainders R_n and r, respectively.

LAN expansion. Write, for all $\eta \in A_n$,

$$\ell_n(\eta) - \ell_n(\eta_0) = \frac{-n\|\eta - \eta_0\|_L^2}{2} + \sqrt{n}W_n(\eta - \eta_0) + R_n(\eta, \eta_0), \tag{2.3}$$

where $[W_n(h), h \in \mathcal{H}]$ is a collection of real random variables verifying that, P_0^n-almost surely, the mapping $h \to W_n(h)$ is linear, and that for all $h \in \mathcal{H}$, we have $W_n(h) \rightsquigarrow \mathcal{N}(0, \|h\|_L^2)$ as $n \to \infty$.

Functional smoothness. Consider $\psi_0^{(1)} \in \mathcal{H}$ and a self-adjoint linear operator $\psi_0^{(2)} : \mathcal{H} \to \mathcal{H}$ and write, for any $\eta \in A_n$,

$$\begin{aligned}\psi(\eta) = \psi(\eta_0) &+ \langle\psi_0^{(1)}, \eta - \eta_0\rangle_L \\ &+ \tfrac{1}{2}\langle\psi_0^{(2)}(\eta - \eta_0), \eta - \eta_0\rangle_L + r(\eta, \eta_0),\end{aligned} \tag{2.4}$$

where there exists a positive constant C_1 such that

$$\big\|\psi_0^{(2)}h\big\|_L \le C_1\|h\|_L \qquad \forall h \in \mathcal{H} \quad \text{and} \quad \big\|\psi_0^{(1)}\big\|_L \le C_1. \tag{2.5}$$

Note that both formulations, on the functional smoothness and on the LAN expansion, are not assumptions since nothing is required yet on $r(\eta, \eta_0)$ or on

$R(\eta, \eta_0)$. This is done in Assumption A. The norm $\|\cdot\|_L$ is typically identified from a local asymptotic normality property of the model at the point η_0. It is thus intrinsic to the considered statistical model. Next, the expansion of ψ around η_0 is in term of the latter norm: since this norm is intrinsic to the model, this can be seen as a canonical choice.

Consider two cases, depending on the value of $\psi_0^{(2)}$ in (2.4). The first case corresponds to a first-order analysis of the problem. It ignores any potential nonlinearity in the functional $\eta \to \psi(\eta)$ by considering a linear approximation with representer $\psi_0^{(1)}$ in (2.4) and shifting any remainder term into r.

Case A1. We set $\psi_0^{(2)} = 0$ in (2.4) and, for all $\eta \in A_n$ and $t \in \mathbb{R}$ define

$$\eta_t = \eta - \frac{t\psi_0^{(1)}}{\sqrt{n}}. \tag{2.6}$$

Case A2. We allow for a nonzero second-order term $\psi_0^{(2)}$ in (2.4). In this case, we need a few more assumptions. One is simply the existence of some posterior convergence rate in $\|\cdot\|_L$-norm. Suppose that, for some sequence $\varepsilon_n = o(1)$ and A_n as in (2.2),

$$\Pi\big[\eta \in A_n; \|\eta - \eta_0\|_L \le \varepsilon_n/2 | Y^n\big] = 1 + o_p(1). \tag{2.7}$$

Next, we assume that the action of the process W_n above can be approximated by an inner-product, with a representer w_n, which will be particularly useful in defining a suitable path η_t enabling to handle second-order terms.

Suppose that there exists $w_n \in \mathcal{H}$ such that, for all $h \in \mathcal{H}$,

$$W_n(h) = \langle w_n, h\rangle_L + \Delta_n(h), \qquad P_0^n\text{-almost surely}, \tag{2.8}$$

where the remainder term $\Delta_n(\cdot)$ is such that

$$\sup_{\eta \in A_n} \big|\Delta_n\big(\psi_0^{(2)}(\eta - \eta_0)\big)\big| = o_p(1) \tag{2.9}$$

and where one further assumes that

$$\big\langle w_n, \psi_0^{(2)}\big(\psi_0^{(1)}\big)\big\rangle_L + \varepsilon_n \|w_n\|_L = o_p(\sqrt{n}). \tag{2.10}$$

Finally set, for all $\eta \in A_n$ and w_n as in (2.8), for all $t \in \mathbb{R}$,

$$\eta_t = \eta - \frac{t\psi_0^{(1)}}{\sqrt{n}} - \frac{t\psi_0^{(2)}(\eta - \eta_0)}{2\sqrt{n}} - \frac{t\psi_0^{(2)} w_n}{2n}. \tag{2.11}$$

ASSUMPTION A. In cases A1 and A2, with η_t defined by (2.6) and (2.11), respectively, assume that for all $t \in \mathbb{R}$, $\eta_t \in S$ for n large enough and that

$$\sup_{\eta \in A_n} \big|t\sqrt{n} r(\eta, \eta_0) + R_n(\eta, \eta_0) - R_n(\eta_t, \eta_0)\big| = o_p(1). \tag{2.12}$$

The suprema in the previous display may not be measurable, in this case one interprets the previous probability statements in terms of outer measure.

We then provide a characterisation of the asymptotic distribution of $\psi(\eta)$. At first read, one may set $\psi_0^{(2)} = 0$ in the next theorem: this provides a first-order result that will be used repeatedly in Sections 4 and 5. The complete statement allows for a second-order analysis via a possibly nonzero $\psi_0^{(2)}$ and will be applied in Section 3.

THEOREM 2.1. *Consider a statistical model* $\{P_\eta^n, \eta \in S\}$, *a real-valued functional* $\eta \to \psi(\eta)$ *and* $\langle \cdot, \cdot \rangle_L, \psi_0^{(1)}, \psi_0^{(2)}, W_n, w_n$ *as defined above. Suppose that Assumption* A *is satisfied, and denote*

$$\hat{\psi} = \psi(\eta_0) + \frac{W_n(\psi_0^{(1)})}{\sqrt{n}} + \frac{\langle w_n, \psi_0^{(2)} w_n \rangle_L}{2n}, \qquad V_{0,n} = \left\| \psi_0^{(1)} - \frac{\psi_0^{(2)} w_n}{2\sqrt{n}} \right\|_L^2.$$

Let Π *be a prior distribution on* η. *Let* A_n *be any measurable set such that* (2.2) *holds. Then for any real* t *with* η_t *as in* (2.11),

$$E^{\Pi}\big[e^{t\sqrt{n}(\psi(\eta)-\hat{\psi})} | Y^n, A_n\big] = e^{o_p(1)+t^2 V_{0,n}/2} \frac{\int_{A_n} e^{\ell_n(\eta_t)-\ell_n(\eta_0)}\, d\Pi(\eta)}{\int_{A_n} e^{\ell_n(\eta)-\ell_n(\eta_0)}\, d\Pi(\eta)}. \tag{2.13}$$

Moreover, if $V_{0,n} = V_0 + o_p(1)$ *for some* $V_0 > 0$ *and if for some possibly random sequence of reals* μ_n, *for any real* t,

$$\frac{\int_{A_n} e^{\ell_n(\eta_t)-\ell_n(\eta_0)}\, d\Pi(\eta)}{\int_{A_n} e^{\ell_n(\eta)-\ell_n(\eta_0)}\, d\Pi(\eta)} = e^{\mu_n t}\big(1 + o_p(1)\big), \tag{2.14}$$

then the posterior distribution of $\sqrt{n}(\psi(\eta) - \hat{\psi}) - \mu_n$ *is asymptotically normal and mean-zero, with variance* V_0.

The proof of Theorem 2.1 is given in Section 6.1.

COROLLARY 1. *Under the conditions of Theorem* 2.1, *if* (2.14) *holds with* $\mu_n = o_p(1)$ *and* $\|\psi_0^{(2)} w_n\|_L = o_p(\sqrt{n})$, *then the posterior distribution of* $\sqrt{n}(\psi(\eta) - \hat{\psi})$ *is asymptotically mean-zero normal, with variance* $\|\psi_0^{(1)}\|_L^2$.

Assumption A ensures that the local behaviour of the likelihood resembles the one in a Gaussian experiment with norm $\|\cdot\|_L$. An assumption of this type is expected, as the target distribution in the BvM theorem is Gaussian. As will be seen in the examples in Sections 3, 4 and 5, A_n is often a well chosen subset of a neighbourhood of η_0, with respect to a given metric, which need not be the LAN norm $\|\cdot\|_L$.

We note that for simplicity here we restrict to approximating paths η_t to η_0 in (2.6) (first-order results) and (2.11) (second-order results) that are linear in the

perturbation. This covers already quite a few interesting models. More generally, some models may be locally *curved* around η_0, with a possibly nonlinear form of approximating paths. A more general statement would possibly have an extra condition to control the curvature. Examining this type of example is left for future work.

The central condition for applying Theorem 2.1 is (2.13). To check this condition, a possible approach is to construct a change of parameter from η to η_t (or some parameter close enough to η_t), which leaves the prior and A_n approximately unchanged. More formally, let ψ_n be an approximation of $\psi_0^{(1)}$ in a sense to be made precise below and let $\Pi^{\psi_n} := \Pi \circ (\tau^{\psi_n})^{-1}$ denote the image measure of Π through the mapping

$$\tau^{\psi_n} : \eta \to \eta - t\psi_n/\sqrt{n}.$$

To check (2.13), one may for instance suppose that the measures Π^{ψ_n} and Π are mutually absolutely continuous and that the density $d\Pi/d\Pi^{\psi_n}$ is close to the quantity $e^{\mu_n t}$ on A_n. This is the approach we follow for various models and priors in the sequel. In particular, we prove that a functional change of variable is possible for various classes of prior distributions. For instance, in density estimation, Gaussian process priors and piecewise constant priors are considered and Propositions 1 and 3 below give a set of sufficient conditions that guarantee (2.13) for each class of priors.

In general, the construction of a *feasible* change of parameterisation heavily depends on the structure of the prior model. We note that this change of parameter approach above only provides a sufficient condition. For some priors, shifted measures may be far from being absolutely continuous, even using approximations of the shifting direction: for such priors, one may have to compare the integrals directly.

REMARK 1. Here, the main focus is on estimation of abstract semiparametric functionals $\psi(\eta)$. Our results also apply to the case of separated semiparametric models where $\eta = (\psi, f)$ and $\psi(\eta) = \psi \in \mathbb{R}$, as considered in [10], with a weak convergence to the normal distribution instead of a strong convergence obtained in [10]. We have $\psi(\eta) - \psi(\eta_0) = \langle \eta - \eta_0, (1, -\gamma)\rangle_L / \tilde{I}_{\eta_0}$ where γ is the least favorable direction and $\tilde{I}_{\eta_0} = \|(1, -\gamma)\|_L^2$; see [10]. We can then choose $\psi_0^{(1)} = (1, -\gamma)/\tilde{I}_{\eta_0}$ in [10]. If $\gamma = 0$ (no loss of information), $\eta_t = (\psi - t\tilde{I}_{\eta_0}^{-1}/\sqrt{n}, f)$ and (2.13) is satisfied if $\pi = \pi_\psi \otimes \pi_f$ with π_ψ positive and continuous at $\psi(\eta_0)$, so that we obtain a similar result as Theorem 1 of [10]. In [10], a slightly weaker version of condition (2.12) is considered; however, the proof of Section 6.1 can be easily adapted—in the case of separated semiparametric models—so that the result holds under the weaker version of (2.12) as well.

REMARK 2. As follows from the proof of Theorem 2.1, $\psi_0^{(1)}$ can be replaced by any element, say $\tilde{\psi}$ of $\mathcal{H}$ such that

$$\langle \tilde{\psi}, \eta - \eta_0 \rangle_L = \langle \psi_0^{(1)}, \eta - \eta_0 \rangle_L, \qquad \|\tilde{\psi}\|_L = \|\psi_0^{(1)}\|_L,$$

where $\tilde{\psi}$ may potentially depend on η. This proves to be useful when considering constraint spaces as in the case of density estimation.

We now apply Theorem 2.1 in the cases of white noise, density and autoregressive models and for various types of functionals and priors.

3. Applications to the white noise model. Consider the model

$$dY^n(t) = f(t)\,dt + n^{-1/2}\,dB(t), \qquad t \in [0,1],$$

where $f \in L^2[0,1]$ and B is standard Brownian motion. Let $(\phi_k)_{k\geq 1}$ be an orthonormal basis for $L^2[0,1] =: L^2$. The model can be rewritten

$$Y_k = f_k + n^{-1/2}\epsilon_k, \qquad f_k = \int_0^1 f(t)\phi_k(t)\,dt, \qquad \epsilon_k \sim \mathcal{N}(0,1) \qquad \text{i.i.d.}, k \geq 1.$$

The likelihood admits a LAN expansion, with $\eta = f$ here, $\|\cdot\|_L = \|\cdot\|_2$ and $R_n = 0$:

$$\ell_n(f) - \ell_n(f_0) = -\frac{n\|f - f_0\|^2}{2} + \sqrt{n}W(f - f_0),$$

where for any $u \in L^2 = \mathcal{H}$ with coefficients $u_k = \int_0^1 u(t)\phi_k(t)\,dt$, we set $W(u) = \sum_{k\geq 1}\epsilon_k u_k$.

In this model, consider the squared-L^2 norm as a functional of f. Set

$$\psi(f) = \|f\|_2^2 = \psi(f_0) + 2\langle f_0, f - f_0\rangle_2 + \|f - f_0\|_2^2,$$
$$\psi_0^{(1)} = 2f_0, \qquad \psi_0^{(2)}h = 2h, \qquad r(f, f_0) = 0.$$

The functional has been extensively studied in the frequentist literature; see [4, 18, 20, 28] and [8] to name but a few, as it is used in many testing problems. The verification of Assumption A and of condition (2.14) is prior-dependent and is considered within the proof of the next theorem.

Suppose that the true function f_0 belongs to the Sobolev class

$$W_\beta := \left\{ f \in L^2, \sum_{k\geq 1} k^{2\beta}\langle f, \phi_k\rangle^2 < \infty \right\}$$

of order $\beta > 1/4$. First, one should note that, while the case $\beta > 1/2$ can be treated using the first-order term of the expansion of the functional only (case A1), the case $1/4 < \beta < 1/2$ requires the conditions from case A2 as the second-order term cannot be neglected. This is related to the fact that the so-called plug-in property in [5]

does not work for $\beta < 1/2$. An analysis based on second-order terms as in Theorem 2.1 is thus required. The case $\beta \leq 1/4$ is interesting too, but one obtains a rate slower than $1/\sqrt{n}$; see, for example, Cai and Low [8] and references therein, and a BvM result in a strict sense does not hold. Although a BvM-type result can be obtained essentially with the tools developed here, its formulation is more complicated and this case will be treated elsewhere. When $\beta > 1/4$, a natural frequentist estimator of $\psi(\eta)$ is

$$\bar{\psi} := \bar{\psi}_n := \sum_{k=1}^{K_n} \left[Y_k^2 - \frac{1}{n} \right] \qquad \text{with } K_n = \lfloor n/\log n \rfloor.$$

Now define a prior Π on f by sampling independently each coordinate f_k, $k \geq 1$ in the following way. Given a density φ on $\mathbb{R}$ and a sequence of positive real numbers (σ_k), set $K_n = \lfloor n/\log n \rfloor$ and

$$f_k \sim \frac{1}{\sigma_k} \varphi\left(\frac{\cdot}{\sigma_k} \right) \qquad \text{if } 1 \leq k \leq K_n \quad \text{and} \quad f_k = 0 \qquad \text{if } k > K_n. \tag{3.1}$$

In particular, we focus on the cases where φ is either the standard Gaussian density or $\varphi(x) = \mathbb{1}_{[-\mathcal{M},\mathcal{M}]}(x)/\mathcal{M}$, $\mathcal{M} > 0$, called respectively *Gaussian* φ and *Uniform* φ.

Suppose that there exists $M > 0$ such that, for any $1 \leq k \leq K_n$,

$$\frac{|f_{0,k}|}{\sigma_k} \leq M \quad \text{and} \quad \sigma_k \geq \frac{1}{\sqrt{n}}. \tag{3.2}$$

THEOREM 3.1. *Suppose the true function f_0 belongs to the Sobolev space W_β of order $\beta > 1/4$. Let the prior Π and K_n be chosen according to* (3.1) *and let $f_0, \{\sigma_k\}$ satisfy* (3.2). *Consider the following choices for φ*:

1. *Gaussian φ. Suppose that as $n \to \infty$,*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{K_n} \frac{\sigma_k^{-2}}{n} = o(1). \tag{3.3}$$

2. *Uniform φ. Suppose $\mathcal{M} > 4 \vee (16M)$ and that for any $c > 0$*

$$\sum_{k=1}^{K_n} \sigma_k e^{-cn\sigma_k^2} = o(1). \tag{3.4}$$

Then, in $P_{f_0}^n$-probability, as $n \to \infty$,

$$\Pi\left(\sqrt{n} \left(\psi(f) - \bar{\psi} - 2\frac{K_n}{n} \right) \middle| Y^n \right) \rightsquigarrow \mathcal{N}(0, 4\|f_0\|_2^2). \tag{3.5}$$

The proof of Theorem 3.1 is given in Section 2.2 of the supplemental article Castillo and Rousseau [15].

Theorem 3.1 gives the BvM theorem for the nonlinear functional $\psi(f) = \int f^2$, up to a (known) bias term $2K_n/n$. Indeed it implies that the posterior distribution of $\psi(f) - \hat{\psi}_n = \psi(f) - \bar{\psi} - 2\frac{K_n}{n}$ is asymptotically Gaussian with mean 0 and variance $4\|f_0\|_2^2/n$ which is the inverse of the efficient information (divided by n). Recall that $\bar{\psi}$ is an efficient estimate when $\beta > 1/4$; see, for instance, [8]. Therefore, even though the posterior distribution of $\psi(\eta)$ does not satisfy the BvM theorem per se, it can be modified a posteriori by recentering with the known quantity $2K_n/n$ to lead to a BvM theorem. The possibility of existence of a Bayesian nonparametric prior leading to a BvM for the functional $\|f\|_2^2$ without any bias term in general is unclear. However, if we restrict our attention to $\beta > 1/2$, a different choice of K_n can be made, in particular $K_n = \sqrt{n}/\log n$ leads to a standard BvM property without bias term.

Condition (3.2) can be interpreted as an undersmoothing condition: the true function should be at least as "smooth" as the prior; for a fixed prior, it corresponds to intersecting the Sobolev regularity constraint on f_0 with a Hölder-type constraint. It is used to verify the concentration of the posterior (2.7); see Lemma 3 of the supplemental article (Castillo and Rousseau [15]) (it is used here mostly for simplicity of presentation and can possibly be slightly improved). For instance, if $\sigma_k \gtrsim k^{-1/4}$ for all $k \leq K_n$, then condition (3.2) is valid for all $f_0 \in W_\beta$, with $\beta > 1/4$. Conditions (3.3) and (3.4) are here to ensure that the prior is hardly modified by the change of parametrisation (2.11), they are verified in particular for any $\sigma_k \gtrsim k^{-1/4}$.

An interesting phenomenon appears when comparing the two examples of priors considered in Theorem 3.1. If $\sigma_k = k^{-\delta}$, for some $\delta \in \mathbb{R}$, condition (3.3) holds for any $\delta \leq 1/4$ in the Gaussian φ case, whereas (3.4) only requires $\delta < 1/2$ in the Uniform φ case, this for any f_0 in $W_{1/4}$ intersected with the Hölder-type space $\{f_0 : |f_{0,k}| \leq Mk^{-\delta}, k \geq 1\}$. One can conclude that *fine details* of the prior (here, the specific form of φ chosen, for given variances $\{\sigma_k^2\}$) really matter for BvM to hold in this case. Indeed, it can be checked that the condition for the Gaussian prior is sharp: while the proof of Theorem 3.1 is an application of the general Theorem 2.1, a completely different proof can be given for Gaussian priors using conjugacy, similar in spirit to [26], leading to (3.3) as a necessary condition. Hence, choosing $\sigma_k \gtrsim k^{-1/4}$ leads to a posterior distribution satisfying the BvM property adaptively over Sobolev balls with smoothness $\beta > 1/4$.

The introduced methodology also allows us to provide conditions under generic smoothness assumptions on φ. For instance, if the density φ of the prior is a Lipschitz function on $\mathbb{R}$, then the conclusion of Theorem 3.1 holds when, as $n \to \infty$,

$$\sum_{k=1}^{K_n} \frac{\sigma_k^{-1}}{n} = o(1). \tag{3.6}$$

This last condition is not sharp in general [compare for instance with the sharp (3.3) in the Gaussian case], but provides a sufficient condition for a variety of prior distributions, including light and heavy tails behaviours. For instance, if $\sigma_k = k^{-\delta}$, then (3.6) asks for $\delta \leq 0$.

4. Application to the density model. The case of functionals of the density is another interesting application of Theorem 2.1. The case of linear functionals of the density has first been considered by [30]. Here, we obtain a broader version of Theorem 2.1 in [30], which weakens the assumptions for the case of linear functionals and also allows for nonlinear functionals.

4.1. *Statement.* Let $Y^n = (Y_1, \ldots, Y_n)$ be independent and identically distributed, having density f with respect to Lebesgue measure on the interval $[0, 1]$. In all of this section, we assume that the true density f_0 belongs to the set $\mathcal{F}_0$ of all densities that are bounded away from 0 and ∞ on $[0, 1]$. Let us consider $A_n = \{f; \|f - f_0\|_1 \leq \varepsilon_n\}$ where ε_n is a sequence decreasing to 0, or any set of the form $A_n \cap \mathcal{F}_n$, as long as $P_0^n \Pi(\mathcal{F}_n^c | Y^n) \to 0$. Define

$$L^2(f_0) = \left\{ \varphi : [0, 1] \to \mathbb{R}, \int_0^1 \varphi(x)^2 f_0(x)\, dx < \infty \right\}.$$

For any φ in $L^2(f_0)$, let us write $F_0(\varphi)$ as shorthand for $\int_0^1 \varphi(x) f_0(x)\, dx$ and set, for any positive density f on $[0, 1]$,

$$\eta = \log f, \qquad \eta_0 = \log f_0, \qquad h = \sqrt{n}(\eta - \eta_0).$$

Following [30], we have the LAN expansion

$$\begin{aligned} \ell_n(\eta) - \ell_n(\eta_0) &= \sqrt{n} F_0(h) + \frac{1}{\sqrt{n}} \sum_{i=1}^n [h(Y_i) - F_0(h)] \\ &= -\frac{1}{2}\|h\|_L^2 + W_n(h) + R_n(\eta, \eta_0), \end{aligned}$$

with the following notation, for any g in $L^2(f_0)$,

$$\|g\|_L^2 = \int_0^1 \big(g - F_0(g)\big)^2 f_0, \; W_n(g) = \mathbb{G}_n g = \frac{1}{\sqrt{n}} \sum_{i=1}^n [g(Y_i) - F_0(g)], \tag{4.1}$$

and $R_n(\eta, \eta_0) = \sqrt{n} P_{f_0} h + \frac{1}{2}\|h\|_L^2$. Note that $\|\cdot\|_L$ is an Hilbertian norm induced by the inner-product $\langle g_1, g_2 \rangle_L = \int g_1 g_2 f_0$ defined on the space $\mathcal{H}_T := \{g \in L^2(P_{f_0}), \int g f_0 = 0\} \subset \mathcal{H} = L^2(f_0)$, the so-called maximal tangent set at f_0.

We consider functionals $\psi(f)$ of the density f, which are differentiable relative to (a dense subset of) the tangent set $\mathcal{H}_T$ with efficient influence function $\tilde{\psi}_{f_0}$; see

[32], Chapter 25. In particular, $\tilde{\psi}_{f_0}$ belongs to $\mathcal{H}_T$, so $F_0(\tilde{\psi}_{f_0}) = 0$. We further assume that $\tilde{\psi}_{f_0}$ is *bounded* on $[0, 1]$. Set

$$\begin{aligned}
&\psi(f) - \psi(f_0) \\
&= \left\langle \frac{f - f_0}{f_0}, \tilde{\psi}_{f_0} \right\rangle_L + \tilde{r}(f, f_0) \qquad (4.2)\\
&= \langle \eta - \eta_0 - F_0(\eta - \eta_0), \tilde{\psi}_{f_0} \rangle_L + \mathcal{B}(f, f_0) + \tilde{r}(f, f_0), \qquad \eta = \log f,
\end{aligned}$$

where $\mathcal{B}(f, f_0)$ is the difference

$$\mathcal{B}(f, f_0) = \int_0^1 \left[\eta - \eta_0 - \frac{f - f_0}{f_0} \right](x) \tilde{\psi}_{f_0}(x) f_0(x)\, dx,$$

and define $r(f, f_0) = \mathcal{B}(f, f_0) + \tilde{r}(f, f_0)$.

THEOREM 4.1. *Let ψ be a differentiable functional relative to the tangent set $\mathcal{H}_T$, with efficient influence function $\tilde{\psi}_{f_0}$ bounded on $[0, 1]$. Let $\tilde{r}$ be defined by (4.2). Suppose that for some $\varepsilon_n \to 0$ it holds*

$$\Pi\big[f : \|f - f_0\|_1 \le \varepsilon_n | Y^n \big] \to 1, \qquad (4.3)$$

in P_0-probability and that, for $A_n = \{f, \|f - f_0\|_1 \le \varepsilon_n\}$,

$$\sup_{f \in A_n} \tilde{r}(f, f_0) = o(1/\sqrt{n}).$$

Set $\eta_t = \eta - \frac{t}{\sqrt{n}} \tilde{\psi}_{f_0} - \log \int_0^1 e^{\eta - (t/\sqrt{n})\tilde{\psi}_{f_0}}$ and assume that in P_0-probability

$$\frac{\int_{A_n} e^{\ell_n(\eta_t) - \ell_n(\eta_0)}\, d\Pi(\eta)}{\int e^{\ell_n(\eta) - \ell_n(\eta_0)}\, d\Pi(\eta)} \to 1. \qquad (4.4)$$

Then, for $\hat{\psi}$ any linear efficient estimator of $\psi(f)$, the BvM theorem holds for the functional ψ. That is, the posterior distribution of $\sqrt{n}(\psi(f) - \hat{\psi})$ is asymptotically Gaussian with mean 0 and variance $\|\tilde{\psi}_{f_0}\|_L^2$, in P_0-probability.

The semiparametric efficiency bound for estimating ψ is $\|\tilde{\psi}_{f_0}\|_L^2$ and linear efficient estimators of ψ are those for which $\hat{\psi} = \psi(f_0) + \mathbb{G}_n(\tilde{\psi}_{f_0})/\sqrt{n} + o_p(1/\sqrt{n})$; see, for example, van der Vaart [32], Chapter 25, so Theorem 4.1 yields the BvM theorem (with best possible limit distribution).

REMARK 3. The L^1-distance between densities in Theorem 4.1 can be replaced by Hellinger's distance $h(\cdot, \cdot)$ up to replacing ε_n by $\varepsilon_n/\sqrt{2}$.

Theorem 4.1 is proved in Section 6 and is deduced from Theorem 2.1 with $\psi_0^{(2)} = 0$ and $\psi_0^{(1)} = \tilde{\psi}_{f_0} - t^{-1}\sqrt{n} \log \int_0^1 e^{\eta - (t/\sqrt{n})\tilde{\psi}_{f_0}}$. The condition

$\sup_{f\in A_n}\tilde{r}(f,f_0)=o(1/\sqrt{n})$, together with (4.3) imply Assumption A. It improves on Theorem 2.1 of [30] in the sense that an L_1 -posterior concentration rate is required instead of a posterior concentration rate in terms of the LAN norm $\|\cdot\|_L$, it is also a generalisation to approximately linear functionals, which include the following examples.

EXAMPLE 4.1 (Linear functionals). Let $\psi(f)=\int_0^1 f(x)a(x)\,dx$, for some bounded function a. Then, writing $\int$ as shorthand for $\int_0^1$,

$$\psi(f)-\psi(f_0)=\left\langle\frac{f-f_0}{f_0},a-\int af_0\right\rangle_L$$

with the efficient influence function $\tilde{\psi}_{f_0}=a-\int af_0$. In this case, $\tilde{r}(f,f_0)=0$.

EXAMPLE 4.2 (Entropy functional). Let $\psi(f)=\int_0^1 f(x)\log f(x)\,dx$, for f bounded away from 0 and infinity. Then

$$\psi(f)-\psi(f_0)=\left\langle\frac{f-f_0}{f_0},\log f_0-\int f_0\log f_0\right\rangle_L+\int f\log\frac{f}{f_0}$$

with the efficient influence function $\tilde{\psi}_{f_0}=\log f_0-\int f_0\log f_0$. In this case, $\tilde{r}(f,f_0)=\int f\log\frac{f}{f_0}$. For the two types of priors considered below, under some smoothness assumptions on f_0, it holds $\sup_{f\in A_n}\tilde{r}(f,f_0)=o(1/\sqrt{n})$.

EXAMPLE 4.3 (Square-root functional). Let $\psi(f)=\int_0^1\sqrt{f(x)}\,dx$, for f a bounded density. Then

$$\psi(f)-\psi(f_0)=\frac{1}{2}\left\langle\frac{f-f_0}{f_0},\frac{1}{\sqrt{f_0}}-\int\sqrt{f_0}\right\rangle_L+\frac{1}{2}\int\frac{\sqrt{f_0}-\sqrt{f}}{\sqrt{f_0}+\sqrt{f}}\frac{f-f_0}{\sqrt{f_0}}$$

with the efficient influence function $\tilde{\psi}_{f_0}=\frac{1}{2}(\frac{1}{\sqrt{f_0}}-\int\sqrt{f_0})$. In this case, $\tilde{r}(f,f_0)=-\int\frac{(\sqrt{f_0}-\sqrt{f})^2}{2\sqrt{f_0}}$. In particular, the remainder term of the functional expansion is bounded by a constant times the square of the Hellinger distance between densities, hence as soon as $\varepsilon_n^2\sqrt{n}=o(1)$, if A_n is written in terms of h (see Remark 3), one has $\sup_{f\in A_n}\tilde{r}(f,f_0)=o(1/\sqrt{n})$.

EXAMPLE 4.4 (Power functional). Let $\psi(f)=\int_0^1 f(x)^q\,dx$, for f a bounded density and $q\geq 2$ an integer. Then

$$\psi(f)-\psi(f_0)=\left\langle\frac{f-f_0}{f_0},qf_0^{q-1}-q\int f_0^q\right\rangle_L+r(f,f_0).$$

The remainder $\tilde{r}(f,f_0)$ is a sum of terms of the form $\int(f-f_0)^{2+s}f_0^{q-2-s}$, for $0\leq s\leq q-2$ an integer. For the two types of priors considered below, $\sup_{f\in A_n}\tilde{r}(f,f_0)=o(1/\sqrt{n})$, under some smoothness assumptions on f_0.

We now consider two families of priors: random histograms and Gaussian process priors. For each family, we provide a key no-bias condition for BvM on functionals to be valid. For each, the idea is based on a certain functional change of variables formula. To simplify the notation, we write $\tilde{\psi} = \tilde{\psi}_{f_0}$ in the sequel.

4.2. *Random histograms.* For any $k \in \mathbb{N}^*$, consider the partition of $[0,1]$ defined by $I_j = [(j-1)/k, j/k)$ for $j = 1, \dots, k$. Denote by

$$\mathcal{H}_k = \left\{ g \in L^2[0,1], g(x) = \sum_{j=1}^{k} g_j \mathbb{1}_{I_j}(x), g_j \in \mathbb{R}, j = 1, \dots, k \right\}$$

the set of all regular histograms with k bins on $[0,1]$. Let $\mathcal{S}_k = \{\omega \in [0,1]^k; \sum_{j=1}^k \omega_j = 1\}$ be the unit simplex in $\mathbb{R}^k$ and denote $\mathcal{H}_k^1$ the subset of $\mathcal{H}_k$ consisting of histograms which are densities on $[0,1]$:

$$\mathcal{H}_k^1 = \left\{ f \in L^2[0,1], f(x) = f_{\omega,k} = k \sum_{j=1}^{k} \omega_j \mathbb{1}_{I_j}(x), (\omega_1, \dots, \omega_k) \in \mathcal{S}_k \right\}.$$

A prior on $\mathcal{H}_k^1$ is completely specified by the distributions of k and of $(\omega_1, \dots, \omega_k)$ given k. Conditionally, on k, we consider a Dirichlet prior on $\omega = (\omega_1, \dots, \omega_k)$:

$$\omega \sim \mathcal{D}(\alpha_{1,k}, \dots, \alpha_{k,k}), \qquad c_1 k^{-a} \le \alpha_{j,k} \le c_2, \tag{4.5}$$

for some fixed constants $a, c_1, c_2 > 0$ and any $1 \le j \le k$.

Consider two situations: either a deterministic number of bins with $k = K_n = o(n)$ or, for π_k a distribution on positive integers,

$$k \sim \pi_k, \qquad e^{-b_1 k \log(k)} \le \pi_k(k) \le e^{-b_2 k \log(k)}, \tag{4.6}$$

for all k large enough and some $0 < b_2 < b_1 < \infty$. Condition (4.6) is verified for instance by the Poisson distribution which is commonly used in Bayesian nonparametric models; see, for instance, [1].

The set $\mathcal{H}_k$ is a closed subspace of $L^2[0,1]$. For any function h in $L^2[0,1]$, consider its projection $h_{[k]}$ in the L^2-sense on $\mathcal{H}_k$. It holds

$$h_{[k]} = k \sum_{j=1}^{k} \left\{ \int_{I_j} h \right\} \mathbb{1}_{I_j}.$$

Lemma 4 in the supplemental article Castillo and Rousseau [15] gathers useful properties on histograms.

Let the functional ψ satisfy (4.2) with bounded efficient influence function $\tilde{\psi}_{f_0} = \tilde{\psi} \neq 0$ and set, for $k \ge 1$,

$$\begin{aligned} \hat{\psi}_k &= \psi(f_{0[k]}) + \frac{\mathbb{G}_n \tilde{\psi}_{[k]}}{\sqrt{n}}, \qquad V_k = \|\tilde{\psi}_{[k]}\|_L^2, \\ \hat{\psi} &= \psi(f_0) + \frac{\mathbb{G}_n \tilde{\psi}}{\sqrt{n}}, \qquad V = \|\tilde{\psi}\|_L^2, \end{aligned} \tag{4.7}$$

with $\|\cdot\|_L, \mathbb{G}_n$ as in (4.1). Finally, for $n \geq 2$, $k \geq 1$, $M > 0$, denote

$$A_{n,k}(M) = \{f \in \mathcal{H}_k^1, h(f, f_{0,[k]}) \leq M\varepsilon_{n,k}\} \qquad \text{with } \varepsilon_{n,k}^2 = \frac{k \log n}{n}. \tag{4.8}$$

In Section 6.3, we shall see that the posterior distribution of k concentrates on a deterministic subset $\mathcal{K}_n$ of $\{1, \ldots, \lfloor n/(\log n)^2 \rfloor\}$ and that under the following technical condition on the weights, as $n \to \infty$,

$$\sup_{k \in \mathcal{K}_n} \sum_{j=1}^{k} \alpha_{j,k} = o(\sqrt{n}), \tag{4.9}$$

the conditional posterior distribution given k, concentrates on the sets $A_{n,k}(M)$. It can then be checked that

$$\begin{aligned}
&\Pi\big[\sqrt{n}(\psi - \hat{\psi}) \leq z | Y^n\big] \\
&\quad = \sum_{k \in \mathcal{K}_n} \Pi\big[k|Y^n\big] \Pi\big[\sqrt{n}(\psi - \hat{\psi}_k) \leq z + \sqrt{n}(\hat{\psi} - \hat{\psi}_k) | Y^n, k\big] + o_p(1) \\
&\quad = \sum_{k \in \mathcal{K}_n} \Pi\big[k|Y^n\big] \Phi\big((z + \sqrt{n}(\hat{\psi} - \hat{\psi}_k))/\sqrt{V_k}\big) + o_p(1).
\end{aligned}$$

The last line expresses that the posterior is asymptotically close to a mixture of normals, and that the mixture reduces to the target law $N(0, V)$ if V_k goes to V and $\sqrt{n}(\hat{\psi} - \hat{\psi}_k)$ to 0, uniformly for k in $\mathcal{K}_n$. The last quantity can also be rewritten

$$\begin{aligned}
\sqrt{n}(\hat{\psi}_k - \hat{\psi}) &= \sqrt{n}\big(\psi(f_{0[k]}) - \psi(f_0)\big) + \mathbb{G}_n(\tilde{\psi}_{[k]} - \tilde{\psi}) \\
&= \sqrt{n} \int (\tilde{\psi} - \tilde{\psi}_{[k]})(f_{0[k]} - f_0) + \mathbb{G}_n(\tilde{\psi}_{[k]} - \tilde{\psi}) + o(1).
\end{aligned}$$

It is thus natural to ask for, and this is satisfied in most examples (see below),

$$\max_{k \in \mathcal{K}_n} \big| \|\tilde{\psi}_{[k]}\|_L^2 - \|\tilde{\psi}\|_L^2 \big| = o_p(1) \quad \text{and} \quad \max_{k \in \mathcal{K}_n} \mathbb{G}_n(\tilde{\psi}_{[k]} - \tilde{\psi}) = o_p(1). \tag{4.10}$$

This leads to the next proposition, proved in Section 6.

PROPOSITION 1. *Let f_0 belong to $\mathcal{F}_0$ and the prior Π be defined by (4.5)–(4.9). Let the prior π_k be either the Dirac mass at $k = K_n \leq n/(\log n)^2$, or the law given in (4.6). Let $\mathcal{K}_n$ be a subset of $\{1, 2, \ldots, n/\log^2 n\}$ such that $\Pi(\mathcal{K}_n | Y^n) = 1 + o_p(1)$.*

Consider estimating a functional $\psi(f)$, with $\tilde{r}$ in (4.2), verifying (4.10) and, for any $M > 0$, with $A_{n,k}(M)$ defined in (4.8),

$$\sup_{k \in \mathcal{K}_n} \sup_{f \in A_{n,k}(M)} \sqrt{n}\tilde{r}(f, f_0) = o_p(1), \tag{4.11}$$

as $n \to \infty$. Additionally, suppose

$$\max_{k \in \mathcal{K}_n} \sqrt{n} \left| \int (\tilde{\psi} - \tilde{\psi}_{[k]})(f_{0[k]} - f_0) \right| = o(1). \tag{4.12}$$

Then the BvM theorem for the functional ψ holds.

The core condition is (4.12), which can be seen as a *no-bias* condition. Condition (4.11) controls the remainder term of the expansion of $\psi(f)$ around f_0. Condition (4.10) is satisfied under very mild conditions: for its first part it is enough that $\inf \mathcal{K}_n$ goes to ∞ with n. For the second part, barely more than this typically suffices, using a simple empirical process argument; see Section 6.

The next theorem investigates the previous conditions under deterministic and random priors on k, for the examples of functionals 4.1 to 4.4.

THEOREM 4.2. *Suppose* $f_0 \in \mathcal{C}^\beta$, *with* $\beta > 0$. *Let two priors* Π_1, Π_2 *be defined by* (4.5)–(4.9) *and the prior on* k *be either the Dirac mass at* $k = K_n = \lfloor n^{1/2}(\log n)^{-2} \rfloor$ *for* Π_1, *or* $k \sim \pi_k$ *given by* (4.6) *for* Π_2. *Then*:

- *Example* 4.1, *linear functionals* $\psi(f) = \int af$, *under the prior* Π_1 *with deterministic* $k = K_n$
  - ◇ *if* $a(\cdot) \in \mathcal{C}^\gamma$ *with* $\gamma + \beta > 1$ *for some* $\gamma > 0$, *then the BvM theorem holds for the functional* $\psi(f)$;
  - ◇ *if* $a(\cdot) = \mathbb{1}_{\cdot \le z}$ *for* $z \in [0, 1]$, *then BvM holds for the functional* $\int \mathbb{1}_{\cdot \le z} f = F(z)$, *the cumulative distribution function of* f.
- *Examples* 4.2–4.3–4.4. *For all* $\beta > 1/2$, *the BvM theorem holds for* $\psi(f)$ *for both priors* Π_1 (*deterministic* k) *and* Π_2 (*random* k).

Theorem 4.2 is proved in Section 6.3. From this proof, it may be noted that different choices of K_n in some range lead to similar results for some examples. For instance, if $\psi(f) = \int \psi f$ and $\psi \in \mathcal{C}^\gamma$, choosing $K_n = \lfloor n/(\log n)^2 \rfloor$ implies that the BvM holds for all $\gamma + \beta > 1/2$.

Obtaining BvM in the case of a prior with random k in Example 4.1 is case-dependent. The answer lies in the respective approximation properties of both f_0 and $\tilde{\psi}_{f_0}$ through the prior (note that a random k prior typically adapts to the regularity of f_0), and the no-bias condition (4.12) may not be satisfied if $\inf \mathcal{K}_n$ is not large enough.

We present below a counterexample where BvM is proved to fail for a large class of true densities f_0 when a prior with random k is chosen.

4.3. *A semiparametric curse of adaptation*: *A counterexample for BvM under random number of bins histogram priors*. Consider a $\mathcal{C}^1$, strictly increasing true function f_0, say

$$f_0' \ge \rho > 0 \qquad \text{on } [0, 1]. \tag{4.13}$$

The following reasoning can be extended to any approximately monotone smooth function on $[0, 1]$. Consider estimation of the linear functional $\psi(f) = \int \psi f$. The BvM theorem is not satisfied if the bias term $\sqrt{n}(\hat{\psi} - \hat{\psi}_k)$ is predominant for all

k's which are asymptotically given mass under the posterior. This will happen if for all such k's,

$$-b_{n,k} = \sqrt{n}\int \psi(f_0 - f_{0[k]}) = \sqrt{n}\int (\psi - \psi_{[k]})(f_0 - f_{0[k]}) \gg 1,$$

as $n \to \infty$. To simplify the presentation, we restrict ourselves to the case of dyadic random histograms; in other words, the prior on k only puts mass on values of $k = 2^p$, $p \geq 0$. Then define $\psi(x)$ as, for $\alpha > 0$,

$$\psi(x) = \sum_{l\geq 0}\sum_{j=0}^{2^l-1} 2^{-l(1/2+\alpha)}\psi_{lj}^H(x), \tag{4.14}$$

where $\psi_{lj}^H(x) = 2^{l/2}\psi_{00}(2^l x - j)$ and $\psi_{00}(x) = -\mathbb{1}_{[0,1/2]}(x) + \mathbb{1}_{(1/2,1]}(x)$ is the mother wavelet of the Haar basis (we omit the scaling function 1 in the definition of ψ).

PROPOSITION 2. *Let f_0 be any function as in (4.13) and α, ψ as in (4.14). Let the prior be as in Theorem 4.2. Then there exists $k_1 > 0$ such that*

$$\Pi\big(k < k_1(n/\log n)^{1/3}|Y^n\big) = 1 + o_P(1)$$

and for all $p \in \mathbb{N}$ such that $2^p := K < k_1(n/\log n)^{1/3}$, the conditional posterior distribution of $\sqrt{n}(\psi(f) - \hat{\psi} - b_{n,k})/\sqrt{V_k}|k = K$ converges in distribution to $\mathcal{N}(0,1)$, in P_0^n-probability, with

$$b_{n,K} \lesssim -\sqrt{n}K^{-\alpha-1}.$$

In particular, the BvM property does not hold if $\alpha < 1/2$.

REMARK 4. For the considered f_0, it can be checked that the posterior even concentrates on values of k such that $k = k_n \asymp (n/\log n)^{1/3}$.

As soon as the regularities of the functional $\psi(f)$ to be estimated and of the true function f_0 are fairly different, taking an *adaptive* prior (with respect to f) can have disastrous effects with a nonnegligible bias appearing in the centering of the posterior distribution. As in the counterexample in Rivoirard and Rousseau [30], the BvM is ruled out because the posterior distribution concentrates on values of k that are too small and for which the bias $b_{n,k}$ is not negligible. Note that for each of these functionals the BvM is violated for a large class of true densities f_0. Some related phenomena in terms of rates are discussed in Knapik et al. [25] for linear functionals and adaptive priors in white noise inverse problems.

Let us sketch the proof of Proposition 2. It is not difficult to show that (see the Supplement), since $f_0 \in \mathcal{C}^1$, the posterior concentrates on the set $\{f : \|f - f_0\|_1 \leq M(n/\log n)^{-1/3}, k \leq k_1(n/\log n)^{1/3}\}$, for some positive M and k_1. Since

Haar wavelets are special cases of (dyadic) histograms, for any $K \geq 1$ the best approximation of ψ within $\mathcal{H}_K$ is

$$\psi_{[K]}(x) = \sum_{l=0}^{p} \sum_{j=0}^{2^l-1} 2^{-l(1/2+\alpha)} \psi_{lj}^H(x).$$

The semiparametric bias $-b_{n,K}$ is equal to $\sqrt{n}\int_0^1 (f_0 - f_{0,[K]})(\psi - \psi_{[K]}) = \sqrt{n}\int_0^1 f_0(\psi - \psi_{[K]})$, which can be written, for any $K \geq 1$,

$$\begin{aligned} -b_{n,K} &= \sqrt{n} \sum_{l>p} \sum_{j=0}^{2^l-1} 2^{-l(1/2+\alpha)} \int_0^1 f_0(x) \psi_{lj}^H(x)\, dx \\ &= \sqrt{n} \sum_{l>p} \sum_{j=0}^{2^l-1} 2^{-l\alpha} \int_{2^{-l}j}^{2^{-l}(j+1/2)} \bigl(f_0(x + 2^{-l}/2) - f_0(x)\bigr)\, dx \\ &\gtrsim \sqrt{n} \sum_{l>p} 2^{-l\alpha} 2^l 2^{-2l} \gtrsim \sqrt{n} K^{-\alpha-1}. \end{aligned}$$

Since $\Pi(k \leq n^{1/3}|Y^n) = 1 + o_p(1)$, we have that $\inf_{k \leq n^{1/3}} -b_{n,k} \to +\infty$ for all $\alpha < 1/2$. Also, the sequence of real numbers $\{V_k\}_{k \geq 1}$ stays bounded, while the supremum $\sup_{1 \leq k \leq n^{1/3}} |\mathbb{G}_n(\tilde{\psi} - \tilde{\psi}_{[k]})|$ is bounded by a constant times $(\log n)^{1/2}$ in probability, by a standard empirical process argument. This implies that

$$E^{\Pi}\bigl[e^{t\sqrt{n}(\psi(f)-\hat{\psi})}|Y^n, B_n\bigr] = \bigl(1 + o(1)\bigr) \sum_{k \in \mathcal{K}_n} e^{t^2 V_k/2 + t\sqrt{n}(\hat{\psi}-\hat{\psi}_k)} \Pi\bigl[k|Y^n\bigr] = o_p(1),$$

so that the posterior distribution is not asymptotically equivalent to $\mathcal{N}(0, \|\tilde{\psi}\|_L^2)$, and there exists M_n going to infinity such that

$$\Pi\bigl[\sqrt{n}|\psi(f) - \hat{\psi}| > M_n|Y^n\bigr] = 1 + o_p(1).$$

4.4. *Gaussian process priors.* We now investigate the implications of Theorem 4.1 in the case of Gaussian process priors for the density f. Consider as a prior on f the distribution on densities generated by

$$f(x) = \frac{e^{W(x)}}{\int_0^1 e^{W(x)}\, dx}, \tag{4.15}$$

where W is a zero-mean Gaussian process indexed by $[0, 1]$ with continuous sample paths. The process W can also be viewed as a random element in the Banach space $\mathbb{B}$ of continuous functions on $[0, 1]$ equipped with the sup-norm $\|\cdot\|_\infty$; see [34] for precise definitions. We refer to [33, 34] and [9] for basic definitions on Gaussian priors and some convergence properties, respectively. Let $K(x, y) = E[W(x)W(y)]$ denote the covariance kernel of the process and let $(\mathbb{H}, \|\cdot\|_{\mathbb{H}})$ denote the reproducing kernel Hilbert space of W.

EXAMPLE 4.5 (Brownian motion released at 0). Consider the distribution induced by

$$W(x) = N + B_x, \qquad x \in [0,1],$$

where B_x is standard Brownian motion and N is an independent $\mathcal{N}(0,1)$ variable. We use it as a prior on w. It can be seen (see [33]) as a random element in the Banach space $\mathbb{B} = (\mathcal{C}^0, \|\cdot\|_\infty)$ and its RKHS is

$$\mathbb{H}^B = \left\{ c + \int_0^\cdot g(u)\,du, c \in \mathbb{R}, g \in L^2[0,1] \right\},$$

a Hilbert space with norm given by $\|c + \int_0^\cdot g(u)\,du\|_{\mathbb{H}^B}^2 = c^2 + \int_0^1 g(u)^2\,du$.

EXAMPLE 4.6 (Riemann–Liouville-type processes). Consider the distribution induced by, for $\alpha > 0$ and $x \in [0,1]$,

$$W^\alpha(x) = \sum_{k=0}^{\lfloor \alpha \rfloor + 1} Z_k x^k + \int_0^x (x-s)^{\alpha - 1/2}\,dB_s,$$

where Z_is are independent standard normal variables and B is an independent Brownian motion. The RKHS $\mathbb{H}^\alpha$ of W^α can be obtained explicitly from the one of Brownian motion, and is nothing but a Sobolev space of order $\alpha + 1$; see [33], Theorem 4.1.

The concentration function of the Gaussian process in $\mathbb{B}$ at $\eta_0 = \log f_0$ is defined for any $\varepsilon > 0$ by (see [34])

$$\varphi_{\eta_0}(\varepsilon) = -\log \Pi\big(\|W\|_\infty \le \varepsilon\big) + \frac{1}{2} \inf_{h \in \mathbb{H} : \|h - \eta_0\|_{\mathbb{B}} < \varepsilon} \|h\|_{\mathbb{H}}^2.$$

In van der Vaart and van Zanten [33], it is shown that the posterior contraction rate for such a prior is closely connected to a solution ε_n of

$$\varphi_{\eta_0}(\varepsilon_n) \le n\varepsilon_n^2, \qquad \eta_0 = \log f_0. \tag{4.16}$$

PROPOSITION 3. *Suppose f_0 verifies $c_0 \le f_0 \le C_0$ on $[0,1]$, for some positive c_0, C_0. Let the prior Π on f be induced via a Gaussian process W as in (4.15) and let $\mathbb{H}$ denote its RKHS. Let $\varepsilon_n \to 0$ verify (4.16). Consider estimating a functional $\psi(f)$, with $\tilde{r}$ in (4.2) verifying*

$$\sup_{f \in A_n} \tilde{r}(f, f_0) = o(1/\sqrt{n}),$$

for A_n such that $\Pi(A_n | Y^n) = 1 + o_p(1)$ and $A_n \subset \{f : h(f, f_0) \le \varepsilon_n\}$. Suppose that $\tilde{\psi}_{f_0}$ is continuous and that there exists a sequence $\psi_n \in \mathbb{H}$ and $\zeta_n \to 0$, such that

$$\|\psi_n - \tilde{\psi}_{f_0}\|_\infty \le \zeta_n \quad \text{and} \quad \|\psi_n\|_{\mathbb{H}} \le \sqrt{n}\zeta_n, \tag{4.17}$$

$$\sqrt{n}\varepsilon_n\zeta_n \to 0. \tag{4.18}$$

Then, for $\hat{\psi}$ any linear efficient estimator of $\psi(f)$, in P_0^n-probability, the posterior distribution of $\sqrt{n}(\psi(f) - \hat{\psi})$ converges to a Gaussian distribution with mean 0 and variance $\|\tilde{\psi}_{f_0}\|_L^2$ and the BvM theorem holds.

The proof is presented in Section 3.2 of Castillo and Rousseau [15]. We now investigate conditions (4.17)–(4.18) for examples of Gaussian priors.

THEOREM 4.3. *Suppose that $\eta_0 = \log f_0$ belongs to $\mathcal{C}^\beta$, for some $\beta > 0$. Let Π_α be the priors defined from a Gaussian process W via (4.15). For Π_1, we take W to be Brownian motion (released at 0) and for Π_2 we take $W = W^\alpha$, a Riemann–Liouville-type process of parameter $\alpha > 0$.*

- *Example 4.1, linear functionals $\psi(f) = \int af$*
  - ◇ *if $a(\cdot) \in \mathbb{H}^B$, then the BvM theorem holds for the functional $\psi(f)$ and prior Π_1. The same holds if $a(\cdot) \in \mathbb{H}^\alpha$ for prior Π_2;*
  - ◇ *if $a(\cdot) \in \mathcal{C}^\mu$, $\mu > 0$, the BvM property holds for prior Π_2 if*

$$\alpha \wedge \beta > \tfrac{1}{2} + (\alpha - \mu) \vee 0.$$

- *Examples 4.3–4.4. Under the same condition as for the linear functional with $\mu = \beta$, the BvM theorem holds for Π_2.*

An immediate illustration of Theorem 4.3 is as follows. Consider prior Π_1 built from Brownian motion. Then for all linear functionals

$$\psi(f) = \int_0^1 x^r f(x)\,dx, \qquad r > \frac{1}{2},$$

the BvM theorem holds. Indeed, $x \to x^r, r > 1/2$ belongs to $\mathbb{H}^B$.

To prove Theorem 4.3, one applies Proposition 3: it is enough to compute bounds for ε_n and ζ_n. This follows from the results on the concentration function for Riemann–Louville-type processes obtained in Theorem 4 in [9]. For linear functionals $\psi(f) = \int af$ and $a \in \mathcal{C}^\mu$, one can take $\varepsilon_n = n^{-\alpha\wedge\beta/(2\alpha+1)}$ and $\zeta_n = n^{-\mu/(2\alpha+1)}$, up to some logarithmic factors. So (4.18) holds if $\alpha \wedge \beta > \frac{1}{2} + (\alpha - \mu) \vee 0$.

The square-root functional is similar to a linear functional with $\mu = \beta$, since the remainder term in the expansion of the functional is of the order of the Hellinger distance. Indeed, since f_0 is bounded away from 0 and ∞, the fact that $w_0 \in \mathcal{C}^\beta$ implies that $f_0 \in \mathcal{C}^\beta$ and $\sqrt{f_0} \in \mathcal{C}^\beta$. For power functionals, the remainder term $r(f, f_0)$ is more complicated but is easily bounded by a linear combination of terms of the type

$$\int (f - f_0)^{2+r} f_0^{q-2-r} \le \|f_0\|_\infty^{q-r-2} \|f - f_0\|_\infty^r \int (f - f_0)^2.$$

Using Proposition 1 in Castillo and Rousseau [15], one obtains that, under the posterior distribution, $\|f - f_0\|_\infty \lesssim 1$ and $\|f - f_0\|_2 \lesssim \varepsilon_n$. So, $\sqrt{n} r(f, f_0) = o(1)$ holds if $\sqrt{n}\varepsilon_n^2 = o(1)$, which is the case since $\alpha \wedge \beta > 1/2$.

5. Application to the nonlinear autoregressive model. Consider an autoregressive model in which one observes $Y_1, \dots, Y_n$ given by

$$Y_{i+1} = f(Y_i) + \epsilon_i, \qquad \epsilon_i \sim \mathcal{N}(0,1) \qquad \text{i.i.d.}, \tag{5.1}$$

where $\|f\|_\infty \le L$ for a fixed given positive constant L and f belongs to a Hölder space $\mathcal{C}^\beta$, $\beta > 0$. This example has been in particular studied by [22] and it is known that $(Y_i, i = 1, \dots, n)$ is an homogeneous Markov chain and that under these assumptions, for all f, there exists a unique stationary distribution Q_f with density q_f with respect to Lebesgue measure. The transition density is $p_f(y|x) = \phi(y - f(x))$. Denoting $r(y) = (\phi(y - L) + \phi(y + L))/2$, the transition density satisfies $p_f(y|x) \asymp r(y)$ for all $x, y \in \mathbb{R}$. Following [22], define the norms, for any $s \ge 2$,

$$\|f - f_0\|_{s,r} = \left(\int_{\mathbb{R}} |f(x) - f_0(x)|^s r(x)\, dx \right)^{1/s}.$$

As in [22], we consider a prior Π on f based on piecewise constant functions. Let us set $a_n = b\sqrt{\log n}$, where $b > 0$ and consider functions f of the form

$$f(x) := f_{\omega,k}(x) = \sum_{j=0}^{k-1} \omega_j \mathbb{1}_{I_j}(x), \qquad I_j = a_n\big([j/k, (j+1)/k] - 1/2\big).$$

A prior on k and on $\omega = (\omega_0, \dots, \omega_{k-1})$ is then specified as follows. First, draw $k \sim \pi_k$, for π_k a law on the integers. Given k, the law $\omega|k$ is supposed to have a Lebesgue density $\pi_{\omega|k}$ with support $[-M, M]^k$ for some $M > 0$. Assume further that these laws satisfy, for $0 < c_2 \le c_1 < \infty$ and $C_1, C_2 > 0$,

$$\begin{aligned} &e^{-c_1 K \log K} \le \pi_k[k > K] \le e^{-c_2 K \log K} \qquad \text{for large } K, \\ &e^{-C_2 k \log k} \lesssim \pi_{\omega|k}(\omega) \le C_1 \qquad \forall \omega \in [-M, M]^k. \end{aligned} \tag{5.2}$$

We consider the squared-weighted-L_2 norm functional $\psi(f) = \int_{\mathbb{R}} f^2(y) q_f(y)\, dy$. As before, define

$$k_n(\beta) = \lfloor (n/\log n)^{1/(2\beta+1)} \rfloor, \qquad \varepsilon_n(\beta) = (n/\log n)^{-\beta/(2\beta+1)}.$$

For all bounded f_0 and all $k > 0$, define

$$\tilde{\omega}^0_{[k]} = (\tilde{\omega}^0_1, \dots, \tilde{\omega}^0_k), \qquad \tilde{\omega}^0_j = \frac{\int_{I_j} f_0(x) q_{f_0}(x)\, dx}{\int_{I_j} q_{f_0}(x)\, dx};$$

these are the weights of the projection of f_0 on the weighted space $L^2(q_{f_0})$. We then have the following sufficient condition for the BvM to be valid.

THEOREM 5.1. *Consider the autoregressive model* (5.1) *and the prior* (5.2). *Assume that* $f_0 \in \mathcal{C}^\beta$, *with* $\beta > 1/2$ *and* $\|f_0\|_\infty < L$, *and assume that* $\pi_{\omega|k}$ *satisfies for all* $t > 0$ *and all* $M_0 > 0$

$$\sup_{\|\omega - \tilde{\omega}^0_{[k]}\|_{2,r} \le M_0 \varepsilon_n(\beta)} \left| \frac{\pi_{\omega|k}(\omega - t\tilde{\omega}^0_{[k]}/\sqrt{n})}{\pi_{\omega|k}(\omega)} - 1 \right| = o(1). \tag{5.3}$$

Then the posterior distribution of $\sqrt{n}(\psi(f) - \hat{\psi})$ *is asymptotically Gaussian with mean* 0 *and variance* V_0, *where*

$$\hat{\psi} = \psi(f_0) + \frac{2}{n}\sum_{i=1}^{n} \epsilon_i f_0(Y_{i-1}) + o_p(n^{-1/2}), \qquad V_0 = 4\|f_0\|^2_{2,q_{f_0}}$$

and the BvM is valid under the distribution associated to f_0 *and any initial distribution* ν *on* $\mathbb{R}$.

Theorem 5.1 is proved in Section 4 of Castillo and Rousseau [15]. The conditions on the prior (5.2) and (5.3) are satisfied in particular when $k \sim \mathcal{P}(\lambda)$ and when given k, the law $\omega|k$ is the independent product of k laws $\mathcal{U}(-M, M)$. Theorem 5.1 is an application of the general Theorem 2.1, with $A_n = \{f_{\omega,k}; k \le k_1 k_n(\beta); \|\omega - \omega^0_{[k]}\|_{2,r} \le M_0\varepsilon_n(\beta)\}$ and Assumption A implied by $\beta > 1/2$. Condition (5.3) is used to prove condition (2.13).

6. Proofs.

6.1. *Proof of Theorem* 2.1. Let the set A_n be as in Assumption A. Set

$$I_n := E\big[e^{t\sqrt{n}(\psi(\eta)-\psi(\eta_0))}|Y^n, A_n\big].$$

For the sake of conciseness, we prove the result in the case where $\psi_0^{(2)} \neq 0$ since the other case is a simpler version of it. Using the LAN expansion (2.3) together with the expansion (2.4) of the functional ψ, one can write

$$I_n = \frac{\int_{A_n} e^{\sqrt{n}t(\langle\psi_0^{(1)}, \eta-\eta_0\rangle_L + (1/2)\langle\psi_0^{(2)}(\eta-\eta_0), \eta-\eta_0\rangle_L) + \ell_n(\eta) - \ell_n(\eta_0) + t\sqrt{n}r(\eta,\eta_0)}\, d\Pi(\eta)}{\int_{A_n} e^{-n\|\eta-\eta_0\|_L^2/2 + \sqrt{n}W_n(\eta-\eta_0) + R_n(\eta,\eta_0)}\, d\Pi(\eta)}.$$

Consider, for any real number t, as defined in (2.11),

$$\eta_t = \eta - \frac{t\psi_0^{(1)}}{\sqrt{n}} - \frac{t}{2\sqrt{n}}\psi_0^{(2)}(\eta - \eta_0) - \frac{t\psi_0^{(2)}w_n}{2n}.$$

Then using (2.9)–(2.10) in Assumption A, on A_n,

$$\begin{aligned}\ell_n(\eta_t) - \ell_n(\eta_0) - \big(\ell_n(\eta) - \ell_n(\eta_0)\big) \\ = -\frac{n}{2}\big[\|\eta_t - \eta_0\|_L^2 - \|\eta - \eta_0\|_L^2\big] + \sqrt{n}\langle w_n, \eta_t - \eta\rangle_L + R_n(\eta_t, \eta_0)\end{aligned}$$

$$
\begin{aligned}
&- R_n(\eta, \eta_0) + o_P(1) \\
&= -t\langle w_n, \psi_0^{(1)} + \psi_0^{(2)} w_n/(2\sqrt{n})\rangle_L - \frac{t^2}{2}\left\|\psi_0^{(1)} + \frac{\psi_0^{(2)} w_n}{2\sqrt{n}}\right\|_L^2 \\
&\quad + \sqrt{n}t\langle \psi_0^{(1)}, \eta - \eta_0\rangle_L \\
&\quad + \frac{t\sqrt{n}}{2}\langle \psi_0^{(2)}(\eta - \eta_0), \eta - \eta_0\rangle_L + R_n(\eta_t, \eta_0) - R_n(\eta, \eta_0) + o_P(1).
\end{aligned}
$$

One deduces that on A_n, from (2.12) in Assumption A,

$$
\begin{aligned}
&\sqrt{n}t\left(\langle \psi_0^{(1)}, \eta - \eta_0\rangle_L + \frac{1}{2}\langle \psi_0^{(2)}(\eta - \eta_0), \eta - \eta_0\rangle_L\right) + \ell_n(\eta) - \ell_n(\eta_0) \\
&\quad + \sqrt{n}t r(\eta, \eta_0) \\
&\quad = \ell_n(\eta_t) - \ell_n(\eta_0) + t\langle w_n, \psi_0^{(1)} + \psi_0^{(2)} w_n/(2\sqrt{n})\rangle_L \\
&\qquad + \frac{t^2}{2}\left\|\psi_0^{(1)} + \frac{\psi_0^{(2)} w_n}{2\sqrt{n}}\right\|_L^2 + o_P(1).
\end{aligned}
$$

We can then rewrite I_n as

$$
I_n = e^{o_P(1) + \frac{t^2}{2}\|\psi_0^{(1)} + \frac{\psi_0^{(2)} w_n}{2\sqrt{n}}\|_L^2 + t\langle w_n, \psi_0^{(1)} + \frac{\psi_0^{(2)} w_n}{2\sqrt{n}}\rangle_L} \frac{\int_{A_n} e^{\ell_n(\eta_t) - \ell_n(\eta_0)}\, d\Pi(\eta)}{\int_{A_n} e^{\ell_n(\eta) - \ell_n(\eta_0)}\, d\Pi(\eta)},
$$

and Theorem 2.1 is proved using condition (2.14), together with the fact that (see Section 1 of Castillo and Rousseau [15]), convergence of Laplace transforms for all t in probability implies convergence in distribution in probability.

6.2. *Proof of Theorem 4.1.* One can define $\psi_0^{(1)} = \tilde{\psi}_{f_0} + c$ for any constant c, since the inner product associated to the LAN norm corresponds to re-centered quantities. In particular, for all $\eta = \log f$

$$
\langle (\tilde{\psi}_{f_0} + c), \eta - \eta_0\rangle_L = \int (\tilde{\psi}_{f_0} - P_{f_0}\tilde{\psi}_{f_0})(\eta - \eta_0) f_0, \qquad \|\tilde{\psi}_{f_0} + c\|_L = \|\tilde{\psi}_{f_0}\|_L.
$$

To check Assumption A, let us write

$$
\psi_0^{(1)} = \tilde{\psi}_{f_0} + \frac{\sqrt{n}}{t}\log\left(\int_0^1 e^{\eta - (t/\sqrt{n})\tilde{\psi}_{f_0}(x)}\, dx\right), \tag{6.1}
$$

which depends on η but is of the form $\tilde{\psi}_{f_0} + c$ (see also Remark 2), and we study $\sqrt{n}t r(\eta, \eta_0) + R_n(\eta, \eta_0) - R_n(\eta_t, \eta_0)$ using Rivoirard and Rousseau's [30] calculations pages 1504–1505. Indeed, writing $h = \sqrt{n}(\eta - \eta_0)$ we have

$$
R_n(\eta, \eta_0) - R_n(\eta_t, \eta_0) = t\langle h, \tilde{\psi}_{f_0}\rangle_L - \frac{t^2}{2}\|\tilde{\psi}_{f_0}\|_L^2 + n\log F\big[e^{-t\tilde{\psi}_{f_0}/\sqrt{n}}\big]
$$

and expanding the last term as in page 1506 of [30] we obtain that

$$\begin{aligned} n\log F\big[e^{-t\tilde{\psi}_{f_0}/\sqrt{n}}\big] &= n\log\Big(1-\frac{t}{n}\langle h,\tilde{\psi}_{f_0}\rangle_L - \frac{t}{\sqrt{n}}\mathcal{B}(f,f_0)+\frac{t^2}{2n}\|\tilde{\psi}_{f_0}\|_L^2 \\ &\quad + \frac{t^2}{2n}(F-F_0)\big(\tilde{\psi}_{f_0}^2\big)+O\big(n^{-3/2}\big)\Big) \\ &= -t\langle h,\tilde{\psi}_{f_0}\rangle_L - t\sqrt{n}\mathcal{B}(f,f_0)+\frac{t^2}{2}\|\tilde{\psi}_{f_0}\|_L^2 \\ &\quad + O\big(\|f-f_0\|_1+n^{-1/2}\big) \\ &= -t\langle h,\tilde{\psi}_{f_0}\rangle_L - t\sqrt{n}\mathcal{B}(f,f_0)+\frac{t^2}{2}\|\tilde{\psi}_{f_0}\|_L^2+o(1) \end{aligned}$$

since $|(F-F_0)(\tilde{\psi}_{f_0}^2)\le\|\tilde{\psi}_{f_0}\|_\infty^2\|f-f_0\|_1\lesssim\varepsilon_n$ on A_n. Finally, this implies that $\sqrt{n}tr(\eta,\eta_0)+R_n(\eta,\eta_0)-R_n(\eta_t,\eta_0)=o(1)$ uniformly over A_n and Assumption A is satisfied.

6.3. *Proof of Theorem 4.2.* The first part of the proof consists in establishing that the posterior distribution on random histograms concentrates (a) given the number of bins k, around the projection $f_{0,[k]}$ of f_0, and (b) globally around f_0 in terms of the Hellinger distance.

More precisely, (a) there exist $c, M>0$ such that

$$P_0\Big[\exists k\le\frac{n}{\log n};\,\Pi\big[f\notin A_{n,k}(M)|Y^n,k\big]>e^{-ck\log n}\Big]=o(1). \tag{6.2}$$

(b) Suppose $f_0\in\mathcal{C}^\beta$ with $0<\beta\le 1$. If $k_n(\beta)=(n/\log n)^{1/(2\beta+1)}$ and $\varepsilon_n(\beta)=k_n(\beta)^{-\beta}$, then for k_1, M large enough,

$$\Pi\big[h(f_0,f)\le M\varepsilon_n(\beta);\,k\le k_1k_n(\beta)|Y^n\big]=1+o_p(1). \tag{6.3}$$

Both results are new. As (a)–(b) are an intermediate step and concern rates rather than BvM per se, their proofs are given in Castillo and Rousseau [15].

We now prove that the BvM holds if there exists $\mathcal{K}_n$ such that $\Pi(\mathcal{K}_n|Y^n)=1+o_p(1)$, and for which

$$\sup_{k\in\mathcal{K}_n}\sqrt{n}|\hat{\psi}-\hat{\psi}_k|=o_p(1),\qquad \sup_{k\in\mathcal{K}_n}|V_k-V|=o_p(1), \tag{6.4}$$

for all $\psi(f)$ satisfying (4.2) with

$$\sup_{k\in\mathcal{K}_n}\sup_{f\in A_{n,k}(M)}\tilde{r}(f;f_0)=o_p(1). \tag{6.5}$$

Consider first the deterministic $k=K_n$ number of bins case. The study of the posterior distribution of $\sqrt{n}(\psi(f)-\hat{\psi})$ is based on a slight modification of the

proof of Theorem 4.1. Instead of taking the true f_0 as basis point for the LAN expansion, we take instead $f_{0,[k]}$. This enables to write the main terms in the LAN expansion completely within $\mathcal{H}_k$.

Let us define $\bar{\psi}_{(k)} := \psi_{[k]} - \int \psi_{[k]} f_{0,[k]} = \tilde{\psi}_{[k]} - \int \tilde{\psi}_{[k]} f_{0,[k]}$ and $\hat{\psi}_k = \psi(f_{0,[k]}) + \frac{1}{\sqrt{n}} W_n(\bar{\psi}_{(k)})$. With the same notation as in Section 4, where indexation by k means that f_0 is replaced by $f_{0,[k]}$ [in $\|\cdot\|_{L,k}$, $R_{n,k}$, etc., where one can note that for $g \in \mathcal{H}_k$, one has $W_{n,k}(g) = W_n(g)$],

$$\begin{aligned} t\sqrt{n}\big(\psi(f) - \hat{\psi}_k\big) + \ell_n(f) - \ell_n(f_{0,[k]}) &= -\frac{n}{2}\left\|\log\frac{f}{f_{0,[k]}} - \frac{t}{\sqrt{n}}\bar{\psi}_{(k)}\right\|_{L,k}^2 \\ &\quad + \sqrt{n} W_n\left(\log\frac{f}{f_{0,[k]}} - \frac{t}{\sqrt{n}}\bar{\psi}_{(k)}\right) \\ &\quad + \frac{t^2}{2}\|\bar{\psi}_{(k)}\|_{L,k}^2 + t\sqrt{n}\mathcal{B}_{n,k} + R_{n,k}(f, f_{0,[k]}). \end{aligned}$$

Let us set $f_{t,k} = f e^{-t\bar{\psi}_{(k)}/\sqrt{n}} / F(e^{-t\bar{\psi}_{(k)}/\sqrt{n}})$. Then, using the same arguments as in Section 4, together with (6.2) and the fact that $\int \bar{\psi}_{(k)} f_{0,[k]} = 0$,

$$\begin{aligned} t\sqrt{n}\big(\psi(f) - \hat{\psi}_k\big) + \ell_n(f) - \ell_n(f_{0,[k]}) &= \frac{t^2}{2}\|\bar{\psi}_{(k)}\|_{L,k}^2 + \ell_n(f_{t,k}) - \ell_n(f_{0,[k]}) + o(1), \end{aligned}$$

so that choosing $A_{n,k} = \{\omega \in \mathcal{S}_k; \|f_{\omega,k} - f_{0[k]}\|_1 \le M\sqrt{k\log n/n}\}$, we have

$$\begin{aligned} &E^{\Pi}\big[e^{t\sqrt{n}(\psi(f)-\hat{\psi}_k)} | Y^n, A_{n,k}\big] \\ &\quad = e^{(t^2/2)\|\bar{\psi}_{(k)}\|_{L,k}^2 + o(1)} \times \frac{\int_{A_{n,k}} e^{\ell_n(f_{t,k}) - \ell_n(f_{0[k]})}\, d\Pi_k(f)}{\int_{A_{n,k}} e^{\ell_n(f) - \ell_n(f_{0[k]})}\, d\Pi_k(f)}, \end{aligned}$$

uniformly over $k = o(n/\log n)$. Within each model $\mathcal{H}_k$, since $f = f_{\omega,k}$, we can express $f_{t,k} = k\sum_{j=1}^k \zeta_j \mathbb{1}_{I_j}$, with

$$\zeta_j = \frac{\omega_j \gamma_j^{-1}}{\sum_{j=1}^k \omega_j \gamma_j^{-1}}, \tag{6.6}$$

where we have set, for $1 \le j \le k$, $\gamma_j = e^{t\bar{\psi}_j/\sqrt{n}}$, and $\bar{\psi}_j := k\int_{I_j} \bar{\psi}_{(k)}$. Denote $S_{\gamma^{-1}}(\omega) = \sum_{j=1}^k \omega_j \gamma_j^{-1}$. Note that (6.6) implies $S_{\gamma^{-1}}(\omega) = S_\gamma(\zeta)^{-1}$. So,

$$\frac{\Pi_k(\omega)}{\Pi_k(\zeta)} = \prod_{j=1}^k e^{t(\alpha_{j,k}-1)\bar{\psi}_j/\sqrt{n}} S_\gamma(\zeta)^{-\sum_{j=1}^k(\alpha_{j,k}-1)}.$$

Let Δ be the Jacobian of the change of variable computed in Lemma 5 of the supplemental article (Castillo and Rousseau [15]). Over the set $A_{n,k}$, it holds

$$\begin{aligned} &d\Pi_k(\omega) \\ &\quad= \prod_{j=1}^{k} e^{t(\alpha_{j,k}-1)\bar{\psi}_j/\sqrt{n}} S_\gamma(\zeta)^{-\sum_{j=1}^{k}(\alpha_{j,k}-1)} \Delta(\zeta)\, d\Pi_k(\zeta) \\ &\quad= S_\gamma(\zeta)^{-\sum_{j=1}^{k}\alpha_{j,k}} e^{t\sum_{j=1}^{k}\alpha_{j,k}\bar{\psi}_j/\sqrt{n}}\, d\Pi_k(\zeta) \\ &\quad= e^{t\sum_{j=1}^{k}\alpha_{j,k}\bar{\psi}_j/\sqrt{n}}\left(1 - \frac{t}{\sqrt{n}}\int_0^1 \bar{\psi}_{(k)}(f - f_0) + O\big(n^{-1}\big)\right)^{\sum_{j=1}^{k}\alpha_{j,k}} d\Pi_k(\zeta), \end{aligned}$$

where we have used that

$$S_{\gamma^{-1}}(\omega) = \int_0^1 e^{-t\bar{\psi}_{(k)}/\sqrt{n}} f = 1 - \frac{t}{\sqrt{n}}\int_0^1 \bar{\psi}_{(k)}(f - f_0) + O\big(n^{-1}\big).$$

Moreover, if $\|\omega - \omega^0\|_1 \le M\sqrt{k\log n}/\sqrt{n}$,

$$\big\|\zeta - \omega^0\big\|_1 \le M\sqrt{k\log n}/\sqrt{n} + \frac{2|t|\|\tilde{\psi}\|_\infty}{\sqrt{n}} \le (M+1)\frac{\sqrt{k\log n}}{\sqrt{n}}$$

and vice versa. Hence, choosing M large enough (independent of k) such that

$$\Pi\big[\|\omega - \omega^0\|_1 \le (M-1)\sqrt{k\log n/n}|Y^n, k\big] = 1 + o_p(1)$$

implies that if $\sum_{j=1}^{k}\alpha_j = o(\sqrt{n})$, noting $\|\bar{\psi}_{(k)}\|_{L,k} = \|\tilde{\psi}_{[k]}\|_L$,

$$E^\Pi\big[e^{t\sqrt{n}(\psi(f)-\hat{\psi}_{(k)})}|Y^n, A_{n,k}\big] = e^{t^2\|\tilde{\psi}_{[k]}\|_L^2/2}\big(1 + o(1)\big). \tag{6.7}$$

The last estimate is for the restricted distribution $\Pi[\cdot|Y^n, A_{n,k}]$, but (6.2) implies that the unrestricted version also follows. Since $\|\tilde{\psi}\|_L^2$ is the efficiency bound for estimating ψ in the density model, (6.4) follows.

Now we turn to the random k case. The previous proof can be reproduced k by k, that is, one decomposes the posterior $\Pi[\cdot|Y^n, B_n]$, for $B_n = \bigcup_{1\le k\le n} A_{n,k} \cap \{f = f_{\omega,k}, k \in \mathcal{K}_n\}$, into the mixture of the laws $\Pi[\cdot|Y^n, B_n, k]$ with weights $\Pi[k|Y^n]$. Combining the assumption on $\mathcal{K}_n$ and (6.2) yields $\Pi[B_n|Y^n] = 1 + o_p(1)$. Now notice that in the present context (6.7) becomes

$$\begin{aligned} E^\Pi\big[e^{t\sqrt{n}(\psi(f)-\hat{\psi}_{(k)})}|Y^n, B_n, k\big] &= E^\Pi\big[e^{t\sqrt{n}(\psi(f)-\hat{\psi}_{(k)})}|Y^n, A_{n,k}, k\big] \\ &= e^{t^2\|\tilde{\psi}_{[k]}\|_L^2/2}\big(1 + o(1)\big), \end{aligned}$$

where it is important to note that the $o(1)$ is uniform in k. This follows from the fact that the proof in the deterministic case holds for any given k less than n and

any dependence in k has been made explicit in that proof. Thus,

$$E^{\Pi}\big[e^{t\sqrt{n}(\psi(f)-\hat{\psi})}|Y^n, B_n\big] = \sum_{k\in\mathcal{K}_n} E^{\Pi}\big[e^{t\sqrt{n}(\psi(f)-\hat{\psi}_k)}|Y^n, A_{n,k}, k\big]\Pi\big[k|Y^n\big]$$
$$= (1+o(1))\sum_{k\in\mathcal{K}_n} e^{t^2V_k/2+t\sqrt{n}(\hat{\psi}_k-\hat{\psi})}\Pi\big[k|Y^n\big].$$

Using (6.4) together with the continuous mapping theorem for the exponential function yields that the last display converges in probability to $e^{t^2V/2}$ as $n\to\infty$, which leads to the BvM theorem.

We apply this to the four examples. First, in the case of Example 4.1 with deterministic $k = K_n$, we have by definition that $\tilde{r}(f, f_0) = 0$ and $\sqrt{n}(\hat{\psi}_{K_n} - \hat{\psi}) = b_{n,K_n} + o_p(1)$ with $b_{n,K_n} = O(\sqrt{n}K_n^{-\beta-\gamma}) = o(1)$ if $\beta+\gamma > 1$, when $a \in \mathcal{C}^\gamma$. On the other hand, if $a(x) = \mathbb{1}_{x\le z}$, for all $\beta > 0$,

$$|b_{n,K_n}| \lesssim \sqrt{n}\left|\int_{\lfloor K_nz\rfloor/K_n}^{z}\big(f_0(x) - kw^0_{\lfloor K_nz\rfloor}\big)\,dx\right| = O\big(\sqrt{n}K_n^{-(\beta+1)}\big) = o(1).$$

We now verify (6.4) together with (6.5) for Examples 4.2, 4.3 and 4.4. We present the proof in the case Example 4.2, since the other two are treated similarly. Set, in the random k case

$$\mathcal{K}_n = \big\{k \in \big[1, k_1k_n(\beta)\big], \exists f \in \mathcal{H}^1_k, h(f, f_0) \le M\varepsilon_n(\beta)\big\},$$

for some k_1, M large enough so that $\Pi[\mathcal{K}_n|Y^n] = 1 + o_p(1)$ from (6.2), with $\varepsilon_n(\beta) = (n/\log n)^{-\beta/(2\beta+1)}$. For $\beta > 1/2$, note that $k\varepsilon_{n,k}^2 \lesssim k\varepsilon_n(\beta)^2 = o(1)$, uniformly over $k \lesssim k_n(\beta)$. In the deterministic case, simply set $\mathcal{K}_n = \{K_n\}$.

First, observe that for $k \in \mathcal{K}_n$, the elements of the set $\{f \in \mathcal{H}^1_k, h(f, f_0) \le M\varepsilon_n(\beta)\}$ are bounded away from 0 and ∞. Indeed, since this is true for f_0, writing the Hellinger distance as a sum over the various bins leads to $\sqrt{f(x)} \ge \sqrt{c_0} - \varepsilon_{n,k}\sqrt{k}$ which implies that $f(x) \ge c_0/2$ for n large enough, since $k\varepsilon_n^2 = o(1)$. Similarly, $\|f\|_\infty \le 2\|f_0\|_\infty$ for n large. Now, by writing $\log(f/f_0) = 1 + (f - f_0)/f_0 + \rho(f - f_0)$, and using that f/f_0 is bounded away from 0 and ∞, one easily checks that $|\tilde{r}(f, f_0)|$ in Example 4.2 is bounded from above by a multiple of $\int_0^1(f - f_0)^2$, which itself is controlled by $h(f, f_0)^2$ for f, f_0 as before. Also $\sqrt{n}\varepsilon_{n,k}^2 = o(1)$ when $\beta > 1/2$, which implies (6.5). It is easy to adapt the above computations to the case where $k = K_n = O(\sqrt{n}/(\log n)^2)$.

Next, we check condition (6.4). Since $\tilde{\psi} = \log f_0 - \psi(f_0)$, under the deterministic k-prior with $k = K_n = \lfloor n^{1/2}(\log n)^{-2}\rfloor$ and $\beta > 1/2$,

$$\left|\int_0^1 \tilde{\psi}(f_0 - f_{0[k]})\right| = \left|\int_0^1 (\tilde{\psi} - \tilde{\psi}_{[k]})(f_0 - f_{0[k]})\right| \lesssim h^2(f_0, f_{0[k]}) = o(1/\sqrt{n}).$$

In that case, the posterior distribution of $\sqrt{n}(\psi(f) - \hat{\psi})$ is asymptotically Gaussian with mean 0 and variance $\|\tilde{\psi}\|_L^2$, so the BvM theorem is valid.

Under the random k-prior, recall from the reasoning above that any f with $h(f, f_0) \leq M\varepsilon_n(\beta)$ is bounded from below and above, so the Hellinger and L^2-distances considered below are comparable. For a given $k \in \mathcal{K}_n$, by definition there exists $f_k^* \in \mathcal{H}_k^1$ with $h(f_0, f_k^*) \leq M\varepsilon_n(\beta)$, so using (6.3),

$$h^2(f_0, f_{0[k]}) \lesssim \int_0^1 (f_0 - f_{0[k]})^2(x)\,dx \leq \int_0^1 (f_0 - f_k^*)^2(x)\,dx \lesssim h^2(f_0, f_k^*) \lesssim \varepsilon_n^2(\beta).$$

This implies, using the same bound as in the deterministic-k case,

$$F_0\big((\tilde{\psi}_{[k]} - \tilde{\psi})^2\big) \lesssim h(f_0, f_{0[k]})^2 = O\big(\varepsilon_n^2(\beta)\big),$$

and that $|F_0(\tilde{\psi}_{[k]}^2) - F_0(\tilde{\psi}^2)| = o(1)$, uniformly over $k \in \mathcal{K}_n$. To control the empirical process part of (6.4), that is the second part of (4.10), one uses, for example, Lemma 19.33 in [32], which provides an upper-bound for the maximum, together with the last display. So, for random k, the BvM theorem is satisfied if $\beta > 1/2$.

SUPPLEMENTARY MATERIAL

Supplement to "A Bernstein–von Mises theorem for smooth functionals in semiparametric models" (DOI: 10.1214/15-AOS1336SUPP; .pdf). In the supplementary material, we state and prove several technical results used in the paper and provide the remaining proofs.

REFERENCES

[1] ARBEL, J., GAYRAUD, G. and ROUSSEAU, J. (2013). Bayesian optimal adaptive estimation using a sieve prior. *Scand. J. Stat.* **40** 549–570. MR3091697

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

[3] BICKEL, P. J. and KLEIJN, B. J. K. (2012). The semiparametric Bernstein–von Mises theorem. *Ann. Statist.* **40** 206–237. MR3013185

[4] BICKEL, P. J. and RITOV, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhyā Ser. A* **50** 381–393. MR1065550

[5] BICKEL, P. J. and RITOV, Y. (2003). Nonparametric estimators which can be "plugged-in". *Ann. Statist.* **31** 1033–1053. MR2001641

[6] BONTEMPS, D. (2011). Bernstein–von Mises theorems for Gaussian regression with increasing number of regressors. *Ann. Statist.* **39** 2557–2584. MR2906878

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

[8] CAI, T. T. and LOW, M. G. (2006). Optimal adaptive estimation of a quadratic functional. *Ann. Statist.* **34** 2298–2325. MR2291501

[9] CASTILLO, I. (2008). Lower bounds for posterior rates with Gaussian process priors. *Electron. J. Stat.* **2** 1281–1299. MR2471287

[10] CASTILLO, I. (2012). A semiparametric Bernstein–von Mises theorem for Gaussian process priors. *Probab. Theory Related Fields* **152** 53–99. MR2875753

[11] CASTILLO, I. (2012). Semiparametric Bernstein–von Mises theorem and bias, illustrated with Gaussian process priors. *Sankhyā* **74** 194–221. MR3021557

[12] CASTILLO, I. (2014). On Bayesian supremum norm contraction rates. *Ann. Statist.* **42** 2058–2091. MR3262477

[13] CASTILLO, I. and NICKL, R. (2013). Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **41** 1999–2028. MR3127856

[14] CASTILLO, I. and NICKL, R. (2014). On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *Ann. Statist.* **42** 1941–1969. MR3262473

[15] CASTILLO, I. and ROUSSEAU, J. (2015). Supplement to "A Bernstein–von Mises theorem for smooth functionals in semiparametric models." DOI:10.1214/15-AOS1336SUPP.

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

[17] DE BLASI, P. and HJORT, N. L. (2009). The Bernstein–von Mises theorem in semiparametric competing risks models. *J. Statist. Plann. Inference* **139** 2316–2328. MR2507993

[18] EFROMOVICH, S. and LOW, M. (1996). On optimal adaptive estimation of a quadratic functional. *Ann. Statist.* **24** 1106–1125. MR1401840

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

[20] GAYRAUD, G. and TRIBOULEY, K. (1999). Wavelet methods to estimate an integrated quadratic functional: Adaptivity and asymptotic law. *Statist. Probab. Lett.* **44** 109–122. MR1706448

[21] GHOSAL, S. (1999). Asymptotic normality of posterior distributions in high-dimensional linear models. *Bernoulli* **5** 315–331. MR1681701

[22] GHOSAL, S. and VAN DER VAART, A. W. (2007). Convergence rates of posterior distributions for noniid observations. *Ann. Statist.* **35** 192–223. MR2332274

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

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

[25] KNAPIK, B. T., SZABÓ, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2012). Bayes procedures for adaptive inference in inverse problems for the white noise model. Available at arXiv:1209.3628.

[26] KNAPIK, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2011). Bayesian inverse problems with Gaussian priors. *Ann. Statist.* **39** 2626–2657. MR2906881

[27] KRUIJER, W. and ROUSSEAU, J. (2013). Bayesian semi-parametric estimation of the long-memory parameter under FEXP-priors. *Electron. J. Stat.* **7** 2947–2969. MR3151758

[28] LAURENT, B. (1996). Efficient estimation of integral functionals of a density. *Ann. Statist.* **24** 659–681. MR1394981

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

[30] RIVOIRARD, V. and ROUSSEAU, J. (2012). Bernstein–von Mises theorem for linear functionals of the density. *Ann. Statist.* **40** 1489–1523. MR3015033

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

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

[33] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2008). Rates of contraction of posterior distributions based on Gaussian process priors. *Ann. Statist.* **36** 1435–1463. MR2418663

[34] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2008). Reproducing kernel Hilbert spaces of Gaussian priors. In *Pushing the Limits of Contemporary Statistics*: *Contributions in Honor of Jayanta K. Ghosh*. *Inst. Math. Stat. Collect.* **3** 200–222. IMS, Beachwood, OH. MR2459226

CNRS–LPMA
UNIVERSITÉS PARIS VI & VII
BÂTIMENT SOPHIE GERMAIN
75205 PARIS CEDEX 13
FRANCE
E-MAIL: ismael.castillo@math.cnrs.fr

CREST–ENSAE
LABORATOIRE DE STATISTIQUES
3 AVENUE PIERRE LAROUSSE
92245 MALAKOFF CEDEX
FRANCE
AND
UNIVERSITÉ PARIS DAUPHINE
PLACE DU MARÉCHAL DELATTRE DE TASSIGNY
75016 PARIS
FRANCE
E-MAIL: rousseau@ceremade.dauphine.fr