The Annals of Statistics
2013, Vol. 41, No. 4, 1999–2028
DOI: 10.1214/13-AOS1133
© Institute of Mathematical Statistics, 2013

NONPARAMETRIC BERNSTEIN–VON MISES THEOREMS IN GAUSSIAN WHITE NOISE

BY ISMAËL CASTILLO[1] AND RICHARD NICKL

CNRS and University of Cambridge

Bernstein–von Mises theorems for nonparametric Bayes priors in the Gaussian white noise model are proved. It is demonstrated how such results justify Bayes methods as efficient frequentist inference procedures in a variety of concrete nonparametric problems. Particularly Bayesian credible sets are constructed that have asymptotically exact $1-\alpha$ frequentist coverage level and whose L^2-diameter shrinks at the minimax rate of convergence (within logarithmic factors) over Hölder balls. Other applications include general classes of linear and nonlinear functionals and credible bands for auto-convolutions. The assumptions cover nonconjugate product priors defined on general orthonormal bases of L^2 satisfying weak conditions.

1. Introduction. Consider observing a random sample $X^{(n)}$ of size n, or at noise level $n^{-1/2}$, drawn from distribution P_f^n, and indexed by some unknown parameter $f \in \mathcal{F}$. The Bayesian paradigm views the sample as having law P_f^n conditionally on f, that is, $X^{(n)}|f \sim P_f^n$, and the law of f is a *prior probability distribution* Π on some σ-field $\mathcal{B}$ of $\mathcal{F}$. The random variable $f|X^{(n)}$ then has a law on $\mathcal{F}$ which is known as the *posterior distribution*, denoted by $\Pi(\cdot|X^{(n)})$. Bayesian inference on f is then entirely based on this posterior distribution—it gives access to point estimates for f, credible sets and tests in a natural way.

It is of interest to analyse the behaviour of $\Pi(\cdot|X^{(n)})$ under the frequentist sampling assumption that $X^{(n)}$ is drawn from $P_{f_0}^n$ for some fixed nonrandom $f_0 \in \mathcal{F}$. If $\mathcal{F}$ is a *finite-dimensional* space, then posterior-based inference has a fundamental justification through the Bernstein–von Mises (BvM) theorem, first discovered by Laplace [26], developed by von Mises [40], and put into the framework of modern parametric statistics by Le Cam [28]. It states that, under mild and universal assumptions on the prior, the posterior distribution approximately equals a normal distribution $N(\hat{f}_n, i(f))$ centered at an efficient estimator $\hat{f}_n$ for f and with a covariance $i(f)$ that attains the Cramér–Rao bound in the statistical model considered: as $n \to \infty$

$$\sup_{B\in\mathcal{B}} \left|\Pi\left(B|X^{(n)}\right) - N\left(\hat{f}_n, i(f_0)\right)(B)\right| \to^{P_{f_0}^n} 0. \tag{1}$$

Received January 2013; revised May 2013.

[1]Supported 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, plug-in property, efficiency.

As a consequence posterior-based inference asymptotically coincides with inference based on standard efficient, $1/\sqrt{n}$-consistent frequentist estimators $\hat{f}_n$ of f, giving a rigorous asymptotic justification of Bayesian methods.

The last decade has seen remarkable activity in the development of *nonparametric* Bayes procedures, where $\mathcal{F}$ is taken to be an infinite-dimensional space, typically consisting of functions or infinite vectors: nonparametric regression, classification, density estimation, normal means, and Gaussian white noise models come to mind, and a variety of nonparametric priors have been devised in the literature for such models. Posteriors can be computed efficiently by algorithms such as MCMC, and they provide broadly applicable Bayesian inferential tools for nonparametric problems. It is natural to ask whether an analogue of (1) can still be proved in such situations, as it would give a general justification for the use of nonparametric Bayes procedures. Although remarkable progress has been made in the understanding of the frequentist properties of nonparametric Bayes procedures (we refer here only to some of the key papers such as [18, 19, 34, 36, 38] and references therein), a fully satisfactory answer to the BvM-question seems not to have been found. A first reason is perhaps that it is not immediately clear what $N(\hat{f}_n, i(f_0))$ should be replaced by in the infinite-dimensional situation—Gaussian distributions over infinite-dimensional spaces $\mathcal{F}$ are much more complex objects, and their existence in the form relevant here depends on the topology that $\mathcal{F}$ is endowed with. Another reason is that the commonly used loss functions in nonparametric statistics (such as L^p-type loss) do not admit $1/\sqrt{n}$-consistent estimators—the LAN-type local approximations of the likelihood function used in the proof of the finite-dimensional BvM theorem are thus not accurate enough in such metrics.

One way around these problems is to weaken the loss function on $\mathcal{F}$ so that $1/\sqrt{n}$-consistent estimation with Gaussian limits is possible even in nonparametric models. For example, in the situation where one observes $X_1, \dots, X_n \sim^{\text{i.i.d.}} P$ on $[0, 1]$, and $P_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ is the empirical measure estimating P, then for any P-Donsker class $\mathcal{H}$ of functions $h : [0, 1] \to \mathbb{R}$,

$$\sup_{h\in\mathcal{H}} \left| \int h d(P_n - P) \right| = \sup_{h\in\mathcal{H}} \left| \frac{1}{n} \sum_{i=1}^n \big(h(X_i) - Eh(X_1)\big) \right| = O_P\big(n^{-1/2}\big), \tag{2}$$

in fact P_n is an efficient estimator for P in the space $l^\infty(\mathcal{H})$ of bounded functions on $\mathcal{H}$, attaining the Cramér–Rao information bound for the fully nonparametric model; cf. 3.1.11 in [39]. Hence one may try to prove a BvM-type result by endowing the parameter space $\mathcal{F}$ with the loss function coming from an $l^\infty(\mathcal{H})$-type space. The purpose of the present paper is to investigate this approach rigorously in the setting of the Gaussian white noise model, and with $\mathcal{H}$ a ball in a suitable Sobolev space defined below. This makes the mathematical analysis tractable without any severe loss of conceptual generality; see below for a discussion of extensions to other models. Our main results will imply that for a large and relevant class of nonparametric product priors Π that satisfy mild assumptions, and which do not require conjugacy, one has

$$\sup_{A\in\mathcal{A}}\bigl|\Pi\bigl(A|X^{(n)}\bigr)-\mathcal{N}\bigl(\hat{f}_n, i(f_0)\bigr)(A)\bigr| \to^{P^n_{f_0}} 0, \tag{3}$$

where $\mathcal{N}$ is a Gaussian measure on $l^\infty(\mathcal{H})$ centered at an efficient estimator $\hat{f}_n$ of f, both to be defined in a precise manner, and where the classes $\mathcal{A}$ consist of measurable subsets of $l^\infty(\mathcal{H})$ that have uniformly smooth boundaries for the measure $\mathcal{N}$. We note that some restrictions on the class $\mathcal{A}$ are necessary as one can show that in the infinite-dimensional situation the Bernstein–von Mises theorem cannot hold uniformly in all Borel sets of $l^\infty(\mathcal{H})$; see after Definition 1 for discussion. Our assumptions apply to priors that produce posteriors which achieve frequentist optimal contraction rates in stronger loss functions (such as L^2-distance) and which resemble the state of the art prior choices in the nonparametric Bayes literature.

Our abstract results only gain relevance through the fact that we can demonstrate their applicability: the general result (3) will be shown to imply that posterior-based credible regions give asymptotically exact frequentist confidence sets in a variety of concrete problems of nonparametric inference. A first important application is to weighted L^2-ellipsoid credible regions for the unknown parameter f, which are shown to have optimal width $O_P(n^{-1/2})$ in $\ell^\infty(\mathcal{H})$-loss and which simultaneously are confidence sets that shrink in L^2-diameter at the minimax rate (within log-factors) over Hölder balls. We further give semiparametric applications to estimation of linear and nonlinear functionals defined on $\mathcal{F}$, and to credible bands for estimating an auto-convolution $f * f$.

A key point in these applications is related to the notion of the "plug-in property" coined by Bickel and Ritov [3]. A nonparametric estimator that is rate-optimal in a standard loss function (such as L^p-loss) is said to have the plug-in property if it simultaneously efficiently estimates, at $1/\sqrt{n}$-rate, a large class of linear functionals. Standard frequentist estimators such as kernel, wavelet and nonparametric maximum likelihood estimators satisfy this property; in fact, one can even prove a corresponding uniform central limit theorem in $l^\infty(\mathcal{H})$ for such estimators; see Kiefer and Wolfowitz [24], Nickl [30], Giné and Nickl [21, 22]. Our results imply that this is also true in the Bayesian situation: the posterior contracts at the optimal rate in L^2-loss and at the same time satisfies a Bernstein–von Mises theorem in $l^\infty(\mathcal{H})$.

Our techniques of proof rely on the structure of the Gaussian white noise model and apply as well, with simple modifications, to fixed design nonparametric regression settings. Our proofs moreover indicate a strategy to obtain BvM results of this kind in general nonparametric sampling models: the idea is to obtain "semiparametric" BvM-results for many fixed linear functionals simultaneously, and to re-construct "nonparametric" norms from these functionals. At least for density estimation such ideas can be shown to work as well, and this is subject of forthcoming research.

There is important work on the BvM phenomenon for nonparametric procedures that needs mentioning. Cox [11] and Freedman [14] have shown the impossibility of a nonparametric BvM result in a strict L^2-setting. Leahu [29] derives interesting results on the possibility and impossibility of BvM-theorems for undersmoothing priors—his negative results will be relevant below. His positive findings are, however, strongly tied to the Gaussian conjugate situation, do not address efficiency questions, and do not give rise to posteriors with the above mentioned "plug-in property." For the related question of obtaining semiparametric BvM-results, general sufficient conditions are given in Castillo [8] and Bickel and Kleijn [2], as well as in Rivoirard and Rousseau [32] for linear functionals of probability density functions. A number of BvM-type results have been obtained for the fixed finite-dimensional posterior with dimension increasing to infinity: Ghosal [16] and Bontemps [6] consider regression with a finite number of regressors, Ghosal [17] and Clarke and Ghosal [9] consider exponential families, and the case of discrete probability distributions is treated in Boucheron and Gassiat [7].

This article is organised as follows: in the next two subsections we define a general notion of the nonparametric BvM phenomenon. In Section 2 we demonstrate that when this phenomenon holds, posterior-based inference is valid from a frequentist point of view in a variety of concrete examples from nonparametric statistics. In Section 3 we prove that for a large class of natural priors on L^2, the BvM phenomenon indeed occurs.

1.1. *The weak nonparametric Bernstein–von Mises phenomenon.* We consider a fixed design Gaussian regression model with known variance, but work with its equivalent white noise formulation to streamline the mathematical development. Let $L^2 := L^2([0, 1])$ be the space of square integrable functions on $[0, 1]$. For $f \in L^2$, dW standard white noise, consider observing

$$dX^{(n)}(t) = f(t)\,dt + \frac{1}{\sqrt{n}}\,dW(t), \qquad t \in [0, 1]. \tag{4}$$

Except in conjugate situations the proof of a Bernstein–von Mises-type result rests typically on the fact that efficient estimation at the rate $1/\sqrt{n}$ is possible. In the nonparametric situation this rules out L^p-type loss functions, but leads one to consider weaker $\ell^\infty(\mathcal{H})$-type norms discussed in (2). For the particular choice of $\mathcal{H}_s$ equal to an order-s Sobolev-ball, we can understand this better by using simple but useful Hilbert space duality arguments in the nested scale of Sobolev spaces $\{H_2^r\}_{r\in\mathbb{R}}$ on $[0, 1]$: we define these in precise detail below, but note for the moment that $H_2^r \subseteq H_2^t, r \geq t$, $H^0 = L^2$, so to weaken the norm beyond L^2 means that we should decrease r to be negative. For $s > 0$ the space H_2^{-s} can be realised in an isometric way as a closed subspace of $l^\infty(\mathcal{H}_s)$, explaining heuristically the connection to the discussion surrounding (2) above. The space should be large enough so that the Gaussian experiment in (4) can be realised as a tight random element in H_2^{-s}. The critical value for this to be the case is $s = 1/2$, and we define in (8) below a (in a certain sense "maximal") Sobolev space H with norm $\|\cdot\|_H$ in which the random trajectory $dX^{(n)}$ defines a tight Gaussian Borel random variable $\mathbb{X}^{(n)}$ with mean f and covariance $n^{-1}I$. That is, if we denote by $\mathbb{W}$ the centered Gaussian Borel random variable on H with covariance I, then (4) can be written as

$$\mathbb{X}^{(n)} = f + \frac{1}{\sqrt{n}}\mathbb{W}, \tag{5}$$

a natural Gaussian shift experiment in the Hilbert space H. One can show moreover that $\mathbb{X}^{(n)}$ is an efficient estimator for f for the loss function of H.

Any (Borel or cylindrical) probability measure on L^2 gives rise to a tight probability measure on H simply by the continuous (Hilbert–Schmidt) injection $L^2 \subset H$. Let thus Π be a prior on L^2, and let

$$\Pi_n = \Pi(\cdot|X^{(n)}) = \Pi(\cdot|\mathbb{X}^{(n)})$$

be the posterior distribution on H given the observed trajectory from (4), or equivalently, from (5). On H and for $z \in H$, define the transformation

$$\tau_z : f \mapsto \sqrt{n}(f - z).$$

Let $\Pi_n \circ \tau_{\mathbb{X}^{(n)}}^{-1}$ be the image of the posterior law under $\tau_{\mathbb{X}^{(n)}}$. The shape of $\Pi_n \circ \tau_{\mathbb{X}^{(n)}}^{-1}$ reveals how the posterior concentrates on $1/\sqrt{n}$-H-neighborhoods of the efficient estimator $\mathbb{X}^{(n)}$. To compare probability distributions on H we may use any metric for weak convergence of probability measures, and we choose the bounded Lipschitz metric here for convenience (it is defined in Section 4.1). Let $\mathcal{N}$ be the standard Gaussian probability measure on H with mean zero and covariance I constructed in Section 1.2 below. It should be distinguished from the standard Gaussian law $N(0, I)$ on $\mathbb{R}^k$, $k \in \mathbb{N}$.

DEFINITION 1. Consider data generated from equation (4) under a fixed function f_0, and denote by $P_{f_0}^n$ the distribution of $\mathbb{X}^{(n)}$. Let β be the bounded Lipschitz metric for weak convergence of probability measures on H. We say that a prior Π satisfies the weak Bernstein–von Mises phenomenon in H if, as $n \to \infty$,

$$\beta(\Pi_n \circ \tau_{\mathbb{X}^{(n)}}^{-1}, \mathcal{N}) \to^{P_{f_0}^n} 0.$$

We note that the fact that the result is phrased in a way in which $\mathcal{N}$ is independent of n is important since β does not induce a uniformity structure for the topology of weak convergence; see the remark on page 413 in [12].

Thus when the weak Bernstein–von Mises phenomenon holds, the posterior necessarily has the approximate shape of an infinite-dimensional Gaussian distribution. Moreover, we require this Gaussian distribution to equal $\mathcal{N}$, the canonical choice in view of efficiency considerations. The covariance of $\mathcal{N}$ is the Cramér–Rao bound for estimating f in the Gaussian shift experiment (5) in H-loss, and we shall see how this carries over to sufficiently regular real-valued functionals $\Psi(f)$; see Section 2.3 below.

One may ask by analogy to the finite-dimensional situation whether a *strong* Bernstein–von Mises phenomenon, where β is replaced by the total variation norm, can be proved. It follows from Theorem 2 in [29] that already in the Gaussian conjugate situation, such a result is impossible unless one restricts to very specific priors (which in particular do not possess the plug-in property that will be needed in the key applications below).

Now with weak instead of total variation convergence, we cannot infer that $\Pi_n \circ \tau_{\mathbb{X}^{(n)}}^{-1}$ and $\mathcal{N}$ are approximately the same for every Borel set in H, but only for sets B that are continuity sets for the probability measure $\mathcal{N}$. For statistical applications of the Bernstein–von Mises phenomenon, one typically needs some uniformity in B, and this is where total variation results would be particularly useful. Weak convergence in H implies that $\Pi_n \circ \tau_{\mathbb{X}^{(n)}}^{-1}$ is close to $\mathcal{N}$ uniformly in certain classes of subsets of H whose boundaries are sufficiently regular relative to the measure $\mathcal{N}$ (see Section 4.1), and we show below how this allows for enough uniformity to deal with a variety of concrete nonparametric statistical problems.

The Bernstein–von Mises phenomenon in Definition 1 will often be complemented by convergence of moments, that is, convergence of the Bochner integrals (e.g., page 100 in [1]) $\int_H f\,d\Pi_n \circ \tau_{\mathbb{X}^{(n)}}^{-1}(f) \to^{P_{f_0}^n} \int_H f\,d\mathcal{N}(f) = 0$ as $n \to \infty$ in H. This implies that the posterior mean $\bar{f}_n$ of Π_n satisfies

$$\|\bar{f}_n - \mathbb{X}^{(n)}\|_H = o_P(n^{-1/2}), \tag{6}$$

so in semiparametric terminology the posterior mean is asymptotically linear in H with respect to $\mathbb{X}^{(n)}$; in particular, $\bar{f}_n$ is an efficient estimator for f.

1.2. *Sobolev spaces and white noise.* Denote by $\langle f, g\rangle = \int_0^1 f(x)\overline{g(x)}\,dx$ the standard inner product on L^2. General order Sobolev spaces will be defined via orthonormal bases of L^2 that satisfy the following weak regularity condition. While notationally it reflects a wavelet-type basis $\{\psi_{lk} : l \geq J_0 - 1, 0 \leq k \leq 2^l - 1\}$ of CDV-type [10] (with notational convention $\psi_{(J_0-1)k} = \phi_k$ for the scaling function), it also includes the trigonometric basis $\psi_{lk}(x) \equiv e_l(x) = e^{2\pi i l x}$ and bases of standard Karhunen–Loève expansions.

DEFINITION 2. Let $S \in \mathbb{N}$. By an S-regular basis $\{\psi_{lk} : l \in \mathcal{L}, k \in \mathcal{Z}_l\}$ of L^2 with index sets $\mathcal{L} \subset \mathbb{Z}$, $\mathcal{Z}_l \subset \mathbb{Z}$ and characteristic sequence a_l we shall mean any of the following:

(a) $\psi_{lk} \equiv e_l$ is S-times differentiable with all derivatives in L^2, $|\mathcal{Z}_l| = 1$, $a_l = \max(2, |l|)$, and $\{e_l : l \in \mathcal{L}\}$ forms an orthonormal basis of L^2.

(b) ψ_{lk} is S-times differentiable with all derivatives in L^2, $\mathcal{L} \subset \mathbb{N}$, $a_l = |\mathcal{Z}_l| = 2^l$, and $\{\psi_{lk} : l \in \mathcal{L}, k \in \mathcal{Z}_l\}$ forms an orthonormal basis of L^2.

Define for $0 \le s < S$ the standard Sobolev spaces as

$$H_2^s := \left\{ f \in L^2([0,1]) : \|f\|_{s,2}^2 := \sum_{l \in \mathcal{L}} a_l^{2s} \sum_{k \in \mathcal{Z}_l} |\langle \psi_{lk}, f \rangle|^2 < \infty \right\},$$

which for the usual wavelet or trigonometric bases are in fact spaces independent of the basis. For $\mathcal{Z}_l' \subset \mathcal{Z}_l$, $\mathcal{L}' \subset \mathcal{L}$ finite we can form linear subspaces

$$V \equiv V_{\mathcal{L}', \mathcal{Z}_l'} = \operatorname{span}\{\psi_{lk} : l \in \mathcal{L}', k \in \mathcal{Z}_l'\}$$

of $H_2^s \subset L^2$, and we denote the L^2-projection of $f \in L^2$ onto V by $\pi_V(f)$.

For $s > 0$ we define the dual space

$$H_2^{-s}([0,1]) := (H_2^s[0,1])^*.$$

Using standard duality arguments (as in Proposition 9.16 in [13]) one shows the following: H_2^{-s} consists precisely of those linear forms L acting on H_2^s for which the $\|L\|_{-s,2}$-norms [defined as above also for negative s, with $\langle \psi_{lk}, L \rangle$ replaced by $L(\psi_{lk})$, noting $\psi_{lk} \in H_2^s$] are finite. In fact the so-defined norm $\|\cdot\|_{-s,2}$ is equivalent to the standard operator norm on $(H_2^s[0,1])^*$. Moreover every $f \in L^2$ gives rise to a continuous linear form on $H_2^s \subset L^2$ by using the $\langle \cdot, \cdot \rangle$ duality, so we can view L^2 as a subspace of H_2^{-s}. By reflexivity of H_2^s one concludes $(H_2^{-s}([0,1]))^* = H_2^s([0,1])$ up to isomorphism, that is, any linear continuous map $K : H_2^{-s} \to \mathbb{R}$ is of the form $K : L \mapsto L(g)$ for some $g \in H_2^s$, and if L itself is a functional coming from integrating against an L^2-function f_L, then $L(g) = \langle g, f_L \rangle$.

To obtain sharp results we also need "logarithmic" Sobolev spaces

$$H_2^{s,\delta} \equiv \left\{ f : \|f\|_{s,2,\delta}^2 := \sum_{l \in \mathcal{L}} \frac{a_l^{2s}}{(\log a_l)^{2\delta}} \sum_{k \in \mathcal{Z}_l} |\langle \psi_{lk}, f \rangle|^2 < \infty \right\}, \qquad \delta \ge 0, s \in \mathbb{R},$$

which are Hilbert spaces satisfying the compact imbeddings $H_2^r \subset H_2^{r,\delta} \subset H_2^s$ for any real valued $s < r$.

For any $f \in H_2^s \subseteq L^2$ ($s \ge 0$) and dW standard white noise, we have a random linear application

$$\mathbb{W} : f \mapsto \int_0^1 f(t)\, dW(t) \sim N(0, \|f\|_2^2). \tag{7}$$

For any $\delta > 1/2$, the $\|\mathbb{W}\|_{-1/2,2,\delta}$-norm converges almost surely since, by Fubini's theorem, for g_{lk} independent $N(0,1)$ variables,

$$E\|\mathbb{W}\|_{-1/2,2,\delta}^2 = \sum_{l \in \mathcal{L}} a_l^{-1} (\log a_l)^{-2\delta} \sum_{k \in \mathcal{Z}_l} E g_{lk}^2 < \infty,$$

so $\mathbb{W} \in H_2^{-1/2,\delta}$ almost surely, measurable for the cylindrical σ-algebra, and by separability of $H_2^{-1/2,\delta}$ also for the Borel σ-algebra (page 374 in [5]). By Ulam's theorem (Theorem 7.1.4 in [12]), $\mathbb{W}$ is thus tight in $H_2^{-1/2,\delta}$. One can show that the spaces

$$(8)\qquad H \equiv H(\delta) \equiv H_2^{-1/2,\delta}, \qquad \|\cdot\|_H \equiv \|\cdot\|_{H(\delta)} \equiv \|\cdot\|_{-1/2,2,\delta}, \qquad \delta > 1/2,$$

are minimal in the considered scale of spaces on which this happens: decreasing δ below $1/2$ would lead to a space in which $\mathbb{W}$ is not tight.

The Gaussian variable $\mathbb{W}$ has mean zero and covariance I diagonal for the L^2-inner product, that is, $E\mathbb{W}(g)\mathbb{W}(h) = \langle g, h\rangle$, for all $g, h \in L^2$. We call the law $\mathcal{N}$ of $\mathbb{W}$ a standard, or canonical, Gaussian probability measure on the Hilbert space H (note that it is the isonormal Gaussian measure for the inner product of L^2 but *not* for the one of H). In the same way the random trajectory $dX^{(n)}$ from (4) defines a tight Gaussian Borel random variable $\mathbb{X}^{(n)}$ on H with mean f and covariance $n^{-1}I$, thus rigorously justifying (5).

We finally define Hölder-type spaces of smooth functions: for $S > s > 0$ and ψ_{lk} a S-regular wavelet basis from Definition 2(b), we set

$$(9)\qquad C^s \equiv \Big\{ f \in C\big([0,1]\big) : \|f\|_{s,\infty} := \sup_{l \in \mathcal{L}, k \in \mathcal{Z}_l} 2^{l(s+1/2)} \big|\langle \psi_{lk}, f\rangle\big| < \infty \Big\}.$$

2. Confidence sets for nonparametric Bayes procedures.

2.1. *Weighted L^2-credible ellipsoids.* Throughout this section H stands for the space $H(\delta)$ from (8) for some arbitrary choice of $\delta > 1/2$. Denote by $B(g, r) = \{f \in H : \|f - g\|_H \le r\}$ the norm ball in H of radius r centered at g. In terms of an orthonormal basis $\{\psi_{lk}\}$ of L^2 from Definition 2 this corresponds to L^2-ellipsoids

$$\Big\{ \{c_{lk}\} : \sum_{l,k} a_l^{-1} \big(\log(a_l)\big)^{-2\delta} \big|c_{lk} - \langle g, \psi_{lk}\rangle\big|^2 \le r^2 \Big\},$$

where coefficients in the tail are downweighted by $a_l^{-1}(\log(a_l))^{-2\delta}$. A frequentist goodness of fit test of a null hypothesis $H_0 : f = f_0$ could, for instance, be based on the test statistic $\|f_0 - \mathbb{X}^{(n)}\|_H$, resembling in nature a Cramér–von Mises-type procedure that has power against arbitrary fixed alternatives $f \in L^2$.

A Bayesian approach consists in using the quantiles of the posterior directly. Given $\alpha > 0$ one solves for $R_n \equiv R(\mathbb{X}^{(n)}, \alpha)$ such that

$$(10)\qquad \Pi\big(f : \|f - T_n\|_H \le R_n/\sqrt{n} | \mathbb{X}^{(n)}\big) = 1 - \alpha,$$

where $T_n = \mathbb{X}^{(n)}$ or, when the posterior mean $\bar{f}_n$ exists, possibly $T_n = \bar{f}_n$. A $\|\cdot\|_H$-ball centred at T_n of radius R_n constitutes a level $(1-\alpha)$-credible set for the posterior distribution. The weak Bernstein–von Mises phenomenon in H implies that this credible ball asymptotically coincides with the exact $(1-\alpha)$-confidence set built using the efficient estimator $\mathbb{X}^{(n)}$ for f.

THEOREM 1. *Suppose the weak Bernstein–von Mises phenomenon in the sense of Definition 1 holds. Given* $0 < \alpha < 1$ *consider the credible set*

$$C_n = \{f : \|f - \mathbb{X}^{(n)}\|_H \le R_n/\sqrt{n}\}, \tag{11}$$

where $R_n \equiv R(\mathbb{X}^{(n)}, \alpha)$ *is such that* $\Pi(C_n|\mathbb{X}^{(n)}) = 1 - \alpha$. *Then, as* $n \to \infty$,

$$P^n_{f_0}(f_0 \in C_n) \to 1 - \alpha \quad \text{and} \quad R_n = O_P(1).$$

If in addition $\|\bar{f}_n - \mathbb{X}^{(n)}\|_H = o_P(n^{-1/2})$, *then the same is true if in the definition of* C_n *the posterior mean* $\bar{f}_n$ *replaces* $\mathbb{X}^{(n)}$.

When available, using further prior knowledge in the construction of the credible set may lead to favourable frequentist properties, such as optimal performance in stronger loss functions.

To see this, consider first the specific but instructive case of a uniform wavelet prior Π on L^2 arising from the law of the random wavelet series

$$U_{\gamma,M} = \sum_{l=J_0-1}^{\infty} \sum_{k=0}^{2^l-1} 2^{-l(\gamma+1/2)} u_{lk} \psi_{lk}(\cdot), \qquad \gamma > 0,$$

where the u_{lk} are i.i.d. uniform on $[-M, M]$ for some $M > 0$, with S-regular CDV-wavelets ψ_{lk}, $S > \max(\gamma, 1/2)$, $J_0 \in \mathbb{N}$. Such priors model functions that lie in a fixed Hölder ball of $\|\cdot\|_{\gamma,\infty}$-radius M, with posteriors $\Pi(\cdot|\mathbb{X}^{(n)})$ contracting about f_0 at the L^2-minimax rate within logarithmic factors if $\|f_0\|_{\gamma,\infty} \le M$; see [23] and also Corollary 3 below.

In this situation it is natural to intersect the credible set C_n with the Hölderian support of the prior (or posterior),

$$C'_n = \{f : \|f\|_{\gamma,\infty} \le M, \|f - \bar{f}_n\|_H \le R_n/\sqrt{n}\}, \tag{12}$$

where R_n is as in (10) with $T_n = \bar{f}_n$. Note that the posterior mean also satisfies $\|\bar{f}_n\|_{\gamma,\infty} \le M$, so that C'_n is a random subset of a Hölder ball that has credibility $\Pi(C'_n|\mathbb{X}^{(n)}) = 1 - \alpha$. Theorem 1 implies the following result.

COROLLARY 1. *Consider observations generated from* (5) *under a fixed function* $f_0 \in C^\gamma$ *with* $\|f_0\|_{\gamma,\infty} < M$. *Let* Π *be the law of* $U_{\gamma,M}$, *let* $\Pi(\cdot|\mathbb{X}^{(n)})$ *be the posterior distribution given* $\mathbb{X}^{(n)}$ *and let* C'_n *be as in* (12). *Then*

$$P^n_{f_0}(f_0 \in C'_n) \to 1 - \alpha$$

as $n \to \infty$ *and the* L^2*-diameter* $|C'_n|_2$ *of* C'_n *satisfies, for some* $\kappa > 0$,

$$|C'_n|_2 = O_P(n^{-\gamma/(2\gamma+1)}(\log n)^\kappa).$$

We consider next the situation of a general series prior Π modelling γ-regular functions, including the important case of Gaussian priors. Let

$$G_\gamma = \sum_{l=J_0-1}^{\infty} \sum_{k=0}^{2^l-1} 2^{-l(\gamma+1/2)} g_{lk} \psi_{lk}(\cdot), \qquad \gamma > 0,$$

where g_{lk} are i.i.d. random variables that possess a bounded positive density φ such that $\mathrm{Var}(g_{lk}) < \infty$, and with S-regular CDV-wavelets ψ_{lk}, $S > \max(\gamma, 1/2)$. Denote by $\Pi_n = \Pi(\cdot|\mathbb{X}^{(n)})$ the posterior distribution from observing $\mathbb{X}^{(n)} \sim P_{f_0}^n$. The idea behind the construction of C'_n can be adapted to this more general situation by taking for $M_n \to \infty$, $M_n = O(\log n)$,

$$\tilde{C}'_n = \{f : \|f\|_{\gamma,2} \le M_n, \|f - \bar{f}_n\|_H \le R_n/\sqrt{n}\}.$$

This parallels the frequentist practice of "undersmoothing," taking into account the fact that we usually do not know a bound for $\|f\|_{\gamma,2}$ in the construction of confidence sets. While this can be shown to work (as in the proof of Corollary 1, assuming $f_0 \in C^\gamma \cap H_2^\gamma$), we wish to avoid such ad hoc methods here and prefer to explicitly use posterior information on the size of $\|f\|_{\gamma,2,1}$: fix $\delta > 0$ arbitrarily, and set

$$C''_n = \{f : \|f\|_{\gamma,2,1} \le M_n + 4\delta, \|\bar{f}_n - f\|_H \le R_n/\sqrt{n}\}, \tag{13}$$

where R_n is as in (10) with $T_n = \bar{f}_n$, and where M_n is defined as follows: for any n and $\delta_n = (\log n)^{-1/4}$,

$$M_n = \inf\{M > 0 : \Pi_n(f : |\|f\|_{\gamma,2,1} - M| \le \delta) \ge 1 - \delta_n\} \tag{14}$$

with the convention that $M_n = \infty$ if the set over which one takes the infimum in (14) is empty.

COROLLARY 2. *Consider observations generated from equation* (5) *under a fixed function $f_0 \in C^\gamma$. Let Π be the law of G_γ, let $\Pi_n = \Pi(\cdot|\mathbb{X}^{(n)})$ be the posterior distribution given $\mathbb{X}^{(n)}$, and let C''_n be as in* (13). *Then*

$$P_{f_0}^n(f_0 \in C''_n) \to 1 - \alpha, \qquad \Pi_n(C''_n) = 1 - \alpha + o_P(1)$$

as $n \to \infty$, and the L^2-diameter $|C''_n|_2$ of C''_n satisfies, for some $\kappa > 0$,

$$|C''_n|_2 = O_P(n^{-\gamma/(2\gamma+1)}(\log n)^\kappa).$$

Additionally, both M_n and R_n occuring in (13) *are bounded in probability.*

These credible sets can be compared to those in [25] constructed in the Gaussian conjugate situation [i.e., for g_{lk} i.i.d. $N(0, 1)$]. Both constructions give rise to credible sets that have frequentist minimax optimal diameter (within log-factors). In contrast to C''_n, however, the credible sets in [25] are conservative in the sense that their asymptotic frequentist coverage probability may exceed the desired level $1-\alpha$.

The purpose of M_n in (14) is to provide a bound on the unknown $\|f\|_{\gamma,2,1}$ using the posterior distribution, similar in spirit to a posterior quantile. Using (14) and Theorem 6 below (with $\sigma_l = 2^{-l(\gamma+1/2)}$), one shows that $\Pi_n(f : \|f - f_0\|_{\gamma,2,1} > (\log n)^{-1/4}) = o_P(1)$ and then also that

$$\|f_0\|_{\gamma,2,1} - 2\delta + o_P(1) \le M_n \le \|f_0\|_{\gamma,2,1} + 2\delta + o_P(1). \tag{15}$$

It is also possible to take $\delta = \delta_n$ in (14). Corollary 2 then still holds, and δ is replaced by δ_n in the previous display, in which case M_n is a consistent estimator of $\|f\|_{\gamma,2,1}$.

2.2. *Credible bands for self-convolutions.* We proceed with a semiparametric example: suppose we are interested in estimating the function

$$f * f = \int_0^1 f(\cdot - t) f(t)\, dt,$$

where addition is *mod*-1 (so the convolution of f with itself on the unit circle). The related problem in density estimation was studied in the papers [15, 20, 30, 31, 33], where it is shown that $f * f$ can be estimated at the $1/\sqrt{n}$-rate even when this is impossible for f. See particularly [15] for applications. Assume f is one-periodic and contained in H_2^s for some $s > 1/2$, and that the posterior is supported in $L^2([0,1)) \equiv L^2_{per}([0,1))$ which, in this subsection, denotes the subspace of L^2 consisting of one-periodic functions. We will assume that the basis used to define H_2^s is such that $(\sum_m |\hat{f}(m)|^2 (1+|m|)^{2s})^{1/2}$ is an equivalent norm on H_2^s (which is the case for CDV- or periodised wavelets and trigonometric bases of L^2).

By standard properties of convolutions $\kappa : f \mapsto f * f$ maps $L^2([0,1))$ into $C([0,1))$, the space of bounded continuous periodic functions on $[0,1)$ equipped with the uniform norm $\|\cdot\|_\infty$. If $\Pi_n = \Pi(\cdot|\mathbb{X}^{(n)})$ with posterior mean $\bar{f}_n \in L^2([0,1))$, we can construct a confidence band for $f * f$ by solving for R_n such that

$$\Pi_n \circ \kappa^{-1}\bigl(g : \|g - \bar{f}_n * \bar{f}_n\|_\infty \le R_n/\sqrt{n}\bigr) = 1 - \alpha \tag{16}$$

with resulting credible band

$$C_n = \bigl\{g : \|g - \bar{f}_n * \bar{f}_n\|_\infty \le R_n/\sqrt{n}\bigr\}. \tag{17}$$

THEOREM 2. *Suppose the weak Bernstein–von Mises phenomenon in the sense of Definition 1 holds, and that $f_0 \in H_2^s$ for some $s > 1/2$. Assume*

$$\|\bar{f}_n - \mathbb{X}^{(n)}\|_H = o_P(n^{-1/2}),$$

and that for some sequence $r_n = o(n^{-1/2})$,

$$\|\bar{f}_n - f_0\|_2^2 = O_P(r_n), \qquad \Pi_n\bigl(f : \|f - f_0\|_2^2 > r_n\bigr) = o_P(1).$$

Let C_n be the credible band from (17) *with R_n as in* (16). *Then, as $n \to \infty$,*

$$P^n_{f_0}(f_0 * f_0 \in C_n) \to 1 - \alpha \quad \textit{and} \quad R_n = O_P(1).$$

If $f_0 \in C^s \cap H^s_2$ for some $s > 1/2$, the priors from Condition 1 below with σ_l and $\gamma = s$ chosen as in Remark 1 are admissible in Theorem 2 with $r_n = n^{-2s/(2s+1)}$; cf. Corollaries 3, 4 below and Section 3.4.

2.3. *Credible sets for functionals.*

2.3.1. *Linear functionals.* Let L be any linear form on L^2 given by

$$L(f) = \langle f, g_L \rangle = \int_0^1 f(t) g_L(t)\, dt, \qquad f \in L^2,$$

where $g_L \in H^s_2$, $s > 1/2$, and $g_L \neq 0$. If $\Pi_n = \Pi(\cdot|\mathbb{X}^{(n)})$ is the posterior, one may construct credible sets for $L(f_0)$ based on the induced law $\Pi^L_n = \Pi_n \circ L^{-1}$ in several ways: for example, one solves for $R_n = R(\mathbb{X}^{(n)}, L, \alpha)$ in

$$\Pi^L_n\big(z : \big|z - L\big(\mathbb{X}^{(n)}\big)\big| \le R_n/\sqrt{n}\big) = 1 - \alpha, \tag{18}$$

which gives rise to the credible set

$$C_n = \big\{z : \big|z - L\big(\mathbb{X}^{(n)}\big)\big| \le R_n/\sqrt{n}\big\} \tag{19}$$

for $L(f)$. An alternative way to build the credible set is discussed below in a more general setting.

THEOREM 3. *Suppose the weak Bernstein–von Mises phenomenon in the sense of Definition 1 holds. Let $L = \langle \cdot, g_L \rangle$ be a linear functional on L^2 where $0 \neq g_L \in H^s_2$, $s > 1/2$. Let $\beta_{\mathbb{R}}$ be the bounded-Lipschitz metric for weak convergence on $\mathbb{R}$, and define $\theta_t : x \mapsto \sqrt{n}(x - t)$ for $t, x \in \mathbb{R}$. Then*

$$\beta_{\mathbb{R}}\big(\Pi^L_n \circ \theta^{-1}_{L(\mathbb{X}^{(n)})}, N\big(0, \|g_L\|_2^2\big)\big) \to^{P^n_{f_0}} 0.$$

Moreover let C_n be the credible region from (19) *with R_n chosen as in* (18). *Then*

$$P^n_{f_0}\big(L(f_0) \in C_n\big) \to 1 - \alpha \quad \textit{and} \quad R_n = O_P(1)$$

as $n \to \infty$. If, in addition, (6) *holds, then the same result holds true if C_n is centered at $L(\bar{f}_n)$ where $\bar{f}_n$ is the posterior mean of $\Pi(\cdot|\mathbb{X}^{(n)})$.*

The induced posterior $\Pi_n \circ L^{-1}$ has the approximate shape of a normal distribution centered at the efficient estimator $L(\mathbb{X}^{(n)})$ of $L(f)$ with variance $\|g_L\|_2^2/n$. This implies in particular that the width of the credible set C_n is asymptotically efficient from the semiparametric perspective; in fact $\|g_L\|_2^2$ is the semiparametric Cramér–Rao bound for estimating $L(f)$ from observations in the Gaussian white noise model (when maintaining standard nonparametric models for f).

The fact that any integral functional $\int f(t)g_L(t)\,dt$, $g_L \in H_2^s$, $s > 1/2$, is covered gives rise to a rich class of examples, such as the moment functionals $\int t^\alpha f(t)\,dt$, $\alpha \in \mathbb{N}$. The restriction to $s > 1/2$ is intrinsic to our methods and cannot be relaxed.

2.3.2. *Smooth nonlinear functionals.* We next consider statistical inference for nonlinear functionals of f_0 that satisfy a good quadratic approximation in L^2 at f_0, and more precisely, we assume that $\Psi : L^2 \to \mathbb{R}$ satisfies

$$\Psi(f_0 + h) - \Psi(f_0) = D\Psi_{f_0}[h] + O(\|h\|_2^2), \tag{20}$$

uniformly in $h \in L^2$ and for some $D\Psi_{f_0} : L^2 \to \mathbb{R}$ linear and continuous that has a (nonzero) L^2-Riesz representer $\dot{\Psi}_{f_0} \in H_2^s$ for some $s > 1/2$. This setting includes several standard examples discussed in more detail at the end of this section, but also the linear functionals discussed above.

Note that now Ψ cannot necessarily be evaluated at $\mathbb{X}^{(n)}$ [think of $\Psi(f) = \|f\|_2^2$]. However, since the posterior is supported in L^2 with probability one, the following Bayesian credible set can be constructed for $\Psi(f)$: for $\Pi_n = \Pi(\cdot|\mathbb{X}^{(n)})$ the posterior distribution, set $\Pi_n^\Psi = \Pi_n \circ \Psi^{-1}$, and solve for the $\alpha/2$ and $1 - \alpha/2$ quantiles μ_n, ν_n of Π_n^Ψ,

$$\Pi_n^\Psi\big((-\infty, \mu_n]\big) = \Pi_n^\Psi\big((\nu_n, +\infty)\big) = \frac{\alpha}{2}. \tag{21}$$

THEOREM 4. *Suppose the weak Bernstein–von Mises phenomenon in the sense of Definition 1 holds. Consider a functional Ψ satisfying (20). Assume moreover either that Ψ is linear, or that for some sequence $r_n = o(n^{-1/2})$,*

$$\Pi_n\big(f : \|f - f_0\|_2^2 > r_n\big) = o_P(1).$$

Let μ_n, ν_n satisfy (21). Then as $n \to \infty$,

$$P_{f_0}^n\big(\Psi(f_0) \in (\mu_n, \nu_n]\big) \to 1 - \alpha.$$

Similar to Theorem 3, the shape of the induced posterior $\Pi_n \circ \Psi^{-1}$ is approximately Gaussian, this time centered at $\Psi(f_0) + \langle \dot{\Psi}_{f_0}/\sqrt{n}, \mathbb{W}\rangle$, and with variance $\|\dot{\Psi}_{f_0}\|_2^2/n$. More precisely, for $\beta_{\mathbb{R}}$ the bounded-Lipschitz metric for weak convergence,

$$\beta_{\mathbb{R}}\big(\Pi_n^\Psi \circ \theta^{-1}_{\Psi(f_0)+\langle\dot{\Psi}_{f_0},\mathbb{W}\rangle/\sqrt{n}}, N\big(0, \|\dot{\Psi}_{f_0}\|_2^2\big)\big) \to^{P_{f_0}^n} 0.$$

In fact the proofs imply that the random quantile μ_n admits the expansion, for Φ_* the distribution function of a $N(0, \|\dot{\Psi}_{f_0}\|^2)$ variable,

$$\mu_n = \Psi(f_0) + \frac{1}{\sqrt{n}}\langle\dot{\Psi}_{f_0}, \mathbb{W}\rangle + \frac{\Phi_*^{-1}(\alpha/2)}{\sqrt{n}} + o_P(1/\sqrt{n})$$

and ν_n likewise, with $\Phi_*^{-1}(1-\frac{\alpha}{2})$ replacing $\Phi_*^{-1}(\frac{\alpha}{2})$. Again, $\|\dot\Psi_f\|_2^2$ is the semiparametric efficiency bound for estimating $\Psi(f)$ in the Gaussian white noise model, which shows that the asymptotic width of the credible set $(\mu_n, \nu_n]$ for $\Psi(f)$ is optimal in the semiparametric sense.

If $f_0 \in C^\gamma$ for some $\gamma > 1/2$, then the priors from Condition 1 below with σ_l chosen as in Remark 1 are admissible in the above theorem with $r_n = n^{-2\gamma/(2\gamma+1)}$; cf. Corollaries 3, 4 and Section 3.4 below.

Examples include the standard quadratic functionals such as $\Psi(f) = \int f^2(t)\,dt$ or composite functionals of the form $\Psi(f) = \int \phi(f(x), x)\,dx$. Some functionals may necessitate some straightforward modifications of our proofs: for instance, $\|f\|_p^p$ requires differentiation on L^p instead of L^2, and for the entropy functional $\int f(t)\log f(t)\,dt$ one assumes $f_0 \geq \zeta > 0$ on $[0, 1]$ and differentiates Ψ on L^∞. In these situations, to control remainder terms, one may use contraction results in $L^p, 2 < p \leq \infty$, instead of L^2, such as the ones in [23]. Our assumption $\gamma > 1/2$ is stronger than the critical assumption $\gamma \geq 1/4$ needed for $1/\sqrt{n}$-estimability of some of these functionals [27], a phenomenon intrinsic to general plug-in procedures.

3. Bernstein–von Mises theorems in white noise. We now develop general tools that allow us to prove that priors satisfy the Bernstein–von Mises phenomenon in the sense of Definition 1, and show how they can be successfully applied to a wide variety of natural classes of product priors.

For $f \in L^2$ consider again observing a random trajectory in the white noise model (4) of law P_f^n, with corresponding expectation operator denoted by E_f^n. Given an orthonormal basis from Definition 2, the white noise model is equivalent to observing the action of $\mathbb{X}^{(n)}$ on the basis, that is,

$$\mathbb{X}_{lk}^{(n)} = \theta_{lk} + \frac{1}{\sqrt{n}}\varepsilon_{lk}, \qquad k \in \mathcal{Z}_l, l \in \mathcal{L},$$

where $\theta_{lk} = \langle f, \psi_{lk}\rangle$, $\varepsilon_{lk} \sim^{\text{i.i.d.}} N(0, 1)$. Let Π be a prior Borel probability distribution on L^2 which induces a prior, also denoted by Π, on infinite sequences $\{\theta_{lk}\} \in l^2$. Let $\Pi(\cdot|\mathbb{X}^{(n)})$ be the posterior distribution, and let $\Pi(\theta_{lk}|\mathbb{X}^{(n)})$ denote the marginal posterior on the coordinate θ_{lk}.

3.1. *Contraction results in $H(\delta)$.* In this subsection we consider priors of the form $\Pi = \bigotimes_{lk} \pi_{lk}$ defined on the coordinates of the orthonormal basis $\{\psi_{lk}\}$, where π_{lk} are probability distributions with Lebesgue density φ_{lk} on the real line. Further assume, for some fixed density φ on the real line,

$$\varphi_{lk}(\cdot) = \frac{1}{\sigma_l}\varphi\left(\frac{\cdot}{\sigma_l}\right) \qquad \forall k \in \mathcal{Z}_l, \text{ with } \sigma_l > 0, \sum_{l,k}\sigma_l^2 < \infty.$$

CONDITION 1. (P1) Suppose that for a finite constant $M > 0$,

$$\sup_{l \in \mathcal{L}, k \in \mathcal{Z}_l} \frac{|\theta_{0,lk}|}{\sigma_l} \le M.$$

(P2) Suppose that φ is such that for some $\tau > M$ and $0 < c_\varphi \le C_\varphi < \infty$

$$\varphi(x) \le C_\varphi \qquad \forall x \in \mathbb{R}, \qquad \varphi(x) \ge c_\varphi \qquad \forall x \in (-\tau, \tau),$$
$$\int_{\mathbb{R}} x^2 \varphi(x)\, dx < \infty.$$

Some discussion of this condition is in order: we allow for a rich variety of base priors φ, such as Gaussian, sub-Gaussian, Laplace, most Student laws, or more generally any law with positive continuous density and finite second moment, but also uniform priors with large enough support. The full prior on f considered here is thus a sum of independent terms over the basis $\{\psi_{lk}\}$, including many, especially non-Gaussian, processes. For Gaussian processes Condition 1 applies simply by verifying that the L^2-basis provided by the Karhunen–Loève expansion of the process satisfies the conditions of Definition 2. This includes in particular Brownian motion: the corresponding φ is then the standard Gaussian density, and $\sigma_l = 1/(\pi(l + \frac{1}{2}))$ are the square-roots of the eigenvalues of the covariance operator. Through condition (P1), this allows for signals $f_0 \equiv (\theta_{0,lk})$ whose coefficients on the basis decrease at least as fast as $1/l$. For primitives of Brownian motion similar remarks apply, with stronger but natural decay restrictions on $\langle f_0, \psi_{lk}\rangle$.

In principle, making the prior rougher allows for more signals through condition (P1), but this may harm the performance of the posterior in stronger loss functions than the one considered in the next theorem. Its proof basically consists of showing that, under $P^n_{f_0}$, the coordinate-wise marginal posterior distributions contract about each "true" coordinate $\langle f_0, \psi_{lk}\rangle$ at rate $1/\sqrt{n}$ with constants independent of k, l.

THEOREM 5. *Consider data generated from equation* (4) *under a fixed function $f_0 \in L^2$ with coefficients $\theta_0 = \{\theta_{0,lk}\} = \{\langle f_0, \psi_{lk}\rangle\}$. Then if the product prior Π and f_0 satisfy Condition* 1, *we have for every $\delta > 1/2$, as $n \to \infty$,*

$$E^n_{f_0} \int \|f - f_0\|^2_{H(\delta)}\, d\Pi\big(f|\mathbb{X}^{(n)}\big) = O\bigg(\frac{1}{n}\bigg).$$

PROOF. We decompose the index set $\mathcal{L}$ into $\mathcal{J}_n := \{l \in \mathcal{L}, \sqrt{n}\sigma_l \ge S_0\}$ and its complement, where S_0 is a fixed positive constant. The quantity we wish to bound equals, by definition of the H-norm and Fubini's theorem,

$$\sum_{l,k} a_l^{-1} (\log a_l)^{-2\delta} E^n_{f_0} \int (\theta_{lk} - \theta_{0,lk})^2\, d\Pi\big(\theta_{lk}|\mathbb{X}^{(n)}\big).$$

Define further $B_{lk}(\mathbb{X}^{(n)}) := \int (\theta_{lk} - \theta_{0,lk})^2 \, d\Pi(\theta_{lk}|\mathbb{X}^{(n)})$ whose $P_{f_0}^n$-expectation we now bound. We write $\mathbb{X} = \mathbb{X}^{(n)}$ and $E = E_{f_0}^n$ throughout the proof to ease notation.

Using the independence structure of the prior we have $\Pi(\theta_{lk}|\mathbb{X}) = \pi_{lk}(\theta_{lk}|\mathbb{X}_{lk})$, and under $P_{f_0}^n$,

$$\begin{aligned} B_{lk}(\mathbb{X}) &= \frac{\int (\theta_{lk} - \theta_{0,lk})^2 e^{-n(\theta_{lk}-\theta_{0,lk})^2/2 + \sqrt{n}\varepsilon_{lk}(\theta_{lk}-\theta_{0,lk})} \varphi_{lk}(\theta_{lk}) \, d\theta_{lk}}{\int e^{-n(\theta_{lk}-\theta_{0,lk})^2/2 + \sqrt{n}\varepsilon_{lk}(\theta_{lk}-\theta_{0,lk})} \varphi_{lk}(\theta_{lk}) \, d\theta_{lk}} \\ &= \frac{1}{n} \frac{\int v^2 e^{-v^2/2+\varepsilon_{lk}v} \varphi((\theta_{0,lk} + n^{-1/2}v)/\sigma_l)/(\sqrt{n}\sigma_l) \, dv}{\int e^{-v^2/2+\varepsilon_{lk}v} \varphi((\theta_{0,lk} + n^{-1/2}v)/\sigma_l)/(\sqrt{n}\sigma_l) \, dv} =: \frac{1}{n} \frac{N_{lk}}{D_{lk}}(\varepsilon_{lk}). \end{aligned}$$

About the indices $l \in \mathcal{J}_n^c$: Taking a smaller integrating set on the denominator makes the integral smaller

$$D_{kl}(\varepsilon_{lk}) \ge \int_{-\sqrt{n}\sigma_l}^{\sqrt{n}\sigma_l} e^{-v^2/2+\varepsilon_{lk}v} \frac{1}{\sqrt{n}\sigma_l} \varphi\left(\frac{\theta_{0,lk} + n^{-1/2}v}{\sigma_l}\right) dv.$$

To simplify the notation we suppose that $\tau > M + 1$. If this is not the case, one multiplies the bounds of the integral in the last display by a small enough constant. The argument of the function φ in the previous display stays in $[-M+1, M+1]$ under (P1). Under assumption (P2) this implies that the value of φ in the last expression is bounded from below by c_φ. Next applying Jensen's inequality with the logarithm function

$$\begin{aligned} \log D_{kl}(\varepsilon_{lk}) &\ge \log(2c_\varphi) - \int_{-\sqrt{n}\sigma_l}^{\sqrt{n}\sigma_l} \frac{v^2}{2} \frac{dv}{2\sqrt{n}\sigma_l} + \varepsilon_{lk} \int_{-\sqrt{n}\sigma_l}^{\sqrt{n}\sigma_l} v \frac{dv}{2\sqrt{n}\sigma_l} \\ &= \log(2c_\varphi) - (\sqrt{n}\sigma_l)^2/6. \end{aligned}$$

Thus, $D_{kl}(\varepsilon_{lk}) \ge 2c_\varphi e^{-(\sqrt{n}\sigma_l)^2/6}$, which is bounded away from zero for indices in $\mathcal{J}_n^c$. Now about the numerator, let us split the integral defining N_{kl} into two parts $\{v : |v| \le \sqrt{n}\sigma_l\}$ and $\{v : |v| > \sqrt{n}\sigma_l\}$. That is, $N_{kl}(\varepsilon_{lk}) = (I) + (II)$. Taking the expectation of the first term and using Fubini's theorem,

$$E(I) = \int_{-\sqrt{n}\sigma_l}^{\sqrt{n}\sigma_l} v^2 e^{-v^2/2} E\big[e^{\varepsilon_{lk}v}\big] \frac{1}{\sqrt{n}\sigma_l} \varphi\left(\frac{\theta_{0,lk} + n^{-1/2}v}{\sigma_l}\right) dv \le 2n\sigma_l^2 C_\varphi/3.$$

The expectation of the second term is bounded by first applying Fubini's theorem as before and then changing variables back,

$$\begin{aligned} E(II) &= \int_{|v|>\sqrt{n}\sigma_l} v^2 e^{-v^2/2} E\big[e^{\varepsilon_{lk}v}\big] \frac{1}{\sqrt{n}\sigma_l} \varphi\left(\frac{\theta_{0,lk} + n^{-1/2}v}{\sigma_l}\right) dv \\ &= \int_{\theta_{0,lk}/\sigma_l+1}^{+\infty} \left(\sqrt{n}\sigma_l u - \sqrt{n}\sigma_l \frac{\theta_{0,lk}}{\sigma_l}\right)^2 \varphi(u) \, du \end{aligned}$$

$$+\int_{-\infty}^{\theta_{0,lk}/\sigma_l-1}\left(\sqrt{n}\sigma_l u-\sqrt{n}\sigma_l\frac{\theta_{0,lk}}{\sigma_l}\right)^2\varphi(u)\,du$$
$$\le 2n\sigma_l^2\left[\frac{\theta_{0,lk}^2}{\sigma_l^2}+\int_{-\infty}^{+\infty}u^2\varphi(u)\,du\right].$$

Thus, using (P1) again, $E(I)+E(II)$ is bounded on $\mathcal{J}_n^c$ by a fixed constant times $n\sigma_l^2$. In particular, there exists a fixed constant independent of n, k, l such that $E(nB_{lk}(X))$ is bounded from above by a constant on $\mathcal{J}_n^c$.

Now about the indices in $\mathcal{J}_n$. For such l, k, using (P1)–(P2), one can find $L_0>0$ depending only on S_0, M, τ such that, for any v in $(-L_0, L_0)$, $\varphi((\theta_{0,lk}+n^{-1/2}v)/\sigma_l)\ge c_\varphi$. Thus the denominator $D_{lk}(\varepsilon_{lk})$ can be bounded from below by

$$D_{lk}(\varepsilon_{lk})\ge c_\varphi\int_{-L_0}^{L_0}e^{-v^2/2+\varepsilon_{lk}v}\frac{1}{\sqrt{n}\sigma_l}\,dv.$$

On the other hand, the numerator can be bounded above by

$$N_{lk}(\varepsilon_{lk})\le C_\varphi\int v^2e^{-v^2/2+\varepsilon_{lk}v}\frac{1}{\sqrt{n}\sigma_l}\,dv.$$

Putting these two bounds together leads to

$$B_{lk}(\varepsilon_{lk})\le\frac{1}{n}\frac{C_\varphi}{c_\varphi}\frac{\int v^2e^{-v^2/2+\varepsilon_{lk}v}\,dv}{\int_{-L_0}^{L_0}e^{-v^2/2+\varepsilon_{lk}v}\,dv}.$$

The last quantity has a distribution independent of l, k. Let us thus show that

$$Q(L_0)=E\left[\frac{\int v^2e^{-(v-\varepsilon)^2/2}\,dv}{\int_{-L_0}^{L_0}e^{-(v-\varepsilon)^2/2}\,dv}\right]$$

is finite for every $L_0>0$, where $\varepsilon\sim N(0,1)$. In the numerator we substitute $u=v-\varepsilon$. Using the inequality $(u+\varepsilon_{lk})^2\le 2v^2+2\varepsilon_{lk}^2$, the second moment of a standard normal variable appears, and this leads to the bound

$$Q(L_0)\le CE\left[\frac{1+\varepsilon^2}{\int_{-L_0}^{L_0}e^{-(v-\varepsilon)^2/2}\,dv}\right]$$

for some finite constant $C>0$. Denote by g the density of a standard normal variable, by Φ its distribution function and $\bar\Phi=1-\Phi$. It is enough to prove that the following quantity is finite:

$$q(L_0):=\int_{-\infty}^{+\infty}\frac{(1+u^2)g(u)}{\bar\Phi(u-L_0)-\bar\Phi(u+L_0)}\,du$$
$$=2\int_0^{+\infty}\frac{(1+u^2)g(u)}{\bar\Phi(u-L_0)-\bar\Phi(u+L_0)}\,du,$$

since the integrand is an even function. Using the standard inequalities

$$\frac{1}{\sqrt{2\pi}}\frac{u^2}{1+u^2}\frac{1}{u}e^{-u^2/2} \le \bar{\Phi}(u) \le \frac{1}{\sqrt{2\pi}}\frac{1}{u}e^{-u^2/2}, \qquad u \ge 1,$$

it follows that for any $\delta > 0$, one can find $M_\delta > 0$ such that, for any $u \ge M_\delta$,

$$(1-\delta)\frac{1}{u}e^{-u^2/2} \le \sqrt{2\pi}\,\bar{\Phi}(u) \le \frac{1}{u}e^{-u^2/2}, \qquad u \ge M_\delta.$$

Set $A_\delta = 2L_0 \vee M_\delta$. Then for $\delta < 1 - e^{-2L_0}$ we deduce

$$\begin{aligned} q(L_0) &\le 2\int_0^{A_\delta} \frac{(1+u^2)g(u)}{\bar{\Phi}(A_\delta - L_0) - \bar{\Phi}(A_\delta + L_0)}\,du \\ &\quad + 2\sqrt{2\pi}\int_{A_\delta}^{+\infty} (u-L_0)\big(1+u^2\big)\frac{e^{(u-L_0)^2/2}g(u)}{1-\delta-e^{-2L_0}}\,du \\ &\le C(A_\delta, L_0) + \frac{2e^{-L_0^2/2}}{1-\delta-e^{-2L_0}}\int_{A_\delta}^{+\infty} u\big(1+u^2\big)e^{-L_0 u}\,du < \infty. \end{aligned}$$

Conclude that $\sup_{l,k} E_{f_0}^n|B_{lk}(\mathbb{X})| = O(1/n)$. Since $\sum_{l,k} a_l^{-1}(\log a_l)^{-2\delta} < \infty$ the result follows. $\square$

For the following theorem note that $\gamma = \delta = 0$ gives $\|\cdot\|_{0,2,0} = \|\cdot\|_2$.

THEOREM 6. *With the notation of Theorem 5, suppose the product prior* Π *and* f_0 *satisfy Condition 1. Then for any real numbers* γ, δ,

$$E_{f_0}^n \int \|f - f_0\|_{\gamma,2,\delta}^2\, d\Pi\big(f|\mathbb{X}^{(n)}\big) = O\bigg(\sum_{l,k} a_l^{2\gamma}(\log a_l)^{-2\delta}\big(\sigma_l^2 \wedge n^{-1}\big)\bigg).$$

PROOF. We only prove $\gamma = \delta = 0$; the general case is the same. With the notation used in the proof of Theorem 5, using Fubini's theorem,

$$\begin{aligned} E_{f_0}^n \int \|f - f_0\|_2^2\, d\Pi\big(f|\mathbb{X}^{(n)}\big) &= \sum_{l,k} E_{f_0}^n \int (\theta_{lk} - \theta_{0,lk})^2\, d\Pi(\theta_{lk}|\mathbb{X}) \\ &= \sum_{l,k} E_{f_0}^n B_{lk}(\mathbb{X}). \end{aligned}$$

In the proof of Theorem 5, the following two bounds have been obtained, with the notation $\mathcal{J}_n := \{l \in \mathcal{L}, \sqrt{n}\sigma_l \ge S_0\}$,

$$\sup_{l \in \mathcal{J}_n, k} E_{f_0}^n B_{lk}(\mathbb{X}) = O\big(n^{-1}\big), \qquad \sup_{l \notin \mathcal{J}_n, k} E_{f_0}^n \sigma_l^{-2} B_{lk}(\mathbb{X}) = O(1).$$

For any $l \in \mathcal{J}_n^c$, by definition of $\mathcal{J}_n$ it holds $\sigma_l^2 < S_0^2 n^{-1}$, thus $\sigma_l^2 \le (1 \vee S_0^2)(\sigma_l^2 \wedge n^{-1})$. Similarly, if $l \in \mathcal{J}_n$, we have $n^{-1} \le (1 \vee S_0^{-2})(\sigma_l^2 \wedge n^{-1})$. $\square$

COROLLARY 3. *Set* $\sigma_l = |l|^{-1/2-\gamma}$ *or* $\sigma_l = 2^{-(1/2+\gamma)l}$ *depending on the chosen S-regular basis of type either* (a) *or* (b). *Suppose that the conditions of Theorem* 6 *are satisfied. Then*

$$E^n_{f_0} \int \|f - f_0\|_2^2 \, d\Pi\big(f|\mathbb{X}^{(n)}\big) = O\big(n^{-2\gamma/(2\gamma+1)}\big).$$

PROOF. For both types of basis $\sum_l |\mathcal{Z}_l|(\sigma_l^2 \wedge n^{-1}) = O(n^{-2\gamma/(2\gamma+1)})$. □

REMARK 1. The previous choice of σ_l entails a regularity condition on f_0 through condition (P1), namely $\sup_k |\theta_{0,lk}| \leq M\sigma_l$. If $\sigma_l = 2^{-(1/2+\gamma)l}$ this amounts to the standard Hölderian condition if one uses a CDV wavelet basis, or a periodised wavelet basis—any f_0 in C^γ from (9) satisfies (P1) for such bases. For other bases similar remarks apply.

COROLLARY 4. *Denote by* $\bar{f}_n := \bar{f}_n(\mathbb{X}^{(n)}) := \int f \, d\Pi(f|\mathbb{X}^{(n)})$ *the posterior mean associated to the posterior distribution. Under the conditions of Theorem* 6,

$$E^n_{f_0} \|\bar{f}_n - f_0\|_2^2 = O\bigg(\sum_{l,k} (\sigma_l^2 \wedge n^{-1})\bigg).$$

PROOF. Apply the Cauchy–Schwarz inequality and Theorem 6. □

3.2. *Convergence of the finite-dimensional distributions*. Consider again the posterior distribution $\Pi_n \equiv \Pi(\cdot|\mathbb{X}^{(n)})$ on L^2 from the beginning of this section (not necessarily arising from a product measure). Let V be any of the finite-dimensional projection subspaces of L^2 defined in Section 1.2, equipped with the L^2-norm, and recall that π_V denotes the orthogonal L^2-projection onto V. For $z \in H(\delta)$, define the transformation

$$T_z \equiv T_{z,V} : f \mapsto \sqrt{n}\pi_V(f - z)$$

from $H(\delta)$ to V, and consider the image measure $\Pi_n \circ T_z^{-1}$. The finite-dimensional space V carries a natural Lebesgue product measure on it.

CONDITION 2. Suppose that $\Pi \circ \pi_V^{-1}$ has a Lebesgue-density $d\Pi_V$ in a neighborhood of $\pi_V(f_0)$ that is continuous and positive at $\pi_V(f_0)$. Suppose also that for every $\delta > 0$ there exists a fixed L^2-norm ball $C = C_\delta$ in V such that, for n large enough, $E^n_{f_0}(\Pi_n \circ T_{f_0}^{-1})(C^c) < \delta$.

This condition requires that the projected prior has a continuous density at $\pi_V(f_0)$ and that the image of the posterior distribution under the finite-dimensional projection onto V concentrates on a $1/\sqrt{n}$-neighborhood of $\pi_V(f_0)$. Let $\|\cdot\|_{\mathrm{TV}}$ denote the total variation norm on the space of finite signed measures on V, and $N(0, I)$ a standard Gaussian measure on V.

THEOREM 7. *Consider data generated from equation* (4) *under a fixed function* f_0, *denote by* $P^n_{f_0}$ *the distribution of* $\mathbb{X}^{(n)}$. *Assume Condition* 2. *Then we have, as* $n \to \infty$,

$$\|\Pi_n \circ T^{-1}_{\mathbb{X}^{(n)}} - N(0, I)\|_{\mathrm{TV}} \to^{P^n_{f_0}} 0.$$

The proof of Theorem 7 is similar to the parametric proof in Chapter 10 in [35], and is omitted. In the special case of product priors relevant for most examples in the present paper, one can also derive the result directly from Theorem 1 in [8]: by independence of the Gaussian coordinate experiments $\langle \psi_{lk}, \mathbb{X}^{(n)} \rangle \equiv \theta_{0,lk} + \frac{1}{\sqrt{n}} \varepsilon_{lk}$, when estimating one or more generally any finite number of the θ_{lk}'s, there is no loss of information with respect to the case where all other θ_{lk}'s would be known. Since the model is LAN with zero remainder, condition (N) in [8] is satisfied, and condition (C) in [8] amounts to asking that the full posterior concentrate at some rate $\varepsilon_n \to 0$ in the L^2-norm (which for product priors is implied by Corollary 3).

3.3. *A BvM-theorem in* $H(\delta)$. Let $\Pi_n = \Pi(\cdot|\mathbb{X}^{(n)})$ be the posterior distribution on L^2. Under the following Condition 3, which depends on a positive real δ' to be specified in the sequel, we will prove that a weak Bernstein–von Mises phenomenon holds true in $H(\delta)$ for any $\delta > 1/2$. For the product priors considered above we will then verify Condition 3 below.

CONDITION 3. Suppose for every $\varepsilon > 0$ there exists a constant $0 < M \equiv M(\varepsilon) < \infty$ independent of n such that, for any $n \geq 1$, some $\delta' > 1/2$,

$$E^n_{f_0} \Pi\left[\left\{ f : \|f - f_0\|^2_{H(\delta')} > \frac{M}{n} \right\} \Big| \mathbb{X}^{(n)} \right] \leq \varepsilon. \tag{22}$$

Assume moreover that the conclusion of Theorem 7 holds true for every V (i.e., the finite-dimensional distributions converge).

On $H(\delta)$ and for $z \in H(\delta)$, define the measurable map

$$\tau_z : f \mapsto \sqrt{n}(f - z).$$

Recalling the definitions from Section 1.2, consider $\Pi_n \circ \tau^{-1}_{\mathbb{X}^{(n)}}$, a Borel probability measure on $H(\delta)$. Let $\mathcal{N}$ be the Gaussian measure on $H(\delta)$ constructed in Section 1.2 above.

THEOREM 8. *Fix* $\delta > \delta' > 1/2$, *and assume Condition* 3 *for such* δ'. *If* β *is the bounded Lipschitz metric for weak convergence of probability measures on* $H(\delta)$, *then as* $n \to \infty$, $\beta(\Pi_n \circ \tau^{-1}_{\mathbb{X}^{(n)}}, \mathcal{N}) \to 0$ *in* $P^n_{f_0}$-*probability.*

PROOF. It is enough to show that for every $\varepsilon > 0$ there exists $N = N(\varepsilon)$ large enough such that for all $n \geq N$,

$$P_{f_0}^n\big(\beta\big(\Pi_n \circ \tau_{\mathbb{X}^{(n)}}^{-1}, \mathcal{N}\big) > 4\varepsilon\big) < 4\varepsilon.$$

Fix $\varepsilon > 0$, and let V_J be the finite-dimensional subspace of L^2 spanned by $\{\psi_{lk} : k \in \mathcal{Z}_l, l \in \mathcal{L}, |l| \leq J\}$, for any integer $J \geq 1$. Writing $\tilde{\Pi}_n$ for $\Pi_n \circ \tau_{\mathbb{X}^{(n)}}^{-1}$ we see from the triangle inequality

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

The middle term converges to zero in $P_{f_0}^n$-probability for every V_J, by convergence of the finite-dimensional distributions (Condition 3 and since the total variation distance dominates β). Next we handle the first term. Set $Q = M = M(\varepsilon^2/4)$, and consider the random subset D of $H(\delta')$ defined as

$$D = \big\{g : \|g + \mathbb{W}\|_{H(\delta')}^2 \leq Q\big\}.$$

Under $P_{f_0}^n$ we have $\tilde{\Pi}_n(D) = \Pi_n(D_n)$, where

$$D_n = \big\{f : \|f - f_0\|_{H(\delta')}^2 \leq Q/n\big\}$$

is the complement of the set appearing in (22). In particular, using Condition 3 and Markov's inequality yields $P_{f_0}^n(\tilde{\Pi}_n(D^c) > \varepsilon/4) \leq \varepsilon^2/\varepsilon = \varepsilon$.

If $Y_n \sim \tilde{\Pi}_n$ (conditional on $\mathbb{X}^{(n)}$), then $\pi_{V_J}(Y_n) \sim \tilde{\Pi}_n \circ \pi_{V_J}^{-1}$. For F any bounded function on $H(\delta)$ of Lipschitz-norm less than one,

$$\begin{aligned}
&\left|\int_{H(\delta)} F\, d\tilde{\Pi}_n - \int_{H(\delta)} F\, d\big(\tilde{\Pi}_n \circ \pi_{V_J}^{-1}\big)\right| \\
&\quad = \big|E_{\tilde{\Pi}_n}\big[F(Y_n) - F\big(\pi_{V_J}(Y_n)\big)\big]\big| \\
&\quad \leq E_{\tilde{\Pi}_n}\big[\big\|Y_n - \pi_{V_J}(Y_n)\big\|_{H(\delta)} 1_D(Y_n)\big] + 2\tilde{\Pi}_n\big(D^c\big),
\end{aligned}$$

where $E_{\tilde{\Pi}_n}$ denotes expectation under $\tilde{\Pi}_n$ (given $\mathbb{X}^{(n)}$). With $y_{lk} = \langle Y_n, \psi_{lk}\rangle$,

$$\begin{aligned}
&E_{\tilde{\Pi}_n}\big[\big\|Y_n - \pi_{V_J}(Y_n)\big\|_{H(\delta)}^2 1_D(Y_n)\big] \\
&\quad = E_{\tilde{\Pi}_n}\Big[\sum_{l>J} a_l^{-1}(\log a_l)^{-2\delta} \sum_k |y_{lk}|^2 1_D(Y_n)\Big] \\
&\quad = E_{\tilde{\Pi}_n}\Big[\sum_{l>J} a_l^{-1}(\log a_l)^{2\delta'-2\delta-2\delta'} \sum_k |y_{lk}|^2 1_D(Y_n)\Big] \\
&\quad \leq (\log a_J)^{2\delta'-2\delta} E_{\tilde{\Pi}_n}\big[\|Y_n\|_{H(\delta')}^2 1_D(Y_n)\big] \\
&\quad \leq 2(\log a_J)^{2\delta'-2\delta}\big[Q + \|\mathbb{W}\|_{H(\delta')}^2\big].
\end{aligned}$$

From the definition of β one deduces

$$\beta(\tilde{\Pi}_n, \tilde{\Pi}_n \circ \pi_{V_J}^{-1}) \le 2\tilde{\Pi}_n(D^c) + \sqrt{2}(\log a_J)^{\delta'-\delta}\sqrt{Q + \|\mathbb{W}\|_{H(\delta')}^2}.$$

Since $a_J \to \infty$ as $J \to \infty$ we conclude that $P_{f_0}^n(\beta(\tilde{\Pi}_n, \tilde{\Pi}_n \circ \pi_{V_J}^{-1}) > \varepsilon) < 2\varepsilon$ for J large enough, combining the previous deviation bound for $\tilde{\Pi}_n(D^c)$ and that $\|\mathbb{W}\|_{H(\delta')}$ is bounded in probability; cf. after (7) above. A similar (though simpler) argument leads to $P_{f_0}^n(\beta(\mathcal{N} \circ \pi_{V_J}^{-1}, \mathcal{N}) > \varepsilon) < \varepsilon$, using again that any random variable with law $\mathcal{N}$ has square integrable Hilbert-norm on $H(\delta')$. This completes the proof. $\square$

3.4. *The BvM theorem for product priors.* Combining Theorems 5, 7 and 8 implies that for product priors the weak Bernstein–von Mises theorem in the sense of Definition 1 holds. The following results can be seen to be uniform ("honest") in all f_0 that satisfy Condition 1 with fixed constant M.

THEOREM 9. *Suppose the assumptions of Theorem 5 are satisfied and that φ is continuous near $\{\theta_{0,lk}\}$ for every $k \in \mathcal{Z}_l, l \in \mathcal{L}$. Let $\delta > 1/2$. Then for β the bounded Lipschitz metric for weak convergence of probability measures on $H(\delta)$ we have, as $n \to \infty$, $\beta(\Pi_n \circ \tau_{\mathbb{X}^{(n)}}^{-1}, \mathcal{N}) \to^{P_{f_0}^n} 0$.*

PROOF. We only need to verify Condition 3 with some $1/2 < \delta' < \delta$ so that we can apply Theorem 8. From Theorem 5 with any such δ' in place of δ, we see that

$$nE_{f_0}^n \int \|f - f_0\|_{H(\delta')}^2 \, d\Pi(f|\mathbb{X}^{(n)}) = O(1), \tag{23}$$

which verifies the first part of Condition 3 for some M large enough using Markov's inequality. The second part follows from verifying Condition 2 to invoke Theorem 7: let V be arbitrary. If V_J is defined as in the proof of Theorem 8, and if J is the smallest integer such that $V \subset V_J$, then

$$\|\pi_V(f - f_0)\|_2^2 \le \|\pi_{V_J}(f - f_0)\|_2^2 \le a_J \log(a_J)^{2\delta'} \|f - f_0\|_{H(\delta')}^2$$

so that the second part of Condition 2 follows from the estimate (23) and again Markov's inequality, for C a fixed norm ball in V of squared diameter of order $a_J \log(a_J)^{2\delta'} M^2$. The first part of Condition 2 follows from the fact that $\Pi \circ T_{f_0}^{-1}$ is a product measure in V with bounded marginals φ_{lk} constant in k, and from the continuity assumption on φ. $\square$

THEOREM 10. *Suppose the assumptions of Theorem 5 are satisfied and that φ is continuous near $\{\theta_{0,lk}\}$ for every $k \in \mathcal{Z}_l, l \in \mathcal{L}$. Let $\delta > 1/2$ be arbitrary, let Y_n be a random variable drawn from $\Pi_n \circ \tau_{\mathbb{X}^{(n)}}^{-1}$ (conditional on $\mathbb{X}^{(n)}$), and let $\bar{f}_n$ be the (Bochner-) mean of the posterior distribution $\Pi(\cdot|\mathbb{X}^{(n)})$. Then $E[Y_n|\mathbb{X}^{(n)}] = \sqrt{n}(\bar{f}_n - \mathbb{X}^{(n)}) \to^{P_{f_0}^n} 0$ in $H(\delta)$ as $n \to \infty$.*

PROOF. Note that

$$
\begin{aligned}
E\big[\|Y_n\|_{H(\delta)}^2|\mathbb{X}^{(n)}\big] &= \int \|h\|_{H(\delta)}^2\, d\Pi_n \circ \tau_{\mathbb{X}^{(n)}}^{-1}(h) \\
&\le 2n \int \|f-f_0\|_{H(\delta)}^2\, d\Pi\big(f|\mathbb{X}^{(n)}\big) + 2\|\mathbb{W}\|_{H(\delta)}^2 \\
&= O_{P_{f_0}^n}(1)
\end{aligned}
$$

by Theorem 5 and since $\|\mathbb{W}\|_{H(\delta)} < \infty$ almost surely, as after (7). Moreover $Y_n \to N$ weakly in $H(\delta)$ in $P_{f_0}^n$-probability, where $N \sim \mathcal{N}$, by Theorem 9. By a standard uniform integrability argument [using that $\{Y_n : n \in \mathbb{N}\}$ has $H(\delta)$-norms with uniformly bounded second moments and converges to N weakly], and arguing as in the last paragraph of Section 4.1 below, we conclude $E[Y_n|\mathbb{X}^{(n)}] \to EN$ in $H(\delta)$ in $P_{f_0}^n$-probability, which implies the result since $EN = 0$. □

4. Remaining proofs.

PROOF OF THEOREM 1. By Corollary 6.8.5 in [5] the image measure $\mathcal{N} \circ (\|\cdot\|_{H(\delta)})^{-1}$ of $\mathcal{N}$ under the norm mapping is absolutely continuous on $[0,\infty)$, so the mapping

$$\Phi : t \mapsto \mathcal{N}\big(B(0,t)\big) = \mathcal{N} \circ \big(\|\cdot\|_{H(\delta)}\big)^{-1}\big([0,t]\big)$$

is uniformly continuous and increasing on $[0,\infty)$. In fact, the mapping is strictly increasing on $[0,\infty)$: using the results on pages 213–214 in [37], it suffices to show that any shell $\{f : s < \|f\|_{H(\delta)} < t\}$, $s < t$, contains an element of the RKHS L^2 of $\mathcal{N}$, which is obvious as L^2 is dense in $H(\delta)$. Thus Φ has a continuous inverse $\Phi^{-1} : [0,1) \to [0,\infty)$. Since Φ is uniformly continuous for every $\epsilon > 0$, there exists $\delta > 0$ small enough such that $|\Phi(t+\delta) - \Phi(t)| < \epsilon$ for every $t \in [0,\infty)$. Now

$$\mathcal{N}\big(\partial_\delta B(0,t)\big) = \mathcal{N}\big(B(0,t+\delta)\big) - \mathcal{N}\big(B(0,t-\delta)\big) = \big|\Phi(t+\delta) - \Phi(t-\delta)\big| < 2\epsilon$$

for $\delta > 0$ small enough, independently of t. Using (26) below we deduce that the balls $\{B(0,t)\}_{0\le t<\infty}$ form a $\mathcal{N}$-uniformity class, and we can thus conclude from Definition 1 and the results in Section 4.1 below that

$$\sup_{0\le t<\infty} \big|\Pi\big(f : \|f - \mathbb{X}^{(n)}\|_{H(\delta)} \le t/\sqrt{n}|\mathbb{X}^{(n)}\big) - \mathcal{N}\big(B(0,t)\big)\big| \to 0$$

in $P_{f_0}^n$-probability, as $n \to \infty$. This combined with (10) gives

$$\mathcal{N}\big(B(0,R_n)\big) = \mathcal{N}\big(B(0,R_n)\big) - \Pi\big(f : \|f - \mathbb{X}^{(n)}\|_{H(\delta)} \le R_n/\sqrt{n}|\mathbb{X}^{(n)}\big) + 1 - \alpha,$$

which converges to $1-\alpha$ as $n \to \infty$ in $P_{f_0}^n$-probability, and thus, by the continuous mapping theorem,

$$R_n \to^{P_{f_0}^n} \Phi^{-1}(1-\alpha) \tag{24}$$

as $n \to \infty$. Now using this last convergence in probability,

$$\begin{aligned} P_{f_0}^n(f_0 \in C_n) &= P_{f_0}^n\big(f_0 \in B\big(\mathbb{X}^{(n)}, R_n/\sqrt{n}\big)\big) = P_{f_0}^n\big(0 \in B(\mathbb{W}, R_n)\big) \\ &= P_{f_0}^n\big(0 \in B\big(\mathbb{W}, \Phi^{-1}(1-\alpha)\big)\big) + o(1) \\ &= \mathcal{N}\big(B\big(0, \Phi^{-1}(1-\alpha)\big)\big) + o(1) \\ &= \Phi\big(\Phi^{-1}(1-\alpha)\big) + o(1) = 1 - \alpha + o(1), \end{aligned}$$

which completes the proof of the first claim. The second claim follows from the same arguments combined with $\|\bar{f}_n - \mathbb{X}^{(n)}\|_H = o_P(n^{-1/2})$ which implies

$$P_{f_0}^n\big(f_0 \in B(\bar{f}_n, R_n/\sqrt{n})\big) - P_{f_0}^n\big(f_0 \in B\big(\mathbb{X}^{(n)}, R_n/\sqrt{n}\big)\big) \to^{P_{f_0}^n} 0$$

as $n \to \infty$, regardless of whether R_n is defined via the centering $T_n = \mathbb{X}^{(n)}$ or $T_n = \bar{f}_n$; cf. (10). $\square$

PROOF OF COROLLARY 1. By Theorems 9 and 10 this prior satisfies the weak Bernstein–von Mises phenomenon in the sense of Definition 1, as well as (6). The proof of coverage of C_n' is thus the same as in Theorem 1 noting that by hypothesis on f_0 the probability $P_{f_0}^n(f_0 \in C_n')$ in question equals

$$P_{f_0}^n\big(\|f_0\|_{\gamma,\infty} \le M, \|f_0 - \bar{f}_n\|_{H(\delta)} \le R_n/\sqrt{n}\big) = P_{f_0}^n\big(f_0 \in B(\bar{f}_n, R_n/\sqrt{n})\big).$$

To control $|C_n'|_2$, pick two arbitrary functions f_1, f_2 in C_n', and let $g = f_1 - f_2$. Then by construction and (24),

$$\|g\|_{\gamma,2,1} \le c\|g\|_{\gamma,\infty} \le 2cM, \qquad \|g\|_{H(\delta)} = O_P\big(n^{-1/2}\big).$$

Choosing J_n such that $2^{J_n} \sim n^{1/(2\gamma+1)}$,

$$\begin{aligned} \|g\|_2^2 &= \sum_{l \ge J_0 - 1} \sum_{k=0}^{2^l - 1} \big|\langle g, \psi_{lk}\rangle\big|^2 \\ &= \sum_{l=J_0-1}^{J_n - 1} l^{2\delta} 2^l 2^{-l} l^{-2\delta} \sum_k \big|\langle g, \psi_{lk}\rangle\big|^2 + \sum_{l=J_n}^{\infty} 2^{-2l\gamma} l^2 2^{2l\gamma} l^{-2} \sum_k \big|\langle g, \psi_{lk}\rangle\big|^2 \\ &\le 2^{J_n} J_n^{2\delta} \|g\|_{H(\delta)}^2 + 2^{-2J_n\gamma} J_n^2 \|g\|_{\gamma,2,1}^2 \\ &= O_P\bigg(\frac{2^{J_n} J_n^{2\delta}}{n} + 2^{-2J_n\gamma} J_n^2\bigg) = O_P\big(n^{-2\gamma/(2\gamma+1)} (\log n)^{\kappa}\big) \end{aligned}$$

with constants independent of g, implying the same bound for $|C_n'|_2^2$. $\square$

PROOF OF COROLLARY 2. By Theorems 9 and 10 this prior satisfies the weak Bernstein–von Mises phenomenon in the sense of Definition 1, as well as (6).

By (15) we have $\|f_0\|_{\gamma,2,1} \le M_n + 2\delta + o_P(1)$ and so $P^n_{f_0}(f_0 \in C''_n) = P^n_{f_0}(f_0 \in B(\bar{f}_n, R_n/\sqrt{n})) + o(1)$. The proof of asymptotic $1-\alpha$-coverage of C''_n is thus the same as in Theorem 1. Likewise, (15) implies $\Pi_n(C''_n) = 1-\alpha+o_P(1)$. To control $|C''_n|_2$, pick two arbitrary functions f_1, f_2 in C''_n and let $g = f_1 - f_2$. Then by (15) we have $\|g\|_{\gamma,2,1} = O(M_n) = O_P(1)$ and by (24) also $\|g\|_{H(\delta)} = O_P(n^{-1/2})$. The rest of the proof is the same as in the previous corollary. $\square$

PROOF OF THEOREM 2. Since $f_0 \in L^1 \cap H^s_2$, we see by Fourier inversion on the circle, the Cauchy–Schwarz inequality, and our assumption on the equivalent Sobolev norm that

$$\begin{aligned}\|f * f_0\|_\infty &\le \sum_m |\hat{f}(m)|(1+|m|)^{-s}(1+|m|)^s|\hat{f}_0(m)| \\ &\le \Bigl(\sum_m |\hat{f}(m)|^2(1+|m|)^{-2s}\Bigr)^{1/2}\Bigl(\sum_m |\hat{f}_0(m)|^2(1+|m|)^{2s}\Bigr)^{1/2} \\ &\le C'\|f\|_{H(\delta)}\end{aligned}$$

for any $\delta > 0$, in particular $f * f_0$, for $f \in H(\delta)$, $f_0 \in H^s_2$, defines a continuous function on [0, 1) (by Fourier inversion), and the mapping $\lambda : f \mapsto 2f * f_0$ is linear and continuous from $H(\delta)$ to $C([0,1))$; this argument is adapted from Theorem 1 in [31]. By Definition 1 and the continuous mapping theorem we thus have $\beta((\Pi_n \circ \tau^{-1}_{\mathbb{X}^{(n)}}) \circ \lambda^{-1}, \mathcal{N} \circ \lambda^{-1}) \to^{P_{f_0}} 0$ as $n \to \infty$, where β is the bounded Lipschitz metric for weak convergence in $C([0,1))$. Moreover from Corollary 6.8.5 in [5] we deduce as in the proof of Theorem 1 that norm balls $\{f : \|f\|_\infty \le t\}_{0 \le t < \infty}$ are $\mathcal{N} \circ \lambda^{-1}$ uniformity classes for weak convergence, and that the mapping $\Phi_\lambda : t \mapsto \mathcal{N} \circ \lambda^{-1}(f : \|f\|_\infty \le t)$ from $[0,\infty)$ to $[0,1)$ is continuous and increasing. In fact, it is strictly increasing, using the results on pages 213–214 in [37] combined with the fact that the RKHS of $\mathbb{W} * f_0$, equal to $L^2 * f_0$, contains functions of arbitrary supremum norm. Denote by Φ^{-1}_λ the continuous inverse of Φ_λ. As in the previous proofs, as $n \to \infty$

$$\mathcal{N} \circ \lambda^{-1}(f : \|f\|_\infty \le R_n) - (\Pi_n \circ \lambda^{-1}) \circ \theta^{-1}_{\mathbb{X}^{(n)} * f_0}(f : \|f\|_\infty \le R_n) \to^{P^n_{f_0}} 0,$$

where $\theta_{\mathbb{X}^{(n)} * f_0} : g \mapsto \sqrt{n}(g - \mathbb{X}^{(n)} * f_0)$ maps $C([0,1)) \to C([0,1))$.

Thus, using the hypotheses on $\bar{f}_n$ and the posterior contraction rate, the decomposition $f * f - g * g = 2(f-g) * g + (f-g) * (f-g)$ and the convolution inequality $\|h * h'\|_\infty \le \|h\|_2\|h'\|_2$, we see

$$\begin{aligned}1-\alpha &= \Pi_n \circ \kappa^{-1}(g : \|g - \bar{f}_n * \bar{f}_n\|_\infty \le R_n/\sqrt{n}) \\ &= \Pi_n(f : \|f * f - \bar{f}_n * \bar{f}_n\|_\infty \le R_n/\sqrt{n}) \\ &\le \Pi_n(f : 2\|(f - \mathbb{X}^{(n)}) * f_0\|_\infty \le R_n/\sqrt{n} + r_n) + o_P(1) \\ &\le \Pi_n(f : 2\sqrt{n}\|(f - \mathbb{X}^{(n)}) * f_0\|_\infty \le R_n + \delta_n) + o_P(1)\end{aligned}$$

with $\delta_n = r_n\sqrt{n} = o(1)$ as $n \to \infty$ by assumption. Using the weak convergence property established above,

$$1-\alpha \le \Phi_\lambda(R_n+\delta_n) + o_P(1) \quad \text{and similarly} \quad 1-\alpha \ge \Phi_\lambda(R_n-\delta_n) + o_P(1).$$

From this we conclude $R_n \to^{P^n_{f_0}} \Phi_\lambda^{-1}(1-\alpha)$ as $n \to \infty$. Now as above,

$$\begin{aligned}
P^n_{f_0}(f_0 * f_0 \in C_n) &= P^n_{f_0}\big(\|f_0 * f_0 - \bar{f}_n * \bar{f}_n\|_\infty \le R_n/\sqrt{n}\big) \\
&= P^n_{f_0}\big(2\|(\bar{f}_n - f_0) * f_0\|_\infty \le R_n/\sqrt{n}\big) + o(1) \\
&= P^n_{f_0}\big(2\sqrt{n}\|(\mathbb{X}^{(n)} - f_0) * f_0\|_\infty \le \Phi_\lambda^{-1}(1-\alpha)\big) + o(1) \\
&= P^n_{f_0}\big(2\|\mathbb{W} * f_0\|_\infty \le \Phi_\lambda^{-1}(1-\alpha)\big) + o(1) \\
&= \Phi_\lambda\big(\Phi_\lambda^{-1}(1-\alpha)\big) + o(1) = 1-\alpha+o(1)
\end{aligned}$$

completing the proof. □

PROOF OF THEOREM 3. The proof is similar to the previous ones, using the continuous mapping theorem for $L : H(\delta) \to \mathbb{R}$, hence ommitted. □

PROOF OF THEOREM 4. The following notation is used in the proof:

$$\theta_n^* = \Psi(f_0) + \left\langle \dot{\Psi}_{f_0}, \frac{\mathbb{W}}{\sqrt{n}} \right\rangle \quad \text{and} \quad \Phi_*(\cdot) = N\big(0, \|\dot{\Psi}_{f_0}\|_2^2\big)\big((-\infty, \cdot]\big).$$

By definition of the quantile μ_n it holds

$$\begin{aligned}
\frac{\alpha}{2} &= \Pi_n \circ \Psi^{-1}\big((-\infty, \mu_n]\big) = \Pi_n\big(\Psi(f) \le \mu_n\big) \\
&= \Pi_n\big(\Psi(f) - \Psi(f_0) \le \mu_n - \Psi(f_0)\big) \\
&= \Pi_n\big(\langle \dot{\Psi}_{f_0}, f - \mathbb{X}^{(n)}\rangle \le \mu_n - \theta_n^* - \big[\Psi(f) - \Psi(f_0) - \langle \dot{\Psi}_{f_0}, f - f_0\rangle\big]\big).
\end{aligned}$$

The assumed contraction of the posterior in a L^2-neighborhood of f_0 at rate r_n together with (20) and the fact that $\sqrt{n}r_n = o(1)$ imply the existence of $\delta_n \to 0$ such that

$$\begin{aligned}
\frac{\alpha}{2} &\le \Pi_n\big(\sqrt{n}\langle \dot{\Psi}_{f_0}, f - \mathbb{X}^{(n)}\rangle \le \sqrt{n}(\mu_n - \theta_n^*) + \delta_n\big) + o_P(1), \\
\frac{\alpha}{2} &\ge \Pi_n\big(\sqrt{n}\langle \dot{\Psi}_{f_0}, f - \mathbb{X}^{(n)}\rangle \le \sqrt{n}(\mu_n - \theta_n^*) - \delta_n\big) + o_P(1).
\end{aligned}$$

Using the continuous mapping theorem and Definition 1,

$$\beta_{\mathbb{R}}\big(\Pi_n \circ \tau_{\mathbb{X}^{(n)}}^{-1} \circ (D\Psi_{f_0})^{-1}, \mathcal{N} \circ (D\Psi_{f_0})^{-1}\big) \to^{P^n_{f_0}} 0$$

as $n \to \infty$. Note that $\mathcal{N} \circ (D\Psi_{f_0})^{-1}$ has distribution function Φ_*. Since the sets $\{(-\infty, t], t \in \mathbb{R}\}$ form a uniformity class for weak convergence towards a normal distribution, we obtain

$$\frac{\alpha}{2} \le \Phi_*\bigl(\sqrt{n}(\mu_n - \theta_n^*) + \delta_n\bigr) + o_P(1), \qquad \frac{\alpha}{2} \ge \Phi_*\bigl(\sqrt{n}(\mu_n - \theta_n^*) - \delta_n\bigr) + o_P(1).$$

From this we deduce $\mu_n = \theta_n^* + \frac{1}{\sqrt{n}}\Phi_*^{-1}(\frac{\alpha}{2}) + o_P(1/\sqrt{n})$. The quantile ν_n expands similarly, with $\Phi_*^{-1}(\frac{\alpha}{2})$ replaced by $\Phi_*^{-1}(1-\frac{\alpha}{2})$. By definition of θ_n^*,

$$\begin{aligned}
&P_{f_0}^n\bigl(\Psi(f_0) \in (\mu_n, \nu_n]\bigr)\\
&\quad = P_{f_0}^n\left(\left\langle \dot{\Psi}_{f_0}, \frac{\mathbb{W}}{\sqrt{n}}\right\rangle \in \left[\frac{\Phi_*^{-1}(\alpha/2)}{\sqrt{n}} + o_P\left(\frac{1}{\sqrt{n}}\right), \right.\right.\\
&\qquad\qquad\qquad\qquad \left.\left. \frac{\Phi_*^{-1}(1-\alpha/2)}{\sqrt{n}} + o_P\left(\frac{1}{\sqrt{n}}\right)\right]\right)\\
&\quad = P_{f_0}^n\bigl(\langle \dot{\Psi}_{f_0}, \mathbb{W}\rangle \in \bigl[\Phi_*^{-1}(\alpha/2), \Phi_*^{-1}(1-\alpha/2)\bigr]\bigr) + o(1)\\
&\quad = 1 - \alpha + o(1),
\end{aligned}$$

completing the proof. $\square$

4.1. *Some weak convergence facts.* Let μ, ν be Borel probability measures on a separable metric space (S, d). We call a family $\mathcal{U}$ of measurable real-valued functions defined on S a μ-uniformity class for weak convergence if for any sequence μ_n of Borel probability measures on S that converges weakly to μ, we also have

$$\sup_{u \in \mathcal{U}} \left| \int_S u(s)(d\mu_n - d\mu)(s) \right| \to 0 \tag{25}$$

as $n \to \infty$. Necessary and sufficient conditions for classes $\mathcal{U}$ of functions or sets $\{1_A : A \in \mathcal{A}\}$ to form uniformity classes are given in Billingsley and Topsøe [4]. For any subset A of S, define $A^\delta = \{x \in S : d(x, A) < \delta\}$ and the δ-boundary of A by $\partial_\delta A = \{x \in S : d(x, A) < \delta, d(x, A^c) < \delta\}$. A family $\mathcal{A}$ of measurable subsets of S is a μ-uniformity class if and only if

$$\lim_{\delta \to 0} \sup_{A \in \mathcal{A}} \mu(\partial_\delta A) = 0; \tag{26}$$

see Theorem 2 in [4]. For classes of functions a similar characterisation is available using moduli of continuity of the involved functions; see Theorem 1 in [4]. In particular the bounded Lipschitz metric

$$\beta(\mu, \nu) = \sup_{u \in \mathrm{BL}(1)} \left| \int_S u(s)(d\mu - d\nu)(s) \right|$$

tests against the class

$$\mathrm{BL}(1) = \left\{ f : S \to \mathbb{R}, \sup_{s \in S} |f(s)| + \sup_{s \neq t, s, t \in S} |f(s) - f(t)|/d(s,t) \le 1 \right\},$$

a uniformity class for any probability measure μ. The metric β metrises weak convergence of probability measures on S ([12], Theorem 11.3.3).

We conclude with the following observation, which was used repeatedly in our proofs: let $\mathcal{P}(S)$ denote the space of Borel probability measures on S, let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, let $\mu_n : (\Omega, \mathcal{A}, \mathbb{P}) \to \mathcal{P}(S)$, $n \in \mathbb{N}$, be random probability measures on S, and let $\mu \in \mathcal{P}(S)$ be fixed. If $\beta(\mu_n, \mu) \to^{\mathbb{P}} 0$ as $n \to \infty$, and if $\mathcal{U}$ is a μ-uniformity class, then the convergence in (25) holds in $\mathbb{P}$-probability, as is easily proved by contradiction and passing to a.s. convergent subsequences. Likewise, if (T, d') is a metric space and $F : S \to T$ a continuous mapping, then $\beta(\mu_n \circ F^{-1}, \mu \circ F^{-1}) \to 0$ in $\mathbb{P}$-probability.

Acknowledgement. We would like to thank the Associate Editor for several fundamental remarks that improved this article substantially. I. Castillo would like to thank the Cambridge Statistical Laboratory as well as Queens' College Cambridge for their hospitality during his visit to Cambridge, where parts of this research were carried out.

REFERENCES

[1] ARAUJO, A. and GINÉ, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables*. Wiley, New York. MR0576407

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

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

[4] BILLINGSLEY, P. and TOPSØE, F. (1967). Uniformity in weak convergence. *Z. Wahrsch. Verw. Gebiete* **7** 1–16. MR0209428

[5] BOGACHEV, V. I. (1998). *Gaussian Measures. Mathematical Surveys and Monographs* **62**. Amer. Math. Soc., Providence, RI. MR1642391

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

[9] CLARKE, B. and GHOSAL, S. (2010). Reference priors for exponential families with increasing dimension. *Electron. J. Stat.* **4** 737–780. MR2678969

[10] COHEN, A., DAUBECHIES, I. and VIAL, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* **1** 54–81. MR1256527

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

[12] DUDLEY, R. M. (2002). *Real Analysis and Probability*. Cambridge Univ. Press, Cambridge.

[13] FOLLAND, G. B. (1999). *Real Analysis: Modern Techniques and Their Applications*, 2nd ed. Wiley, New York. MR1681462

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

[15] FREES, E. W. (1994). Estimating densities of functions of observations. *J. Amer. Statist. Assoc.* **89** 517–525. MR1294078

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

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

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

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

[20] GINÉ, E. and MASON, D. M. (2007). On local U-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Statist*. **35** 1105–1145. MR2341700

[21] GINÉ, E. and NICKL, R. (2008). Uniform central limit theorems for kernel density estimators. *Probab. Theory Related Fields* **141** 333–387. MR2391158

[22] GINÉ, E. and NICKL, R. (2009). Uniform limit theorems for wavelet density estimators. *Ann. Probab.* **37** 1605–1646. MR2546757

[23] GINÉ, E. and NICKL, R. (2011). Rates on contraction for posterior distributions in L^r-metrics, $1 \le r \le \infty$. *Ann. Statist*. **39** 2883–2911. MR3012395

[24] KIEFER, J. and WOLFOWITZ, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions. *Z. Wahrsch. Verw. Gebiete* **34** 73–85. MR0397974

[25] 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

[26] LAPLACE, P. S. (1810). Mémoire sur les formules qui sont fonctions de très grands nombres et sur leurs applications aux probabilités. *Oeuvres de Laplace* **12** 301–345.

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

[28] LE CAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York. MR0856411

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

[30] NICKL, R. (2007). Donsker-type theorems for nonparametric maximum likelihood estimators. *Probab. Theory Related Fields* **138** 411–449. Correction ibid. **141** 331–332.

[31] NICKL, R. (2009). On convergence and convolutions of random signed measures. *J. Theoret. Probab.* **22** 38–56. MR2472004

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

[33] SCHICK, A. and WEFELMEYER, W. (2004). Root n consistent density estimators for sums of independent random variables. *J. Nonparametr. Stat*. **16** 925–935. MR2094747

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

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

[36] 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

[37] 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

[38] VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2009). Adaptive Bayesian estimation using a Gaussian random field with inverse gamma bandwidth. *Ann. Statist*. **37** 2655–2675. MR2541442

[39] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*: *With Applications to Statistics*. Springer, New York. MR1385671

[40] VON MISES, R. (1931). *Wahrscheinlichkeitsrechnung*. Deuticke, Vienna.

CNRS
LABORATOIRE PROBABILITÉS
ET MODÈLES ALÉATOIRES
LPMA
UNIVERSITÉS PARIS VI & 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