The Annals of Statistics
2011, Vol. 39, No. 6, 2883–2911
DOI: 10.1214/11-AOS924

RATES OF CONTRACTION FOR POSTERIOR DISTRIBUTIONS IN L^r-METRICS, $1 \le r \le \infty$

BY EVARIST GINÉ AND RICHARD NICKL

University of Connecticut and University of Cambridge

The frequentist behavior of nonparametric Bayes estimates, more specifically, rates of contraction of the posterior distributions to shrinking L^r-norm neighborhoods, $1 \le r \le \infty$, of the unknown parameter, are studied. A theorem for nonparametric density estimation is proved under general approximation-theoretic assumptions on the prior. The result is applied to a variety of common examples, including Gaussian process, wavelet series, normal mixture and histogram priors. The rates of contraction are minimax-optimal for $1 \le r \le 2$, but deteriorate as r increases beyond 2. In the case of Gaussian nonparametric regression a Gaussian prior is devised for which the posterior contracts at the optimal rate in all L^r-norms, $1 \le r \le \infty$.

1. Introduction. In finite-dimensional statistical models the Bernstein–von Mises theorem provides a frequentist justification of the use of Bayesian methods. In the case of infinite-dimensional models, consistency properties in weak metrics hold under relatively mild conditions; see Schwartz [28]. Consistency in stronger metrics was considered by Barron, Schervish and Wasserman [1] and by Ghosal, Ghosh and Ramamoorthi [9], and, shortly after, Ghosal, Ghosh and van der Vaart [10] and Shen and Wasserman [30] developed techniques that allow us to prove frequentist rates of contraction of the posterior to the true infinite-dimensional parameter in the Hellinger metric, if the prior is suitably chosen according to the structure of the nonparametric problem at hand. This led to further progress recently; we refer to [11, 12, 32, 34] and the references therein.

This literature has been successful in generalizing the scope of these techniques to a variety of different statistical models, and has naturally focussed on consistency and rates of contraction results in the *Hellinger* distance. For instance, if p_0 is the unknown density to be estimated, and if $\Pi(\cdot|X_1, \dots, X_n)$ is the posterior based on a prior Π and a sample $X_1, \dots, X_n$ with joint law P_0^n, results of the kind

$$\Pi\big(p : h(p, p_0) \ge \varepsilon_n | X_1, \dots, X_n\big) \to 0 \qquad \text{in } P_0^n \text{ probability} \tag{1}$$

were established, where $h^2(f, g) = \int(\sqrt{f} - \sqrt{g})^2$ is the Hellinger metric and where $\varepsilon_n \to 0$. Such posterior contraction results are known to imply the same frequentist consistency rate ε_n, also in the metric h, for the associated formal Bayes estimators.

Received March 2011; revised September 2011.
MSC2010 subject classifications. Primary 62G20; secondary 62G07, 62G08.
Key words and phrases. Rate of contraction, posterior, nonparametric hypothesis testing.

In this article we investigate the question of how to generalize results of this kind to more general loss-functions than the Hellinger metric, with a particular focus on L^r-norms, $1 \le r \le \infty$. Such results are of interest for a variety of reasons, for example, the construction of simultaneous confidence bands, or for plug-in procedures that require control of nonparametric remainder terms (e.g., in the proof of the Bernstein–von Mises theorem in semiparametric models in Castillo [6]). They are also of interest with a view on a more unified understanding of nonparametric Bayes procedures that complements the existing L^r-type results for standard frequentist methods.

The main challenge in extending the theory to the L^r-case, except for specific conjugate situations discussed below, rests in generalizing the Le Cam–Birgé testing theory for the Hellinger metric to more general situations. A main ingredient of the proof of a result of the kind (1) is that, in testing problems of the form

$$H_0 : p = p_0 \quad \text{against} \quad H_A : p \in \{p : h(p, p_0) \ge \varepsilon_n\}, \tag{2}$$

universal tests with concentration bounds on type-II errors of the type $e^{-Cn\varepsilon_n^2}$ exist, under assumptions on the size, or entropy, of the "alternative" space defining H_A. This fact is rooted in the subtle connection between nonparametric testing problems and the Hellinger metric as highlighted in the work of Le Cam [21] and Birgé [2]. A main contribution of this article is the development of a new approach to testing problems of the kind (2) based on concentration properties of linear centered kernel-type density estimators, derived from empirical process techniques. While this approach can only be used if one has sufficient control of the approximation properties of the support of the prior, it can be generalized to arbitrary L^r-metrics, including the supremum norm $\|f\|_\infty = \sup_x |f(x)|$. The concentration properties of these tests depend on the geometry of the L^r-norm and deteriorate as $r \to \infty$, which is, in a sense, dual to the fact that the minimax testing rate in the sense of Ingster [20] approaches the minimax rate of estimation as $r \to \infty$.

While our main results can be viewed as "abstract" in that they replace the entropy conditions in [10] for sieve sets $\mathcal{P}_n$ by general approximation-theoretic conditions (see Theorems 2 and 3 below), our findings become most transparent by considering specific examples, selected in an attempt to reflect the spectrum of situations that can arise in Bayesian nonparametrics: In Section 2 we study the "ideal" situation of a simple uniform wavelet prior on a Hölder ball, the "supersmooth" situation of mixtures of normals, the case of random histograms based on a Dirichlet process where no uniform bound on the L^∞-norm of the support of the prior is available, as well as Gaussian process priors of the kind studied in [32]. The general conclusion is that if f_0 is α-smooth, then the rate of contraction obtained in the L^r-norm for a posterior based on an adequately chosen prior of smoothness α is, up to $\log n$ factors, and with $\bar{r} = \max(2, r)$,

$$\left(\frac{1}{n}\right)^{(\alpha - 1/2 + 1/\bar{r})/(2\alpha+1)}. \tag{3}$$

So as soon as $r \le 2$ our proof retrieves the minimax optimal rate, but for $r > 2$ the rate deteriorates by a genuine power of n. As α approaches infinity this effect becomes more lenient and vanishes in the limit.

We currently have no proof of the fact that our general theorem gives the right rate for Bayesian posteriors if $r > 2$—similar problems are known with nonparametric maximum likelihood estimators in L^r-metrics (cf. the proof of Proposition 6 in [27]). While we do not settle the issue of optimality of our rates for $r > 2$ in this article, we also prove in Theorem 1 below that in nonparametric Gaussian regression the minimax rate of contraction can be obtained by certain diagonal Gaussian wavelet priors, in all L^r-norms simultaneously. We believe that this result is closely tied to the fact that the posterior is then itself Gaussian, and conjecture that our rates cannot be substantially improved in the nonconjugate situation.

2. Main results. Let $\mathcal{P}$ be a class of probability densities on $[0, 1]$ or $\mathbb{R}$, and let $X_1, \dots, X_n$ be a random sample drawn from some unknown probability density p_0 with joint law the first n coordinate projections of the infinite product probability measure $P_0^{\mathbb{N}}$. Suppose one is given a prior probability distribution Π defined on some σ-algebra $\mathcal{B}$ of $\mathcal{P}$. The posterior is the random probability measure

$$\Pi(B|X_1, \dots, X_n) = \frac{\int_B \prod_{i=1}^n p(X_i)\, d\Pi(p)}{\int_{\mathcal{P}} \prod_{i=1}^n p(X_i)\, d\Pi(p)}, \qquad B \in \mathcal{B}.$$

We wish to analyze contraction properties of the posterior distribution under certain regularity conditions on Π and p_0, and these regularity properties can be conveniently characterized by wavelet theory.

2.1. *Function spaces and wavelets.* For $T = \mathbb{R}$ or $T = [0, 1]$, $f : T \mapsto \mathbb{R}$, we shall write $\|f\|_\infty = \sup_{x \in T} |f(x)|$, the norm on the space $C(T)$ of bounded continuous real-valued functions defined on T. We shall use wavelet theory throughout; see [19, 26]. Let ϕ, ψ be the scaling function and wavelet of a multiresolution analysis of the space $L^2(T)$ of square integrable real-valued functions on T. We shall say that the wavelet basis is S-*regular* if ϕ, ψ are S-times continuously differentiable on T. For instance we can take Daubechies wavelets on $T = \mathbb{R}$ of sufficiently large order N (see [26]) and define the translated scaling functions and wavelets

$$\phi_k = \phi(\cdot - k), \qquad \psi_{\ell k} = 2^{\ell/2} \psi\bigl(2^\ell (\cdot) - k\bigr), \qquad \ell \in \mathbb{N} \cup \{0\}, k \in \mathbb{Z}, \tag{4}$$

which form an orthonormal basis of $L^2(\mathbb{R})$.

For $T = [0, 1]$ we consider the orthonormal wavelet bases of $L^2([0, 1])$ constructed in Theorem 4.4 of Cohen, Daubechies and Vial [8]. Each such basis is built from a Daubechies scaling function ϕ and its corresponding wavelet ψ, of order N, starting at a fixed resolution level J_0 such that $2^{J_0} \ge 2N$ (see Theorem 4.4 in [8]): the $\psi_{\ell k}, \phi_k$ that are supported in the interior of $[0, 1]$ are all kept, and suitable boundary corrected wavelets are added, so that the $\{\phi_k, \psi_{\ell k} : 0 \le k <$

$2^\ell, \ell \in \mathbb{N}, \ell > J_0\}$ still form an orthonormal basis for $L^2([0,1])$. While formula (4) now only applies to the "interior" wavelets, one can still write $\phi_{jk} = 2^{j/2}\phi_k(2^j\cdot)$ for every k, $j \geq J_0$; cf. page 73 in [8] and also after Condition 1 below.

DEFINITION 1. Let $T = [0,1]$ or $T = \mathbb{R}$, and let $1 \leq p, q \leq \infty$, $0 \leq s < S$, $s \in \mathbb{R}$, $S \in \mathbb{N}$. Let ϕ, ψ be bounded, compactly supported S-regular scaling function and wavelet, respectively, and denote by $\alpha_k(f) = \int_T \phi_k f$ and $\beta_{\ell k}(f) = \int_T \psi_{\ell k} f$ the wavelet coefficients of $f \in L^p(T)$. The Besov space $B^s_{pq}(T)$ is defined as the set of functions $\{f \in L^p(T) : \|f\|_{s,p,q} < \infty\}$ where

$$\|f\|_{s,p,q} := \|\alpha_{(\cdot)}(f)\|_p + \left(\sum_{\ell=0}^{\infty} \left(2^{\ell(s+1/2-1/p)}\|\beta_{\ell(\cdot)}(f)\|_p\right)^q\right)^{1/q}$$

with the obvious modification in case $q = \infty$.

REMARK 1. We note the following standard embeddings/identifications we shall use (cf. [19, 26]): for $\mathcal{C}^s(T)$ the Hölder (-Zygmund in case s integer) spaces on T, we have $B^s_{\infty\infty}(T) = \mathcal{C}^s(T)$. Moreover $B^s_{22}(T) = H^s(T)$ where $H^s(T)$ are the standard L^2-Sobolev spaces. We also have the "Sobolev-type" imbeddings $B^s_{rq}(T) \subset B^{s-1/r+1/t}_{tq}(T)$ for $t \geq r$, $1 \leq q \leq \infty$. Finally, if $T = [0,1]$, then $C^\alpha(T) \subset B^\alpha_{r\infty}(T)$ for every $r \leq \infty$, where $C^\alpha(T) = \{f : T \mapsto \mathbb{R} : \|f\|_{\alpha,\infty} < \infty\}$, with $\|f\|_{\alpha,\infty} := \sum_{k=0}^{\alpha}\|f^{(k)}\|_\infty$, $\alpha \in \mathbb{N}$.

2.2. *Uniform wavelet series.* Let us consider first the case where an a priori upper bound on the Hölder norm $\|p_0\|_{\alpha,\infty,\infty}$ is available, so that the prior can be chosen to have bounded support in $\mathcal{C}^\alpha([0,1])$. An example is obtained, for example, by uniformly distributing wavelet coefficients on a Hölder ball. Let $\{\phi_k, \psi_{\ell k}\}$ be a N-regular CDV-wavelet basis for $L^2([0,1])$, let $u_{\ell k}$ be i.i.d. $U(-B, B)$ random variables, and define, for $\alpha < N$, the random wavelet series

$$U_\alpha(x) = \sum_k u_{0k}\phi_k(x) + \sum_{\ell=J_0}^{\infty}\sum_k 2^{-\ell(\alpha+1/2)} u_{\ell k}\psi_{\ell k}(x), \tag{5}$$

which has trajectories in $\mathcal{C}^\alpha([0,1]) \subset L^r([0,1])$, $1 \leq r \leq \infty$, almost surely (in view of Definition 1 and Remark 1). Since moreover $\|U_\alpha\|_\infty \leq C(B, \alpha, \psi)$, and since the exponential map has bounded derivatives on bounded subsets of $\mathbb{R}$, the same applies to the random density

$$p^{U,\alpha}(x) := \frac{e^{U_\alpha(x)}}{\int_0^1 e^{U_\alpha(y)}\,dy},$$

whose induced law on $C([0,1])$ we denote by Π^α. Our general results below imply the following proposition, which, since p_0 is bounded away from zero, implies the same contraction rate in Hellinger distance h. Note moreover that the result for $2 < r < \infty$ could be obtained from interpolation properties of L^r-spaces.

PROPOSITION 1. *Let $X_1, \dots, X_n$ be i.i.d. on $[0, 1]$ with density p_0 satisfying $\|\log p_0\|_{\alpha,\infty} \le B$. Let $1 \le r \le \infty$, $\bar{r} = \max(2, r)$, $r^* = \min(r, 2)$, and suppose $\alpha \ge 1 - 1/r^*$. Then there exist finite positive constants $M, \eta = \eta(\alpha, r)$ such that, as $n \to \infty$,*

$$\begin{aligned}\Pi^\alpha\{p \in \mathcal{P} : \|p - p_0\|_r \ge M n^{-(\alpha-1/2+1/\bar{r})/(2\alpha+1)} (\log n)^\eta | X_1, \dots, X_n\} \\ \to^{P_0^{\mathbb{N}}} 0.\end{aligned} \tag{6}$$

2.3. *Dirichlet mixtures.* Consider first, as in [9, 12, 13], a normal mixture prior Π, defined as follows: for φ the standard normal density, set:

(-) $p_{F,\sigma} = \int_{\mathbb{R}} \sigma^{-1} \varphi((\cdot - y)/\sigma)\, dF(y)$,
(-) $F \sim D_\alpha$ the Dirichlet-process with base measure $\alpha = \alpha(\mathbb{R})\bar{\alpha}$, $\alpha(\mathbb{R}) < \infty$ and $\bar{\alpha}$ a probability measure,
(-) $\sigma \sim G$, where G is a probability distribution with compact support in $(0, \infty)$.

PROPOSITION 2. *Let $X_1, \dots, X_n$ be i.i.d. on $\mathbb{R}$ with density p_{F_0,σ_0} where $\sigma_0 > 0$ and where F_0 is supported in $[-k_0, k_0]$, $k_0 > 0$. Suppose that G has a positive continuous density in a neighborhood of σ_0, and that the base measure α has compact support and a continuous density on an interval containing $[-k_0, k_0]$. Then there exist finite positive constants M, η such that*

$$\Pi^\alpha\left\{p \in \mathcal{P} : \|p - p_0\|_\infty \ge M \frac{(\log n)^\eta}{\sqrt{n}} \Big| X_1, \dots, X_n\right\} \to^{P_0^{\mathbb{N}}} 0 \qquad \textit{as } n \to \infty. \tag{7}$$

Consider next a random histogram based on a Dirichlet process, similar to the priors studied in [29]: for $j \in \mathbb{N}$ let Dir_j be a Dirichlet-distribution on the 2^j-dimensional unit simplex, with all parameters equal to one. Consider the dyadic random histogram with resolution level j

$$\sum_{k=1}^{2^j} \alpha_{jk} 2^j 1\left\{\left(\frac{k-1}{2^j}, \frac{k}{2^j}\right]\right\}(x), \qquad \{a_{jk}\} \sim \mathrm{Dir}_j, \qquad x \in [0, 1],$$

and denote its law on the space of probability densities by Π_j. Note that this prior is not concentrated uniformly (in j) on bounded densities (despite the densities in the support being uniformly bounded for fixed j).

PROPOSITION 3. *Let $X_1, \dots, X_n$ be i.i.d. on $[0, 1]$ with density $p_0 \in \mathcal{C}^\alpha([0, 1])$, $0 < \alpha \le 1$, satisfying $p_0 > 0$ on $[0, 1]$. Let j_n be such that $2^{j_n} \sim (n/\log n)^{1/(2\alpha+1)}$, let $1 \le r \le \infty$, $\bar{r} = \max(2, r)$ and let either $\alpha > 1/2$ or $r = 1$. Then for some $M, \eta = \eta(\alpha, r)$, as $n \to \infty$*

$$\begin{aligned}\Pi_{j_n}\{p \in \mathcal{P} : \|p - p_0\|_r \ge M n^{-(\alpha-1/2+1/\bar{r})/(2\alpha+1)} (\log n)^\eta | X_1, \dots, X_n\} \\ \to^{P_0^{\mathbb{N}}} 0.\end{aligned} \tag{8}$$

2.4. *Gaussian process priors.* We now study a variety of Gaussian process priors that were considered in the nonparametric Bayes literature recently; see [32, 34] for references. To reduce technicalities we shall restrict ourselves to integrated Brownian motions, but see also the remark below.

DEFINITION 2. Let $B(t) = B_{1/2}(t)$, $t \in [0,1]$, be a (sample-continuous version of) standard Brownian motion. For $\alpha > 1$, $\alpha \in \{n - 1/2 : n \in \mathbb{N}\}$, setting $\{\alpha\} = \alpha - [\alpha]$, $[\alpha]$ being the integer part of α, B_α is defined as the $[\alpha]$-fold integral

$$B_\alpha(t) = \int_0^t \int_0^{t_{[\alpha]-1}} \cdots \int_0^{t_2} \int_0^{t_1} B(s)\,ds\,dt_1 \cdots dt_{[\alpha]-1}$$
$$= \frac{1}{([\alpha]-1)!} \int_0^t (t-s)^{[\alpha]-1} B(s)\,ds, \qquad t \in [0,1],$$

where for $[\alpha] = 1$ the multiple integral is understood to be only $\int_0^t B(s)\,ds$.

Following [23, 32], and as before Proposition 1, we would like to define our prior on densities as the probability law of the random process

$$\frac{e^{B_\alpha}}{\int_0^1 e^{B_\alpha(t)}\,dt}, \tag{9}$$

but we must make two corrections: first, since $B_\alpha^{(k)}(0) = 0$ a.s., $k \le [\alpha]$, would impose unwanted conditions on the value at zero of the density, we should release B_α at zero, that is, take $\bar{B}_\alpha := \sum_{k=0}^{[\alpha]} Z_k t^k / k! + B_\alpha$, where Z_k are i.i.d. $N(0,1)$ variables independent of B_α; see [32]. In order to deal with bounded densities, we introduce a second modification to (9), and define our prior (on the Borel sets of $C([0,1])$) as

$$\Pi = \mathcal{L}\left(\frac{e^{\bar{B}_\alpha}}{\int_0^1 e^{\bar{B}_\alpha(t)}\,dt} \,\middle|\, \|\bar{B}_\alpha\|_\infty \le c \right), \tag{10}$$

where c is a fixed arbitrary positive constant. This prior works as follows: if $A \subset C([0,1])$ is a measurable set of continuous densities on $[0,1]$, then

$$\Pi(A) = \Pr\left\{ e^{\bar{B}_\alpha} \Big/ \int e^{\bar{B}_\alpha} \in A, \|\bar{B}_\alpha\|_\infty \le c \right\} \Big/ \Pr\{\|\bar{B}_\alpha\|_\infty \le c\},$$

and clearly the denominator is strictly positive for all $c > 0$; see Proposition 7 below.

PROPOSITION 4. *Let* $1 \le r \le \infty$, $\bar{r} = \max(r,2)$, $\alpha \in \{n - 1/2, n \in \mathbb{N}\}$ *and assume* (a) $p_0 \in \mathcal{C}^\alpha([0,1])$, *and* (b) p_0 *is bounded and bounded away from zero, say*, $2\|\log p_0\|_\infty \le c < \infty$. *Let* Π *be the prior defined by* (10) *where* α *is as in* (a) *and c is as in* (b). *Then, if X_i are i.i.d. with common law P_0 of density p_0, there exists $M < \infty$ s.t.*

$$\Pi\{p \in \mathcal{P} : \|p - p_0\|_r \geq M n^{-(\alpha-1/2+1/\bar{r})/(2\alpha+1)} (\log n)^{(1/2)1_{\{r=\infty\}}} | X_1, \dots, X_n\} \to 0$$

in $P_0^{\mathbb{N}}$-probability as $n \to \infty$.

As remarked before Proposition 1, a contraction result in the Hellinger distance follows as well, and the case $2 < r < \infty$ could be obtained from interpolation.

The result in Proposition 4 extrapolates to fractional multiple integrals of Brownian motion (Riemann–Liouville processes) of any real valued index $\alpha > 1/2$, and it also extends to the related fractional Brownian motion processes (see, e.g., [32] for definitions), but, for conciseness and clarity of exposition, we refrain from carrying out these extensions.

2.5. *Sharp rates in the Gaussian conjugate situation.* We currently have no proof that the rates obtained in the previous subsections are optimal for these priors as soon as $r > 2$. While we conjecture that Bayesian posteriors may suffer from suboptimal contraction rates in density estimation problems in L^r-loss, $r > 2$, we finally show here that in the much simpler conjugate situation of nonparametric regression with Gaussian errors, sharp rates in all L^r norms can be obtained at least for certain diagonal wavelet priors. The proof of this result follows from a direct analysis of the posterior distribution, available in closed form due to conjugacy.

Given a noise level $1/\sqrt{n}$, $n \in \mathbb{N}$, we observe

$$dY^{(n)}(t) = f(t)\,dt + \frac{1}{\sqrt{n}}\,dB(t), \qquad t \in [0, 1], \tag{11}$$

for $f = f_0 \in L^2([0, 1])$, where B is Brownian motion on $[0, 1]$. This model is well known to be asymptotically equivalent to nonparametric regression with fixed, equally-spaced design and Gaussian errors.

Consider priors on $L^2([0, 1])$ defined on a S-regular CDV-wavelet basis as

$$\Pi = \mathcal{L}\left(\sum_{k=0}^{N} g_k \phi_k + \sum_{\ell=J_0}^{\infty} \sum_{k=0}^{2^\ell-1} \sqrt{\mu_\ell}\, g_{\ell k} \psi_{\ell k}\right) \tag{12}$$

in $L^2([0, 1])$, with the g's i.i.d. $N(0, 1)$ and with $\mu_\ell = \ell^{-1} 2^{-\ell(2\alpha+1)}$ $\forall \ell \geq J_0$. Such a prior is designed for α-smooth f_0. As is easily seen, the series in (12) converges uniformly almost surely.

THEOREM 1. *Let $0 < \alpha < S$, and let Π be the Gaussian prior on $L^2([0, 1])$ defined by* (12) *based on a CDV wavelet basis of $L^2([0, 1])$ of smoothness at least S. Let $f_0 \in \mathcal{C}^\alpha([0, 1])$, let $\varepsilon_n = (n/\log n)^{-\alpha/(2\alpha+1)}$ and suppose we observe*

$dY_0^{(n)}(t) = f_0(t)\,dt + dB(t)/\sqrt{n}$. Then there exists $C < \infty$ and $M_0 < \infty$ depending only on the wavelet basis, α and $\|f_0\|_{\alpha,\infty,\infty}$ such that, for every $M_0 \le M < \infty$, and for all $1 \le r \le \infty$, $n \in \mathbb{N}$,

$$E_{Y_0^{(n)}} \Pi\big(f : \|f - f_0\|_r > M\varepsilon_n | Y_0^{(n)}\big) \le n^{-C^2(M-M_0)^2}. \tag{13}$$

This rate of convergence is sharp (in case $r < \infty$ up to the $\log n$-term) in view of the usual minimax lower bounds and since the contraction rate implies the same rate of convergence for the formal Bayes estimator $E_\Pi(f|Y_0^{(n)})$ to f_0 (using Anderson's lemma and the fact that the posterior is a random Gaussian measure on $L^2([0,1])$, as inspection of the proof shows). One may even apply the usual thresholding techniques to the posterior mean to obtain a Bayesian rate adaptive estimator of f_0 by proceeding as in [17, 25].

3. General contraction theorems for density estimates in L^r-loss, $1 \le r \le \infty$. We shall, in our main results, use properties of various approximation schemes in function spaces, based on integrating a localized kernel-type function $K_j(x, y)$ against functions p, $K_j(p) = \int K_j(\cdot, y)p(y)\,dy$. Let, in slight abuse of notation, for $T \subseteq \mathbb{R}$, $L^1(\mu_w) = L^1(T, \mathcal{B}, \mu_w)$, $w \ge 0$ be the space of μ_w-integrable functions, $d\mu_w(t) = (1 + |t|)^w\,dt$, normed by $\|f\|_{\mu_w} = \int_T |f(t)|(1 + |t|)^w\,dt$. Recall the notion of p-variation of a function (e.g., as before Lemma 1 in [17]).

CONDITION 1. Let $T = \mathbb{R}$ or $T = [0, 1]$. The sequence of operators $K_j(x, y) = 2^j K(2^j x, 2^j y)$; $x, y \in T$, $j \ge 0$, is called an admissible approximating sequence if it satisfies one of the following conditions:

(a) (convolution kernel case): $K(x, y) = K(x - y)$, where $K \in L^\infty(T)$ is of bounded p-variation for some finite $p \ge 1$, right (or left) continuous, and satisfies $\|K\|_{\mu_w} < \infty$ for some $w > 2$.

(b) (multiresolution projection case): $K(x, y) = \sum_k \phi(x - k)\phi(y - k)$, the sum extending over any subset of $\mathbb{Z}$, where $\phi \in L^1 \cap L^\infty$ has bounded p-variation for some finite $p \ge 1$ and satisfies, in addition, $\sup_{x\in\mathbb{R}} \sum_k |\phi_k(x)| < \infty$ as well as $|K(x, y)| \le \Phi(|x - y|)$ for every $x, y \in T$ and some function $\Phi \in L^\infty(\mathbb{R})$ for which $\|\Phi\|_{\mu_w} < \infty$ for some $w > 2$.

(c) (multiresolution case, $T = [0, 1]$): $K(x, y) = \sum_k \phi_k(x)\phi_k(y)$ is the projection kernel of a Cohen–Daubechies–Vial (CDV) wavelet basis.

Condition (a) is a standard assumption on kernels, condition (b) is satisfied for most wavelet basis on $\mathbb{R}$, such as Daubechies, Meyer or spline wavelets, by using standard wavelet theory (e.g., [19]). For part (c) we note the following: as in the case of the whole line, an orthonormal basis of $V_j = \{\phi_{jk} = 2^{j/2}\phi_k(2^j\cdot)\}$ is obtained from 2^{j-J_0}-fold dilates of the basic linear span V_{J_0}, for every $j \geq J_0$ (page 73 in [8]). In this case, V_j has dimension 2^j, and a basis consists of: (i) N left edge functions $\phi^0_{jk}(x) = 2^{j/2}\phi^0_k(2^j x)$, $k = 0, \dots, N-1$, where ϕ^0_k is a modification of ϕ, which is still bounded and of bounded support; (ii) N right edge functions $\phi^1_{jk}(x) = 2^{j/2}\phi^1_k(2^j x)$, $k = 0, \dots, N-1$, ϕ^1_k also modifications of ϕ bounded and of bounded support, and then the $2^j - N$ "interior" usual translations of dilations of ϕ, ϕ_{jk}, $k = N, \dots, 2^j - N - 1$. The projection kernel $K_j(x, y) = K^0_j(x, y) + K^1_j(x, y) + \tilde{K}_j(x, y)$ corresponds to the projection onto the three orthogonal components of V_j (the linear spans, respectively, of the left edge functions $\phi^0_{j,k}$, the right edge functions ϕ^1_k, and the interior functions ϕ_{jk}). The first two spaces have dimension N and the third, $2^j - 2N$. By Lemma 8.6 in [19], there exist bounded, compactly supported nonnegative functions Φ such that $\tilde{K}(x, y) \leq \Phi(|x - y|)$, for all x, y. We call this function a majorizing kernel of the interior part of K.

Let X_i be i.i.d. with law P_0 and density p_0.

THEOREM 2. *Let $T = [0, 1]$ or $T = \mathbb{R}$, let $\mathcal{P} = \mathcal{P}(T)$ be a set of probability densities on T, and let Π_n be priors defined on some σ-algebra of $\mathcal{P}$ for which the maps $p \mapsto p(x)$ are measurable for all $x \in T$. Let $1 \leq r \leq \infty$ and let $\varepsilon_n \to 0$ as $n \to \infty$ be a sequence of positive numbers such that $\sqrt{n}\varepsilon_n \to \infty$ as $n \to \infty$. Let*

$$\delta_n = \varepsilon_n (n\varepsilon_n^2)^{1/2-1/(2r)}\gamma_n \tag{14}$$

for some sequence γ_n satisfying $\gamma_n \geq 1$ $\forall n$. Let J_n be any sequence satisfying $2^{J_n} \leq cn\varepsilon_n^2$ for some fixed $0 < c < \infty$, and let K_j be an admissible approximator sequence. Let $\mathcal{P}_n$ be a sequence of subsets of

$$\{p \in \mathcal{P} : \|K_{J_n}(p) - p\|_r \leq C(K)\delta_n, \|p\|_{\mu_w} \leq D\}, \tag{15}$$

where $C(K)$ is a constant that depends only on the operator kernel K, D is a fixed constant, and where $w > (2 - r)/r$ if $r < 2$, $w = 0$ if $r \geq 2$.

Assume there exists $C > 0$ such that, for every n large enough:

(1) $\Pi_n(\mathcal{P} \setminus \mathcal{P}_n) \leq e^{-(C+4)n\varepsilon_n^2}$ and
(2) $\Pi_n\{p \in \mathcal{P} : -P_0 \log \frac{p}{p_0} \leq \varepsilon_n^2, P_0(\log \frac{p}{p_0})^2 \leq \varepsilon_n^2\} \geq e^{-Cn\varepsilon_n^2}$.

Let $p_0 \in L^r(T)$ be s.t. $\|K_{J_n}(p_0) - p_0\|_r = O(\delta_n)$ and s.t. $\|p_0\|_{\mu_w} < \infty$ if $T = \mathbb{R}$, $1 \leq r < 2$. If $\delta_n \to 0$ as $n \to \infty$, then there exists $M < \infty$ such that

$$\Pi_n\{p \in \mathcal{P} : \|p - p_0\|_r \geq M\delta_n | X_1, \dots, X_n\} \to 0 \qquad \text{as } n \to \infty \tag{16}$$

in $P_0^{\mathbb{N}}$-probability.

Note that the moment condition in (15) is void if $r \geq 2$ or if $T = [0, 1]$. If $r = 1$ the rate can be taken to be $\delta_n = \varepsilon_n$ or, more generally, $\delta_n = \gamma_n\varepsilon_n$. For $r = \infty$ one only has at best $\delta_n = \sqrt{n}\varepsilon_n^2$, which is always slower than ε_n (since $\sqrt{n}\varepsilon_n \to \infty$). In case $1 < r < \infty$ the rate interpolates between these two rates without, however, requiring $p_0 \in L^\infty$.

In the case where p_0 is bounded, and if it is known that the posterior concentrates on a fixed sup-norm ball with probability approaching one, we can refine the rates in the above theorem for $1 < r < \infty$, and retrieve the (in applications of the theorem often optimal) rate ε_n for $1 \le r \le 2$. The following theorem can be applied with $\gamma_n = 1$ $\forall n$, in which case conditions (a) and (b) require the rate ε_n to be fast enough (which in applications typically entails that a minimal degree of smoothness of p_0 has to be assumed).

THEOREM 3. *Let $T, \mathcal{P}, \Pi_n$ be as in Theorem 2. Let $1 < r < \infty$, and let $\varepsilon_n \to 0$ as $n \to \infty$ be a sequence of positive numbers such that $\sqrt{n}\varepsilon_n \to \infty$ as $n \to \infty$. Let $\bar{r} = \max(r, 2)$, and set*

$$\delta_n = \varepsilon_n (n\varepsilon_n^2)^{1/2-1/\bar{r}} \gamma_n \tag{17}$$

for some sequence $\gamma_n \ge 1$. Assume either:

(a) *that $1 < r < 2$ and that $\varepsilon_n = O(\gamma_n (n\varepsilon_n^2)^{1/r-1})$ or*
(b) *that $2 \le r < \infty$ and that $\varepsilon_n^2 = O(\gamma_n/\sqrt{n})$.*

Let $J_n, \mathcal{P}_n$ be defined as in Theorem 2, assume that conditions (1) *and* (2) *in that theorem are satisfied, and that, in addition,*

(3) *there exists $0 < B < \infty$ such that*

$$\Pi_n(p \in \mathcal{P} : \|p\|_\infty > B | X_1, \dots, X_n) \to 0$$

as $n \to \infty$ in $P_0^{\mathbb{N}}$-probability.

Let $p_0 \in L^\infty(T)$ be s.t. $\|K_{J_n}(p_0) - p_0\|_r = O(\delta_n)$ and such that $\|p_0\|_{\mu_w} < \infty$ for some $w > (2-r)/r$ if $T = \mathbb{R}$, $1 \le r < 2$. If $\delta_n \to 0$ as $n \to \infty$, then there exists $M < \infty$ s.t.

$$\Pi_n\{p \in \mathcal{P} : \|p - p_0\|_r \ge M\delta_n | X_1, \dots, X_n\} \to 0 \qquad \text{as } n \to \infty \tag{18}$$

in $P_0^{\mathbb{N}}$-probability.

3.1. *L^r-norm inequalities.* A main step in the proof of Theorems 2 and 3 [see (30) below] is the construction of nonparametric tests for L^r-alternatives, $1 \le r \le \infty$, that have sufficiently good exponential bounds on the type-two errors. For this we first derive sharp concentration inequalities for L^r-norms of centered density estimators. It is convenient to observe that the degree of concentration of a kernel-type density estimator around its expectation in L^r depends on r, as can already be seen from comparing the known cases $r = 1, \infty$ in [14, 16] for kernel estimators and [17] for wavelets. These results are derived from Talagrand's inequality

[31] for empirical processes: let $X_1, \dots, X_n$ be i.i.d. with law P on a measurable space $(S, \mathcal{S})$, let $\mathcal{F}$ be a P-centered (i.e., $\int f\, dP = 0$ for all $f \in \mathcal{F}$) countable class of real-valued measurable functions on S, uniformly bounded by the constant U, and set $\|H\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |H(f)|$ for any $H : \mathcal{F} \to \mathbb{R}$. Let σ be any positive number such that $\sigma^2 \geq \sup_{f \in \mathcal{F}} E(f^2(X))$, and set $V := n\sigma^2 + 2UE\|\sum_{j=1}^n f(X_j)\|_{\mathcal{F}}$. Then, Bousquet's [5] version of Talagrand's inequality, with constants, is as follows (see Theorem 7.3 in [5]): for every $x \geq 0, n \in \mathbb{N}$,

$$\Pr\left\{\left\|\sum_{j=1}^n f(X_j)\right\|_{\mathcal{F}} \geq E\left\|\sum_{j=1}^n f(X_j)\right\|_{\mathcal{F}} + \sqrt{2Vx} + Ux/3\right\} \leq 2e^{-x}. \tag{19}$$

This applies to our situation as follows: let $X_1, \dots, X_n$ be i.i.d. with density p_0 on T with respect to Lebesgue measure λ, $dP_0 = p_0 d\lambda$, and let $\hat{p}_n(j) = \frac{1}{n}\sum_{i=1}^n K_j(\cdot, X_i)$ be a kernel-type estimator with K_j as in Condition 1. Its expectation equals $P_0^n \hat{p}_n(j)(x) = EK_j(x, X) = K_j(p_0)(x)$, and we wish to derive sharp exponential bounds for the quantity $\|\hat{p}_n(j) - K_j(p_0)\|_r$ for $1 \leq r \leq \infty$. In case $r = \infty$ this can be achieved by studying the empirical process indexed by

$$\mathcal{K} = \{K_j(x, \cdot) - K_j(p_0)(x) : x \in T\},$$

and in case $r < \infty$ we shall view $\hat{p}_n(j) - P_0^n \hat{p}_n(j)$ as a sample average of the centered $L^r(T)$-valued random variables $K_j(\cdot, X_i) - K_j(p_0)$, and reduce the problem to an empirical process as follows: let s be conjugate to r, that is, $1 = 1/s + 1/r$. By the Hahn–Banach theorem, the separability of $L^r(T)$ implies that there is a countable subset B_0 of the unit ball B of $L^s(T)$ such that

$$\|H\|_r = \sup_{f \in B_0}\left|\int_{\mathbb{R}} H(t) f(t)\, dt\right|$$

for all $H \in L^r(T)$. We thus have $\|\hat{p}_n(j) - P_0^n \hat{p}_n(j)\|_r = \|P_n - P_0\|_{\mathcal{K}}$, where $P_n = \sum_{i=1}^n \delta_{X_i}/n$ is the empirical measure, and where

$$\mathcal{K} = \left\{x \mapsto \int_T f(t) K_j(t, x)\, dt - \int_T f(t) K_j(p_0)(t)\, dt : f \in B_0\right\}.$$

To apply (19) with the countable class $\mathcal{K}$ we need to find suitable bounds for the envelope $U \geq \sup_{k \in \mathcal{K}} |k(x)|$ and the weak variances $\sigma^2 \geq \sup_{k \in \mathcal{K}} Ek^2(X)$. We will also apply (19) in the case $r = \infty$, and note that the corresponding empirical process suprema are over countable subsets B_0 of T, by the continuity property of K in the convolution kernel case, and by finiteness of the p-variation of the scaling function in the wavelet case (Remark 2 in [17]).

3.1.1. *Envelope and variance bounds for $\mathcal{K}$.* We first consider Condition 1(a), the convolution kernel case: let us write in abuse of notation $K_j(\cdot) = 2^j K(2^j \cdot)$ and

$f = \delta_y$, $y \in B_0 \subset T$ for $r = \infty$. (One naturally replaces L^s by the Banach space of finite signed measures if $r = \infty$ in the arguments below.) The class $\mathcal{K}$ then equals

$$\mathcal{K} = \{x \mapsto K_j * f(x) - E(K_j * f(X)) : f \in B_0\}.$$

The bound for the envelope is seen to be of size $2^{j(1-1/r)}$: by Hölder's inequality

$$\|K_j * f\|_\infty \le \|K_j\|_r \|f\|_s \le C(K, r) 2^{j(1-1/r)} \equiv U, \tag{20}$$

a bound that remains true when $r = \infty$ since $|2^j K(2^j(x-y))| \le \|K\|_\infty 2^j$. To bound the variances, for densities $p_0 \in L^r$, we have

$$E(K_j * f)(X)^2 \le \|p_0\|_r \|K_j * f\|_{2s}^2 \le C'(K, r)\|p_0\|_r 2^{j(1-1/r)} \equiv \sigma^2 \tag{21}$$

from Hölder's inequality and since $\|K_j * f\|_{2s}$, for $f \in L^s$ is bounded up to constants by $2^{j(1/2-1/2r)}$, by using Young's inequality $\|h * g\|_t \le \|h\|_p \|g\|_q$ for $1 + 1/t = 1/p + 1/q$, $1 \le p, q, t \le \infty$.

The last estimate can be refined if p_0 is known to be bounded, where we recall that $\bar{r} = \max(r, 2)$, to yield

$$E(K_j * f)(X)^2 \le C(p_0) 2^{j(1-2/\bar{r})} \equiv \sigma^2, \tag{22}$$

where $C(\cdot)$ is bounded on uniformly bounded sets of densities. To see this, consider first $r \ge 2$ and thus $s \le 2$: then Young's inequality gives, as above,

$$E(K_j * f)(X)^2 \le \|p_0\|_\infty \|K_j * f\|_2^2 \le C\|p_0\|_\infty 2^{j(1-2/r)} = \sigma^2.$$

If $1 < r < 2$, then $p_0 \in L^\infty \cap L^1 \subset L^{s/(s-2)}$, so by Hölder's inequality

$$E(K_j * f)(X)^2 \le \|K_j * f\|_s^2 \|p_0\|_{s/(s-2)} \le C(p_0)\|K_j\|_1^2 \|f\|_s^2 \le C(p_0, K).$$

For Condition 1(b), so in the multiresolution case for $T = \mathbb{R}$, the arguments as in (a) and obvious modifications give the same bounds for U, σ in view of the estimate $|\int_{\mathbb{R}} K_j(x, y) f(y)\,dy| \le \Phi_j * |f|(x)$, which allows us to compare wavelet projections to convolutions and proceed as above.

For Condition 1(c), note that, by the comments following the statement of Condition 1, the projection kernels have the form $K_j = K_j^0 + K_j^1 + \tilde{K}_j$ where $\tilde{K}_j(x, t) = 2^j \tilde{K}(2^j t, 2^j x)$ with $\tilde{K}$ majorized by a convolution kernel. Therefore the envelope and variance bounds for the previous two cases apply as well to this "interior part" of the kernel. For the boundary part,

$$K_j^i(x, t) = \sum_{k=0}^{N-1} 2^j \phi_k^i(2^j x)\phi_k^i(2^j t), \qquad i = 0, 1,\ j \ge J_0, \tag{23}$$

with N finite and ϕ_k^i bounded and with bounded support, it is immediate to check, just using Hölder's inequality, that for $f \in B_0$,

$$\left\| 2^j \phi_k^i(2^j x) \int_0^1 \phi_k^i(2^j t) f(t)\,dt \right\|_\infty \le \|\phi_k^i\|_\infty \|\phi_k^i\|_r 2^{j(1-1/r)}, \qquad 1 \le r \le \infty,$$

and that

$$2^{2j}E(\phi_k^i(2^jX))^2\left(\int_0^1 |\phi_k^i(2^jt)||f(t)|\,dt\right)^2 \le \|p_0\|_r\|\phi_k^i\|_{2s}^2\|\phi_k^i\|_r^2 2^{j(1-1/r)}$$

for $p_0 \in L^r$, with the refinement $\|p_0\|_\infty\|\phi_k^i\|_2^2\|\phi_k^i\|_r^2 2^{j(1-2/\bar{r})}$ if $\|p_0\|_\infty < \infty$. This shows that the bounds for U, σ^2 from (a), (b) apply to (c) as well.

3.1.2. *Application of Talagrand's inequality.* To apply Talagrand's inequality we need a bound on the moment of the supremum of the empirical process involved, provided in the following lemma, known for the cases $r = \infty$ (see [14, 17, 25]) and, implicitly, $1 \le r \le 2$ (see [15]). As the proof is standard but somewhat lengthy it is given in the supplementary file for this paper, [18].

LEMMA 1. *Assume Condition* 1(a), (b) *or* (c) *and that* $p_0 \in L^r(T)$. *If* $1 \le r < 2$ *in the cases* (a) *or* (b), *assume further that* $p_0 \in L^1(\mu_s)$ *for some* $s > (2-r)/r$. *Then, if* $1 \le r < \infty$, *there exists* L_r *such that, for all* $j \ge 0$ *if* $r \le 2$, *and for all* j *such that* $2^j < n$ *for* $r > 2$, *we have*

$$E\|n(P_n - P_0)\|_{\mathcal{K}} = E\left\|\sum_{i=1}^n \bigl(K_j(\cdot, X_i) - EK_j(\cdot, X)\bigr)\right\|_r \le L_r\sqrt{2^j n}. \tag{24}$$

If $r = \infty$, *for* p_0 *and* Φ *bounded, there exists a constant* L_∞ *such that for all* j *satisfying* $2^j j < n$ *we have*

$$E\|n(P_n - P_0)\|_{\mathcal{K}} = E\left\|\sum_{i=1}^n \bigl(K_j(\cdot, X_i) - EK_j(\cdot, X)\bigr)\right\|_\infty \le L_\infty\sqrt{2^j jn}. \tag{25}$$

We are now ready to apply (19): for $V = n\sigma^2 + 2UE\|\hat{p}_n(j) - E\hat{p}_n(j)\|_r$ we have the bound

$$\Pr\left\{n\|\hat{p}_n(j) - P_0^n\hat{p}_n(j)\|_r \ge nE\|\hat{p}_n(j) - P_0^n\hat{p}_n(j)\|_r + \sqrt{2Vx} + \frac{Ux}{3}\right\} \le 2e^{-x}.$$

This can be further simplified, using the standard inequalities $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$, $\sqrt{ab} \le (a+b)/2$, to

$$\begin{aligned}&\Pr\{n\|\hat{p}_n(j) - P_0^n\hat{p}_n(j)\|_r \ge \tfrac{3}{2}nE\|\hat{p}_n(j) - P_0^n\hat{p}_n(j)\|_r + \sqrt{2n\sigma^2x} + \tfrac{7}{3}Ux\}\\ &\qquad \le 2e^{-x}.\end{aligned}$$

Combining the moment estimate Lemma 1 with (20) and (21), we obtain, for $2^j j(r) < n$ with $j(\infty) = j$ and $j(r) = 1$ for $r < \infty$,

$$\begin{aligned}&\Pr\{n\|\hat{p}_n(j) - P_0^n\hat{p}_n(j)\|_r\\ &\qquad \ge C\bigl(\sqrt{2^j nj(r)} + \sqrt{n2^{j(1-1/r)}\|p_0\|_r x} + 2^{j(1-1/r)}x\bigr)\} \le 2e^{-x}\end{aligned} \tag{26}$$

for some constant C, and in the case where $\|p_0\|_\infty < \infty$ we have, analogously, from (22),

$$\Pr\{n\|\hat{p}_n(j) - P_0^n \hat{p}_n(j)\|_r \geq C\big(\sqrt{2^j n j(r)} + \sqrt{n 2^{j(1-2/\bar{r})}\|p_0\|_\infty x} + 2^{j(1-1/r)}x\big)\} \leq 2e^{-x}. \tag{27}$$

If we take $\varepsilon_n, \delta_n, 2^{j_n} \sim n\varepsilon_n^2$ as in Theorems 2, 3, and if $\|p_0\|_r$ is bounded by a fixed constant B, then the choice $x = Ln\varepsilon_n^2$ gives for every L and $M = M(L, K, B)$ large enough, after some simple computations using the conditions on ε_n, δ_n from the theorem, that

$$nM\delta_n \geq C\big(\sqrt{2^{j_n} j_n(r) n} + \sqrt{\|p_0\|_r n 2^{j_n(1-1/r)} Ln\varepsilon_n^2} + 2^{j_n(1-1/r)} Ln\varepsilon_n^2\big)$$

and, likewise, if $\|p_0\|_\infty$ is bounded by a fixed constant, the corresponding choice of δ_n, M also satisfies

$$nM\delta_n \geq C\big(\sqrt{2^{j_n} j_n(r) n} + \sqrt{C(p_0) n 2^{j_n(1-2/\bar{r})} Ln\varepsilon_n^2} + 2^{j_n(1-1/r)} Ln\varepsilon_n^2\big).$$

Moreover for $\|p_0\|_r \geq \zeta > 0$ we have

$$n\|p_0\|_r \geq C\big(\sqrt{2^{j_n} j_n(r) n} + \sqrt{\|p_0\|_r n 2^{j_n(1-1/r)} Ln\varepsilon_n^2} + 2^{j_n(1-1/r)} Ln\varepsilon_n^2\big)$$

from some index n_0 onwards that depends only on C, ζ.

Using these inequalities in (26), (27), we conclude that in both cases, for every $0 < L < \infty$ we can find a large enough $M(L, K, B)$ such that

$$\Pr\{n\|\hat{p}_n(j_n) - P_0^n \hat{p}_n(j_n)\|_r \geq Mn\delta_n\} \leq 2e^{-Ln\varepsilon_n^2} \tag{28}$$

and, likewise, for n large enough,

$$\Pr\{n\|\hat{p}_n(j_n) - P_0^n \hat{p}_n(j_n)\|_r \geq n\|p_0\|_r/3\} \leq 2e^{-Ln\varepsilon_n^2}. \tag{29}$$

3.2. *Proof of Theorems* 2 *and* 3. Using the small ball estimate from condition (2), it suffices to construct tests (indicator functions) $\phi_n = \phi_n(X_1, \dots, X_n; p_0)$ such that

$$P_0^n \phi_n \to 0 \qquad \text{as } n \to \infty \quad \text{and} \qquad \sup_{p \in \mathcal{P}_n : \|p - p_0\|_r \geq M\delta_n} P^n(1 - \phi_n) \leq 2e^{-(C+4)n\varepsilon_n^2} \tag{30}$$

for n large enough; see the proof of Theorem 2.1 in [10].

Consider first Theorem 2. Let $\hat{p}_n$ be a kernel-type density estimator based on an i.i.d. sample $X_1, \dots, X_n$ of common law P_0, $n \in \mathbb{N}$, at resolution J_n. For M_0, a constant to be chosen below, set $T_n = \|\hat{p}_n - p_0\|_r$ and $\phi_n = I(T_n > M_0\delta_n)$. Note that ϕ_n is the (indicator of the) rejection region of a natural test of the hypothesis $H_0 : p = p_0$. Then we have

$$\begin{aligned} P_0^n \phi_n &= P_0^n \{\|\hat{p}_n - p_0\|_r > M_0 \delta_n\} \\ &\le P_0^n \{\|\hat{p}_n - P_0^n \hat{p}_n\|_r > M_0 \delta_n - \|P_0^n \hat{p}_n - p_0\|_r\}. \end{aligned}$$

Since $\|K_{J_n}(p_0) - p_0\|_r \le c' \delta_n$ for some $c' > 0$ by assumption, we have for all n large enough, $P_0^n \phi_n \le P_0^n \{\|\hat{p}_n - P_0^n \hat{p}_n\|_r > (M_0 - c')\delta_n\}$. Then using inequality (28), we have for some constant L_1 for some constant L_1, choosing M_0 large enough, that, as $n \to \infty$,

$$P_0^n \phi_n \le 2e^{-L_1 n \varepsilon_n^2} \to 0. \tag{31}$$

Let now p be a density in $\mathcal{P}_n$ such that $\|p - p_0\|_r \ge M \delta_n$ (the alternatives). Set $dP(x) = p(x)\,dx$. We have, from the triangle inequality,

$$\begin{aligned} P^n(1 - \phi_n) &= P^n \{\|\hat{p}_n - p_0\|_r \le M_0 \delta_n\} \\ &\le P^n \{\|\hat{p}_n - P^n \hat{p}_n\|_r \ge \|p - p_0\|_r - M_0 \delta_n - \|P^n \hat{p}_n - p\|_r\} \\ &\le P^n \bigl\{\|\hat{p}_n - P^n \hat{p}_n\|_r \ge \|p - p_0\|_r - \bigl(M_0 + C(K)\bigr)\delta_n\bigr\} \end{aligned} \tag{32}$$

since by assumption on $\mathcal{P}_n$, $\sup_{p \in \mathcal{P}_n} \|P^n \hat{p}_n - p\|_r \le C(K)\delta_n$, uniformly in $p \in \mathcal{P}_n$.

To complete the estimation of the last probability, we consider first $r > 1$. For those $p \in \mathcal{P}_n$ satisfying $\|p\|_r \ge 2\|p_0\|_r$ we have $\|p - p_0\|_r \ge \|p\|_r/2 \ge \|p_0\|_r$, and, using inequality (29) for $p_0 = p$, we deduce, that for all $L > 0$, there exists $n_0 \in \mathbb{N}$ such that for all $n \ge n_0$,

$$\begin{aligned} &\sup_{p \in \mathcal{P}_n : \|p\|_r \ge 2\|p_0\|_r} P^n(1 - \phi_n) \\ &\quad \le \sup_{p \in \mathcal{P}_n, \|p\|_r \ge 2\|p_0\|_r} P^n \left\{ \|\hat{p}_n - P^n \hat{p}_n\|_r > \frac{\|p\|_r}{3} \right\} \\ &\quad \le 2e^{-L n \varepsilon_n^2}. \end{aligned} \tag{33}$$

For those $p \in \mathcal{P}_n$ for which $\|p\|_r < 2\|p_0\|_r$, we apply (28) with $p = p_0$ and use as well $\|p - p_0\|_r \ge M \delta_n$ to obtain that for all $L > 0$ there exists M large enough such that

$$\begin{aligned} &\sup_{p \in \mathcal{P}_n : \|p\|_r < 2\|p_0\|_r, \|p - p_0\|_r \ge M\delta_n} P^n(1 - \phi_n) \\ &\quad \le \sup_{p \in \mathcal{P}_n : \|p\|_r < 2\|p_0\|_r, \|p - p_0\|_r \ge M\delta_n} P^n \{\|\hat{p}_n - P^n \hat{p}_n\|_r \\ &\qquad\qquad > \bigl(M - M_0 - C(K)\bigr)\delta_n\} \\ &\quad \le 2e^{-L n \varepsilon_n^2}. \end{aligned} \tag{34}$$

We conclude from (32) and (33) that for any $L > 0$ there exists $n_L < \infty$ such that

$$\sup_{p\in\mathcal{P}_n : \|p-p_0\|_r \ge M\delta_n} P^n(1-\phi_n) \le 2e^{-Ln\varepsilon_n^2}. \tag{35}$$

Now (31) and (35) prove (30) if $r > 1$. If $r = 1$ the above case distinction is not necessary as $\|p\|_1 = 1$ always holds, so that the proof of the second case applies with the full supremum over $\{p \in \mathcal{P}_n : \|p - p_0\|_1 \ge M\delta_n\}$. This completes the proof of Theorem 2.

To prove Theorem 3 we argue similarly, and only have to slightly modify the derivation of the error probabilities of the tests: when it is known that the posterior concentrates on a fixed sup-norm ball of radius B, then we can restrict the alternatives in (30) further to densities bounded by B, and, using (28) with $p = p_0$ and the present choice of δ_n, we also obtain

$$\begin{aligned}
&\sup_{p\in\mathcal{P}_n : \|p\|_\infty \le B, \|p-p_0\|_r \ge M\delta_n} P^n(1-\phi_n) \\
&\quad\le \sup_{p\in\mathcal{P}_n : \|p\|_\infty \le B, \|p-p_0\|_r \ge M\delta_n} P^n\{\|\hat{p}_n - P^n\hat{p}_n\|_r > (M - M_0 - C(K))\delta_n\} \\
&\quad\le 2e^{-Ln\varepsilon_n^2}.
\end{aligned}$$

4. Remaining proofs.

4.1. *Proofs of Propositions 1, 2 and 3.*

PROOF OF PROPOSITION 1. Since $\|U_\alpha\|_\infty \le C$ almost surely for some fixed constant $C = C(B, \alpha, \psi)$, we infer $\|p^{U,\alpha}\|_{\alpha,r,\infty} \le D(B, \alpha, \psi)$ almost surely for $1 \le r \le \infty$. In particular the prior is supported in a ball of bounded densities, hence so is the posterior, and we can attempt to apply Theorems 2 (for $r = 1, \infty$) and 3 for $(1 < r < \infty)$, which we shall do with the choice $\varepsilon_n = (n/\log n)^{-\alpha/(2\alpha+1)}$.

We verify the small ball estimate in the second condition in Theorem 2. By Lemma 3.1 in [32] we can lower bound the prior probability in question by $\Pr\{\|\log p_0 - U_\alpha\|_\infty \le c\varepsilon_n\}$ for some constant $c > 0$. Since

$$\|h\|_\infty \le C(\phi, \psi) \max\Bigl(\sup_k |\alpha_k(h)|, \sum_\ell \sup_k 2^{\ell/2}|\beta_{\ell k}(h)|\Bigr)$$

for any continuous function h on [0, 1] and some constant $C(\phi, \psi)$, we can lower bound the last probability, writing $\alpha_k, \beta_{\ell k}$ for the wavelet coefficients of $\log p_0$, by

$$\begin{aligned}
&\Pr\Bigl\{\max\Bigl(\sup_{k=0,\dots,N} |\alpha_k - u_{0k}|, \sum_\ell \sup_k 2^{\ell/2}\bigl|\beta_{\ell k} - 2^{-\ell(\alpha+1/2)}u_{\ell k}\bigr|\Bigr) \le c'\varepsilon_n\Bigr\} \\
&\quad= \Pr\Bigl\{\max_k |\alpha_k - u_{0k}| \le c'\varepsilon_n\Bigr\} \Pr\Bigl\{\sum_{\ell \ge J_0} \max_{k \le 2^\ell} 2^{\ell/2}|\beta_{\ell k} - 2^{-\ell(\alpha+1/2)}u_{\ell k}| \le c'\varepsilon_n\Bigr\},
\end{aligned}$$

where N, J_0 depend only on the wavelet basis (see before Definition 1). Since $|\alpha_k| \le B$ and since the u_{0k} are $U(-B, B)$, the first probability exceeds $(c'\varepsilon_n/2B)^{N+1} = e^{-(N+1)\log(2B/c'\varepsilon_n)}$ which is bounded below by $e^{-c\log(1/\varepsilon_n)}$ for some $c > 0$ that depends only on B, α and the wavelet basis. For the second probability set $b_{\ell k} \equiv 2^{\ell(\alpha+1/2)}\beta_{\ell k}$, $\ell \ge J_0$, and $M(J) \equiv \sum_{\ell=J_0}^{J}\sum_{k=0}^{2^\ell-1} 1 \le 2 \cdot 2^J$, and note that $|b_{\ell k}| \le \|\log p_0\|_{\alpha,\infty} \le B$. Choosing $J = J_n \ge J_0$ large enough and of order $\varepsilon_n \simeq 2^{-J\alpha}$, this probability is bounded below by

$$\begin{aligned}
&\Pr\left\{\sum_{\ell=J_0}^{J} 2^{-\ell\alpha}\sup_k|b_{\ell k} - u_{\ell k}| \le c'\varepsilon_n - C(\psi, B)2^{-J\alpha}\right\}\\
&\quad \ge \Pr\Bigl\{\max_{\ell\le J}\max_{k\le 2^\ell}|b_{\ell k} - u_{\ell k}| \le c''\varepsilon_n\Bigr\}\\
&\quad = \prod_{\ell\le J}\prod_{k\le 2^\ell}\Pr\{|b_{\ell k} - u_{\ell k}| \le c''\varepsilon_n\} \ge \left(\frac{c''\varepsilon_n}{2B}\right)^{M(J)}\\
&\quad \ge e^{-c'''\log(1/\varepsilon_n)/\varepsilon_n^{1/\alpha}}
\end{aligned}$$

for n large enough and some $c''' > 0$ that depends only on B, α and the wavelet basis. Summarizing we have, by definition of ε_n, that the Π^α probability in condition (2) of Theorem 2 is bounded from below by

$$\Pr\{\|\log p_0 - U_\alpha\|_\infty \le c\varepsilon_n\} \ge e^{-c\log(1/\varepsilon_n)}e^{-c'''\log(1/\varepsilon_n)/\varepsilon_n^{1/\alpha}} \ge e^{-Cn\varepsilon_n^2} \tag{36}$$

for some C that depends only on B, α and the wavelet basis, which proves that condition (2) holds.

We next verify the bias condition with $\mathcal{P}_n = \operatorname{supp}(\Pi)$ so that $\Pi(\mathcal{P}\setminus\mathcal{P}_n) = 0$. We bound the L^r-norm of the approximation errors of any element in $\mathcal{P}_n$ by a constant times δ_n, where we take γ_n equal to $\log n$ to a sufficiently large power chosen below. Since $2^{J_n} \ge cn\varepsilon_n^2 \ge cn^{1/(2\alpha+1)}$ we have, using $B^0_{r1}([0,1]) \subset L^r([0,1])$ and $p \in \mathcal{C}^\alpha([0,1])$,

$$\|K_{J_n}(p) - p\|_r \le c\sum_{\ell=J_n}^{\infty} 2^{\ell(1/2-1/r)}\left(\sum_{k=1}^{2^j}|\beta_{\ell k}(p)|^r\right)^{1/r} \le c'(B,r)\sum_{\ell=J_n}^{\infty} 2^{-\ell\alpha},$$

which is $O(\varepsilon_n)$, so the bias condition is satisfied for some $C(K)$ large enough, both for $\mathcal{P}_n$, as well as for p_0.

Finally condition (c) from Theorem 2 and (a), (b) from Theorem 3, as well as $\delta_n \to 0$, are verified for this choice of ε_n and under the conditions on α, r, except for the cases $\alpha = 0$ or $\alpha = 1/2, r = \infty$, where the result trivially follows from δ_n being bounded from below by a constant multiple of $\log n$ (and as the prior is supported in a L^r-bounded set). $\square$

PROOF OF PROPOSITION 2. We apply Theorem 2 with $r = \infty$. We have from the proof of Theorem 5.1 in [13] that for $\varepsilon_n = (\log n)^\kappa/\sqrt{n}$, $\kappa \geq 1$, the small-ball estimate in condition (2) of Theorem 2 is satisfied. Choose γ_n in such a way that δ_n equals $(\log n)^\eta/\sqrt{n}$ where $\eta > \kappa$. For the bias, we take $\mathcal{P}_n$ to be the support of Π and consider a Meyer-wavelet basis and the wavelet projection onto it, with $2^{J_n} = c(\log n)^{2\kappa}$, where c is a large enough constant that depends on $\inf\{\sigma : \sigma \in \operatorname{supp}(G)\}$, and apply Proposition 4 in [25] with $s = 2$ and suitable $\tilde{c}_0$, to see that $\|K_{J_n}(p_{F,\sigma}) - p_{F,\sigma}\|_\infty = o(1/n)$ uniformly in the support of Π. A more detailed proof is in the supplementary file [18]. $\square$

PROOF OF PROPOSITION 3. Taking $\varepsilon_n = M'(n/\log n)^{-\alpha/(2\alpha+1)}$, and noting $\varepsilon_n^{-1/\alpha} = O(n\varepsilon_n^2)$, we can take J_n such that $2^{j_n} \leq 2^{J_n} \leq cn\varepsilon_n^2$ for every n, some $c > 0$. Taking $K(x, y)$ equal to the Haar wavelet projection kernel (CDV-wavelet of regularity $S = 0$), we conclude that $\|K_{J_n}(p) - p\|_r = 0$ Π_{j_n}-a.s. $\forall n$, so condition (1) in Theorem 2 is satisfied with $\mathcal{P}_n$ equal to the support of Π_{j_n}. The small ball estimate (2) follows, as in the proof of Theorem 1 ([29], pages 636 and 637, with $k_0 = 2^{j_n}$, and approximating p_0 by $K_{j_n}(p_0)$ s.t. $\|K_{j_n}(p_0) - p_0\|_1 \leq \varepsilon_n/2$ for M' large enough), and from the second inequality in (36). The bias condition for p_0 is satisfied by standard approximation properties of Haar wavelets. The result now follows from first applying Theorem 2 with $r = 1, \infty$ and then using the conclusion that the posterior concentrates on a $\|\cdot\|_\infty$ neighborhood of p_0 to invoke Theorem 3 for the cases $1 < r < \infty$. $\square$

4.2. *Proof of Proposition* 4. We shall construct subsets of $\mathcal{P}$ on which we can control the approximation errors from (15). We define Hölder spaces. For $\alpha, \tau \geq 0$ positive real numbers, define the norm $\|f\|_{\alpha,\infty,\tau} := \sum_{k=0}^{[\alpha]} \|f^{(k)}\|_\infty + H(\alpha, \tau, f)$ where

$$H(\alpha, \tau, f) = \sup_{0<t<1} \frac{\sup_{h:|h|\leq t, x+h\in[0,1]} \sup_{x\in[0,1]} |f^{(k)}(x+h) - f^{(k)}(x)|}{t^{\{\alpha\}}(\log t^{-1})^\tau},$$

and where we take $\|f^{(k)}\|_\infty = \infty$ if $f^{(k)}$ does not exist. Define, moreover, $C^{\alpha,\infty,\tau}([0, 1]) := \{f : [0, 1] \to \mathbb{R} : \|f\|_{\alpha,\infty,\tau} < \infty\}$. The case $\tau = 0$ specialises to the strict α-Hölder case $C^\alpha([0, 1])$.

In case $1 \leq r < \infty$, we shall use approximation theoretic properties of the reproducing kernel Hilbert spaces (RKHSs) of $B_\alpha, \bar{B}_\alpha$, which are Sobolev spaces. Recall that the RKHS $\mathbb{H}(1/2)$ of Brownian motion on $[0, 1]$ is the space of absolutely continuous functions that are zero at zero and whose first derivatives are in $L^2([0, 1])$, equipped with the inner product $\langle f, g\rangle_{\mathbb{H}(1/2)} = \int_0^1 f'g'$. Then, the RKHS of integrated Brownian motion B_α is

$$\mathbb{H}(\alpha) = \left\{\int_0^t \int_0^{[\alpha]-1} \cdots \int_0^{t_1} f(s)\, ds\, dt_1 \cdots dt_{[\alpha]-1} : f \in \mathbb{H}(1/2)\right\}$$

with inner product $\langle f, g\rangle_{\mathbb{H}(\alpha)} = \int_0^1 f^{([\alpha]+1)} g^{([\alpha]+1)}$. Finally, $f \in \bar{\mathbb{H}}(\alpha)$, the RKHS of $\bar{B}_\alpha$, iff $f = P_{[\alpha]} + g$ where P_α is a polynomial of degree $[\alpha]$ and $g \in \mathbb{H}(\alpha)$, and note that $P_{[\alpha]}(t) = \sum_{i=0}^{[\alpha]} f^{(i)}(0)t^i/i!$; the inner product in $\bar{\mathbb{H}}(\alpha)$ is $\langle f, g\rangle_{\bar{\mathbb{H}}(\alpha)} = \sum_{i=0}^{[\alpha]} f^{(i)}(0)g^{(i)}(0) + \int_0^1 f^{([\alpha]+1)} g^{([\alpha]+1)}$; see, for example, [33]. The spaces $\bar{\mathbb{H}}(\alpha)$ are precisely the Sobolev spaces $H^{\alpha+1/2}$, and other equivalent norms may be used below.

We will also require the following definition. For a B-valued Gaussian random vector W, B a Banach space, and for $w \in B$, the "concentration function" $\phi_w^W(\varepsilon)$ of W at w is defined as

$$e^{-\phi_w^W(\varepsilon)} = \Pr\{\|W - w\| < \varepsilon\}. \tag{37}$$

The following result is a consequence of Borell's isoperimetric inequality [4], and is essentially contained in the proof of Theorem 2.1 in [32].

PROPOSITION 5. *Let $\alpha \in \{n - 1/2 : n \in \mathbb{N}\}$, denote by $\bar{\mathbb{H}}_1(\alpha)$ the unit ball of $\bar{\mathbb{H}}(\alpha)$ and let $B^1 = \{f \in C([0,1]) : \|f\|_\infty \le 1\}$. Let ε_n satisfy $\phi_0^{\bar{B}_\alpha}(\varepsilon_n) \le n\varepsilon_n^2$ for all n. Then the released integrated Brownian motion process $\bar{B}_\alpha$ has a version, that we continue denoting by $\bar{B}_\alpha$, such that for every $C > 0$, $D > 0$,*

$$\Pr\{\bar{B}_\alpha \notin M_n\bar{\mathbb{H}}_1(\alpha) + \varepsilon_n B^1\} \le De^{-(C+4)n\varepsilon_n^2},$$

where $M_n = M_n(C, D) = -2\Phi^{-1}(De^{-(C+4)n\varepsilon_n^2}) \simeq \sqrt{n}\varepsilon_n$ and Φ is the standard normal distribution function.

PROOF. Borell's inequality (e.g., Theorem 4.3.3 in [3]) implies

$$\Pr\{\bar{B}_\alpha \notin M_n\bar{\mathbb{H}}_1(\alpha) + \varepsilon_n B^1\} \le 1 - \Phi(a_n + M_n), \tag{38}$$

where a_n solves the equation $\Phi(a_n) = \Pr\{\|\bar{B}_\alpha\|_\infty \le \varepsilon_n\} \ge e^{-n\varepsilon_n^2}$. It then follows ($C + 4 > 1$) that $a_n \ge -M_n/2$, which implies

$$1 - \Phi(a_n + M_n) \le \Phi(-M_n/2) = De^{-(C+4)n\varepsilon_n^2}. \qquad \square$$

In particular, taking $D = \Pr\{\|\bar{B}_\alpha\|_\infty \le c\}$ for any $c > 0$, this proposition gives

$$\Pr\{\bar{B}_\alpha \notin M_n\bar{\mathbb{H}}_1(\alpha) + \varepsilon_n B^1 \mid \|\bar{B}_\alpha\|_\infty \le c\} \le e^{-(C+4)n\varepsilon_n^2} \tag{39}$$

with M_n depending on C and c, and of the order $\sqrt{n}\varepsilon_n$.

In case $r = \infty$ we need a different result that reflects the almost sure Hölder regularity of the trajectories of B_α.

PROPOSITION 6. *For all $\alpha \in \{n - 1/2 : n \in \mathbb{N}\}$, integrated Brownian motion has a version, that we continue denoting by B_α, with almost all its sample paths in $C^{\alpha,\infty,1/2}([0,1])$ and for every $D > 0$ there exist $t_\alpha < \infty$ and $L_\alpha < \infty$ such that*

$$\Pr\{\|B_\alpha\|_{\alpha,\infty,1/2} \ge t\} \le De^{-L_\alpha t^2}, \qquad t \ge t_\alpha. \tag{40}$$

The same is true for the processes $\bar{B}_\alpha = \sum_{k=0}^{[\alpha]+1} Z_k t^k/k! + B_\alpha$, *that is*,

$$\Pr\{\|\bar{B}_\alpha\|_{\alpha,\infty,1/2} \geq t\} \leq De^{-L_\alpha t^2}, \qquad t \geq t_\alpha, \tag{41}$$

for possibly different $L_\alpha(D)$ *and* $t_\alpha(D)$, *for all* $D > 0$.

PROOF. By a classical result of Lévy (see also Theorem IV.5 in [7]) Brownian motion $B_{1/2}$ has a version in $C^{1/2,\infty,1/2}([0,1])$. Since, for $\alpha > 1$, by the definitions,

$$\|B_\alpha\|_{\alpha,\infty,1/2} = \|B_\alpha\|_\infty + \|B'_\alpha\|_{\alpha-1,\infty,1/2} = \|B_\alpha\|_\infty + \|B_{\alpha-1}\|_{\alpha-1,\infty,1/2},$$

and $\|B_\alpha\|_\infty < \infty$ a.s., induction extends the result to all $\alpha \in \{n - 1/2 : n \in \mathbb{N}\}$.

For $0 < \alpha < 1$, Theorem III.6 in [7] shows that the norms $\|f\|_{\alpha,\infty,1/2}$ and $\|f\|^{(d)}_{\alpha,\infty,1/2}$ are equivalent, where $\|f\|^{(d)}_{\alpha,\infty,1/2}$ is defined as

$$\begin{aligned} \|f\|^{(d)}_{\alpha,\infty,1/2} &:= \|(y_i^f, y_{j,k}^f)\|_{\alpha,\infty,\infty} \\ &= \sup\left\{|y_0^f|, |y_1^f|, \max_{k,j} \frac{2^{\alpha j}}{\sqrt{j \log 2}}|y_{j,k}^f|\right\} \end{aligned} \tag{42}$$

with

$$\begin{aligned} y_0^f &= f(0), \\ y_1^f &= 3^{-1/2}\big(f(1) - f(0)\big), \\ y_{j,k}^f &= (3 \cdot 2^J)^{-1/2}\left[f\left(\frac{2k-1}{2^{j+1}}\right) - \frac{1}{2}\left(f\left(\frac{k}{2^j}\right) + f\left(\frac{k-1}{2^j}\right)\right)\right] \end{aligned} \tag{43}$$

for $k = 1, \dots, 2^j$, $j = 0, 1, \dots$. Obviously, $\|\cdot\|^{(d)}_{\alpha,\infty,1/2}$ is a supremum norm on a sequence space; more specifically, it is the sup of the absolute values of a countable number of linear functionals on the space $C^{\alpha,\infty,1/2}([0,1])$ (linear combinations of point evaluations). Hence Lemma 3.1 and inequality (3.2) in [22] (this last inequality even with $\pi^2/2$ replaced by 2) apply to $\|B_\alpha\|_{\alpha,\infty,1/2}$, giving (40) for $D = 1$. For $D < 1$, take $t'_\alpha \geq t_\alpha$ such that $D \geq e^{-(L_\alpha/2)(t'_\alpha)^2}$ and $L'_\alpha = L_\alpha/2$. If $\alpha > 1$, then the result follows by applying these inequalities to the $C^{\{\alpha\},\infty,1/2}$-norm of the $[\alpha]$th derivative of the process and to the sup norms of the process and of its derivatives of order smaller than $[\alpha]$. Since (40) is obviously true for the processes $Z_k t^k$, it is true as well for $\bar{B}_\alpha$ possibly with a different constant, which gives (41). □

Again, taking $D = \Pr\{\|\bar{B}_\alpha\|_\infty \leq c\}$, for any $c > 0$, this proposition gives

$$\Pr\{\|\bar{B}_\alpha\|_{\alpha,\infty,1/2} \geq t \,|\, \|\bar{B}_\alpha\|_\infty \leq c\} \leq e^{-L_\alpha t^2}, \qquad t \geq t_\alpha, \tag{44}$$

L_α and t_α depending on c.

These two consequences of Borell's inequality imply that the integrated Brownian motions concentrate on suitable subsets of $C([0,1])$, and the following lemma achieves the same for the normalized trajectories of the processes $e^{\bar{B}_\alpha(t,\omega)}$.

LEMMA 2. *Let $\alpha \in \{n - 1/2 : n \in \mathbb{N}\}$, and let K_j be a CDV-projection kernel of regularity $\alpha + 1/2$, at resolution $j \geq 0$.*

(1) (*Case $1 \leq r < \infty$.*) *Let $f \in \{M_n \bar{\mathbb{H}}_1(\alpha) + \varepsilon_n B^1, \|f\|_\infty \leq c\}$, where $\bar{\mathbb{H}}_1(\alpha)$ is the unit ball of the RKHS of $\bar{B}_\alpha$ and set $p = e^f / \int_0^1 e^f$. Then, for $\bar{r} = \max(2, r)$ and some $C > 0$,*

$$\|K_j(p) - p\|_r \leq C\big(M_n 2^{-j(\alpha+1/\bar{r})} + \varepsilon_n\big).$$

(2) (*Case $r = \infty$.*) *Let f satisfy $\|f\|_\infty \leq c$ and $\|f\|_{\alpha,\infty,1/2} \leq L\sqrt{n}\varepsilon_n$, and let p be as above. Then, for some $C > 0$,*

$$\|K_j(p) - p\|_\infty \leq C\sqrt{n}\varepsilon_n 2^{-j\alpha}\sqrt{j}.$$

PROOF. We first consider $1 \leq r < \infty$. Since $\|f\|_\infty \leq c$ we have $e^{-c} \leq \int_0^1 e^f \leq e^c$ so, $\int K_j(x,y)(\cdot)(y)\,dy$ being a linear operator, it suffices to bound $\|K_j(e^f) - e^f\|_r$. Writing $f = f_1 + f_2$ with $f_1 \in M_n \bar{\mathbb{H}}_1(\alpha)$ and $f_2 \in \varepsilon_n B^1$, we see that $\|f_2\|_\infty \leq \varepsilon_n < c$, $\|f_1\|_\infty \leq c + \varepsilon_n < 2c$, and in particular, $|e^{f_2(x)} - e^{f_2(y)}| \leq e^c |f_2(x) - f_2(y)|$. Note also that, for some constant $C(K) < \infty$, $\|2^{-j}K_j(x, x + 2^{-j}\cdot)\|_1 \leq C(K)$. Then we have

$$\begin{aligned}
&|K_j(e^f) - e^f|(x) \\
&\quad = \left| \int 2^{-j} K_j(x, x + 2^{-j}u)\big(e^{(f_1+f_2)(x+2^{-j}u)} - e^{(f_1+f_2)(x)}\big)\,du \right| \\
&\quad \leq \left| e^{f_2(x)} \int 2^{-j} K_j(x, x + 2^{-j}u)\big(e^{f_1(x+2^{-j}u)} - e^{f_1(x)}\big)\,du \right| \\
&\qquad + \left| \int 2^{-j} K_j(x, x + 2^{-j}u) e^{f_1(x+u2^{-j})}\big(e^{f_2(x+2^{-j}u)} - e^{f_2(x)}\big)\,du \right| \\
&\quad \leq e^c |K_j(e^{f_1})(x) - e^{f_1}(x)| + 2e^{3c} \sup_x \|2^{-j} K_j(x, x + 2^{-j}\cdot)\|_1 \varepsilon_n.
\end{aligned}$$

The $L^r([0,1])$-norm of the second term is bounded by a fixed constant times ε_n, and it remains to control the $L^r([0,1])$-norm of the first term in the bound. Note that the Sobolev space $\bar{\mathbb{H}}(\alpha) = H^{\alpha+1/2}$ is contained in the Besov space $B_{22}^{\alpha+1/2}([0,1])$, which itself is continuously imbedded into the Besov space $B_{\bar{r}2}^{\alpha+1/2-1/2+1/\bar{r}}([0,1]) = B_{\bar{r}2}^{\alpha+1/\bar{r}}([0,1])$; cf. Remark 1. We conclude, for some constant C', that $\|K_j(e^{f_1}) - e^{f_1}\|_r \leq C' \|f_1\|_{\bar{\mathbb{H}}(\alpha)} 2^{-j(\alpha+1/\bar{r})}$ from the approximation properties of wavelet projections on Besov spaces (Definition 1). This establishes the bound in the first part of the lemma.

For the case $r=\infty$, note that, f being bounded by c, the chain rule gives that there exists $C(c,\alpha)$ such that

$$\|e^f\|_{\alpha,\infty,1/2} \le C(c,\alpha)(\|f\|_{\alpha,\infty,1/2}+1). \tag{45}$$

We conclude from a standard bias bound for wavelet projections that $\|K_j(e^f)-e^f\|_\infty \le c(\|f\|_{\alpha,\infty,1/2}+1)2^{-j\alpha}\sqrt{j}$ which, in view of $e^{-c}\le \int_0^1 e^f \le e^c$ gives the overall inequality. $\square$

The choice $j=J_n$ with $2^{J_n}\sim n\varepsilon_n^2$, relevant in Theorems 2 and 3, gives, for p satisfying the hypotheses of the previous proposition, the bounds

$$\|K_{J_n}(p)-p\|_r \le C\big((n\varepsilon_n^2)^{-\alpha}+\varepsilon_n\big) \qquad \text{for } 1\le r\le 2 \tag{46}$$

and

$$\|K_{J_n}(p)-p\|_r \le C\big(\sqrt{n}\varepsilon_n(n\varepsilon_n^2)^{-(\alpha+1/r)}+\varepsilon_n\big) \qquad \text{for } 2<r<\infty \tag{47}$$

as well as

$$\|K_{J_n}(p)-p\|_\infty \le C\sqrt{n}\varepsilon_n(n\varepsilon_n^2)^{-\alpha}\sqrt{\log(n\varepsilon_n^2)}. \tag{48}$$

The last auxiliary fact that we will require about B_α is a small ball probability estimate, concretely an upper bound for the concentration function $\phi_w^{\bar B_\alpha}(\varepsilon)$ as ε approaches zero.

PROPOSITION 7. *Let $B_\alpha, \alpha\in\{n-1/2 : n\in\mathbb{N}\}$ be integrated Brownian motion, considered as a Gaussian vector taking values in the Banach space $C([0,1])$, and let $w\in C^\alpha([0,1])$. Then, $\phi_w^{\bar B_\alpha}(\varepsilon)=O(\varepsilon^{-1/\alpha})$, and the same is true for $\phi_w^{B_\alpha}$ if we further assume $w^{(k)}(0)=0$, $k\le[\alpha]$.*

PROOF. Since $B_\alpha = W_{2\alpha}$ in [24] and it also equals a constant times R_α in [32], this proposition simply combines Theorem 2.1 in [24] and Theorem 4.3 in [32]. $\square$

This result applies to the "conditional" concentration function: if $\|w_0\|_\infty\le c/2$ and $\varepsilon\le c/2$, then

$$\begin{aligned}
&\Pr\{\|\bar B_\alpha - w_0\|_\infty<\varepsilon \mid \|\bar B_\alpha\|_\infty\le c\}\\
&\quad=\frac{\Pr\{\|\bar B_\alpha - w_0\|_\infty<\varepsilon, \|\bar B_\alpha\|_\infty\le c\}}{\Pr\{\|\bar B_\alpha\|_\infty\le c\}}\\
&\quad=\frac{e^{-\phi_{w_0}^{\bar B_\alpha}(\varepsilon)}}{\Pr\{\|\bar B_\alpha\|_\infty\le c\}}.
\end{aligned} \tag{49}$$

We are now in a position to apply Theorems 2 and 3 to prove Proposition 4. To ease notation define $I(w)=e^w/\int_0^1 e^{w(t)}\,dt$, $w\in C([0,1])$, and record that, for

$\|w\|_\infty \leq c$,

$$|I(w)| \leq L(|w|+1), \tag{50}$$

where L depends only on c.

Set $w_0 = \log p_0$, so that, since $\|w_0\|_\infty \leq c/2$ and p_0 is a density, hence $p_0 = I(w_0)$, Lemma 3.1 in [32] gives that if $p = I(w)$ for $w = \bar{B}_\alpha(\omega)$ for some $\omega \in \Omega$, and $\|w\|_\infty \leq c$, then $-P_0 \log \frac{p}{p_0} \leq R\|w - w_0\|_\infty^2$ and $P_0(\log \frac{p}{p_0})^2 \leq R\|w - w_0\|_\infty^2$ for some $R < \infty$ (that depends on c). Hence, for any $\varepsilon > 0$ such that $R^{-1/2}\varepsilon < c/2$,

$$\begin{aligned} &\Pi\left\{p \in \mathcal{P} : -P_0 \log \frac{p}{p_0} \leq \varepsilon^2, P_0\left(\log \frac{p}{p_0}\right)^2 \leq \varepsilon^2\right\} \\ &\qquad \geq \Pr\{\|\bar{B}_\alpha - w_0\|_\infty \leq R^{-1/2}\varepsilon \,|\, \|\bar{B}_\alpha\|_\infty \leq c\}. \end{aligned} \tag{51}$$

Since w_0 is in $C^\alpha([0,1])$, it follows from Proposition 7 that $\phi_{w_0}^{\bar{B}_\alpha}(\varepsilon) = O(\varepsilon^{-1/\alpha})$ as $\varepsilon \to 0$, say, there exist c_1 large enough and $\varepsilon_1 > 0$ such that

$$\phi_{w_0}^{\bar{B}_\alpha}(\varepsilon) \leq c_1 \varepsilon^{-1/\alpha} \qquad \text{for all } \varepsilon \leq \varepsilon_1.$$

Then we have, for $\varepsilon_n = (c_1/n)^{\alpha/(2\alpha+1)}$, from some n on, both

$$\phi_{w_0}^{\bar{B}_\alpha}(R^{-1/2}\varepsilon_n) \leq c_1 R^{1/(2\alpha)} \varepsilon_n^{-1/\alpha} \quad \text{and} \quad \phi_{w_0}^{\bar{B}_\alpha}(\varepsilon_n) \leq n\varepsilon_n^2.$$

Hence, for these n, by (49),

$$\Pr\{\|\bar{B}_\alpha - w_0\|_\infty \leq R^{-1/2}\varepsilon_n \,|\, \|\bar{B}_\alpha\|_\infty \leq c\} \geq e^{-Cn\varepsilon_n^2}, \tag{52}$$

where $C = c_1 R^{1/(2\alpha)}$. This proves condition (2) in Theorems 2, 3 for these C, ε_n.

To proceed with the verification of the conditions of Theorem 2, take $\mathcal{P}_n = \{I(w) : w \in \{M_n \bar{\mathbb{H}}_1(\alpha) + \varepsilon_n B^1\}\}$ if $r < \infty$ and $\mathcal{P}_n = \{I(w) : \|w\|_{\alpha,\infty,1/2} \leq \sqrt{(C+4)/L_\alpha}\sqrt{n}\varepsilon_n\}$ if $r = \infty$, and note that condition (1) in Theorem 2 is satisfied for these choices in view of Propositions 5 and 6; see (39) and (44). The bias condition is satisfied for the above choice of ε_n, $\gamma_n = 1$ if $r < \infty$ and $\gamma_n = \sqrt{\log n}$ if $r = \infty$, in view of Lemma 2; cf. also (46), (47), (48). Finally the additional restrictions on ε_n in Theorems 2 and 3 are also satisfied, unless $\alpha = 1/2, r = \infty$. In this case the rate of contraction δ_n exceeds a constant multiple times $\sqrt{\log n}$, so that the result follows trivially from the fact that the prior is supported in a sup-norm bounded set.

4.3. *Proof of Theorem* 1. Observing $Y^{(n)}$ is equivalent to observing its action, on the basis,

$$\begin{aligned} y_k &= \int_0^1 \phi_k(t)\, dY^{(n)}(t) = \langle f, \phi_k\rangle + \frac{1}{\sqrt{n}} \int_0^1 \phi_k(t)\, dB(t) \\ &:= \theta_k + \frac{1}{\sqrt{n}} g_k, \qquad k = 0, \ldots, N-1, \end{aligned} \tag{53}$$

$$
\begin{aligned}
y_{\ell k} &= \int_0^1 \psi_{\ell k}(t)\, dY^{(n)}(t) \\
&= \langle f, \psi_{\ell k}\rangle + \frac{1}{\sqrt{n}} \int_0^1 \psi_{\ell k}(t)\, dB(t) \qquad (54)\\
&:= \theta_{\ell k} + \frac{1}{\sqrt{n}} g_{\ell k}, \qquad k = 0, \dots, 2^\ell - 1, \ell \ge J_0,
\end{aligned}
$$

with the variables g_k, $g_{\ell k}$ all i.i.d. $N(0,1)$. The observed process, still denoted by $Y^{(n)}$, can thus be viewed as a random element $Y^{(n)} = (y_k, y_{\ell k})^t$ of ℓ_2, where y_k is $N(\theta_k, 1/n)$, and $y_{\ell k}$ is $N(\theta_{\ell k}, 1/n)$, all independent. Likewise the function f_0 to be estimated becomes the vector $\theta_0 = (\theta_k^0, \theta_{\ell k}^0)^t$ of the coefficients of its wavelet expansion, that is, $\theta_k^0 = \langle f_0, \phi_k\rangle$ and $\theta_{\ell k}^0 = \langle f_0, \phi_{\ell k}\rangle$, and any prior Π on L_2 maps onto a prior, still denoted by Π, on the parameter space $\theta = (\theta_k, \theta_{\ell k})^t \in \ell_2$.

The posterior $\Pi(\cdot|Y^{(n)})$ is then the law of θ given the observed process $Y^{(n)}$. Standard results on Gaussian measures on ℓ^2 imply that if the prior Π on ℓ_2 is a centered Gaussian vector of trace class covariance Σ, then the posterior probability law given $Y^{(n)}$, $\hat\Pi_n^Y = \hat\Pi^{Y^{(n)}}$, is also Gaussian, with mean $\hat\theta(Y) = E_\Pi(\theta|Y^{(n)}) = \Sigma(\Sigma + I/n)^{-1} Y^{(n)} = \Sigma(\Sigma + I/n)^{-1}(y_k; y_{\ell k})^t$ and with covariance $\Sigma|Y^{(n)} = \Sigma(n\Sigma + I)^{-1}$; see, for example, Theorem 3.2 in [35]. We will drop the superindex (n) from the processes $Y^{(n)}$ and $Y_0^{(n)}$ from now on to expedite notation.

The posterior $\hat\Pi_n^Y$ gives rise to a Gaussian measure on $L_2([0,1])$ by simply "undoing" the isometry, that is, by taking the law of the random wavelet series in $L^2([0,1])$ with coefficients drawn from $\hat\Pi_n^Y$ equal to

$$
\begin{aligned}
X &= \sum_{k=0}^{N-1} \biggl[\frac{1}{1 + 1/n} y_k + \biggl(\frac{1}{n+1}\biggr)^{1/2} \bar g_k\biggr]\phi_k \\
&\quad + \sum_{\ell=J_0}^{\infty} \sum_{k=0}^{2^\ell - 1} \biggl[\frac{\mu_\ell}{\mu_\ell + 1/n} y_{\ell k} + \biggl(\frac{\mu_\ell}{n\mu_\ell + 1}\biggr)^{1/2} \bar g_{\ell k}\biggr]\psi_{\ell k} \\
&= E_{\Pi_n}(f|Y) + \sum_{k=0}^{N} \biggl(\frac{1}{n+1}\biggr)^{1/2} \phi_k \bar g_k \\
&\quad + \sum_{\ell=J_0}^{\infty} \sum_{k=0}^{2^\ell - 1} \biggl(\frac{\mu_\ell}{n\mu_\ell + 1}\biggr)^{1/2} \psi_{\ell k} \bar g_{\ell k},
\end{aligned}
$$

where the $\bar g$ variables are i.i.d. $N(0,1)$, and y_k, $y_{\ell k}$ are, as defined above, the integrals of the wavelet basis functions with respect to $dY(t)$. Under $dY_0(t) = f_0(t)\, dt + dB(t)/\sqrt{n}$, we have $y_k = \langle f_0, \phi_k\rangle + g_k/\sqrt{n}$, $y_{\ell k} = \langle f_0, \phi_{\ell k}\rangle + g_{\ell k}/\sqrt{n}$, where the $g_k, g_{\ell k}$ are again i.i.d. $N(0,1)$, independent of the variables $\bar g$. So, the posterior given Y_0 integrates the $\bar g$ variables, and E_{Y_0} integrates the g variables, and we have

$$
\begin{aligned}
&E_{Y_0}\hat{\Pi}_n^{Y_0}\{\|f-f_0\|_\infty > M\varepsilon_n\}\\
&\quad = \Pr\Biggl\{\Biggl\|\sum_{k=0}^{N-1}\biggl[\frac{-1/n}{1+1/n}\langle f_0,\phi_k\rangle\\
&\qquad\qquad + \frac{1}{\sqrt{n}(1+1/n)}g_k + \biggl(\frac{1}{n+1}\biggr)^{1/2}\bar{g}_k\biggr]\phi_k\\
&\qquad + \sum_{\ell=J_0}^{\infty}\sum_{k=0}^{2^\ell-1}\biggl[\frac{-1/n}{\mu_\ell+1/n}\langle f_0,\psi_{\ell k}\rangle \qquad (55)\\
&\qquad\qquad + \frac{\mu_\ell}{\sqrt{n}(\mu_\ell+1/n)}g_{\ell k}\\
&\qquad\qquad + \biggl(\frac{\mu_\ell}{n\mu_\ell+1}\biggr)^{1/2}\bar{g}_{\ell k}\biggr]\psi_{\ell k}\Biggr\|_\infty > M\varepsilon_n\Biggr\}\\
&\quad = \Pr\{\|E_{Y_0^{(n)}}(E_{\Pi_n}(f|Y_0)-f_0)+G\|_\infty > M\varepsilon_n\},
\end{aligned}
$$

where G is the centered Gaussian process

$$
\begin{aligned}
G(t) = {}&\sum_{k=0}^{N-1}\biggl[\frac{1}{\sqrt{n}(1+1/n)}g_k + \biggl(\frac{1}{n+1}\biggr)^{1/2}\bar{g}_k\biggr]\phi_k(t)\\
&+ \sum_{\ell=J_0}^{\infty}\sum_{k=0}^{2^\ell-1}\biggl[\frac{\mu_\ell}{\sqrt{n}(\mu_\ell+1/n)}g_{\ell k} + \biggl(\frac{\mu_\ell}{n\mu_\ell+1}\biggr)^{1/2}\bar{g}_{\ell k}\biggr]\psi_{\ell k}(t)
\end{aligned}
$$

and

$$
\begin{aligned}
E_{Y_0}(E_{\Pi_n}(f|Y_0)-f_0) = {}&\sum_{k=0}^{N-1}\frac{-1/n}{1+1/n}\langle f_0,\phi_k\rangle\phi_k\\
&+ \sum_{\ell=J_0}^{\infty}\sum_{k=0}^{2^\ell-1}\frac{-1/n}{\mu_\ell+1/n}\langle f_0,\psi_{\ell k}\rangle\psi_{\ell k}.
\end{aligned}
$$

It suffices to prove the theorem for $r=\infty$. We will apply Borell's [4] inequality (a consequence thereof, in fact, equation (3.2) in [22], page 57) to the probability in (55), and for this we need to estimate $\|E(E_{\Pi_n}(f|Y_0)-f_0)\|_\infty$, $E\|G\|_\infty$ and $\|E(G^2(\cdot))\|_\infty$.

Choose $J_n \geq J_0$ such that $2^{J_n} \simeq (n/\log n)^{1/(2\alpha+1)}$. Since $f_0 \in \mathcal{C}^\alpha([0,1])$ and $\|\sum_k|\psi_{\ell k}|\|_\infty \leq C2^{\ell/2}$, we obtain

$$
\Biggl\|\sum_{k=0}^{N-1}\frac{-1/n}{1+1/n}\langle f_0,\phi_k\rangle\phi_k\Biggr\|_\infty \leq \Biggl\|\sum_{k=0}^{N-1}|\phi_k|\Biggr\|_\infty\frac{C}{n+1} \leq \frac{C_1}{n}
$$

and

$$\left\|\sum_{\ell=J_0}^{\infty}\sum_{k=0}^{2^\ell-1}\frac{-1/n}{\mu_\ell+1/n}\langle f_0,\psi_{\ell,k}\rangle\psi_{\ell,k}\right\|_\infty \le \sum_{\ell=J_0}^{\infty}\left\|\sum_{k=0}^{2^\ell-1}|\psi_{\ell k}|\right\|_\infty \frac{C2^{-\ell(\alpha+1/2)}}{n\mu_\ell+1}$$
$$\le C'\left(\sum_{\ell=J_0}^{J_n}\frac{2^{-\ell\alpha}}{n\mu_\ell}+\sum_{\ell=J_n+1}^{\infty}2^{-\ell\alpha}\right)$$
$$\le C_2\left(\frac{\log n}{n}\right)^{\alpha/(2\alpha+1)},$$

where C_1 and C_2 depend only on the wavelet basis, α and $\|f_0\|_{\alpha,\infty}$. Collecting the last two sets of inequalities yields the bound

$$\|E_{Y_0}(E_{\Pi_n}(f|Y_0)-f_0)\|_\infty \le \bar{C}_1\left(\frac{\log n}{n}\right)^{\alpha/(2\alpha+1)} \tag{56}$$

for some $\bar{C}_1<\infty$. To bound $E\|G\|_\infty$, recall that for any sequence of centered normal random variables Z_j,

$$E\max_{1\le j\le N}|Z_j|\le C\sqrt{\log N}\max_{j\le N}(EZ_j^2)^{1/2}, \tag{57}$$

where C is a universal constant. Therefore, from the definitions of J_n,μ_ℓ,

$$E\left\|\sum_k\left[\frac{1}{\sqrt{n}(1+1/n)}g_k+\left(\frac{1}{n+1}\right)^{1/2}\bar{g}_k\right]\phi_k\right\|_\infty$$
$$\le\left\|\sum_k|\phi_k|\right\|_\infty\left(\frac{1}{n(1+1/n)^2}+\frac{1}{n+1}\right)^{1/2}E\max_k|g_k|$$
$$=O\left(\frac{1}{\sqrt{n}}\right)$$

and, using $\mu_\ell\lesssim n^{-1}$ for $\ell\ge J_n$,

$$E\left\|\sum_{\ell=J_0}^{\infty}\sum_{k=0}^{2^\ell-1}\left[\frac{\mu_\ell}{\sqrt{n}(\mu_\ell+1/n)}g_{\ell k}+\left(\frac{\mu_\ell}{n\mu_\ell+1}\right)^{1/2}\bar{g}_{\ell k}\right]\psi_{\ell k}\right\|_\infty$$
$$\le C'\sum_{\ell=J_0}^{\infty}2^{\ell/2}E\max_{k\le 2^\ell}|g_{\ell k}|\left(\frac{\mu_\ell^2}{n(\mu_\ell+1/n)^2}+\frac{\mu_\ell}{n\mu_\ell+1}\right)^{1/2}$$
$$\le C''\sum_{\ell=J_0}^{\infty}(\ell 2^\ell)^{1/2}\left(\frac{\mu_\ell^2}{n(\mu_\ell+1/n)^2}+\frac{\mu_\ell}{n\mu_\ell+1}\right)^{1/2}$$

$$
\begin{aligned}
&\le C''\Bigg(2\sum_{\ell=J_0}^{J_n}\sqrt{\frac{2^\ell\ell}{n}}+\sum_{\ell>J_n}\sqrt{2^\ell\ell n}\mu_\ell+\sum_{\ell>J_n}\sqrt{2^\ell\ell\mu_\ell}\Bigg)\\
&\le C'''\Bigg(\sqrt{\frac{2^{J_n}J_n}{n}}+2^{-J_n\alpha}\Bigg)\le D\bigg(\frac{\log n}{n}\bigg)^{\alpha/(2\alpha+1)}.
\end{aligned}
$$

Conclude

$$
E\|G\|_\infty\le\bar C_2\bigg(\frac{\log n}{n}\bigg)^{\alpha/(2\alpha+1)} \tag{58}
$$

for some $\bar C_2<\infty$. Finally,

$$
\begin{aligned}
EG^2(t)&=\sum_{k=0}^{N-1}\bigg(\frac{1}{n(1+1/n)^2}+\frac{1}{n+1}\bigg)\phi_k^2(t)\\
&\quad+\sum_{\ell=J_0}^{\infty}\sum_{k=0}^{2^\ell-1}\bigg(\frac{\mu_\ell^2}{n(\mu_\ell+1/n)^2}+\frac{\mu_\ell}{n\mu_\ell+1}\bigg)\psi_{\ell k}^2(t) \qquad (59)\\
&\le C\bigg(\frac{1}{n}+\frac{2^{J_n}}{n}+2^{-J_n(2\alpha+1)}\bigg)\le C_3\frac{2^{J_n}}{n}.
\end{aligned}
$$

So, setting $\varepsilon_n=(n/\log n)^{-\alpha/(2\alpha+1)}$, the estimates (56), (58) and (59) together with inequality (3.2) on page 57 of [22], give

$$
\begin{aligned}
&\Pr\{\|E_{Y_0}(E_{\Pi_n}(f|Y_0)-f_0)+G\|_\infty>M\varepsilon_n\}\\
&\quad\le\Pr\{\|G\|_\infty-E\|G\|_\infty>M\varepsilon_n-\|E(E_{\Pi_n}(f|Y_0)-f_0)\|_\infty-E\|G\|_\infty\}\\
&\quad\le\Pr\{\|G\|_\infty-E\|G\|_\infty>(M-\bar C_1-\bar C_2)\varepsilon_n\} \qquad (60)\\
&\quad\le\exp\bigg(-\frac{(M-\bar C_1-\bar C_2)^2\varepsilon_n^2}{C_3^2 2^{J_n}/n}\bigg).
\end{aligned}
$$

Collecting (55) and (60) and taking into account that $\varepsilon_n^2\simeq 2^{J_n}J_n/n$ completes the proof.

Acknowledgments. E. Giné's work was carried out during a sabbatical leave at the M.I.T. Mathematics Department and on a visit at the University of Cambridge Statistical Laboratory, and he is grateful for the hospitality of these institutions. R. Nickl would like to thank the Cafés Bräunerhof and Florianihof in Vienna for their continued hospitality.

We are further grateful to Ismael Castillo, Vladimir Koltchinskii, Natesh Pillai, Catia Scricciolo and Aad van der Vaart for valuable conversations about the subject of this article. We also thank the Associate Editor and two referees for substantial and influential reports on preliminary versions of this manuscript.

SUPPLEMENTARY MATERIAL

Supplement to "Rates of contraction for posterior distributions in L^r-metrics, $1 \le r \le \infty$" (DOI: 10.1214/11-AOS924SUPP; .pdf). This supplement contains a detailed proof of Lemma 1 and an expanded proof of Proposition 2 from the mentioned article.

REFERENCES

[1] BARRON, A., SCHERVISH, M. J. and WASSERMAN, L. (1999). The consistency of posterior distributions in nonparametric problems. *Ann. Statist.* **27** 536–561. MR1714718

[2] BIRGÉ, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Gebiete* **65** 181–237. MR0722129

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

[4] BORELL, C. (1975). The Brunn–Minkowski inequality in Gauss space. *Invent. Math.* **30** 207–216. MR0399402

[5] BOUSQUET, O. (2003). Concentration inequalities for sub-additive functions using the entropy method. In *Stochastic Inequalities and Applications*. *Progress in Probability* **56** 213–247. Birkhäuser, Basel. MR2073435

[6] CASTILLO, I. (2011). A semiparametric Bernstein–von Mises theorem for Gaussian process priors. *Probab. Theory Related Fields.* To appear.

[7] CIESIELSKI, Z., KERKYACHARIAN, G. and ROYNETTE, B. (1993). Quelques espaces fonctionnels associés à des processus gaussiens. *Studia Math.* **107** 171–204. MR1244574

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

[9] GHOSAL, S., GHOSH, J. K. and RAMAMOORTHI, R. V. (1999). Posterior consistency of Dirichlet mixtures in density estimation. *Ann. Statist.* **27** 143–158. MR1701105

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

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

[12] GHOSAL, S. and VAN DER VAART, A. (2007). Posterior convergence rates of Dirichlet mixtures at smooth densities. *Ann. Statist.* **35** 697–723. MR2336864

[13] GHOSAL, S. and VAN DER VAART, A. W. (2001). Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. *Ann. Statist.* **29** 1233–1263. MR1873329

[14] GINÉ, E. and GUILLOU, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 907–921. MR1955344

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

[16] GINÉ, E. and NICKL, R. (2008). Adaptation on the space of finite signed measures. *Math. Methods Statist.* **17** 113–122. MR2429123

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

[18] GINÉ, E. and NICKL, R. (2011). Supplement to "Rates of contraction for posterior distributions in L^r-metrics, $1 \le r \le \infty$." DOI:10.1214/11-AOS924SUPP.

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

[20] INGSTER, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Math. Methods Statist.* **2** 85–114. MR1257978

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

[22] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces*: *Isoperimetry and Processes*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete* (3) [*Results in Mathematics and Related Areas* (3)] **23**. Springer, Berlin. MR1102015

[23] LENK, P. J. (1991). Towards a practicable Bayesian nonparametric density estimator. *Biometrika* **78** 531–543. MR1130921

[24] LI, W. V. and LINDE, W. (1998). Existence of small ball constants for fractional Brownian motions. *C. R. Acad. Sci. Paris Sér. I Math.* **326** 1329–1334. MR1649147

[25] LOUNICI, K. and NICKL, R. (2011). Global uniform risk bounds for wavelet deconvolution estimators. *Ann. Statist.* **39** 201–231. MR2797844

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

[27] NICKL, R. (2007). Donsker-type theorems for nonparametric maximum likelihood estimators. *Probab. Theory Related Fields* **138** 411–449. MR2299714

[28] SCHWARTZ, L. (1965). On Bayes procedures. *Z. Wahrsch. Verw. Gebiete* **4** 10–26. MR0184378

[29] SCRICCIOLO, C. (2007). On rates of convergence for Bayesian density estimation. *Scand. J. Statist.* **34** 626–642. MR2368802

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

[31] TALAGRAND, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505–563. MR1419006

[32] 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

[33] 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

[34] 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

[35] ZHAO, L. H. (2000). Bayesian aspects of some nonparametric problems. *Ann. Statist.* **28** 532–552. MR1790008

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CONNECTICUT
STORRS, CONNECTICUT 06269-3009
USA
E-MAIL: gine@math.uconn.edu

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