

ON THE BERNSTEIN–VON MISES PHENOMENON FOR NONPARAMETRIC BAYES PROCEDURES

BY ISMAËL CASTILLO¹ AND RICHARD NICKL

CNRS and University of Cambridge

We continue the investigation of Bernstein–von Mises theorems for nonparametric Bayes procedures from [*Ann. Statist.* **41** (2013) 1999–2028]. We introduce multiscale spaces on which nonparametric priors and posteriors are naturally defined, and prove Bernstein–von Mises theorems for a variety of priors in the setting of Gaussian nonparametric regression and in the i.i.d. sampling model. From these results we deduce several applications where posterior-based inference coincides with efficient frequentist procedures, including Donsker– and Kolmogorov–Smirnov theorems for the random posterior cumulative distribution functions. We also show that multiscale posterior credible bands for the regression or density function are optimal frequentist confidence bands.

1. Introduction. The Bernstein–von Mises (BvM) theorem constitutes a powerful and precise tool to study Bayes procedures from a frequentist point of view. It gives universal conditions on the prior under which the posterior distribution has the approximate shape of a normal distribution. The theorem is well understood in finite-dimensional models (see [30] and [35]), but involves some delicate conceptual and mathematical issues in the infinite-dimensional setting. There exists a Donsker-type BvM theorem for the cumulative distribution function based on Dirichlet process priors, see Lo [31], and this carries over to a variety of closely related nonparametric situations, including quantile inference and censoring models, where Bernstein–von Mises results are available: see [8, 9, 21, 26, 27] and [22]. The proofs of these results rely on a direct analysis of the posterior distribution, which is explicitly given in these settings (and typically of Dirichlet form).

When considering general priors that model potentially smoother nonparametric objects such as densities or regression functions, the BvM phenomenon appears to be much less well understood. Notably, Freedman [14] has shown that in a basic Gaussian conjugate ℓ_2 -sequence space setting, the BvM theorem does not hold true in generality; see also the related recent contributions [24, 28]. In contrast, in the recent paper [4], nonparametric BvM theorems have been proved in a topology that

Received October 2013; revised March 2014.

¹Supported in part by ANR Grants “Banhdits” ANR-2010-BLAN-0113-03 and “Calibration” ANR-2011-BS01-010-01.

MSC2010 subject classifications. Primary 62G20; secondary 62G15, 62G08.

Key words and phrases. Bayesian inference, posterior asymptotics, multiscale statistics.

is weaker than the one of ℓ_2 , and it was shown that such results can be useful for several nonparametric problems, including the ℓ_2 -setting, when applied with care. An important consequence is that, in contrast to the finite-dimensional situation, whether a nonparametric posterior credible set is a frequentist confidence set or not depends in a possibly quite subtle way on the *geometry* of the set.

The results in [4] are confined to the most basic nonparametric model—Gaussian white noise—and strongly rely on Hilbert space techniques. The main novelties of the present paper are: (a) extensions of the results in [4] to the i.i.d. sampling model and (b) the derivation of sharp Bernstein–von Mises results in spaces whose geometry resembles an ℓ_∞ -type space and whose norms are strong enough to allow one to deduce some fundamental new applications to posterior credible bands and Kolmogorov–Smirnov type results. Our results are based on mathematical tools developed recently in Bayesian nonparametrics, particularly the papers [3, 5] and also [32]. These give sub-Gaussian estimates on fixed (semi-parametric) functionals of posterior distributions over well-chosen events in the support of the posterior, which in turn can be used to control the supremum-type norms relevant in our context via concentration properties of maxima of sub-Gaussian variables.

Let us outline some applications of our results: consider a prior distribution Π on a family \mathcal{F} of probability densities f , such as a random Dirichlet histogram or a Gaussian series prior on the log-density. Let $\Pi(\cdot|X_1, \dots, X_n)$ be the posterior distribution obtained from observing $X_1, \dots, X_n \sim^{\text{i.i.d.}} f$. It is of interest to study the induced posterior distribution on the cumulative distribution function F of f . Making the “frequentist” assumption $X_i \sim^{\text{i.i.d.}} P_0$, the stochastic fluctuations of F around the empirical distribution function $F_n(\cdot) = (1/n) \sum_{i=1}^n 1_{[0, \cdot]}(X_i)$ under the posterior distribution will be shown to be approximately those of a P_0 -Brownian bridge G_{P_0} : under the law $P_0^{\mathbb{N}}$ of (X_1, X_2, \dots) the distributional approximation ($n \rightarrow \infty$)

$$(1) \quad \sqrt{n}(F - F_n)|X_1, \dots, X_n \approx G_{P_0}$$

holds true, in a sense to be made fully precise below (Corollary 1). This parallels Lo’s [31] results for the Dirichlet process and can be used to validate Bayesian Kolmogorov–Smirnov tests and credible bands from a frequentist point of view. Note, however, that unlike the results in [31], our techniques are not at all based on any conjugate analysis and open the door to the derivation of Bernstein–von Mises results in general settings of Bayesian nonparametrics. We also note that (1) is comparable to central limit theorems $\sqrt{n}(F_n^b - F_n) \rightarrow G_{P_0}$ in $P_0^{\mathbb{N}}$ -probability for bootstrapped empirical measures F_n^b ; see the classical paper [18]. This illustrates how BvM theorems are in some sense the Bayesian versions of bootstrap consistency results.

Our results also have important applications for inference on the more difficult functional parameter f itself. For instance, we will show that certain $1 - \alpha$ posterior credible sets for a density or regression function are also frequentist optimal, asymptotically exact level $1 - \alpha$ confidence bands.

Before we explain these applications in detail it is convenient to shed some more light on our general setting. The spaces in which we derive BvM-type results are in principle abstract and dictated by the applications we have in mind. They are, however, connected to the frequentist literature on nonparametric multiscale inference, as developed in the papers [10–13, 33], where also many further references can be found. This connection gives a further motivation for our general setting as well as heuristics for the inference procedures we suggest here. Let us thus explain some main ideas behind the multiscale approach in the simple regression framework of observing a signal in Gaussian white noise

$$(2) \quad dX^{(n)}(t) = f(t) dt + \frac{1}{\sqrt{n}} dW(t), \quad t \in [0, 1], n \in \mathbb{N},$$

which can also be written $\mathbb{X}^{(n)} = f + \mathbb{W}/\sqrt{n}$, with \mathbb{W} a standard white noise; see (13) below for details. The i.i.d. sampling model, which will be treated below, gives rise to similar intuitions after replacing $\mathbb{X}^{(n)} - f$ by $P_n - P$ where $P_n = (1/n) \sum_{i=1}^n \delta_{X_i}$ is the empirical measure from a sample from law P with density f . One introduces a double-indexed family of linear multiscale functionals

$$f \mapsto 2^{l/2} \int_0^1 \psi(2^l x - k) f(x) dx \equiv \langle f, \psi_{lk} \rangle,$$

where l is a scaling parameter which has $O(2^l)$ associated location indices k . The prototypical example that we will focus on is to take a Haar wavelet $\psi = \mathbb{1}_{(0,1/2]} - \mathbb{1}_{(1/2,1]}$, or a more general wavelet function ψ generating a frame or orthonormal basis $\{\psi_{lk}\}$ of L^2 . The projection of $\mathbb{X}^{(n)} - f$ onto the first $\leq J$ scales gives rise to random variables

$$\sqrt{n} \langle \mathbb{X}^{(n)} - f, \psi_{lk} \rangle = \langle \mathbb{W}, \psi_{lk} \rangle \equiv g_{lk} \sim N(0, 1), \quad k, l \leq J,$$

and the maximum over all these statistics scaled by \sqrt{l}

$$(3) \quad Z_J \equiv \sqrt{n} \max_{l \leq J, k} \frac{|\langle \mathbb{X}^{(n)} - f, \psi_{lk} \rangle|}{\sqrt{l}} = \max_{l \leq J, k} \frac{|g_{lk}|}{\sqrt{l}},$$

has a canonical distribution under the null hypothesis $H_0 = \{f\}$. The quantity Z_J is often called a *multiscale statistic*, and the quantiles of its distribution are used to test hypotheses on f . One can also construct confidence sets C_n by simply taking C_n to consist of all those f that satisfy simultaneously all the linear constraints

$$\frac{|\langle \mathbb{X}^{(n)} - f, \psi_{lk} \rangle|}{\sqrt{l}} \leq c_n \quad \forall k, l,$$

where c_n are suitable constants chosen in dependence of the distribution of Z_J . Intersecting these linear restrictions with further qualitative information about f , such as smoothness or shape constraints, can be shown to give optimal frequentist confidence sets (as, e.g., in Propositions 1 and 4 below).

A key challenge in the multiscale approach is of course the analysis of the distribution of the random variables Z_J . One approach is to re-center Z_J by a quantity of order \sqrt{J} and to use extreme value theory to obtain a Gumbel approximation of the distribution of these random variables. The slow convergence rates (as $J \rightarrow \infty$) of such limit theorems are often not satisfactory; see, for example, [19]. Instead we shall introduce certain sequence spaces in which direct Gaussian asymptotics can be obtained for multiscale statistics (without re-centring). This allows for faster convergence rates (by using standard Berry–Esseen bounds for the central limit theorem). It is also naturally compatible with a Bayesian approach to multiscale inference: one distributes independent random variables across the scales l and locations k , corresponding to a random series prior common in Bayesian nonparametrics. The posterior distribution then allows one effectively to “bootstrap” the law of Z_J , and our BvM-results in multiscale spaces will give a full frequentist justification of this approach.

Let us illustrate the last point in a key example involving a histogram prior $\Pi_L, L \in \mathbb{N}$, equal to the law of the random probability density

$$(4) \quad f \sim \sum_{k=0}^{2^L-1} h_k \mathbb{1}_{I_k^L}, \quad I_0^L = [0, 2^{-L}], I_k^L = (k2^{-L}, (k+1)2^{-L}), k \geq 1,$$

where the h_k are drawn from a $\mathcal{D}(1, \dots, 1)$ -Dirichlet distribution on the unit simplex of \mathbb{R}^{2^L} . Let $\Pi(\cdot | X_1, \dots, X_n)$ denote the resulting posterior distribution based on observing X_1, \dots, X_n i.i.d. from density f . For any sequence (w_l) such that $w_l/\sqrt{l} \uparrow \infty$ as $l \rightarrow \infty$ and for standard Haar wavelets

$$\psi_{-10} = \mathbb{1}_{[0,1]}, \quad \psi_{lk} = 2^{l/2}(\mathbb{1}_{(k/2^l, (k+1/2)/2^l]} - \mathbb{1}_{((k+1/2)/2^l, (k+1)/2^l]}),$$

with indices $l \in \mathbb{N} \cup \{-1, 0\}, k = 0, \dots, 2^l - 1$, define

$$(5) \quad C_n \equiv \left\{ f : \max_{k,l \leq L} \frac{|\langle f - P_n, \psi_{lk} \rangle|}{w_l} \leq \frac{R_n}{\sqrt{n}} \right\},$$

where $\langle P_n, \psi_{lk} \rangle = n^{-1} \sum_{i=1}^n \psi_{lk}(X_i)$ are the empirical wavelet coefficients and where $R_n = R(\alpha, X_1, \dots, X_n)$ are random constants chosen such that

$$\Pi(C_n | X_1, \dots, X_n) = 1 - \alpha, \quad 0 < \alpha < 1.$$

Any set C_n satisfying the identity in the last display is a *posterior credible set* of level $1 - \alpha$, or simply a $(1 - \alpha)$ -credible set. Note that in this example the posterior distribution, and hence R_n , can be explicitly computed due to conjugacy of the Dirichlet distribution under multinomial sampling (i.e., counting observation points in dyadic bins I_k^L).

PROPOSITION 1. *Consider the random histogram prior Π from (4) where $L = L_n$ is such that $2^{L_n} \sim (n/\log n)^{1/2(\gamma+1)}$. Let C_n be as in (5). Suppose X_1, \dots, X_n*

are i.i.d. from law P_0 with density f_0 satisfying the Hölder condition

$$\sup_{x,y \in [0,1], x \neq y} \frac{|f_0(x) - f_0(y)|}{|x - y|^\gamma} < \infty, \quad 1/2 < \gamma \leq 1.$$

Then we have as $n \rightarrow \infty$,

$$(6) \quad P_0^{\mathbb{N}}(f_0 \in C_n) \rightarrow 1 - \alpha.$$

Moreover, if $u_n = w_{L_n} / \sqrt{L_n}$, then (6) remains true with C_n replaced by

$$\bar{C}_n = C_n \cap \{f : |\langle f, \psi_{lk} \rangle| \leq u_n 2^{-l(\gamma+1/2)} \forall k, l\},$$

and the diameter $|\bar{C}_n|_\infty = \sup\{\|f - g\|_\infty : f, g \in \bar{C}_n\}$, satisfies

$$(7) \quad |\bar{C}_n|_\infty = O_{P_0^{\mathbb{N}}}\left(\left(\frac{\log n}{n}\right)^{\gamma/(2\gamma+1)} u_n\right).$$

We conclude that the $(1 - \alpha)$ -credible set C_n is an exact asymptotic frequentist $(1 - \alpha)$ -confidence set. Following the multiscale approach, the same is true for \bar{C}_n obtained from intersecting C_n with a γ -Hölder constraint (expressed through the decay of the Haar wavelet coefficients). The L^∞ -diameter of \bar{C}_n shrinks at the optimal rate if the true density f_0 is also γ -Hölder (noting $u_n \rightarrow \infty$ as slowly as desired). For the proof see Section 4.2.

A summary of this article is as follows: in the next section we introduce the multiscale framework and the statistical sampling models and show how to construct efficient frequentist estimators in them. In Section 3 we introduce the Bayesian approach, formulate a general notion of a nonparametric Bernstein–von Mises phenomenon in multiscale spaces and prove that the phenomenon occurs for a variety of relevant nonparametric prior distributions, including Gaussian series priors and random histograms. In Section 4 we discuss statistical applications to Donsker–Kolmogorov–Smirnov theorems and credible bands. Section 5 contains the proofs.

2. The general framework. We use the usual notation for $L^p = L^p([0, 1])$ -spaces of integrable functions, and we denote by ℓ_p the usual sequence spaces. The usual supremum norm is denoted by $\|\cdot\|_\infty$. Throughout we consider an S -regular, $S \geq 0$, wavelet basis

$$(8) \quad \{\psi_{lk} : l \geq J_0 - 1, k = 0, \dots, 2^l - 1\}, \quad J_0 \in \mathbb{N} \cup \{0\},$$

of $L^2([0, 1])$ (by convention we denote the usual “scaling function” φ as the first wavelet $\psi_{(J_0-1)0}$). We restrict to Haar wavelets ($S = J_0 = 0$), periodised wavelet bases ($J_0 = 0, S > 0$) or boundary corrected wavelet bases ($S > 0, J_0 = J_0(S)$ large enough, see [7]). Functions $f \in L^2$ generate double-indexed sequences $\{\langle f, \psi_{lk} \rangle = \int_0^1 f \psi_{lk}\}$, and conversely any sequence (x_{lk}) generates wavelet series of (possibly generalised) functions $\sum_{k,l} x_{lk} \psi_{lk}$ on $[0, 1]$.

We define Hölder-type spaces C^s of continuous functions on $[0, 1]$:

$$(9) \quad C^s([0, 1]) = \left\{ f \in C([0, 1]) : \|f\|_{s,\infty} := \sup_{l,k} 2^{l(s+1/2)} |\langle \psi_{lk}, f \rangle| < \infty \right\}.$$

When the wavelets are regular enough, this norm characterises the scale of Hölder (–Zygmund when $s \in \mathbb{N}$) spaces. Otherwise we work with the spaces defined through decay of the multiscale coefficients, which still contain the classical s -Hölder spaces by standard results in wavelet theory.

Convergence in distribution of random variables $X_n \rightarrow^d X$ in a metric space (S, d) can be metrised by metrising weak convergence of the induced laws $\mathcal{L}(X_n)$ to $\mathcal{L}(X)$ on S . For convenience we work with the bounded-Lipschitz metric β_S : let μ, ν be probability measures on (S, d) , and define

$$(10) \quad \beta_S(\mu, \nu) \equiv \sup_{F: \|F\|_{BL} \leq 1} \left| \int_S F(x) (d\mu(x) - d\nu(x)) \right|,$$

$$\|F\|_{BL} = \sup_{x \in S} |F(x)| + \sup_{x \neq y, x, y \in S} \frac{|F(x) - F(y)|}{d(x, y)}.$$

2.1. *Multiscale spaces.* For monotone increasing weighting sequences $w = (w_l : l \geq J_0 - 1), w_l \geq 1$, we define multiscale sequence spaces

$$(11) \quad \mathcal{M} \equiv \mathcal{M}(w) \equiv \left\{ x = \{x_{lk}\} : \|x\|_{\mathcal{M}(w)} \equiv \sup_l \frac{\max_k |x_{lk}|}{w_l} < \infty \right\}.$$

The space $\mathcal{M}(w)$ is a nonseparable Banach space (it is isomorphic to ℓ_∞). The (weighted) sequences in $\mathcal{M}(w)$ that vanish at infinity form a separable closed subspace for the same norm

$$(12) \quad \mathcal{M}_0 = \mathcal{M}_0(w) = \left\{ x \in \mathcal{M}(w) : \lim_{l \rightarrow \infty} \max_k \frac{|x_{lk}|}{w_l} = 0 \right\}.$$

We notice that $w_l \geq 1$ implies $\|x\|_{\mathcal{M}} \leq \|x\|_{\ell_2}$ so that \mathcal{M} always contains ℓ_2 . For suitable divergent weighting sequences (w_l) , these spaces contain objects that are much less regular than ℓ_2 -sequences, such as a Gaussian white noise dW . The action of dW on $\{\psi_{lk}\}$ generates an i.i.d. sequence g_{lk} of standard $N(0, 1)$'s, hence whether dW defines a Gaussian Borel random variable \mathbb{W} in \mathcal{M}_0 or not depends entirely on the weighting function w .

DEFINITION 1. Call a sequence (w_l) admissible if $w_l/\sqrt{l} \uparrow \infty$ as $l \rightarrow \infty$.

PROPOSITION 2. Let $\mathbb{W} = (\int \psi_{lk} dW : l, k) = (g_{lk}), g_{lk} \sim N(0, 1)$, be a Gaussian white noise. For $\omega = (\omega_l) = \sqrt{l}$ we have $E\|\mathbb{W}\|_{\mathcal{M}(\omega)} < \infty$. If $w = (w_l)$ is admissible, then \mathbb{W} defines a tight Gaussian Borel probability measure in the space $\mathcal{M}_0(w)$.

PROOF. Since there are 2^l i.i.d. standard Gaussians $g_{lk} = \langle \psi_{lk}, dW \rangle$ at the l th level, we have from a standard bound $E \max_k |g_{lk}| \leq C\sqrt{l}$ for some universal constant C . The Borell–Sudakov–Tsirelson inequality (e.g., [29]) applied to the maximum at the l th level gives, for any M large enough,

$$\begin{aligned} & \Pr\left(\sup_l l^{-1/2} \max_k |g_{lk}| > M\right) \\ & \leq \sum_l \Pr\left(\max_k |g_{lk}| - E \max_k |g_{lk}| > \sqrt{l}M - E \max_k |g_{lk}|\right) \\ & \leq 2 \sum_l \exp\{-c(M - C)^2 l\}. \end{aligned}$$

Now using $E[X] \leq K + \int_K^\infty \Pr[X \geq t] dt$ for any real-valued random variable X and any $K \geq 0$, one obtains that $\|\mathbb{W}\|_{\mathcal{M}(\omega)}$ has finite expectation.

It now also follows immediately from the definition of the space $\mathcal{M}_0(w)$ that for any sequence $w_l/\sqrt{l} \uparrow \infty$, we have $\mathbb{W} \in \mathcal{M}_0$ almost surely. Since the latter is a separable complete metric space, \mathbb{W} is a tight Gaussian Borel random variable in it (e.g., page 374 in [1]). \square

REMARK 1 (Admissible sequences w). Assuming admissibility of w is necessary if one wants to show that \mathbb{W} is tight in $\mathcal{M}(w)$. Since weak convergence of probability measures on a complete metric space implies tightness of the limit distribution, it is in particular impossible, as will be relevant below, to converge weakly towards \mathbb{W} in $\mathcal{M}(w)$ without assuming admissibility of w . To prove that admissibility is necessary, suppose on the contrary that \mathbb{W} were tight in $\mathcal{M}(\omega)$ for some sequence $\omega_l \sim \sqrt{l}$, hence defining a Radon Gaussian measure in that space. Then by Theorem 3.6.1 in [1] the topological support of \mathbb{W} equals the completion of the RKHS ℓ_2 in the norm of the ambient Banach space $\mathcal{M}(\omega)$, which is $\mathcal{M}_0(\omega)$. Since

$$\lim_{J \rightarrow \infty} \frac{\max_k |g_{Jk}|}{\sqrt{J}} = \sqrt{2 \log 2} \neq 0$$

almost surely we have $\mathbb{W} \notin \mathcal{M}_0(\omega)$, a contradiction, so \mathbb{W} cannot be tight. The cylindrically-defined law of \mathbb{W} is in fact a “degenerate” Gaussian measure in $\mathcal{M}(\omega)$ that does (assuming the continuum hypothesis) not admit an extension to a Borel measure on $\mathcal{M}(\omega)$; see Definition 3.6.2 and Proposition 3.11.5 in [1]. It has further unusual properties: \mathbb{W} has a “hole.” That is, for some $c > 0$, $\|\mathbb{W}\|_{\mathcal{M}(\omega)} \in [c, \infty)$ almost surely (see [6]), and depending on finer properties of the sequence ω , the distribution of $\|\mathbb{W}\|_{\mathcal{M}}$ may not be absolutely continuous, and its absolutely continuous part may have infinitely many modes; see [23].

2.2. Nonparametric statistical models.

2.2.1. *Nonparametric regression.* For $f \in L^2$ consider observing a trajectory in the white noise model (2) which is a natural surrogate for a fixed design non-

parametric regression model with Gaussian errors. By Proposition 2 and since any $f \in L^2$ has wavelet coefficients $\{f_{lk}\} \in \ell_2 \subset \mathcal{M}_0(w)$, equation (2) makes rigorous sense as the tight Gaussian shift experiment

$$(13) \quad \mathbb{X}^{(n)} = f + \frac{1}{\sqrt{n}}\mathbb{W}, \quad n \in \mathbb{N},$$

in $\mathcal{M}_0(w)$ for any admissible (w_l) . We denote the law $\mathcal{L}(\mathbb{X}^{(n)})$ by P_f^n . Then

$$(14) \quad \sqrt{n}(\mathbb{X}^{(n)} - f) = \mathbb{W} \quad \text{in } \mathcal{M}_0,$$

and one deduces that $\mathbb{X}^{(n)}$ is an efficient estimator of f in \mathcal{M}_0 .

2.2.2. *The i.i.d. sampling setting.* Consider next the situation where we observe X_1, \dots, X_n i.i.d. from law P with density f on $[0, 1]$. Then a natural estimate of $\langle f, \psi_{lk} \rangle$ is given by $P_n \psi_{lk} \equiv \langle P_n, \psi_{lk} \rangle = \frac{1}{n} \sum_{i=1}^n \psi_{lk}(X_i)$. By the central limit theorem, for k, l fixed and as $n \rightarrow \infty$, the random variable $\sqrt{n}(P_n - P)(\psi_{lk})$ converges in distribution to

$$(15) \quad \mathbb{G}_P(\psi_{lk}) \sim N(0, \text{Var}_P(\psi_{lk}(X_1))).$$

In analogy to the white noise process \mathbb{W} , the process \mathbb{G}_P arising from (15) can be rigorously defined as the Gaussian process indexed by the Hilbert space

$$L^2(P) \equiv \left\{ f : [0, 1] \rightarrow \mathbb{R} : \int_0^1 f^2 dP < \infty \right\}$$

with covariance function $\mathbb{E}[\mathbb{G}_P(g)\mathbb{G}_P(h)] = \int_0^1 (g - Pg)(h - Ph) dP$. We call \mathbb{G}_P the *P-white bridge process*. An analogue of Proposition 2, and of the remark after it, holds true for \mathbb{G}_P whenever P has a bounded density.

PROPOSITION 3. *Proposition 2 holds true for the P-white bridge \mathbb{G}_P replacing \mathbb{W} whenever P has a bounded density on $[0, 1]$.*

PROOF. The proof is exactly the same, using the standard bounds

$$\text{Var}(\mathbb{G}_P(\psi_{lk})) \leq \|f\|_\infty, \quad E \max_k |\mathbb{G}_P(\psi_{lk})| \leq C \|f\|_\infty^{1/2} \sqrt{l},$$

where f denotes the density of P . \square

Any P with bounded density f has coefficients $\langle f, \psi_{lk} \rangle \in \ell_2 \subset \mathcal{M}_0(w)$. We would like to formulate a statement such as

$$\sqrt{n}(P_n - P) \rightarrow^d \mathbb{G}_P \quad \text{in } \mathcal{M}_0,$$

as $n \rightarrow \infty$, paralleling (14) in the Gaussian white noise setting. The fluctuations of $\sqrt{n}(P_n - P)(\psi_{lk})/\sqrt{l}$ along k are stochastically bounded for l such that $2^l \leq n$,

but are unbounded for high frequencies. Thus the empirical process $\sqrt{n}(P_n - P)$ will not define an element of \mathcal{M}_0 for every admissible sequence w . In our non-parametric setting we can restrict to frequencies at levels $l, 2^l \leq n$ and introduce an appropriate "projection" $P_n(j)$ of the empirical measure P_n onto V_j via

$$(16) \quad \langle P_n(j), \psi_{lk} \rangle = \begin{cases} \langle P_n, \psi_{lk} \rangle, & \text{if } l \leq j, \\ 0, & \text{if } l > j, \end{cases}$$

which defines a tight random variable in \mathcal{M}_0 . The following theorem shows that $P_n(j)$ estimates P efficiently in \mathcal{M}_0 if j is chosen appropriately. Note that the natural choice $j = L_n$ such that

$$2^{L_n} \sim N^{1/(2\gamma+1)},$$

where $N = n$ (if $\gamma > 0$) or $N = n/\log n$ (if $\gamma \geq 0$), is possible.

THEOREM 1. *Let $w = (w_l)$ be admissible. Suppose P has density f in $C^\gamma([0, 1])$ for some $\gamma \geq 0$. Let j_n be such that*

$$\sqrt{n}2^{-j_n(\gamma+1/2)}w_{j_n}^{-1} = o(1), \quad \frac{2^{j_n}j_n}{n} = O(1).$$

Then we have, as $n \rightarrow \infty$,

$$\sqrt{n}(P_n(j_n) - P) \rightarrow^d \mathbb{G}_P \quad \text{in } \mathcal{M}_0(w).$$

3. The nonparametric Bayes approach. In both regression or density estimation one constructs a prior probability distribution from which the function f is drawn, and given the observations $X = X^{(n)}$, equal to either $\mathbb{X}^{(n)} \sim P_f^n$ or X_1, \dots, X_n i.i.d. from density f , one computes the posterior distribution $\Pi(\cdot|X)$ of f . Under appropriate conditions the wavelet coefficient sequence associated to a posterior draw $f \sim \Pi(\cdot|X)$ will give rise to a random variable in \mathcal{M}_0 . If $T_n = T_n(X)$ is an efficient estimator of f in \mathcal{M}_0 , such as $\mathbb{X}^{(n)}$ or $P_n(j)$ from the previous subsections, then one can ask, following [4], for a Bernstein-von Mises type result: assuming $X \sim P_{f_0}$ for some fixed f_0 , do we have

$$(17) \quad \mathcal{L}(\sqrt{n}(f - T_n)|X) \rightarrow \mathcal{L}(\mathbb{G}) \quad \text{weakly in } \mathcal{M}_0(w) \text{ as } n \rightarrow \infty,$$

with P_{f_0} -probability close to one? Here, depending on the sampling model considered, \mathbb{G} equals either \mathbb{W} or \mathbb{G}_{P_0} , $dP_0(x) = f_0(x)dx$ and P_{f_0} stands, in slight abuse of notation, for the law $P_{f_0}^n$ of $\mathbb{X}^{(n)}$ or the law $P_0^{\mathbb{N}}$ of (X_1, X_2, \dots) .

To make such a statement rigorous we will metrize weak convergence of laws in $\mathcal{M}_0(w)$ via $\beta_{\mathcal{M}_0(w)}$ from (10), and view the prior Π on the functional parameter $f \in L^2$ as a prior on sequence space ℓ_2 under the wavelet isometry $L^2 \cong \ell_2$ [arising from an arbitrary but fixed wavelet basis (8)].

DEFINITION 2. Let w be admissible, let Π be a prior and $\Pi(\cdot|X)$ the corresponding posterior distribution on $\ell_2 \subset \mathcal{M}_0 = \mathcal{M}_0(w)$, obtained from observations X in the white noise or i.i.d. sampling model. Let $\tilde{\Pi}_n$ be the image measure of $\Pi(\cdot|X)$ under the mapping

$$\tau : f \mapsto \sqrt{n}(f - T_n),$$

where $T_n = T_n(X)$ is an estimator of f in \mathcal{M}_0 . Then we say that Π satisfies the weak Bernstein–von Mises phenomenon in \mathcal{M}_0 with centring T_n if, for $X \sim P_{f_0}$ and fixed f_0 , as $n \rightarrow \infty$,

$$\beta_{\mathcal{M}_0}(\tilde{\Pi}_n, \mathcal{N}) \xrightarrow{P_{f_0}} 0,$$

where \mathcal{N} is the law in \mathcal{M}_0 of \mathbb{W} or of \mathbb{G}_{P_0} , $f_0 \in L^\infty$, respectively.

REMARK 2. If convergence of moments (Bochner-integrals) $E[\tilde{\Pi}_n|X] \xrightarrow{P_{f_0}} E\mathcal{N} = 0$ occurs in the above limit, then we deduce

$$(18) \quad \|\tilde{f}_n - T_n\|_{\mathcal{M}_0} = o_{P_{f_0}}(1/\sqrt{n}),$$

where $\tilde{f}_n = E(f|X)$ is the posterior mean. If T_n is an efficient estimator of $f \in \mathcal{M}_0$, then (18) implies that \tilde{f}_n is so too.

In [4], Bernstein–von Mises theorems are proved in certain negative Sobolev spaces $H(\delta)$, $\delta > 1/2$, and various applications of such results are presented. A multiscale BvM result in \mathcal{M}_0 for a prior $\{f_{lk}\}$ implies a weak BvM for the prior $\sum_{k,l} f_{lk}\psi_{lk}$ in $H(\delta)$, as the following result shows. In particular all the applications from [4] carry over to the present setting.

PROPOSITION 4. *Suppose the weak Bernstein–von Mises phenomenon holds true in $\mathcal{M}_0(w)$ with (w_l) such that $\sum_l w_l^2 l^{-2\delta} < \infty$ for some $\delta > 0$. Then the weak Bernstein–von Mises phenomenon holds in $H(\delta)$.*

PROOF. The norm of $H(\delta)$ is given by (see [4], Section 1.2),

$$\begin{aligned} \|f\|_{H(\delta)}^2 &= \sum_l 2^{-l} l^{-2\delta} \sum_k |\langle f, \psi_{lk} \rangle|^2 \\ &\leq \sup_l w_l^{-2} \max_k |\langle f, \psi_{lk} \rangle|^2 \sum_l w_l^2 l^{-2\delta} \\ &\leq C \|f\|_{\mathcal{M}_0(w)}^2, \end{aligned}$$

so that the result follows from the continuous mapping theorem. \square

While the above notions of the BvM phenomenon will be shown below to be useful and feasible in nonparametric settings, there are other ways to formulate

BvM-type statements. For instance, one may investigate how the classical BvM theorem in finite-dimensions extends to parameter spaces of dimension that increases with n ; see, [2, 15, 25] for results in this direction.

Throughout the rest of this section $\mathcal{M}_0 = \mathcal{M}_0(w)$ is the space defined in (12), with w an admissible sequence as in Definition 1.

3.1. *Bernstein–von Mises theorems in $\mathcal{M}_0(w)$: Gaussian regression case.* In the white noise model (13) natural priors for f are obtained from distributing random coefficients on the ψ_{lk} 's.

CONDITION 1. Consider product priors Π arising from random functions

$$f(x) = \sum_l \sigma_l \sum_k \phi_{lk} \psi_{lk}(x), \quad x \in [0, 1],$$

where the ϕ_{lk} are i.i.d. from probability density $\varphi : \mathbb{R} \rightarrow [0, \infty)$ satisfying

$$(E) \quad \exists a, C > 0 \quad \forall x \in \mathbb{R}, \quad \varphi(x) \leq C e^{-ax^2},$$

and where $\sigma_l = 2^{-l(\alpha+(1/2))}$, $\alpha > 0$, ensuring in particular that $f \in L^2$ almost surely.

For $\mathbb{X}^{(n)} \sim P_{f_0}^n$ and f_0 with wavelet coefficients $\{\langle f_0, \psi_{lk} \rangle\} \in \ell_2$, we assume moreover that there exists a finite constant $M > 0$ such that

$$(P1) \quad \sup_{l,k} \frac{|\langle f_0, \psi_{lk} \rangle|}{\sigma_l} \leq M,$$

and that there exists $\tau > M$, $c_\varphi > 0$ such that

$$(P2) \quad \text{on } (-\tau, \tau) \text{ the density } \varphi \text{ is continuous and satisfies } \varphi \geq c_\varphi.$$

If $f_0 \in C^\beta$, $\beta > 0$, then (P1) is satisfied as soon as $\alpha \leq \beta$ (so any prior that matches the regularity of f_0 , or that “undersmooths,” can be used).

REMARK 3. Condition 1 allows for a sub-Gaussian density φ . Strictly subexponential tails could be allowed too if the weighting sequence w satisfies an additional constraint: Theorem 2 below holds true for exponential-power densities $\varphi(x) \asymp e^{-|x|^p}$ and $1 < p < 2$, provided $w_l/l^{1/p} \uparrow \infty$.

Any prior satisfying Condition 1 defines a Borel probability measure on L^2 (using separability of the latter space), and the resulting posterior distribution also defines an element of $L^2 \cong \ell_2 \subset \mathcal{M}_0$.

THEOREM 2. Suppose Π satisfies Condition 1, and let $\Pi(\cdot|\mathbb{X}^{(n)})$ be the posterior distribution in \mathcal{M}_0 arising from observing (13) for some fixed $f_0 \in C^\beta$, $\beta > 0$. Then Π satisfies the weak Bernstein–von Mises theorem in the sense of Definition 2 in the space $\mathcal{M}_0 = \mathcal{M}_0(w)$ for any admissible w , with \mathcal{N} equal to the law of \mathbb{W} , and with centring T_n equal to $\mathbb{X}^{(n)}$ or equal to the posterior mean $E(f|\mathbb{X}^{(n)})$.

3.2. *Bernstein–von Mises theorems in $\mathcal{M}_0(w)$: Sampling model case.* Let us now turn to the situation where one observes a sample $X_i \sim^{\text{i.i.d.}} P$,

$$(X_1, \dots, X_n) \equiv X^{(n)},$$

from law P with *bounded* probability density f on $[0, 1]$. We define multiscale priors Π on some space \mathcal{F} of probability density functions f giving rise to absolutely continuous probability measures. Let

$$\mathcal{F} := \bigcup_{0 < \rho \leq D < \infty} \mathcal{F}(\rho, D) := \bigcup_{0 < \rho \leq D < \infty} \left\{ f : [0, 1] \rightarrow [\rho, D], \int_0^1 f = 1 \right\}.$$

In the following we assume that the “true” density f_0 belongs to $\mathcal{F}_0 := \mathcal{F}(\rho_0, D_0)$, for some $0 < \rho_0 \leq D_0 < \infty$.

We consider various classes of priors on densities and two possible values for a cut-off parameter L_n . For $\alpha > 0$, let $j_n = j_n(\alpha)$ and $l_n = l_n(\alpha)$ be the *largest* integers such that

$$(19) \quad 2^{j_n} \leq n^{1/(2\alpha+1)}, \quad 2^{l_n} \leq \left(\frac{n}{\log n} \right)^{1/(2\alpha+1)},$$

and set, in slight abuse of notation, either

$$(20) \quad L_n = j_n \quad (\forall n \geq 1) \quad \text{or} \quad L_n = l_n \quad (\forall n \geq 1).$$

(S) Priors on log-densities. Given a multiscale wavelet basis $\{\psi_{lk}\}$ from (8), consider the prior Π induced by, for any $x \in [0, 1]$ and L_n as in (20),

$$(21) \quad T(x) = \sum_{l \leq L_n} \sum_{k=0}^{2^l-1} \sigma_l \alpha_{lk} \psi_{lk}(x),$$

$$(22) \quad f(x) = \exp\{T(x) - c(T)\}, \quad c(T) = \log \int_0^1 e^{T(x)} dx,$$

where α_{lk} are i.i.d. random variables of continuous probability density $\varphi : \mathbb{R} \rightarrow [0, \infty)$. We consider the choices

$$(S1) \quad \varphi(x) = \varphi_H(x),$$

$$(S2) \quad \varphi(x) = \varphi_G(x) = e^{-x^2/2} / \sqrt{2\pi},$$

where φ_H is any density such that $\log \varphi_H$ is Lipschitz on \mathbb{R} . We call this the log-Lipschitz case. For instance, the α_{lk} ’s can be Laplace-distributed or have heavier tails. To simplify some proofs we restrict to a specific form of density: for a given $0 \leq \tau < 1$ and $x \in \mathbb{R}$, and c_τ a normalising constant, suppose φ_H takes the form

$$(23) \quad \varphi_{H,\tau}(x) = c_\tau \exp\{-(1 + |x|)^{1-\tau}\}.$$

Suppose the prior parameters σ_l satisfy, for $\alpha > 1/2$ and $0 < r < \alpha - 1/4$,

$$(24) \quad \begin{aligned} \sigma_l &= 2^{-l(\alpha+1/2)} && \text{(log-Lipschitz-case),} \\ \sigma_l &= 2^{-l(r+(1/2))} && \text{(Gaussian-case).} \end{aligned}$$

(H) *Random histograms density priors.* Associated to the regular dyadic partition of $[0, 1]$ at level $L \in \mathbb{N} \cup \{0\}$, given by $I_0^L = [0, 2^{-L}]$ and $I_k^L = (k2^{-L}, (k + 1)2^{-L}]$ for $k = 1, \dots, 2^L - 1$, is a natural notion of histogram

$$\mathcal{H}_L = \left\{ h \in L^\infty[0, 1], h(x) = \sum_{k=0}^{2^L-1} h_k \mathbb{1}_{I_k^L}(x), h_k \in \mathbb{R}, k = 0, \dots, 2^L - 1 \right\}.$$

Let $\mathcal{S}_L = \{\omega \in [0, 1]^{2^L}; \sum_{k=0}^{2^L-1} \omega_k = 1\}$ be the unit simplex in \mathbb{R}^{2^L} . Further denote \mathcal{H}_L^1 the subset of \mathcal{H}_L consisting of histograms which are densities on $[0, 1]$ with L equally spaced dyadic knots. Let \mathcal{H}^1 be the set of all histograms which are densities on $[0, 1]$.

A simple way to specify a prior Π on \mathcal{H}_L^1 is to set $L = L_n$ deterministic and to fix a distribution for $\omega_L := (\omega_0, \dots, \omega_{2^L-1})$. Set $L = L_n$ as defined in (20). Choose some fixed constants $a, c_1, c_2 > 0$, and let

$$(25) \quad L = L_n, \quad \omega_L \sim \mathcal{D}(\alpha_0, \dots, \alpha_{2^L-1}), \quad c_1 2^{-La} \leq \alpha_k \leq c_2,$$

for any admissible index k , where \mathcal{D} denotes the Dirichlet distribution on \mathcal{S}_L . Unlike those suggested by the notation, the coefficients α of the Dirichlet distribution are allowed to depend on L_n , so that $\alpha_k = \alpha_{k, L_n}$.

The priors (S), (H) above are ‘‘multiscale’’ priors where high frequencies are ignored, corresponding to truncated series priors considered frequently in the non-parametric Bayes literature. The resulting posterior distributions $\Pi(\cdot | X^{(n)})$ attain minimax optimal contraction rates up to logarithmic terms in Hellinger and L^2 -distance [5, 32, 37] and L^∞ -distance [3]. Clearly other priors are of interest as well, for instance, priors without or with random high-frequency cut-off or Dirichlet mixtures of normals etc. While our current proofs do not cover such situations, one can note that our proof strategy via simultaneous control of many linear functionals is applicable in such situations as well. Generalising the scope of our techniques is an interesting direction of future research.

The projection $P_n(j)$ as in (16), with the choice $j = L_n$ from (20), defines a tight random variable in \mathcal{M}_0 . For $z \in \mathcal{M}_0$, the map $\tau_z: f \mapsto \sqrt{n}(f - z)$ maps $\mathcal{M}_0 \rightarrow \mathcal{M}_0$, and we can define the shifted posterior $\Pi(\cdot | X^{(n)}) \circ \tau_{P_n(L_n)}^{-1}$. The following theorem shows that the above priors satisfy a weak BvM theorem in \mathcal{M}_0 in the sense of Definition 2, with efficient centring $P_n(L_n)$; cf. Theorem 1. Denote the law $\mathcal{L}(\mathbb{G}_{P_0})$ of \mathbb{G}_{P_0} from Proposition 3 by \mathcal{N} .

THEOREM 3. *Let $\mathcal{M}_0 = \mathcal{M}_0(w)$ for any admissible $w = (w_l)$. Let $X^{(n)} = (X_1, \dots, X_n)$ i.i.d. from law P_0 with density $f_0 \in \mathcal{F}_0$. Let Π be a prior on the set of probability densities \mathcal{F} , that is:*

- (1) either of type (S), in which case one assumes $\log f_0 \in C^\alpha$ for some $\alpha > 1$,
- (2) or of type (H), and one assumes $f_0 \in C^\alpha$ for some $1/2 < \alpha \leq 1$.

Suppose the prior parameters satisfy (20), (24) and (25). Let $\Pi(\cdot|X^{(n)})$ be the induced posterior distribution on \mathcal{M}_0 . Then, as $n \rightarrow \infty$,

$$(26) \quad \beta_{\mathcal{M}_0}(\Pi(\cdot|X^{(n)}) \circ \tau_{P_n(L_n)}^{-1}, \mathcal{N}) \rightarrow^{P_0^{\mathbb{N}}} 0.$$

4. Some applications.

4.1. *Donsker’s theorem for the posterior cumulative distribution function.* Whenever a prior on f satisfies the weak Bernstein–von Mises phenomenon in the sense of Definition 2, we can deduce from the continuous mapping theorem a BvM for integral functionals $L_g(f) = \int_0^1 g(x)f(x) dx$ simultaneously for many g ’s satisfying bounds on the decay of their wavelet coefficients. More precisely a bound $\sum_k | \langle g, \psi_{lk} \rangle | \leq c_l$ for all l combined with a weak BvM for (w_l) such that $\sum c_l w_l < \infty$ is sufficient. Let us illustrate this in a key example $g_t = 1_{[0,t]}$, $t \in [0, 1]$, where we can derive results paralleling the classical Donsker theorem for distribution functions and its BvM version for the Dirichlet process proved in [31]. With the applications we have in mind, and to simplify some technicalities, we restrict to situations where the posterior $f|X$ is supported in L^2 , and where the centring T_n in Definition 2 is contained in L^2 (resp., equals $\mathbb{X}^{(n)}$). In this case the primitives

$$F(t) = \int_0^t f(x) dx, \quad \mathbb{T}_n(t) = \int_0^t T_n(x) dx \quad \left(\text{resp., } \int_0^t dX^{(n)}(x) \right), \quad t \in [0, 1],$$

define random variables in the separable space $C([0, 1])$ of continuous functions on $[0, 1]$, and we can formulate a BvM result in that space. Different centring (such as the empirical distribution function) are discussed below.

THEOREM 4. *Let Π be a prior supported in $L^2([0, 1])$, and suppose the weak Bernstein–von Mises phenomenon in the sense of Definition 2 holds true in $\mathcal{M}_0(w)$ for some sequence (w_l) such that $\sum_l w_l 2^{-l/2} < \infty$, and with centring T_n either equal to $\mathbb{X}^{(n)}$ or such that $T_n \in L^2$. For $f \sim \Pi(\cdot|X)$ (conditional on X) define the posterior cumulative distribution function*

$$(27) \quad F(t) = \int_0^t f(x) dx, \quad t \in [0, 1].$$

Let G be a Brownian motion $(G(t) : t \in [0, 1])$ in the white noise model or a P_0 -Brownian bridge $(G(t) \equiv G_{P_0}(t) : t \in [0, 1])$, $dP_0(x) = f_0(x) dx$, $f_0 \in L^\infty$, in the sampling model. If $X \sim P_{f_0}$ for some fixed f_0 , then as $n \rightarrow \infty$,

$$(28) \quad \beta_{C([0,1])}(\mathcal{L}(\sqrt{n}(F - \mathbb{T}_n)|X), \mathcal{L}(G)) \rightarrow^{P_{f_0}} 0,$$

$$(29) \quad \beta_{\mathbb{R}}(\mathcal{L}(\sqrt{n}\|F - \mathbb{T}_n\|_\infty|X), \mathcal{L}(\|G\|_\infty)) \rightarrow^{P_{f_0}} 0.$$

PROOF. The mapping

$$(30) \quad L : \{h_{lk}\} \mapsto L_t(\{h_{lk}\}) := \sum_{l,k} h_{lk} \int_0^t \psi_{lk}(x) dx, \quad t \in [0, 1],$$

is linear and continuous from $\mathcal{M}_0(w)$ to $L^\infty([0, 1])$ since, for $0 < c < C < \infty$,

$$\begin{aligned} \left| \sum_{l,k} h_{lk} \int_0^t \psi_{lk}(x) dx \right| &\leq \sum_{l,k} |h_{lk}| |\langle 1_{[0,t]}, \psi_{lk} \rangle| \\ &\leq c \sup_{l,k} \frac{|h_{lk}|}{w_l} \sum_l w_l 2^{-l/2} \\ &\leq C \|h\|_{\mathcal{M}_0}, \end{aligned}$$

where we have used $\sup_{t \in [0,1]} \sum_k |\langle 1_{[0,t]}, \psi_{lk} \rangle| \leq c 2^{-l/2}$, shown, for example, as in the proof of Lemma 3 in [16]. Also, L coincides with the primitive map on any function $h \in L^2([0, 1])$ with wavelet coefficients $\{h_{lk}\} \in \ell_2$, since then

$$L(\{h_{lk}\}) = \sum_{l,k} h_{lk} \langle 1_{[0,t]}, \psi_{lk} \rangle = \langle h, 1_{[0,t]} \rangle = \int_0^t h(x) dx, \quad t \in [0, 1],$$

in view of Parseval's identity. Moreover, if \mathbb{G} is a tight Gaussian random variable in \mathcal{M}_0 , then the linear transformation $L(\mathbb{G})$ is a tight Gaussian random variable in $C([0, 1])$, equal in law to a Brownian motion or a P_0 -Brownian bridge for our choice $\mathbb{G} = \mathbb{W}$ or $\mathbb{G} = \mathbb{G}_{P_0}$, respectively, after checking the identity of the corresponding reproducing kernel Hilbert spaces (cf. [36], and using again that L equals the primitive map on L^2). The displays (28)–(29) now follow from Definition 2, the continuous mapping theorem applied to L and $L \circ \|\cdot\|_\infty^{-1}$, respectively, and noting that $L(f), L(T_n)$ take values in the closed subspace $C([0, 1])$ of $L^\infty([0, 1])$ under the maintained assumptions. (Although not used here, it can in fact be checked that the general inclusion $\text{Im}(L) \subset C([0, 1])$ holds true.) \square

COROLLARY 1. *Let Π be a prior of type (S) or (H), and suppose the conditions of Theorem 3 are satisfied. Let $F_n(t) = (1/n) \sum_{i=1}^n 1_{[0,t]}(X_i), t \in [0, 1]$, be the empirical distribution function based on a sample X_1, \dots, X_n from law P_0 , and let F be a cumulative distribution function induced by $\Pi(\cdot|X_1, \dots, X_n)$ as in (27). Then, as $n \rightarrow \infty$,*

$$\begin{aligned} \beta_{L^\infty([0,1])}(\mathcal{L}(\sqrt{n}(F - F_n)|X), \mathcal{L}(G_{P_0})) &\rightarrow P_0^{\mathbb{N}} 0, \\ \beta_{\mathbb{R}}(\mathcal{L}(\sqrt{n}\|F - F_n\|_\infty|X), \mathcal{L}(\|G_{P_0}\|_\infty)) &\rightarrow P_0^{\mathbb{N}} 0. \end{aligned}$$

PROOF. By Theorems 3 and 4 the result is true with F_n replaced by the primitive \mathbb{T}_n of $P_n(L_n)$. As in the proof leading to Remark 9 in [16] one shows $\|\mathbb{T}_n - F_n\|_\infty = o_P(1/\sqrt{n})$, and hence the result follows from the triangle inequality. (To avoid measurability issues we note that the result holds for convergence

in distribution in $L^\infty([0, 1])$ in the generalised sense of empirical processes (as in [18]), or in the space of càdlàg functions on $[0, 1]$.) \square

Returning to the general setting of Theorem 4, a natural credible band for F is to take C_n, R_n such that, with L the map defined in (30),

$$(31) \quad C_n = \{F : \|F - \mathbb{T}_n\|_\infty \leq R_n/\sqrt{n}\}, \quad \Pi \circ L^{-1}(C_n|X) = 1 - \alpha.$$

The proof of the following result implies in particular that C_n asymptotically coincides with the usual Kolmogorov–Smirnov confidence band. The result is true also with centring $\mathbb{T}_n = F_n$ (in which case the proof requires minor modifications related to the remarks at the end of the proof of Corollary 1).

COROLLARY 2. *Under the conditions of Theorem 4, let $X \sim P_{f_0}, F_0 = \int_0^\cdot f_0(t) dt$ and C_n as in (31). Then we have, as $n \rightarrow \infty$,*

$$P_{f_0}(F_0 \in C_n) \rightarrow 1 - \alpha \quad \text{and} \quad R_n \rightarrow^{P_{f_0}} \text{const.}$$

PROOF. The proof is similar to Theorem 1 in [4], replacing $H(\delta)$ there by $C([0, 1])$ (a separable Banach space): the function Φ in that proof is strictly increasing: any shell $\{g \in C([0, 1]) : s < \|g\|_\infty < t\}, 0 \leq s < t$, contains an element of the RKHS (see [36]) of Brownian motion [in the case of the white noise model (2)] or of the P_0 -Brownian bridge (in the case of the i.i.d. sampling model). Using also Theorem 1 in the sampling model case, all arguments from the proof of Theorem 1 in [4] go through. \square

REMARK 4. Equation (31) [resp., (32) below] reads conditionally on the existence of such a positive real R_n . More generally, one may take a generalised quantile in (31) [resp., in (32)]. Then C_n has credibility $1 - \alpha$ asymptotically, and one can check that the previous corollary [resp., Theorem 5] continues to hold.

4.2. Confidence bands for f . Given a posterior distribution $\Pi(\cdot|X)$ on the parameter f of a regression or sampling model, we can incorporate the multiscale approach to construct confidence sets for f in a Bayesian way. We take an efficient centring T_n [e.g., $\mathbb{X}^{(n)}, P_n(L)$ from above or, when appropriate, the posterior mean $E(f|X)$] and, given $\alpha > 0$ and admissible w , choose R_n and the credible region C_n in such a way that

$$(32) \quad C_n = \left\{ f : \sup_{l,k} \frac{|(f - T_n, \psi_{lk})|}{w_l} \leq \frac{R_n}{\sqrt{n}} \right\}, \quad \Pi(C_n|X) = 1 - \alpha.$$

THEOREM 5. *Let $w = (w_l)$ be admissible. Suppose the weak Bernstein–von Mises phenomenon holds true in $\mathcal{M}_0(w)$ with prior Π and centring T_n . Let C_n be as in (32). Then, as $n \rightarrow \infty$,*

$$P_{f_0}(f_0 \in C_n) \rightarrow 1 - \alpha, \quad R_n \rightarrow^{P_{f_0}} \text{const.}$$

PROOF. The proof is the same as the one of Theorem 1 in [4], replacing $H(\delta)$ there by $\mathcal{M}_0(w)$, and using also Theorem 1 in the sampling model case. \square

The previous theorem can be used to control low frequencies of the estimation error, and following the multiscale approach one needs to employ further qualitative information about f_0 to control high frequencies. In the present case, if we assume $f_0 \in C^\gamma$ for some $\gamma > 0$, we can define, for $u_n = w_{j_n}/\sqrt{j_n}$ and j_n such that $2^{j_n} \sim (n/\log n)^{1/(2\gamma+1)}$, the confidence set

$$(33) \quad \bar{C}_n = \bar{C}_n(\gamma) = C_n \cap \{f : \|f\|_{C^\gamma} \leq u_n\}.$$

The following result combined with Theorems 3 and 5 implies in particular Proposition 1 from the Introduction.

PROPOSITION 5. Under the conditions of Theorem 5 suppose $X \sim P_{f_0}$ where $f_0 \in C^\gamma$ ($[0, 1]$). Then, with \bar{C}_n as in (33), and as $n \rightarrow \infty$,

$$P_{f_0}(f_0 \in \bar{C}_n) \rightarrow 1 - \alpha \quad \text{and} \quad |\bar{C}_n|_\infty = O_{P_{f_0}}((n/\log n)^{-\gamma/(2\gamma+1)} u_n).$$

PROOF. For n large enough such that $u_n \geq \|f_0\|_{C^\gamma}$ we have as $n \rightarrow \infty$

$$P_{f_0}(f_0 \in \bar{C}_n) = P_{f_0}(f_0 \in C_n) \rightarrow 1 - \alpha$$

in view of Theorem 5. Moreover, for $h = f - g$, $f, g \in C_n$ arbitrary,

$$\|h\|_{\mathcal{M}(w)} \leq \|f - T_n\|_{\mathcal{M}(w)} + \|g - T_n\|_{\mathcal{M}(w)} = O\left(\frac{R_n}{\sqrt{n}}\right) = O_{P_{f_0}}\left(\frac{1}{\sqrt{n}}\right).$$

The estimate on $|\bar{C}_n|_\infty$ now follows from

$$\|h\|_\infty \leq \sum_l 2^{l/2} \max_k |\langle h, \psi_{lk} \rangle|$$

combined with the bound

$$\begin{aligned} \sum_{l \leq j_n} 2^{l/2} \max_k |\langle h, \psi_{lk} \rangle| &= \sum_{l \leq j_n} 2^{l/2} \sqrt{l} \frac{w_l}{\sqrt{l}} w_l^{-1} \max_k |\langle h, \psi_{lk} \rangle| \\ &\lesssim \sqrt{\frac{2^{j_n} j_n}{n} \frac{w_{j_n}}{\sqrt{j_n}}} R_n \\ &= O_{P_{f_0}}\left(\left(\frac{\log n}{n}\right)^{\gamma/(2\gamma+1)} u_n\right), \end{aligned}$$

and with

$$\begin{aligned} \sum_{l > j_n} 2^{l/2} \max_k |\langle h, \psi_{lk} \rangle| &= \sum_{l > j_n} 2^{-l\gamma} 2^{l(\gamma+1/2)} \max_k |\langle h, \psi_{lk} \rangle| \\ &\leq \|h\|_{C^\gamma} 2^{-j_n\gamma} \\ &= O_{P_{f_0}}\left(\left(\frac{\log n}{n}\right)^{\gamma/(2\gamma+1)} u_n\right), \end{aligned}$$

completing the proof. \square

REMARK 5 (Optimal diameter, undersmoothing, adaptation). The confidence bands from Propositions 1 and 5 have diameter equal to the L^∞ -minimax rate over Hölder balls multiplied with an under-smoothing penalty u_n , common in frequentist constructions of confidence bands; see [20] and, more recently, [17]. If the BvM phenomenon holds for all admissible sequences w (as in the examples above), then this sequence can be taken to diverge at an arbitrarily slow rate.

If a quantitative a priori bound $\|f_0\|_{C^\gamma} < B$ is available, then in the setting of Theorem 2 one could use a uniform wavelet prior [with $\varphi = \mathbb{1}_{[-B,B]}/(2B)$, for some $B > 0$] concentrating on a Hölder ball of radius B (as in Corollary 1, [4]). The set \bar{C}_n from (33) (even with u_n replaced by B) is then an *exact* level $1 - \alpha$ posterior credible set, consisting of the intersection of two hyper-rectangles in sequence space, and Proposition 5 applies to give the precise frequentist asymptotics of \bar{C}_n .

We can also obtain adaptive confidence bands by using a bandwidth choice \hat{j}_n as in [17] to estimate γ by $\hat{\gamma}$ under a self-similarity constraint on f , corresponding to an empirical Bayes-type selection of γ . More Bayesian approaches to *adaptive* confidence sets are subject of current research; see, for example, the recent contribution [34].

5. Proofs.

5.1. *Proof of Theorem 1.* For J to be chosen below, let V_J be the subspace of $\mathcal{M}(w)$ consisting of the scales $l \leq J$, and let $\pi_{V_J}(P)$ be the projection of f onto V_J . We have by definition of the Hölder space C^γ and assumption

$$\begin{aligned} \|P - \pi_{V_{j_n}}(P)\|_{\mathcal{M}_0} &= \sup_{l > j_n} \frac{\max_k |\langle f, \psi_{lk} \rangle|}{w_l} \\ (34) \qquad \qquad \qquad &\lesssim w_{j_n}^{-1} 2^{-j_n(\gamma+1/2)} \\ &= o\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

so that this term is negligible in the limit distribution. Writing β for $\beta_{\mathcal{M}_0}$ and $\sqrt{n}(P_n(j_n) - \pi_{V_{j_n}}(P)) = v_n$, it suffices to show that

$$\begin{aligned} \beta(\mathcal{L}(v_n), \mathcal{L}(\mathbb{G}_P)) &\leq \beta(\mathcal{L}(v_n), \mathcal{L}(v_n) \circ \pi_{V_J}^{-1}) \\ (35) \qquad \qquad \qquad &+ \beta(\mathcal{L}(v_n) \circ \pi_{V_J}^{-1}, \mathcal{L}(\mathbb{G}_P) \circ \pi_{V_J}^{-1}) \\ &+ \beta(\mathcal{L}(\mathbb{G}_P), \mathcal{L}(\mathbb{G}_P) \circ \pi_{V_J}^{-1}) \end{aligned}$$

converges to zero under $P^{\mathbb{N}}$. Let $\varepsilon > 0$ be given. The second term is less than $\varepsilon/3$ for every J fixed and n large enough by the multivariate central limit theorem

applied to

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\psi_{lk}(X_i) - E_P \psi_{lk}(X)), \quad k, l \leq J,$$

noting that eventually $j_n > J$. For the first term, by definition of β ,

$$\begin{aligned} & \beta(\mathcal{L}(v_n), \mathcal{L}(v_n) \circ \pi_{V_J}^{-1}) \\ (36) \quad & \leq E \|\sqrt{n}(\pi_{V_{j_n}} - \pi_{V_J})(P_n - P)\|_{\mathcal{M}_0} \\ & \leq \left[\max_{J < l \leq j_n} \frac{\sqrt{l}}{w_l} \right] E \max_{J < l \leq j_n} l^{-1/2} \max_k |\langle \sqrt{n}(P_n - P), \psi_{lk} \rangle|. \end{aligned}$$

Thus for J large enough this term can be made smaller than $\varepsilon/3$ if we can show that the expectation is bounded by a fixed constant. For M a large enough constant, this expectation is bounded above by M plus

$$\begin{aligned} & \int_M^\infty P \left(\max_{J \leq l \leq j_n} l^{-1/2} \max_k |\langle \sqrt{n}(P_n - P), \psi_{lk} \rangle| > u \right) du \\ & \leq \sum_{J \leq l \leq j_n, k} \int_M^\infty P(|\langle \sqrt{n}(P_n - P), \psi_{lk} \rangle| > \sqrt{l}u) du \\ & \leq \sum_{J \leq l \leq j_n} 2^l \int_M^\infty e^{-Clu} du \lesssim e^{-C'JM}, \end{aligned}$$

where the second inequality follows from an application of Bernstein's inequality (e.g., [29]) together with the bounds $P\psi_{lk}^2 \leq \|f\|_\infty$ and $\sqrt{l}\|\psi_{lk}\|_\infty \leq \sqrt{l}2^{l/2} = O(\sqrt{n})$ for $l \leq j_n$, using the assumption on j_n .

For the third Gaussian term we argue similarly, replacing v_n by \mathbb{G}_P in (36) and using that $E \sup_l \max_k |\mathbb{G}_P(\psi_{lk})|/\sqrt{l} < \infty$ by Proposition 2.

5.2. *A tightness criterion in \mathcal{M}_0 .* The following proposition considers general random posterior measures $\Pi(\cdot|X)$ in the setting of Definition 2.

PROPOSITION 6. *Let $\pi_{V_J}, J \in \mathbb{N}$, be the projection operator onto the finite-dimensional space spanned by the ψ_{lk} 's with scales up to $l \leq J$. Let $f \sim \Pi(\cdot|X)$, $T_n = T_n(X)$, let $\tilde{\Pi}_n$ denote the laws of $\sqrt{n}(f - T_n)$ conditionally on X and let \mathcal{N} equal the Gaussian probability measure on $\mathcal{M}_0(w)$ given by either \mathbb{W} or \mathbb{G}_P from P with bounded density.*

Assume that the finite-dimensional distributions converge, that is,

$$(37) \quad \beta_{V_J}(\tilde{\Pi}_n \circ \pi_{V_J}^{-1}, \mathcal{N} \circ \pi_{V_J}^{-1}) \xrightarrow{P_{f_0}} 0 \quad \text{as } n \rightarrow \infty,$$

and that for some sequence $\bar{w} = (\bar{w}_l) \uparrow \infty$, $\bar{w}_l/\sqrt{l} \geq 1$,

$$\begin{aligned}
 E[\|f - T_n\|_{\mathcal{M}_0(\bar{w})}|X] &= E\left[\sup_l \bar{w}_l^{-1} \max_k |\langle f - T_n, \psi_{lk} \rangle| |X\right] \\
 (38) \qquad \qquad \qquad &= O_{P_{f_0}}\left(\frac{1}{\sqrt{n}}\right).
 \end{aligned}$$

Then, for any w such that $w_l/\bar{w}_l \uparrow \infty$ we have, as $n \rightarrow \infty$,

$$\beta_{\mathcal{M}_0(w)}(\tilde{\Pi}_n, \mathcal{N}) \xrightarrow{P_{f_0}} 0.$$

REMARK 6. Inspection of the proof shows that the result still holds true if $f \sim \Pi(\cdot|X)$ is replaced by $f \sim \tilde{\Pi}(\cdot|X)$ for random measures $\tilde{\Pi}(\cdot|X)$ s.t.

$$\beta_{\mathcal{M}_0}(\tilde{\Pi}(\cdot|X), \Pi(\cdot|X)) \xrightarrow{P_{f_0}} 0$$

as $n \rightarrow \infty$. Likewise, the posterior can be replaced by the conditional posterior $\Pi^{D_n}(\cdot|X)$ for any sequence of sets D_n such that $\Pi(D_n|X) \xrightarrow{P_{f_0}} 1$.

PROOF. Let us write $\beta = \beta_{\mathcal{M}_0(w)}$ and decompose

$$\beta(\tilde{\Pi}_n, \mathcal{N}) \leq \beta(\tilde{\Pi}_n, \tilde{\Pi}_n \circ \pi_{V_j}^{-1}) + \beta(\tilde{\Pi}_n \circ \pi_{V_j}^{-1}, \mathcal{N} \circ \pi_{V_j}^{-1}) + \beta(\mathcal{N}, \mathcal{N} \circ \pi_{V_j}^{-1}).$$

The second term converges to zero by (37). The third term too, arguing as at the end of the proof of Theorem 1 (and using Proposition 2 or 3). For the first term let $f \sim \Pi(\cdot|X)$ conditional on X . Then using (38) we can bound the β -distance by the expectation of the norm and thus by

$$\begin{aligned}
 E[\|\sqrt{n}(id - \pi_{V_j})(f - T_n)\|_{\mathcal{M}(w)}|X] \\
 \leq \left[\sup_{l>J} \frac{\bar{w}_l}{w_l}\right] E\left[\sup_{l>J} \bar{w}_l^{-1} \max_k |\sqrt{n}\langle f - T_n, \psi_{lk} \rangle| |X\right] \\
 \leq \sup_{l>J} \frac{\bar{w}_l}{w_l} \times O_{P_{f_0}}(1),
 \end{aligned}$$

which can be made as small as desired for J large enough but fixed. \square

5.3. Proof of Theorem 2. We choose integers $j = j_n \rightarrow \infty$ such that

$$(39) \quad \sigma_j^{-1} = 2^{j(\alpha+1/2)} \sim \sqrt{n} \quad \text{and note} \quad \sigma_l \lesssim \frac{1}{\sqrt{n}} \quad \forall l > j.$$

Conditional on $\mathbb{X}^{(n)}$, let $f \sim \Pi(\cdot|\mathbb{X}^{(n)})$ and, for π_{V_j} the projection operator onto V_j , consider the decomposition in $\mathcal{M}_0(w)$, under P_{f_0} ,

$$\begin{aligned}
 \sqrt{n}(f - \mathbb{X}^{(n)}) &= \sqrt{n}(\pi_{V_j}(f) - \pi_{V_j}(\mathbb{X}^{(n)})) + \sqrt{n}(f - \pi_{V_j}(f)) \\
 &\quad + \sqrt{n}(\pi_{V_j}(f_0) - f_0) + (\pi_{V_j}(\mathbb{W}) - \mathbb{W}) \\
 &= \text{I} + \text{II} + \text{III} + \text{IV}.
 \end{aligned}$$

We verify the conditions of Proposition 6 above for the laws $\mathcal{L}(\sqrt{n}(f - \mathbb{X}^{(n)})|X) = \tilde{\Pi}_n$ and for the choice $\bar{w}_l = \sqrt{l}$. From Theorem 7 in [4], with Condition 2 verified

in the proof of Theorem 9 of that paper, we derive condition (37). Next we verify that (38) is satisfied for each of the terms I, II, III, IV, separately. That is, we check that each term has bounded $\mathcal{M}(\bar{w})$ -norm in expectation (and apply Markov's inequality).

(IV) We have as in the proof of Proposition 2 that

$$E \sup_{k,l} l^{-1/2} |\mathbb{W}(\psi_{lk})| \leq C < \infty.$$

(III) This term is nonrandom and we have by Condition 1 and definition of σ_l , and some constant $0 < M < \infty$,

$$\sqrt{n} \sup_{l>j,k} l^{-1/2} |\langle f_0, \psi_{lk} \rangle| \lesssim M \sqrt{n} \sup_{l>j} l^{-1/2} \sigma_l \lesssim M/\sqrt{j}.$$

(II) For E the iterated expectation under P_{f_0} and $\Pi(\cdot|X)$, we can bound

$$E \sup_{l>j,k} l^{-1/2} |\langle f, \psi_{lk} \rangle| \leq \sum_{l>j} l^{-1/2} E \max_k |\langle f, \psi_{lk} \rangle|.$$

Denote $f_{lk} := \langle f, \psi_{lk} \rangle$, $f_{0,lk} := \langle f_0, \psi_{lk} \rangle$ and $\varepsilon_{lk} := \langle \mathbb{W}, \psi_{lk} \rangle$. An application of Jensen's inequality yields, for any $t > 0$,

$$E \max_k |f_{lk}| \leq \frac{1}{t} \log \sum_k E(e^{t f_{lk}} + e^{-t f_{lk}}).$$

It is now enough to bound the Laplace transform $E[e^{s f_{lk}}]$ for $s = t, -t$. Both cases are similar, so we focus on $s = t$,

$$\begin{aligned} E[e^{t f_{lk}}] &= E \frac{\int e^{t(f_{0,lk} + (v/\sqrt{n}))} e^{-(v^2/2) + \varepsilon_{lk} v} (1/(\sqrt{n}\sigma_l)) \varphi((f_{0,lk} + (v/\sqrt{n}))/\sigma_l) dv}{\int e^{-(v^2/2) + \varepsilon_{lk} v} (1/(\sqrt{n}\sigma_l)) \varphi((f_{0,lk} + (v/\sqrt{n}))/\sigma_l) dv} \\ &=: E \frac{N_{lk}(t)}{D_{lk}}. \end{aligned}$$

To bound the denominator D_{lk} from below, one applies the same technique as in [4], proof of Theorem 5. One first restricts the integral to $(-\sqrt{n}\sigma_l, \sqrt{n}\sigma_l)$. Next one notices, using (P1), that over this interval the argument of φ lies in a compact set, and hence the function φ can be bounded below by a constant, using (P2). Next one applies Jensen's inequality to obtain

$$D_{lk} \gtrsim e^{-(1/2) \int_{-\sqrt{n}\sigma_l}^{\sqrt{n}\sigma_l} (v^2/2) dv / (\sqrt{n}\sigma_l)} \gtrsim e^{-C}.$$

To bound the numerator $N_{lk}(t)$ one splits the integral into a part N_1 on $\mathcal{A} := \{v : |f_{0,lk} + v/\sqrt{n}| \leq \sigma_l\}$ and a part N_2 on its complement \mathcal{A}^c . First

$$\begin{aligned} EN_1 &\leq e^{t\sigma_l} E \int_{\mathcal{A}} e^{-v^2/2 + \varepsilon_{lk} v} \varphi\left(\frac{f_{0,lk} + (v/\sqrt{n})}{\sigma_l}\right) dv \\ &\leq e^{t\sigma_l} \int_{\mathcal{A}} \varphi\left(\frac{f_{0,lk} + (v/\sqrt{n})}{\sigma_l}\right) dv \leq e^{t\sigma_l}, \end{aligned}$$

using the definition of \mathcal{A} and Fubini's theorem. On the other hand, the term N_2 , setting $w = f_{0,lk} + v/\sqrt{n}$ and using condition (E), is bounded by

$$\begin{aligned} EN_2 &\leq \int_{(-1,1)^c} e^{t\sigma_l w} E(e^{-(n/2)(w\sigma_l - f_{0,lk})^2 + \varepsilon_{lk}\sqrt{n}(w\sigma_l - f_{0,lk})})\varphi(w) dw \\ &\leq \int_{(-1,1)^c} e^{t\sigma_l w} \varphi(w) dw \lesssim e^{d(\sigma_l t)^2}, \end{aligned}$$

for some $d > 0$. Conclude, setting $t = \sigma_l^{-1}l^{1/2}$, that

$$\begin{aligned} E \max_k |f_{lk}| &\leq \frac{1}{t} \log(2^l [C e^{t\sigma_l} + C e^{d(\sigma_l t)^2}]) \\ &\lesssim \frac{l}{t} + \sigma_l + \frac{1}{t}(\sigma_l t)^2 \lesssim \sigma_l l^{1/2}. \end{aligned}$$

This gives the overall bound,

$$\sum_{l>j} 2^{-l(1/2+\alpha)} \lesssim 2^{-j(1/2+\alpha)} = O(1/\sqrt{n}).$$

(I) For the frequencies $l \leq j_n$ one proves, as in Lemma 1 in [3], for some constant $C > 0$, the sub-Gaussian bound

$$(40) \quad E_{f_0} E(e^{t\sqrt{n}(f_{lk} - X_{lk})} | X) \leq C e^{t^2/2}.$$

[All that is needed here is φ bounded away from zero and infinity on a compact set, and that $(f_{0,lk} + v/\sqrt{n})/\sigma_l$ is bounded by a fixed constant, true for the l 's relevant here.] Then, by a standard application of Markov's inequality to sub-Gaussian random variables, writing \Pr for the law with expectation $E_{f_0} E(\cdot | X)$, we have for all $v > 0$ and universal constants C, C' that

$$\Pr(\sqrt{n}|f_{lk} - X_{lk}| > v) \leq C' e^{-Cv^2}.$$

We then bound, for M a fixed constant

$$\begin{aligned} E_{f_0} E\left(\sup_{l \leq j} l^{-1/2} \max_k \sqrt{n}|f_{lk} - X_{lk}| | X\right) \\ \leq M + \int_M^\infty \Pr\left(\sup_{l \leq j, k} l^{-1/2} \max_k \sqrt{n}|f_{lk} - X_{lk}| > u\right) du. \end{aligned}$$

The tail integral can be further bounded as follows:

$$\begin{aligned} \sum_{l \leq j, k} \int_M^\infty \Pr(\sqrt{n}|f_{lk} - X_{lk}| > \sqrt{lu}) du \\ \lesssim \sum_{l \leq j} 2^l \int_M^\infty e^{-Clu^2} du \lesssim \sum_{l \leq j} 2^l e^{-CM^2 l} \leq \text{const} \end{aligned}$$

for M large enough. This completes the proof of the BvM with centring $T_n = \mathbb{X}^{(n)}$. From weak convergence toward \mathcal{N} of the posterior measures and uniform integrability (as one can uniformly bound $1 + \varepsilon$ -moments by the same arguments as above), we deduce as in Theorem 10 in [4] that $\sqrt{n}(E(f|X) - \mathbb{X}^{(n)}) \rightarrow E\mathcal{N} = 0$ in \mathcal{M}_0 in probability, so that the posterior mean can replace $\mathbb{X}^{(n)}$ as the centring, completing the proof.

5.4. *Proof of Theorem 3.* For h a positive function in L^2 , denote

$$c(h) = \log \int_0^1 h(u) du,$$

so that $he^{-c(h)}$ becomes a density on $[0, 1]$. Also, for any element g of $L^2(P_0)$, denote $\|g\|_L^2 := P_0(g - P_0g)^2 = \int_0^1 (g - \int_0^1 g)^2 dP_0$, where $\|\cdot\|_L$ is a norm on the subspace of $L^2(P_0)$ consisting of P_0 -centered functions. For simplicity of notation within the proof, we denote $X = X^{(n)}$.

Let ρ_n the rate in Lemma 4, where we take $M_n = (\log n) \wedge (w_{L_n}/\sqrt{L_n})^{1/2} \rightarrow \infty$. For ε_n, C , respectively, the rate and constant in Lemma 3, we set

$$D_n = \left\{ f = e^{T-c(T)}, \|f - f_0\|_\infty \leq \rho_n, \max_{l \leq K, k} |\langle T, \psi_{lk} \rangle| \leq C\sqrt{n}\varepsilon_n \right\},$$

where the part involving the maximum in the definition of D_n is only needed for the prior (S2), and where K is a large enough integer. Combining Lemmas 3 and 4, we have $E_{f_0} \Pi[D_n|X] \rightarrow 1$. We also note that for any $l > K$ and any k , the functions ψ_{lk} are orthogonal to constants in L^2 .

We apply Proposition 6 and the remark after it, with the posterior conditioned on D_n , using the decomposition, for $L = L_n$ and writing $\pi_{V_L}(P_n)$ for $\pi_{V_L}(P_n(L))$,

$$\begin{aligned} \tilde{Y}_n &= \sqrt{n}(f - P_n(L)) \\ &= \sqrt{n}(\pi_{V_L}(f) - \pi_{V_L}(P_n)) + \sqrt{n}(f - \pi_{V_L}(f)) =: Y_n + r_n. \end{aligned}$$

Thus to prove (26) it suffices to show (i) that $\tilde{Y}_n - Y_n$ is asymptotically negligible and to check the conditions of Proposition 6, that is, (ii) that (38) holds for Y_n , and (iii) that finite-dimensional convergence (37) occurs.

(i) The term r_n is zero in the case of the histogram prior (H), by definition of the prior and orthogonality of the Haar basis. To check that r_n is negligible for the log-density priors (S), let us write $f = f_0 + (f - f_0)$ and study separately $\pi_{V_L^c} f_0$ and $\pi_{V_L^c}(f - f_0)$. For both choices of L , we have $L \geq l_n$, so

$$\begin{aligned} \sqrt{n}\|\pi_{V_L^c} f_0\|_{\mathcal{M}_0} &\leq \sqrt{n} \sup_{l > l_n} w_l^{-1} \max_k |\langle f_0, \psi_{lk} \rangle| \\ &\lesssim \sqrt{n}(l_n^{1/2}/w_{l_n}) \sup_{l > l_n} l^{-1/2} 2^{-l((1/2)+\alpha)} = o(1), \end{aligned}$$

using that $f_0 \in C^\alpha$, admissibility of w and the definition of l_n . Also,

$$\begin{aligned} & \sqrt{n} \int \|\pi_{V_L^c}(f - f_0)\|_{\mathcal{M}_0} d\Pi^{D_n}(f|X) \\ &= \sqrt{n} \int \sup_{l>L} w_l^{-1} \max_k |\langle f - f_0, \psi_{lk} \rangle| d\Pi^{D_n}(f|X) \\ &\leq \sqrt{n} \sup_{l>L} w_l^{-1} \|\psi_{lk}\|_1 \int \|f - f_0\|_\infty d\Pi^{D_n}(f|X) \\ &\lesssim \sqrt{n}(L^{1/2}/w_L)L^{-1/2}2^{-L/2}M_n(2^L L/n)^{1/2} = o(1), \end{aligned}$$

using $\|\psi_{lk}\|_1 \lesssim 2^{-l/2}$ and Lemma 4 with $M_n \rightarrow \infty$ as defined above.

(ii) To control Y_n , a key ingredient is a bound on the following exponential moment restricted to D_n . Below we prove that for universal constants c_1, c_2 and $|s| \leq \sqrt{l}$, for any $l \leq L$ and k ,

$$(41) \quad \int e^{s\sqrt{n}\langle f - P_n, \psi_{lk} \rangle} d\Pi^{D_n}(f|X) \leq c_1 e^{c_2 s^2} \Pi(D_n|X)^{-1}.$$

Suppose for now that (41) is established. Then aiming at checking (38) with $\bar{w}_l = \sqrt{l}$, we can use it in the study of

$$\sqrt{n} \|\pi_{V_L}(f - P_n)\|_{\mathcal{M}_0(\sqrt{l})} = \sqrt{n} \max_{l \leq L} l^{-1/2} \max_k |\langle f - P_n, \psi_{lk} \rangle|$$

in expectation under $\Pi^{D_n}(\cdot|X)$. Denoting E and \Pr , respectively, for expectation and probability under $\Pi^{D_n}(\cdot|X)$, for any $M > 0$, with $\mathcal{M}_0 = \mathcal{M}_0(\sqrt{l})$,

$$\begin{aligned} \sqrt{n} E \|\pi_{V_L}(f - P_n)\|_{\mathcal{M}_0} &\leq M + \int_M^\infty \Pr[\sqrt{n} \|\pi_{V_L}(f - P_n)\|_{\mathcal{M}_0} > u] du \\ &\leq M + \sum_{l < L, k} \int_M^\infty \Pr[\sqrt{n} |\langle f - P_n, \psi_{lk} \rangle| > \sqrt{l}u] du. \end{aligned}$$

An application of Markov’s inequality for $u > 0$ leads to

$$\Pr[\sqrt{n} |\langle f - P_n, \psi_{lk} \rangle| > \sqrt{l}u] \leq e^{-lu} E[e^{\sqrt{ln}|\langle f - P_n, \psi_{lk} \rangle|}].$$

Combining the last two bounds with (41) leads to

$$\begin{aligned} \sqrt{n} E \|\pi_{V_L}(f - P_n)\|_{\mathcal{M}_0} &\lesssim M + \sum_{l < L} 2^l e^{c_2 l} \Pi(D_n|X)^{-1} \int_M^\infty e^{-lu} du \\ &\lesssim M + \Pi(D_n|X)^{-1} \sum_{l < L} l^{-1} e^{[l \log 2 + lc_2 - lM]}. \end{aligned}$$

For M large enough, the last display is bounded by $M + C \Pi(D_n|X)^{-1}$. Since $\Pi(D_n|X)^{-1} \rightarrow 1$ in probability, one obtains

$$(42) \quad \sqrt{n} E \|\pi_{V_L}(f - P_n)\|_{\mathcal{M}_0(\sqrt{l})} = O_{P_{f_0}}(1).$$

Combining (42) with Markov's inequality and Proposition 6, we see that the BvM result will follow from (41), and from convergence of finite-dimensional distributions that we check in point (iii) below.

Now we check (41), in two steps. First, Lemma 1 below enables us to incorporate the term $\langle f - P_n, \psi_{lk} \rangle$ into the likelihood coming from Bayes' formula applied to $d\Pi^{D_n}(f|X)$ and reduces the problem to a change of measure with respect to the prior. Next, this change of measure is handled below.

Let us now apply Lemma 1 below to $\Pi_n = \Pi^{D_n}$ for Π , one of the considered priors. Set $\gamma_n = \psi_{lk}$. First note that $\|\tilde{\gamma}_n\|_L^2 = \int \psi_{lk}^2 f_0 \lesssim \|f_0\|_\infty$, which is bounded by assumption for $f_0 \in \mathcal{F}_0$. Next note that, for $l \leq L$,

$$\|\tilde{\gamma}_n\|_\infty \lesssim \|\psi_{lk}\|_\infty \lesssim 2^{l/2} \lesssim 2^{L/2}.$$

For h the Hellinger distance we have $h(f, f_0)^2 \lesssim \|f - f_0\|_2^2 \leq \|f - f_0\|_\infty^2$ valid for $f_0 \in \mathcal{F}_0$. Hence on D_n we have that $h(f, f_0) \lesssim \rho_n$. Since $\alpha > 1/2$, we have $2^{L/2} \log n \leq \rho_n^{-1}$, so one can apply Lemma 1 with $a_n = \rho_n$ and deduce

$$(43) \quad \int e^{s\sqrt{n}\langle f - P_n, \psi_{lk} \rangle} d\Pi^{D_n}(f|X) \lesssim \frac{e^{Cs^2} \int_{D_n} e^{\ell_n(f_s) - \ell_n(f_0)} d\Pi(f)}{\Pi(D_n|X) \int e^{\ell_n(f) - \ell_n(f_0)} d\Pi(f)}.$$

Now we are ready to change variables in the last ratio. For each of the examples of priors considered, we show that this ratio is bounded from above by a constant as $n \rightarrow \infty$.

We start with case (S). By definition, the quantity f_s is a function of $\log f - s\tilde{\gamma}_n/\sqrt{n}$. Next, notice that any constant in this expression vanishes due to the subtraction of the renormalising constant $c(\cdot)$. In particular, the expression is a function of $T - s\gamma_n/\sqrt{n}$, where T is defined in (21). The law of T is induced by a finite product of probability measures, via the distributions of the coordinates of T over $\{\psi_{lk}\}$ with $l < L$. Since $\gamma_n = \psi_{lk}$, only one coordinate of the product measure defining T is affected by the subtraction of $s\gamma_n/\sqrt{n}$. The next step is to change variables in the numerator of the ratio above by shifting the corresponding coordinate by s/\sqrt{n} .

For (S1), the change in density on this coordinate can be measured by

$$(44) \quad \frac{\varphi_H(\cdot/\sigma_l)}{\varphi_H((\cdot - s/\sqrt{n})/\sigma_l)},$$

whose logarithm is bounded above in absolute value by $1/(\sqrt{n}\sigma_l) \lesssim 1$, since by assumption, $\log \varphi_H$ is Lipschitz and using (24) combined with $l \leq L$.

In case (S2), the prior on each coordinate is Gaussian, and if θ_{lk} denotes the integrating variable with respect to the coordinate l, k (corresponding to integrating out the law of $\langle T, \psi_{lk} \rangle$) in the considered ratio of integrals, we have

$$(45) \quad \log \frac{\varphi_G(\theta_{lk}/\sigma_l)}{\varphi_G((\theta_{lk} - s/\sqrt{n})/\sigma_l)} = \frac{1}{n\sigma_l^2} - \frac{s}{\sqrt{n}\sigma_l^2} \theta_{lk}.$$

Recall that we work on the set D_n , on which we have the following inequalities: $\|\log(f/f_0)\|_2 \leq \|\log(f/f_0)\|_\infty \lesssim \rho_n$, using that f_0 is bounded from below.

Moreover, note that by definition of T , and if $g := \log f$, $g_0 := \log f_0$, it holds $\langle g - g_0, \psi_{lk} \rangle = \langle T - c(T) - g_0, \psi_{lk} \rangle$. Since $c(T)$ is a constant, and ψ_{lk} are orthogonal to constants for $l \geq K$, K large enough, we deduce, if $g_{0,lk} := \langle g_0, \psi_{lk} \rangle$, that on D_n we have $(\theta_{lk} - g_{0,lk})^2 \lesssim \rho_n^2$, as soon as $l \geq K$. So, for $K \leq l \leq L$, we have

$$|(45)| \lesssim \frac{1}{n\sigma_l^2} + \frac{\rho_n |s|}{\sqrt{n}\sigma_l^2} + \frac{|g_{0,lk}| |s|}{\sqrt{n}\sigma_l^2} \lesssim 1 + \frac{\rho_n}{\sqrt{n}\sigma_l^2} \sqrt{l} + \frac{|g_{0,lk}|}{\sigma_l} \sqrt{l}.$$

Since $\log f_0$ belongs to C^α by assumption and with (24), the last term in the last display is at most a constant. We also have $\sqrt{l}\rho_n \lesssim \sqrt{n}\sigma_l^2$ using (24) in the Gaussian case, thus the previous display is at most a constant on D_n . Now we are left with the indexes such that $l \leq K$. For those, by definition of the set D_n , (45) is in absolute value less than $(n^{-1} + \varepsilon_n)\sqrt{l}\sigma_l^{-2}$. Since $l \leq K$ with K fixed, the last expression is bounded, which yields (41).

Finally, the case of the histogram prior (H) is treated by studying the effect of the change of variables on the Dirichlet distribution. The argument is similar to [3], Section 4.4 and is omitted.

(iii) Convergence of finite-dimensional distributions (37). This can be seen to consist of establishing BvM results for the projected law of the posterior distribution on any fixed finite-dimensional subspace $V = \text{Vect}\{\psi_{lk}, (l, k) \in \mathcal{T}\}$, with \mathcal{T} a finite admissible set of indexes. By Cramér–Wold, this is the same as showing a BvM for estimating the linear functional $\langle f, \psi_{\mathcal{T}} \rangle_2$, with $\psi_{\mathcal{T}} := \sum_{(l,k) \in \mathcal{T}} t_{l,k} \psi_{lk}$ and $t_{l,k} \in \mathbb{R}$. Denote by $\pi_{\mathcal{T}}$ the mapping, for any finite set of indices \mathcal{T} ,

$$\pi_{\mathcal{T}}: f \rightarrow \langle f, \psi_{\mathcal{T}} \rangle_2.$$

Then it is enough to show that, for any finite \mathcal{T} ,

$$\beta_V(\Pi(\cdot|X) \circ \tau_{P_n}^{-1} \circ \pi_{\mathcal{T}}^{-1}, N(0, \|\psi_{\mathcal{T}}\|_L^2)) \rightarrow 0,$$

as $n \rightarrow \infty$. Since \mathcal{T} is finite, the supremum-norm $\|\psi_{\mathcal{T}}\|_\infty$ is bounded. Thus the techniques of [5] can be used for the considered priors.

In the case of histogram priors (H), the previous display follows from the section on random histograms in [5], applied to dyadic histograms. The functional $\pi_{\mathcal{T}}$ above is linear, so the no-bias condition in [5] amounts to check, with $g_{[K_n]}$ the L^2 projection of a given function g in L^2 onto the space of regular dyadic histograms of level K_n , that $\sqrt{n} \int (\psi_{\mathcal{T}} - \psi_{\mathcal{T},[K_n]})(f_0 - f_{0,[K_n]}) = o(1)$. But $\psi_{\mathcal{T}}$ is a dyadic histogram of fixed meshwidth, thus $\psi_{\mathcal{T},[K_n]} = \psi_{\mathcal{T}}$ for large enough n , since $K_n = L_n \rightarrow \infty$, so this trivially holds. Finally, since $K_n = L_n \rightarrow \infty$, the variances $\int (\psi_{\mathcal{T},[K_n]} - \int \psi_{\mathcal{T},[K_n]} f_0)^2 f_0$ converge to $\int (\psi_{\mathcal{T}} - \int \psi_{\mathcal{T}} f_0)^2 f_0$.

In the case of log-density priors (S), one applies the general result on density estimation in [5] (Theorem 4.1). The set A_n in that statement should be replaced by the set D_n defined above. Since D_n is contained in $A_n = \{f : \|f - f_0\|_1 \leq \rho_n\}$ and $\Pi(D_n|X)$ tends to 1 in probability, the proof of that Theorem goes through without further changes. It thus suffices to verify that the ratio of integrals in the former theorem from [5] holds when the functional $f \rightarrow \langle f, \psi_{\mathcal{T}} \rangle_2$ is considered.

Note that this is the same as proving that the ratio on the right-hand side of (43) goes to 1, with ψ_{lk} replaced by $\psi_{\mathcal{T}}$ and now $f_s = f e^{-t\psi_{\mathcal{T}} - c(f e^{-t\psi_{\mathcal{T}}})}$. Since only a finite number of ψ_{lk} s are involved in the sum defining $\psi_{\mathcal{T}}$, the ratio involved in the change of variables tends to 1 in probability: in the case of log-Lipschitz densities, one uses a finite number of times the bound $1/(\sqrt{n}\sigma_l)$ for the logarithm of (44), which is of the order $1/\sqrt{n}$ because l is now bounded. For the Gaussian density case, one argues similarly.

We conclude with the following auxiliary results: for $L_n = l_n$ these are Lemmas 3, 6, 9 and Theorems 2, 3 in [3], and the case $L_n = j_n$ is proved in the same way. Let $h(f, g)$ denote the Hellinger distance between two given densities f, g , and write $\ell_n(f) = (1/n) \sum_{i=1}^n \log f(X_i)$ for $f > 0$.

LEMMA 1. *Let f_0 belong to \mathcal{F}_0 . Let $\{a_n\}$ be a sequence of reals such that $na_n^2 \geq 1$ for any $n \geq 1$. Let $\{\Pi_n\}$ be a collection of priors on densities restricted to the set $\{f, h(f, f_0) \leq a_n\}$. Let $\{\gamma_n\}$ be an arbitrary sequence in $L^\infty[0, 1]$. Set $\tilde{\gamma}_n := \gamma_n - P_0\gamma_n$. Suppose, for some $m > 0$ and all $n \geq 1$,*

$$\|\tilde{\gamma}_n\|_L \leq m, \quad \|\tilde{\gamma}_n\|_\infty \leq (4a_n \log(n + 1))^{-1}.$$

Then there exist $C > 0$ depending on $m, \|f_0\|_\infty$ only such that for any $n \geq 1$ and $|t| \leq \log n$, with $W_n(\gamma_n) = \sqrt{n}(P_n - P_0)\gamma_n$,

$$E^{\Pi_n} [e^{t\sqrt{n}(f-f_0, \gamma_n)_2} | X^{(n)}] \leq e^{Ct^2 + tW_n(\gamma_n)} \frac{\int e^{\ell_n(f_i) - \ell_n(f_0)} d\Pi_n(f)}{\int e^{\ell_n(f) - \ell_n(f_0)} d\Pi_n(f)},$$

where f_i is defined by $\log f_i = \log f - t\tilde{\gamma}_n/\sqrt{n} - c(f e^{-t\tilde{\gamma}_n/\sqrt{n}})$.

LEMMA 2. *Let f, f_0 be two densities such that f_0 is bounded away from infinity. Let g be an element of L^∞ such that $h(f, f_0)\|g\|_\infty \leq C_1$ and $\|g\|_2 \leq C_2$, for some constants $C_1, C_2 > 0$. Then*

$$|(P - P_0)g^2| \leq C_1^2 + C_1\sqrt{4C_2\|f_0\|_\infty + C_1^2}.$$

LEMMA 3. *Let $f_0 \in \mathcal{F}_0$, and suppose $\log f_0 \in C^\alpha, \alpha > 1$. Let $\varphi = \varphi_G$ and σ_l satisfy (24), L_n be as defined in (20) and the prior Π be defined by (22). Then there exists $\nu > 0$ such that if $\varepsilon_n = (\log n)^\nu n^{-\alpha/(2\alpha+1)}$, for $C > 0$ large enough and any fixed given integer K ,*

$$E_{f_0} \Pi \left[\max_{\lambda \leq K, \mu} |\langle T, \psi_{\lambda\mu} \rangle| \leq C\sqrt{n}\varepsilon_n | X^{(n)} \right] \rightarrow 1.$$

Finally, for any $\alpha > 0$ and $n \geq 2$, let us set $\varepsilon_{n,\alpha}^* := (n/\log n)^{-\alpha/(2\alpha+1)}$.

LEMMA 4. *Let Π be of the type (S) or (H) with L_n as in (20) and suppose (24) and (25) are, respectively, satisfied for the corresponding prior. Suppose f_0 belongs to $C^\alpha, 1/2 < \alpha \leq 1$ in the case of prior (H) and $\log f_0 \in C^\alpha, \alpha > 1$ for*

priors (S). Then, as $n \rightarrow \infty$,

$$(46) \quad E_{f_0} \Pi[f : \|f - f_0\|_\infty > \rho_n | X^{(n)}] \rightarrow 0,$$

where, for an arbitrary sequence $M_n \rightarrow \infty$, $\rho_n^2 = M_n^2 L_n 2^{L_n} / n$. That is, $\rho_n = M_n \varepsilon_{n,\alpha}^*$ if $L_n = l_n$ and $\rho_n = M_n \varepsilon_{n,\alpha}^* \sqrt{\log n}$ if $L_n = j_n$.

Acknowledgments. The authors thank two anonymous referees and the Associate Editor for comments that helped to improve the presentation of the paper. Richard Nickl is grateful to the LPMA at Université Paris VII Denis Diderot for its hospitality during a visit in October–November 2012 where this research was initiated.

REFERENCES

- [1] BOGACHEV, V. I. (1998). *Gaussian Measures*. Amer. Math. Soc., Providence, RI. [MR1642391](#)
- [2] BONTEMPS, D. (2011). Bernstein–von Mises theorems for Gaussian regression with increasing number of regressors. *Ann. Statist.* **39** 2557–2584. [MR2906878](#)
- [3] CASTILLO, I. (2014). On Bayesian supremum norm contraction rates. *Ann. Statist.* To appear. Available at [arXiv:1304.1761v2](#).
- [4] CASTILLO, I. and NICKL, R. (2013). Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *Ann. Statist.* **41** 1999–2028. [MR3127856](#)
- [5] CASTILLO, I. and ROUSSEAU, J. (2013). A general Bernstein–von Mises theorem in semiparametric models. Preprint. Available at [arXiv:1305.4482](#).
- [6] CIESIELSKI, Z., KERKYACHARIAN, G. and ROYNETTE, B. (1993). Quelques espaces fonctionnels associés à des processus Gaussiens. *Studia Math.* **107** 171–204. [MR1244574](#)
- [7] COHEN, A., DAUBECHIES, I. and VIAL, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* **1** 54–81. [MR1256527](#)
- [8] CONTI, P. L. (1999). Large sample Bayesian analysis for Geo/G/1 discrete-time queueing models. *Ann. Statist.* **27** 1785–1807. [MR1765617](#)
- [9] CONTI, P. L. (2004). Approximated inference for the quantile function via Dirichlet processes. *Metron* **62** 201–222. [MR2102100](#)
- [10] DAVIES, P. L. and KOVAC, A. (2001). Local extremes, runs, strings and multiresolution. *Ann. Statist.* **29** 1–65. [MR1833958](#)
- [11] DAVIES, P. L., KOVAC, A. and MEISE, M. (2009). Nonparametric regression, confidence regions and regularization. *Ann. Statist.* **37** 2597–2625. [MR2541440](#)
- [12] DÜMBGEN, L. and SPOKOINY, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124–152. [MR1833961](#)
- [13] DÜMBGEN, L. and WALTHER, G. (2008). Multiscale inference about a density. *Ann. Statist.* **36** 1758–1785. [MR2435455](#)
- [14] FREEDMAN, D. (1999). On the Bernstein–von Mises theorem with infinite-dimensional parameters. *Ann. Statist.* **27** 1119–1140. [MR1740119](#)
- [15] GHOSAL, S. (1999). Asymptotic normality of posterior distributions in high-dimensional linear models. *Bernoulli* **5** 315–331. [MR1681701](#)
- [16] GINÉ, E. and NICKL, R. (2009). Uniform limit theorems for wavelet density estimators. *Ann. Probab.* **37** 1605–1646. [MR2546757](#)
- [17] GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. [MR2604707](#)
- [18] GINÉ, E. and ZINN, J. (1990). Bootstrapping general empirical measures. *Ann. Probab.* **18** 851–869. [MR1055437](#)

- [19] HALL, P. (1979). On the rate of convergence of normal extremes. *J. Appl. Probab.* **16** 433–439. [MR0531778](#)
- [20] HALL, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.* **20** 675–694. [MR1165587](#)
- [21] HJORT, N. L. and PETRONE, S. (2007). Nonparametric quantile inference using Dirichlet processes. In *Advances in Statistical Modeling and Inference* 463–492. World Sci. Publ., Hackensack, NJ. [MR2416129](#)
- [22] HJORT, N. L. and WALKER, S. G. (2009). Quantile pyramids for Bayesian nonparametrics. *Ann. Statist.* **37** 105–131. [MR2488346](#)
- [23] HOFFMANN-JØRGENSEN, J., SHEPP, L. A. and DUDLEY, R. M. (1979). On the lower tail of Gaussian seminorms. *Ann. Probab.* **7** 319–342. [MR0525057](#)
- [24] JOHNSTONE, I. M. (2010). High dimensional Bernstein–von Mises: Simple examples. In *Borrowing Strength: Theory Powering Applications—a Festschrift for Lawrence D. Brown*. *Inst. Math. Stat. Collect.* **6** 87–98. IMS, Beachwood, OH. [MR2798513](#)
- [25] KATO, K. (2013). Quasi-Bayesian analysis of nonparametric instrumental variables models. *Ann. Statist.* **41** 2359–2390. [MR3127869](#)
- [26] KIM, Y. (2006). The Bernstein–von Mises theorem for the proportional hazard model. *Ann. Statist.* **34** 1678–1700. [MR2283713](#)
- [27] KIM, Y. and LEE, J. (2004). A Bernstein–von Mises theorem in the nonparametric right-censoring model. *Ann. Statist.* **32** 1492–1512. [MR2089131](#)
- [28] LEAHU, H. (2011). On the Bernstein–von Mises phenomenon in the Gaussian white noise model. *Electron. J. Stat.* **5** 373–404. [MR2802048](#)
- [29] LEDOUX, M. (2001). *The Concentration of Measure Phenomenon*. *Mathematical Surveys and Monographs* **89**. Amer. Math. Soc., Providence, RI. [MR1849347](#)
- [30] LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York. [MR0856411](#)
- [31] LO, A. Y. (1983). Weak convergence for Dirichlet processes. *Sankhyā Ser. A* **45** 105–111. [MR0749358](#)
- [32] RIVOIRARD, V. and ROUSSEAU, J. (2012). Bernstein–von Mises theorem for linear functionals of the density. *Ann. Statist.* **40** 1489–1523. [MR3015033](#)
- [33] SCHMIDT-HIEBER, J., MUNK, A. and DÜMBGEN, L. (2013). Multiscale methods for shape constraints in deconvolution: Confidence statements for qualitative features. *Ann. Statist.* **41** 1299–1328. [MR3113812](#)
- [34] SZABÓ, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2013). Frequentist coverage of adaptive nonparametric Bayesian credible sets. Preprint. Available at [arXiv:1310.4489](#).
- [35] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge Univ. Press, Cambridge. [MR1652247](#)
- [36] 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](#)
- [37] 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](#)

CNRS—LABORATOIRE PROBABILITÉS
 ET MODÈLES ALÉATOIRES
 UNIVERSITÉS PARIS VI AND VII
 BÂTIMENT SOPHIE GERMAIN
 75205 PARIS CEDEX 13
 FRANCE
 E-MAIL: ismael.castillo@upmc.fr

STATISTICAL LABORATORY
 DEPARTMENT OF PURE MATHEMATICS
 AND MATHEMATICAL STATISTICS
 UNIVERSITY OF CAMBRIDGE
 CB3 0WB CAMBRIDGE
 UNITED KINGDOM
 E-MAIL: r.nickl@statslab.cam.ac.uk