Working Paper Series
Department of Economics
University of Verona

Sharp sup-norm Bayesian curve estimation

Catia Scricciolo

WP Number: 4 March 2016

ISSN: 2036-2919 (paper), 2036-4679 (online)

Sharp sup-norm Bayesian curve estimation

Catia Scricciolo*

Department of Economics, University of Verona, Via Cantarane 24, 37129 Verona, Italy

Abstract

Sup-norm curve estimation is a fundamental statistical problem and, in principle, a premise for the construction of confidence bands for infinite-dimensional parameters. In a Bayesian framework, the issue of whether the sup-norm-concentration-of-posterior-measure approach proposed by Giné and Nickl (2011), which involves solving a testing problem exploiting concentration properties of kernel and projection-type density estimators around their expectations, can yield minimax-optimal rates is herein settled in the affirmative beyond conjugate-prior settings obtaining sharp rates for common prior-model pairs like random histograms, Dirichlet Gaussian or Laplace mixtures, which can be employed for density, regression or quantile estimation.

Keywords: McDiarmind's inequality, Nonparametric hypothesis testing, Posterior distributions, Sup-norm rates

1. Introduction

The study of the frequentist asymptotic behaviour of Bayesian nonparametric (BNP) procedures has initially focused on the Hellinger or L^1-distance loss, see Shen and Wasserman (2001) and Ghosal *et al.* (2000), but an extension and generalization of the results to L^r-distance losses, $1 \leq r \leq \infty$, has been the object of two recent contributions by Giné and Nickl (2011) and Castillo (2014). Sup-norm estimation has particularly attracted attention as it may constitute the premise for the construction of confidence bands whose geometric structure can be easily visualized and interpreted. Furthermore, as shown in the example of Section 3.2, the study of sup-norm posterior contraction rates for density estimation can be motivated as being an intermediate step for the final assessment of convergence rates for quantile estimation.

While the contribution of Castillo (2014) has a more prior-model specific flavour, the article by Giné and Nickl (2011) aims at a unified understanding of the drivers of the asymptotic behaviour of BNP procedures by developing a new approach to the involved testing problem constructing nonparametric tests that have good exponential bounds on the type-one and type-two error probabilities that rely on concentration properties of kernel and projection-type density estimators around their expectations.

Even if Giné and Nickl (2011)'s approach can only be useful if a fine control of the approximation properties of the prior support is possible, it has the merit of replacing the entropy condition for sieve sets with approximating conditions. However, the result, as presented in their Theorem 2 (Theorem 3), can only retrieve minimax-optimal rates for L^r-losses when $1 \leq r \leq 2$, while rates deteriorate by a genuine power of n, in fact $n^{1/2}$, for $r > 2$. Thus, the open question remains whether their approach can give the right rates for $2 < r \leq \infty$ for non-conjugate priors and sub-optimal rates are possibly only an artifact of the proof. We herein settle this issue in the affirmative by refining their result and proof and showing in concrete examples that this approach retrieves the right rates.

The paper is organized as follows. In Section 2, we state the main result whose proof is postponed to Appendix A. Examples concerning different statistical settings like density and quantile estimation are presented in Section 3.

*Corresponding author.
Email address: catia.scricciolo@univr.it (Catia Scricciolo)

2. Main result

In this section, we describe the set-up and present the main contribution of this note. Let $((\mathcal{X}, \mathscr{A}, P), P \in \mathcal{P})$ be a collection of probability measures on a measurable space $(\mathcal{X}, \mathscr{A})$ that possess densities with respect to some σ-finite dominating measure μ. Let Π_n be a sequence of priors on $(\mathcal{P}, \mathscr{B})$, where $\mathscr{B}$ is a σ-field on $\mathcal{P}$ for which the maps $x \mapsto p(x)$ are jointly measurable relative to $\mathscr{A} \otimes \mathscr{B}$. Let $X_1, \ldots, X_n$ be i.i.d. (independent, identically distributed) observations from a common law $P_0 \in \mathcal{P}$ with density p_0 on $\mathcal{X}$ with respect to μ, $p_0 = \mathrm{d}P_0/\mathrm{d}\mu$. For a probability measure P on $(\mathcal{X}, \mathscr{A})$ and an $\mathscr{A}$-measurable function $f : \mathcal{X} \to \mathbb{R}^k$, $k \geq 1$, let Pf denote the integral $\int f \, \mathrm{d}P$, where, unless otherwise specified, the set of integration is understood to be the whole domain. When this notation is applied to the empirical measure $\mathbb{P}_n$ associated with a sample $X^{(n)} := (X_1, \ldots, X_n)$, namely the discrete uniform measure on the sample values, this yields $\mathbb{P}_n f = n^{-1} \sum_{i=1}^n f(X_i)$. For each $n \in \mathbb{N}$, let $\hat{p}_n(j)(\cdot) = n^{-1} \sum_{i=1}^n K_j(\cdot, X_i)$ be a kernel or projection-type density estimator based on $X_1, \ldots, X_n$ at resolution level j, with K_j as in Definition (1) below. Its expectation is then equal to $P_0^n \hat{p}_n(j)(\cdot) = P_0 K_j(\cdot, X_1) = K_j(p_0)(\cdot)$, where we have used the notation $K_j(p_0)(\cdot) = \int K_j(\cdot, y) p_0(y) \, \mathrm{d}y$. In order to refine Giné and Nickl (2011)'s result, we use concentration properties of $\|\hat{p}_n(j) - K_j(p_0)\|_1$ around its expectation by applying McDiarmind's inequality for bounded differences functions.

The following definition, which corresponds to Condition 5.1.1 in Giné and Nickl (2015), is essential for the main result.

Definition 1. Let $\mathcal{X} = \mathbb{R}$, $\mathcal{X} = [0, 1]$ or $\mathcal{X} = (0, 1]$. The sequence of operators

$$K_j(x, y) := 2^j K(2^j x, 2^j y), \quad x, y \in \mathcal{X}, \quad j \geq 0,$$

is called *an admissible approximating sequence* if it satisfies one of the following conditions:

a) convolution kernel case, $\mathcal{X} = \mathbb{R}$: $K(x, y) = K(x - y)$, where $K \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$, integrates to 1 and is of bounded p-variation for some finite $p \geq 1$ and right (left)-continuous;

b) multi-resolution projection case, $\mathcal{X} = \mathbb{R}$: $K(x, y) = \sum_{k \in \mathbb{Z}} \phi(x - k)\phi(y - k)$, with K_j as above or $K_j(x, y) = K(x, y) + \sum_{\ell=0}^{j-1} \sum_k \psi_{lk}(x)\psi_{lk}(y)$, where $\phi, \psi \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ define an S-regular wavelet basis, have bounded p-variation for some $p \geq 1$ and are uniformly continuous, or define the Haar basis, see Chapter 4, *ibidem*;

c) multi-resolution case, $\mathcal{X} = [0, 1]$: $K_{j,bc}(x, y)$ is the projection kernel at resolution j of a Cohen-Daubechies-Vial (CDV) wavelet basis, see Chapter 4, *ibidem*;

d) multi-resolution case, $\mathcal{X} = (0, 1]$: $K_{j,per}(x, y)$ is the projection kernel at resolution j of the periodization of a scaling function satisfying b), see (4.126) and (4.127), *ibidem*.

Remark 1. A useful property of S-regular wavelet bases is the following: there exists a non-negative measurable function $\Phi \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ such that $|K(x, y)| \leq \Phi(|x - y|)$ for all $x, y \in \mathbb{R}$, that is, K is dominated by a bounded and integrable convolution kernel Φ.

In order to state the main result, we recall that a sequence of positive real numbers L_n is *slowly varying* at ∞ if, for each $\lambda > 0$, it holds that $\lim_{n\to\infty}(L_{[\lambda n]}/L_n) = 1$. Also, for $s \geq 0$, let $L^1(\mu_s)$ be the space of μ_s-integrable functions, $\mathrm{d}\mu_s(x) := (1 + |x|)^s \mathrm{d}x$, equipped with the norm $\|f\|_{L^1(\mu_s)} := \int |f(x)|(1 + |x|)^s \, \mathrm{d}x$.

Theorem 1. *Let ϵ_n and J_n be sequences of positive real numbers such that $\epsilon_n \to 0$, $n\epsilon_n^2 \to \infty$ and $2^{J_n} = O(n\epsilon_n^2)$. For each $r \in \{1, \infty\}$ and a slowly varying sequence $L_{n,r} \to \infty$, let $\epsilon_{n,r} := L_{n,r}\epsilon_n$. Suppose that, for K as in Definition* (1)*, with K^2, Φ^2 and p_0 that integrate $(1 + |x|)^s$ for some $s > 1$ in cases a) and b),*

$$\|K_{J_n}(p_0) - p_0\|_r = O(\epsilon_{n,r}) \tag{1}$$

and, for a constant $C > 0$, sets $\mathcal{P}_n \subseteq \{P \in \mathcal{P} : \|K_{J_n}(p) - p\|_r \leq C_K \epsilon_{n,r}\}$, where $C_K > 0$ only depends on K, we have

(i) $\Pi_n(\mathcal{P} \setminus \mathcal{P}_n) \leq \exp(-(C + 4)n\epsilon_n^2)$,

(ii) $\Pi_n(P \in \mathcal{P} : -P_0 \log(p/p_0) \leq \epsilon_n^2, \; P_0 \log^2(p/p_0) \leq \epsilon_n^2) \geq \exp(-Cn\epsilon_n^2)$.

Then, for sufficiently large $M_r > 0$,

$$P_0^n\Pi_n(P \in \mathcal{P} : \|p - p_0\|_r \geq M_r\epsilon_{n,r} \mid X^{(n)}) \to 0. \quad (2)$$

If the convergence in (2) *holds for $r \in \{1, \infty\}$, then, for each $1 < s < \infty$. $P_0^n\Pi_n(P \in \mathcal{P} : \|p - p_0\|_s \geq M_s\bar{\epsilon}_n \mid X^{(n)}) \to 0$, where $\bar{\epsilon}_n := (L_{n,1} \vee L_{n,\infty})\epsilon_n$.*

The assertion, whose proof is reported in Appendix A, is an in-probability statement that the posterior mass outside a sup-norm ball of radius a large multiple M of ϵ_n is negligible. The theorem provides the same sufficient conditions for deriving sup-norm posterior contraction rates that are minimax-optimal, up to logarithmic factors, as in Giné and Nickl (2011). Condition (ii), which is mutuated from Ghosal *et al.* (2000), is the essential one: the prior concentration rate is the only determinant of the posterior contraction rate at densities p_0 having sup-norm approximation error of the same order against a kernel-type approximant, provided the prior support is almost the set of densities with the same approximation property.

3. Examples

In this section, we apply Theorem 1 to some prior-model pairs used for (conditional) density or regression estimation, including random histograms, Dirichlet Gaussian or Laplace mixtures, that have been selected in an attempt to reflect cases for which the issue of obtaining sup-norm posterior rates was still open. We do not consider Gaussian priors or wavelets series because these examples have been successfully worked out in Castillo (2014) taking a different approach. We furthermore exhibit an example with the aim of illustrating that obtaining sup-norm posterior contraction rates for density estimation can be motivated as being an intermediate step for the final assessment of convergence rates for estimating single quantiles.

3.1. Density estimation

Example 1 (Random dyadic histograms)**.** For $J_n \in \mathbb{N}$, consider a partition of $[0, 1]$ into 2^{J_n} intervals (*bins*) of equal length $A_{1,2^{J_n}} = [0, 2^{-J_n}]$ and $A_{j,2^{J_n}} = ((j-1)2^{-J_n}, j2^{-J_n}]$, $j = 2, \ldots, 2^{J_n}$. Let $\mathrm{Dir}_{2^{J_n}}$ denote the Dirichlet distribution on the $(2^{J_n} - 1)$-dimensional unit simplex with all parameters equal to 1. Consider the random histogram

$$\sum_{j=1}^{2^{J_n}} w_{j,2^{J_n}} 2^{J_n} 1_{A_{j,2^{J_n}}}(\cdot), \quad (w_{1,2^{J_n}}, \ldots, w_{2^{J_n},2^{J_n}}) \sim \mathrm{Dir}_{2^{J_n}}.$$

Denote by $\Pi_{2^{J_n}}$ the induced law on the space of probability measures with Lebesgue density on $[0, 1]$. Let $X_1, \ldots, X_n$ be i.i.d. observations from a density p_0 on $[0, 1]$. Then, the Bayes' density estimator, that is the posterior expected histogram, has expression

$$\hat{p}_n(x) = \sum_{j=1}^{2^{J_n}} \frac{1 + N_{l(x)}}{2^{J_n} + n} 2^{J_n} 1_{A_{j,2^{J_n}}}(x), \quad x \in [0, 1],$$

where $l(x)$ identifies the bin containing x, i.e., $A_{l(x),2^{J_n}} \ni x$, and $N_{l(x)}$ stands for the number of observations falling into $A_{l(x),2^{J_n}}$. Let $C^\alpha([0, 1])$ denote the class of Hölder continuous functions on $[0, 1]$ with exponent $\alpha > 0$. Let $\epsilon_{n,\alpha} := (n/\log n)^{-\alpha/(2\alpha+1)}$ be the minimax rate of convergence over $(C^\alpha([0, 1]), \|\cdot\|_\infty)$.

Proposition 1. *Let $X_1, \ldots, X_n$ be i.i.d. observations from a density $p_0 \in C^\alpha([0, 1])$, with $\alpha \in (0, 1]$, satisfying $p_0 > 0$ on $[0, 1]$. Let J_n be such that $2^{J_n} \sim \epsilon_{n,\alpha}^{1/\alpha}$. Then, for sufficiently large $M > 0$, $P_0^n\Pi_{2^{J_n}}(P : \|p - p_0\|_\infty \geq M\epsilon_{n,\alpha} \mid X^{(n)}) \to 0$. Consequently, $P_0^n\|\hat{p}_n - p_0\|_\infty \asymp \epsilon_{n,\alpha}$.*

The first part of the assertion, which concerns posterior contraction rates, immediately follows from Theorem (1) combined with the proof of Proposition 3 of Giné and Nickl (2011), whose result, together with that of Theorem 3 in Castillo (2014), is herein improved to the minimax-optimal rate $(n/\log n)^{-\alpha/(2\alpha+1)}$ for every $0 < \alpha \leq 1$. The second part of the assertion, which concerns convergence rates for the histogram density estimator, is a consequence of Jensen's inequality and convexity of $p \mapsto \|p - p_0\|_\infty$, combined with the fact that the prior $\Pi_{2^{J_n}}$ is supported on densities uniformly bounded above by 2^{J_n} and that the proof of Theorem 1 yields the exponential order $\exp(-Bn\epsilon_{n,\alpha}^2)$ for the convergence of the posterior probability of the complement of an $(M\epsilon_{n,\alpha})$-ball around p_0, in symbols, $P_0^n\|\hat{p}_n - p_0\|_\infty < M\epsilon_{n,\alpha} + 2^{J_n}P_0^n\Pi_{2^{J_n}}(P : \|p - p_0\|_\infty \geq M\epsilon_{n,\alpha} \mid X^{(n)}) \leq M\epsilon_{n,\alpha} + 2^{J_n}\exp(-Bn\epsilon_{n,\alpha}^2)$, whence $P_0^n\|\hat{p}_n - p_0\|_\infty = O(\epsilon_{n,\alpha})$.

Example 2 (Dirichlet-Laplace mixtures)**.** Consider, as in Scricciolo (2011), Gao and van der Vaart (2015), a Laplace mixture prior Π thus defined. For $\varphi(x) := \frac{1}{2}\exp(-|x|)$, $x \in \mathbb{R}$, the density of a Laplace (0, 1) distribution, let

- $p_G(\cdot) := \int \varphi(\cdot - \theta)\,dG(\theta)$ denote a mixture of Laplace densities with mixing distribution G,
- $G \sim D_\alpha$, the Dirichlet process with base measure $\alpha := \alpha_{\mathbb{R}}\bar{\alpha}$, for $0 < \alpha_{\mathbb{R}} < \infty$ and $\bar{\alpha}$ a probability measure on $\mathbb{R}$.

Proposition 2. *Let $X_1, \ldots, X_n$ be i.i.d. observations from a density p_{G_0}, with G_0 supported on a compact interval $[-a, a]$. If α has support on $[-a, a]$ with continuous Lebesgue density bounded below away from 0 and above from ∞, then, for sufficiently large $M > 0$, $P_0^n\Pi(P : \|p - p_0\|_\infty \geq M(n/\log n)^{-3/8} \mid X^{(n)}) \to 0$. Consequently, for the Bayes' estimator $\hat{p}_n(\cdot) = \int p_G(\cdot)\Pi(dG \mid X^{(n)})$ we have $P_0^n\|\hat{p}_n - p_0\|_\infty \asymp (n/\log n)^{-3/8}$.*

Proof. It is known from Proposition 4 in Gao and van der Vaart (2015) that the small-ball probability estimate in condition (ii) of Theorem 1 is satisfied for $\epsilon_n = (n/\log n)^{-3/8}$. For the bias condition, we take $\mathcal{P}_n$ to be the support of Π and show that, for $2^{J_n} \sim \epsilon_n^{-1/3} = (n/\log n)^{1/8}$ and any symmetric density K with finite second moment, we have $\|K_{J_n}(p_G) - p_G\|_\infty = O(\epsilon_n)$ uniformly over the support of Π. Indeed, by applying Lemma 1 with $\beta = 2$, for each $x \in \mathbb{R}$ it results $|K_{J_n}(p_G)(x) - p_G(x)|^2 \leq \|K_{J_n}(p_G) - p_G\|_2^2 \leq \int |\tilde{\varphi}(t)|^2|\tilde{K}(2^{-J_n}t) - 1|^2\,dt \sim (2\pi)^{-1}(B_\varphi^2 \times I_2[\tilde{K}])(2^{2J_n})^{-3}$, which implies that both conditions (1) and (i) are satisfied. The assertion on the Bayes' estimator follows from the same arguments laid out for random histograms together with the fact that $p_G \leq 1/2$ uniformly in G. □

Example 3 (Dirichlet-Gaussian mixtures)**.** Consider, as in Ghosal and van der Vaart (2001, 2007), Shen *et al.* (2013), Scricciolo (2014), a Gaussian mixture prior $\Pi \times G$ thus defined. For ϕ the standard normal density, let

- $p_{F,\sigma}(\cdot) := \int \phi_\sigma(\cdot - \theta)\,dF(\theta)$ denote a mixture of Gaussian densities with mixing distribution F,
- $F \sim D_\alpha$, the Dirichlet process with base measure $\alpha := \alpha_{\mathbb{R}}\bar{\alpha}$, for $0 < \alpha_{\mathbb{R}} < \infty$ and $\bar{\alpha}$ a probability measure on $\mathbb{R}$, which has continuous and positive density $\alpha'(\theta) \propto e^{-b|\theta|^\delta}$ as $|\theta| \to \infty$, for some constants $0 < b < \infty$ and $0 < \delta \leq 2$,
- $\sigma \sim G$ which has continuous and positive density g on $(0, \infty)$ such that, for constants $0 < C_1, C_2, D_1, D_2 < \infty$, $0 \leq s, t < \infty$,
$$C_1\sigma^{-s}\exp(-D_1\sigma^{-1}\log^t(1/\sigma) \leq g(\sigma) \leq C_2\sigma^{-s}\exp(-D_2\sigma^{-1}\log^t(1/\sigma))$$
for all σ in a neighborhood of 0.

Let $C^\beta(\mathbb{R})$ denote the class of Hölder continuous functions on $\mathbb{R}$ with exponent $\beta > 0$. Let $\epsilon_{n,\beta} := (n/\log n)^{-\beta/(2\beta+1)}$ be the minimax rate of convergence over $(C^\beta(\mathbb{R}), \|\cdot\|_\infty)$. For any real $\beta > 0$, let $\lfloor\beta\rfloor$ stand for the largest integer strictly smaller than β.

Proposition 3. *Let $X_1, \ldots, X_n$ be i.i.d. observations from a density $p_0 \in L^\infty(\mathbb{R}) \cap C^\beta(\mathbb{R})$ such that condition (ii) is satisfied for $\epsilon_{n,\beta}$. Then, for sufficiently large $M > 0$, $P_0^n(\Pi \times G)((F, \sigma) : \|p_{F,\sigma} - p_0\|_\infty \geq M\epsilon_{n,\beta} \mid X^{(n)}) \to 0$.*

Proof. Let $K \in L^1(\mathbb{R})$ be a convolution kernel such that

- $\int x^k K(x)\,dx = \mathbf{1}_{\{0\}}(k)$, $k = 0, \ldots, \lfloor\beta\rfloor$, and $\int |x|^\beta|K(x)|\,dx < \infty$,
- the Fourier transform $\tilde{K}$ has $\text{supp}(\tilde{K}) \subseteq [-1, 1]$.

Let $2^{J_n} \sim \epsilon_{n,\beta}^{1/\beta}$. For every $x \in \mathbb{R}$, $|K_{J_n}(p_0)(x) - p_0(x)| \leq C_1(2^{-J_n})^\beta \lesssim \epsilon_{n,\beta}$, where the constant $C_1 \propto (1/\lfloor\beta\rfloor!)\int |x|^\beta|K(x)|\,dx$ does not depend on x. Thus, $\|K_{J_n}(p_0) - p_0\|_\infty = O(\epsilon_{n,\beta})$. For the bias condition, let $\underline{\sigma}_n := E(n\epsilon_{n,\beta}^2)^{-1}(\log n)^\psi$, with $1/2 < \psi < t$ and a suitable constant $0 < E < \infty$. For every $\sigma \geq \underline{\sigma}_n$ and uniformly in F,

$$\begin{aligned}\|K_{J_n}(p_{F,\sigma}) - p_{F,\sigma}\|_\infty &= \sup_{x\in\mathbb{R}}\left|\int\int K_{J_n}(u)[\phi_\sigma(x - v - u) - \phi_\sigma(x - v)]\,du\,dF(v)\right| \\ &\leq \frac{1}{2\pi}\sup_{x\in\mathbb{R}}\int\int |e^{-it(x-v)}||\tilde{\phi}_\sigma(t)||\tilde{K}(2^{-J_n}t) - 1|\,dt\,dF(v) \\ &\leq \frac{1}{\pi}\int_{|t|>2^{J_n}}|\tilde{\phi}_\sigma(t)|\,dt \lesssim \underline{\sigma}_n^{-1}\exp(-(\rho\underline{\sigma}_n 2^{J_n})^2) \lesssim n^{-1} < \varepsilon_{n,\beta}\end{aligned}$$

because $(\underline{\sigma}_n 2^{J_n})^2 \propto (\log n)^{2\psi} \gtrsim (\log n)$ as $\psi > 1/2$. Now, $G(\sigma < \underline{\sigma}_n) \lesssim \underline{\sigma}_n^{-s}\exp(-[D_2\underline{\sigma}_n^{-1}\log^t(1/\underline{\sigma}_n)]) \lesssim \exp(-(C + 4)n\epsilon_n^2)$ because $\psi < t$, which implies that the remaining mass condition (ii) is satisfied. □

Remark 2. Conditions on the density p_0 under which assumption (ii) of Theorem 1 is satisfied can be found, for instance, in Shen *et al.* (2013) and Scricciolo (2014).

3.2. Quantile estimation

For $\tau \in (0,\,1)$, consider the problem of estimating the τ-quantile q_0^τ of the population distribution function F_0 from observations $X_1,\,\ldots,\,X_n$. For any (possibly unbounded) interval $I \subseteq \mathbb{R}$ and function g on I, define the Hölder norm as

$$\|g\|_{C^\alpha(I)} := \sum_{k=0}^{\lfloor\alpha\rfloor} \|g^{(k)}\|_{L^\infty(I)} + \sup_{x,\,y\in I:\,x\neq y} \frac{|g^{\lfloor\alpha\rfloor}(x) - g^{\lfloor\alpha\rfloor}(y)|}{|x-y|^{\alpha-\lfloor\alpha\rfloor}}.$$

Let $C^0(I)$ denote the space of continuous and bounded functions on I and $C^\alpha(I,\,R) := \{g \in C^0(I) \,:\, \|g\|_{C^\alpha(I)} \leq R\}$, $R > 0$.

Proposition 4. *Suppose that, given $\tau \in (0,\,1)$, there are constants $r,\,\zeta > 0$ so that $p_0(\cdot + q_0^\tau) \in C^\alpha([-\zeta,\,\zeta],\,R)$ and*

$$\inf_{[q_0^\tau-\zeta,\,q_0^\tau+\zeta]} p_0(x) \geq r. \tag{3}$$

Consider a prior Π concentrated on probability measures having densities $p(\cdot+q_0^\tau) \in C^\alpha([-\zeta,\,\zeta],\,R)$. If, for sufficiently large M, the posterior probability $P_0^n\Pi(P \,:\, \|p-p_0\|_\infty \geq M\epsilon_{n,\alpha} \mid X^{(n)}) \to 0$, then, there exists $M' > 0$ so that $P_0^n\Pi(|q^\tau - q_0^\tau| \geq M'\epsilon_{n,\alpha}^{1+1/\alpha} \mid X^{(n)}) \to 0$.

Proof. We preliminarily make the following remark. Let $F(x) := \int_{-\infty}^x p(y)\,\mathrm{d}y$, $x \in \mathbb{R}$. For $\tau \in (0,\,1)$, let q^τ be the τ-quantile of F. By Lagrange's theorem, there exists a point q_*^τ between q^τ and q_0^τ so that $F(q^\tau) - F(q_0^\tau) = p(q_*^\tau)(q^\tau - q_0^\tau)$. Consequently,

$$0 = \tau - \tau = \int_{-\infty}^{q^\tau} p(x)\,\mathrm{d}x - \int_{-\infty}^{q_0^\tau} p_0(x)\,\mathrm{d}x = \int_{q_0^\tau}^{q^\tau} p(x)\,\mathrm{d}x + \int_{-\infty}^{q_0^\tau} [p(x) - p_0(x)]\,\mathrm{d}x = p(q_*^\tau)(q^\tau - q_0^\tau) + [F(q_0^\tau) - F_0(q_0^\tau)].$$

If $p(q_*^\tau) > 0$, then

$$q^\tau - q_0^\tau = -\frac{[F(q_0^\tau) - F_0(q_0^\tau)]}{p(q_*^\tau)} = -\frac{[F(q_0^\tau) - \tau]}{p(q_*^\tau)}. \tag{4}$$

In order to upper bound $|q^\tau - q_0^\tau|$, by appealing to relationship (4), we can separately control $|F(q_0^\tau) - F_0(q_0^\tau)|$ and $p(q_*^\tau)$. Let the kernel function $K \in L^1(\mathbb{R})$ be such that

- $\int x^k K(x)\mathrm{d}x = 1_{\{0\}}(k)$, $k = 0,\,\ldots,\,\lfloor\alpha\rfloor + 1$, and $\int |x|^{\alpha+1}|K(x)|\,\mathrm{d}x < \infty$,
- its Fourier transform $\tilde{K}$ has $\mathrm{supp}(\tilde{K}) \subseteq [-1,\,1]$.

By Lemma 5.2 in Dattner *et al.* (2013),

$$\sup_{p_0(\cdot+q_0^\tau)\in C^\alpha([-\zeta,\,\zeta],\,R)} \left| \int_{-\infty}^{q_0^\tau} [K_b * p_0 - p_0](x)\,\mathrm{d}x \right| \leq D b^{\alpha+1}, \tag{5}$$

with $D := [R/(\lfloor\alpha\rfloor + 1)! + 2\zeta^{-(\alpha+1)}] \int |x|^{\alpha+1}|K(x)|\,\mathrm{d}x$. Write

$$F(q_0^\tau) - F_0(q_0^\tau) = \int_{-\infty}^{q_0^\tau} [K_b * p_0 - p_0](x)\,\mathrm{d}x + \int_{-\infty}^{q_0^\tau} [K_b * (p - p_0)](x)\,\mathrm{d}x + \int_{-\infty}^{q_0^\tau} [p - K_b * p](x)\,\mathrm{d}x =: T_1 + T_2 + T_3.$$

By inequality (5), we have $|T_1| = O(b^{\alpha+1})$. By the same reasoning, $|T_3| = O(b^{\alpha+1})$. We now consider T_2. Taking into account that $\int K(x)\,\mathrm{d}x = 1$ and

$$T_2 := [K_b*(F-F_0)](q_0^\tau) = \int \frac{1}{b} K\left(\frac{q_0^\tau - u}{b}\right)(F-F_0)(u)\,\mathrm{d}u = -\int K(z)(F-F_0)(q_0^\tau - bz)\,\mathrm{d}z = \int K(z)(F_0-F)(q_0^\tau - bz)\,\mathrm{d}z,$$

for some point ξ between $q_0^\tau - bz$ and q_0^τ (clearly, ξ depends on q_0^τ, z, b),

$$\begin{aligned}T_2 = [K_b * (F - F_0)](q_0^\tau) \mp (F_0 - F)(q_0^\tau) &= \int K(z)[(F_0 - F)(q_0^\tau - bz) - (F_0 - F)(q_0^\tau)]\,\mathrm{d}z + (F_0 - F)(q_0^\tau)\\ &= \int K(z)(-bz)[D^1(F_0 - F)(\xi)]\,\mathrm{d}z + (F_0 - F)(q_0^\tau)\\ &= (-b)\int zK(z)[(p_0 - p)(\xi)]\,\mathrm{d}z + (F_0 - F)(q_0^\tau).\end{aligned}$$

Then, $F(q_0^\tau) - F_0(q_0^\tau) = T_1 + T_3 + (-b)\int zK(z)[(p_0 - p)(\xi)]\,\mathrm{d}z - [(F - F_0)(q_0^\tau)]$, which implies that $2[F(q_0^\tau) - F_0(q_0^\tau)] = T_1 + T_3 + (-b)\int zK(z)[(p_0 - p)(\xi)]\,\mathrm{d}z$. It follows that $2|F(q_0^\tau) - F_0(q_0^\tau)| \le |T_1| + |T_3| + b\|p_0 - p\|_\infty \int |z||K(z)|\,\mathrm{d}z$. Taking into account that $\int |z||K(z)|\,\mathrm{d}z < \infty$, $|T_1| = O(b^{\alpha+1})$ and $|T_3| = O(b^{\alpha+1})$, choosing $b = O(\epsilon_{n,\alpha}^{1/\alpha})$, we have $|F(q_0^\tau) - F_0(q_0^\tau)| \lesssim |T_1| + |T_3| + b\|p_0 - p\|_\infty \lesssim b^{\alpha+1} + b\|p_0 - p\|_\infty \lesssim \epsilon_{n,\alpha}^{1+1/\alpha}$. If $\|p - p_0\|_\infty \lesssim \epsilon_{n,\alpha}$ then, under condition (3), $p(q_*^\tau) > r - \eta > 0$ for every $0 < \eta < r$. In fact, for any interval $I \supseteq [q_0^\tau - \zeta,\, q_0^\tau + \zeta]$ that includes the point q^τ so that it also includes the intermediate point q_*^τ between q^τ and q_0^τ, for any $\eta > 0$ we have $\eta \gtrsim \|p - p_0\|_\infty \ge \sup_I |p(x) - p_0(x)| \ge |p(\tilde{x}) - p_0(\tilde{x})|$ for every $\tilde{x} \in I$. It follows that $p(q_*^\tau) > p_0(q_*^\tau) - \eta \ge \inf_{x \in [q_0^\tau - \zeta,\, q_0^\tau + \zeta]} p_0(x) - \eta \ge r - \eta$. Conclude the proof by noting that, in virtue of (4), $P_0^n \Pi(P : \|p - p_0\|_\infty < M\epsilon_{n,\alpha} \mid X^{(n)}) \le P_0^n \Pi(|q^\tau - q_0^\tau| < M' \epsilon_{n,\alpha}^{1+1/\alpha} \mid X^{(n)})$. The assertion then follows. □

Remark 3. Proposition 4 considers local Hölder regularity of p_0, which seems natural for estimating single quantiles. Clearly, requirements on p_0 are automatically satisfied if p_0 is globally Hölder regular and, in this case, the minimax-optimal sup-norm rate is $\epsilon_{n,\alpha} = (n/\log n)^{-\alpha/(2\alpha+1)}$ so that the rate for estimating single quantiles is $\epsilon_{n,\alpha}^{1+1/\alpha} = (n/\log n)^{-(\alpha+1)/(2\alpha+1)}$. The conditions on the random density p are automatically satisfied if the prior is concentrated on probability measures possessing globally Hölder regular densities.

Appendix A. Proof of Theorem 1

Proof. Using the remaining mass condition (i) and the small-ball probability estimate (ii), by the proof of Theorem 2.1 in Ghosal *et al.* (2000), it is enough to construct, for each $r \in \{1, \infty\}$, a test $\Psi_{n,r}$ for the hypothesis

$$H_0 : P = P_0 \quad vs. \quad H_1 : \{P \in \mathcal{P}_n : \|p - p_0\|_r \ge M_r \epsilon_{n,r}\},$$

with $M_r > 0$ large enough, where $\Psi_{n,r} \equiv \Psi_{n,r}(X^{(n)}; P_0) : \mathcal{X}^n \to \{0, 1\}$ is the indicator function of the rejection region of H_0, such that

$$P_0^n \Psi_{n,r} \to 0 \quad \text{as } n \to \infty \quad \text{and} \quad \sup_{P \in \mathcal{P}_n :\, \|p - p_0\|_r \ge M_r \epsilon_{n,r}} P^n(1 - \Psi_{n,r}) \le \exp(-K_r M_r^2 n \epsilon_{n,r}^2) \ \text{ for sufficiently large } n,$$

where $K_r M_r^2 \ge (C + 4)$, the constant $C > 0$ being that appearing in (i) and (ii). By assumption (1), there exists a constant $C_{0,r} > 0$ such that $\|P_0^n \hat{p}_n - p_0\|_r = \|K_{J_n}(p_0) - p_0\|_r \le C_{0,r} \epsilon_{n,r}$. Define $T_{n,r} := \|\hat{p}_n - p_0\|_r$. For a constant $M_{0,r} > C_{0,r}$, define the event $A_{n,r} := (T_{n,r} > M_{0,r} \epsilon_{n,r})$ and the test $\Psi_{n,r} := 1_{A_{n,r}}$. For

- $r = 1$, the triangular inequality $T_{n,1} \le \|\hat{p}_n - P_0^n \hat{p}_n\|_1 + \|P_0^n \hat{p}_n - p_0\|_1$ implies that, when $T_{n,1} > M_{0,1} \epsilon_{n,1}$, $\|\hat{p}_n - P_0^n \hat{p}_n\|_1 \ge T_{n,1} - \|P_0^n \hat{p}_n - p_0\|_1 > M_{0,1} \epsilon_{n,1} - \|P_0^n \hat{p}_n - p_0\|_1 \ge (M_{0,1} - C_{0,1}) \epsilon_{n,1}$;
- $r = \infty$, we have $|\hat{p}_n(x) - p_0(x)| \le |\hat{p}_n(x) - P_0^n \hat{p}_n(x)| + |P_0^n \hat{p}_n(x) - p_0(x)| \le \|\hat{p}_n - P_0^n \hat{p}_n\|_1 + \|P_0^n \hat{p}_n - p_0\|_\infty$ for every $x \in \mathbb{R}$. It follows that $T_{n,\infty} \le \|\hat{p}_n - P_0^n \hat{p}_n\|_1 + \|P_0^n \hat{p}_n - p_0\|_\infty$, which implies that, when $T_{n,\infty} > M_{0,\infty} \epsilon_{n,\infty}$, $\|\hat{p}_n - P_0^n \hat{p}_n\|_1 \ge T_{n,\infty} - \|P_0^n \hat{p}_n - p_0\|_\infty > M_{0,\infty} \epsilon_{n,\infty} - \|P_0^n \hat{p}_n - p_0\|_\infty \ge (M_{0,\infty} - C_{0,\infty}) \epsilon_{n,\infty}$.

Let $h : \mathcal{X}^n \to [0, 2]$ be the function defined as $h(X^{(n)}) := \|\hat{p}_n - P_0^n \hat{p}_n\|_1$. Thus, for each $r \in \{1, \infty\}$, when $T_{n,r} > M_{0,r} \epsilon_{n,r}$, the inequality $h(X^{(n)}) > (M_{0,r} - C_{0,r}) \epsilon_{n,r}$ holds. Therefore, to control the type-one error probability, it is enough to bound above the probability on the right-hand side of the following display

$$P_0^n \Psi_{n,r} \le P_0^n (h(X^{(n)}) > (M_{0,r} - C_{0,r}) \epsilon_{n,r}), \tag{A.1}$$

which can be done using McDiarmind's inequality, McDiarmid (1989). Given any $x^{(n)} := (x_1, \ldots, x_n) \in \mathcal{X}^n$, for each $1 \le i \le n$, let x_i be the ith component of $x^{(n)}$ and $x_i' := (x_i + \delta)$ a perturbation of the ith variable with $\delta \in \mathbb{R}$ so that $x_i' \in \mathcal{X}$. Letting e_i be the canonical vector with all zeros except for a 1 in the ith position, the vector with the perturbed ith variable can be expressed as $x^{(n)} + \delta e_i$. If

(a) the function h has *bounded differences*: for some non-negative constants $c_1, \ldots, c_n$,

$$\sup_{x^{(n)},\, x_i'} |h(x^{(n)}) - h(x^{(n)} + \delta e_i)| \le c_i, \quad 1 \le i \le n,$$

(b) $P_0^n h(X^{(n)}) = O(\epsilon_n)$,

then, for $C := \sum_{i=1}^n c_i^2$, by McDiarmind's bounded differences inequality,

$$\forall\, t > 0, \quad P_0^n(|h(X^{(n)}) - P_0^n h(X^{(n)})| \ge t) \le 2\exp\left(-2t^2/C\right).$$

We show that (a) and (b) are verified.

(a) Using the inequality $||a| - |b|| \le |a - b|$, setting $\Phi = K$ under condition a) of Definition (1),

$$\begin{aligned}\forall\, i \in \{1, \ldots, n\}, \quad \sup_{x^{(n)},\, x_i'} |h(x^{(n)}) - h(x^{(n)} + \delta e_i)| &= \sup_{x^{(n)},\, x_i'} \left| \int \left[\left| \frac{1}{n}\sum_{i=1}^n K_{J_n}(x,\, x_i) - K_{J_n}(p_0)(x) \right| \right.\right. \\ &\qquad \left.\left. - \left| \frac{1}{n}\sum_{i \ne i'}^n K_{J_n}(x,\, x_i) + \frac{1}{n} K_{J_n}(x,\, x_i') - K_{J_n}(p_0)(x) \right| \right] \mathrm{d}x \right| \\ &\le \sup_{x_i,\, x_i'} \frac{1}{n} \|K_{J_n}(\cdot,\, x_i) - K_{J_n}(\cdot,\, x_i')\|_1 \le \frac{2}{n}\|\Phi\|_1.\end{aligned}$$

Hence, h has bounded differences with $c_i = 2\|\Phi\|_1/n$, $1 \le i \le n$.

(b) By Theorem 5.1.5 in Giné and Nickl (2015), $P_0^n h(X^{(n)}) \le L\sqrt{2^{J_n}/n} = O(\epsilon_n)$, with the following upper bounds for the constant L:

- under conditions a) and b) of Definition (1), setting $\Phi = K$ in case a), $L \le \sqrt{2/(s-1)}\, \|\Phi^2\|_{L^1(\mu_s)}^{1/2} \|p_0\|_{L^1(\mu_s)}^{1/2}$;
- under conditions c) and d), $L \le C(\phi)(1 \vee \|p_0\|_{1/2})^{1/2}$, where the constant $C(\phi)$ only depends on ϕ.

For $\alpha \in (0,\, 1)$, taking $t = \sqrt{2}\alpha(M_{0,r} - C_{0,r})\epsilon_{n,r}$,

$$P_0^n(|h(X^{(n)}) - P_0^n h(X^{(n)})| \ge \sqrt{2}\alpha(M_{0,r} - C_{0,r})\epsilon_{n,r}) \le 2\exp\left(-\alpha^2(M_{0,r} - C_{0,r})^2 n\epsilon_{n,r}^2/\|\Phi\|_1^2\right).$$

By (b), there exists a constant $L' \ge L$ so that $P_0^n h(X^{(n)}) \le L'\epsilon_n = (L'/L_{n,r})\epsilon_{n,r}$. Hence, $|h(X^{(n)}) - P_0^n h(X^{(n)})| \ge h(X^{(n)}) - P_0^n h(X^{(n)}) \ge h(X^{(n)}) - (L'/L_{n,r})\epsilon_{n,r}$. Thus, for sufficiently large $L_{n,r}$ so that $[(M_{0,r} - C_{0,r}) - (L'/L_{n,r})] \ge \sqrt{2}\alpha(M_{0,r} - C_{0,r})$,

$$\begin{aligned}P_0^n(\|\hat{p}_n - P_0^n\hat{p}_n\|_1 \ge (M_{0,r} - C_{0,r})\epsilon_{n,r}) &\le P_0^n(|h(X^{(n)}) - P_0^n h(X^{(n)})| \ge \sqrt{2}\alpha(M_{0,r} - C_{0,r})\epsilon_{n,r}) \\ &\le 2\exp\left(-\alpha^2(M_{0,r} - C_{0,r})^2 n\epsilon_{n,r}^2/\|\Phi\|_1^2\right).\end{aligned}$$

We now provide an exponential upper bound on the type-two error probability. For $r \in \{1,\, \infty\}$, let $P \in \mathcal{P}_n$ be such that $\|p - p_0\|_r \ge M_r\epsilon_{n,r}$. For

- $r = 1$, when $T_{n,1} \le M_{0,1}\epsilon_{n,1}$,

$$\|p - p_0\|_1 \le \|p - P^n\hat{p}_n\|_1 + \|\hat{p}_n - P^n\hat{p}_n\|_1 + T_{n,1} \le \|p - P^n\hat{p}_n\|_1 + \|\hat{p}_n - P^n\hat{p}_n\|_1 + M_{0,1}\epsilon_{n,1},$$

- $r = \infty$, when $T_{n,\infty} \leq M_{0,\infty}\epsilon_{n,\infty}$,

$$\forall\, x \in \mathcal{X}, \quad |p(x) - p_0(x)| \leq \|p - P^n\hat{p}_n\|_\infty + \|\hat{p}_n - P^n\hat{p}_n\|_1 + T_{n,\infty} \leq \|p - P^n\hat{p}_n\|_\infty + \|\hat{p}_n - P^n\hat{p}_n\|_1 + M_{0,\infty}\epsilon_{n,\infty},$$

which implies that $\|p - p_0\|_\infty \leq \|p - P^n\hat{p}_n\|_\infty + \|\hat{p}_n - P^n\hat{p}_n\|_1 + M_{0,\infty}\epsilon_{n,\infty}$.

Summarizing, for $r \in \{1, \infty\}$, when $T_{n,r} \leq M_{0,r}\epsilon_{n,r}$, we have $\|p - p_0\|_r \leq \|p - P^n\hat{p}_n\|_r + \|\hat{p}_n - P^n\hat{p}_n\|_1 + M_{0,r}\epsilon_{n,r}$. If $\sup_{P\in\mathcal{P}_n} \|p - P^n\hat{p}_n\|_r = \sup_{P\in\mathcal{P}_n} \|p - K_{J_n}(p)\|_r \leq C_K\epsilon_{n,r}$, we have $\|\hat{p}_n - P^n\hat{p}_n\|_1 \geq \|p - p_0\|_r - \|p - P^n\hat{p}_n\|_r - M_{0,r}\epsilon_{n,r} \geq [M_r - (C_K + M_{0,r})]\epsilon_{n,r}$. Using, as before, McDiarmind's inequality with P playing the same role as P_0, we get that for a constant $\alpha \in (0, 1)$ small enough and $[M_r - (C_K + M_{0,r})] > 0$,

$$\sup_{P\in\mathcal{P}_n:\,\|p-p_0\|_r\geq M_r\epsilon_{n,r}} P^n(1 - \phi_{n,r}) = P^n(\|\hat{p}_n - p_0\|_r \leq M_{0,r}\epsilon_{n,r}) = P(\|\hat{p}_n - P^n\hat{p}_n\|_1 \geq [M_r - (C_K + M_{0,r})]\epsilon_{n,r})$$

$$\leq 2\exp\left(-\alpha^2[M_r - (C_K + M_{0,r})]^2 n\epsilon_{n,r}^2/\|\Phi\|_1^2\right).$$

We need that $\alpha^2[M_r - (C_K + M_{0,r})]^2/\|\Phi\|_1^2 \geq (C + 4)$, which implies that $[M_r - (C_K + M_{0,r})] \geq \alpha^{-1}\|\Phi\|_1\sqrt{C+4}$. This concludes the proof of the first assertion.

If the convergence in (2) holds for $r = 1$ and $r = \infty$, then the last assertion of the statement follows from the interpolation inequality: for every $1 < s < \infty$, $\|p - p_0\|_s \leq \max\{\|p - p_0\|_1, \|p - p_0\|_\infty\}$. □

Appendix B. Auxiliary results for Proposition 3

Following Parzen (1962), Watson and Leadbetter (1963), we adopt the subsequent definition.

Definition 2. *The Fourier transform or characteristic function of a Lebesgue probability density function p on $\mathbb{R}$, denoted by $\tilde{p}$, is said to* decrease algebraically of degree $\beta > 0$ *if*

$$\lim_{|t|\to\infty} |t|^\beta |\tilde{p}(t)| = B_p, \quad 0 < B_p < \infty.$$

The following lemma is essentially contained in the first theorem of section 3B in Watson and Leadbetter (1963).

Lemma 1. *Let $p \in \mathbb{L}^2(\mathbb{R})$ be a probability density with characteristic function that decreases algebraically of degree $\beta > 1/2$. Let $h \in \mathbb{L}^1(\mathbb{R})$ have Fourier transform $\tilde{h}$ satisfying*

$$I_\beta[\tilde{h}] := \int \frac{|1 - \tilde{h}(t)|^2}{|t|^{2\beta}}\mathrm{d}t < \infty. \tag{B.1}$$

Then, $\delta^{-2(\beta-1/2)}\|p - p * h_\delta\|_2^2 \to (2\pi)^{-1}B_p^2 \times I_\beta[\tilde{h}]$ as $\delta \to 0$.

Proof. Since $p \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, then $\|p * h_\delta\|_q \leq \|p\|_q\|h_\delta\|_1 < \infty$, for $q = 1, 2$. Thus, $p * h_\delta \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. It follows that $(p - p * h_\delta) \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. Hence,

$$\|p - p * h_\delta\|_2^2 = \frac{\delta^{2\beta-1}}{2\pi}\left\{B_p^2 \times I_\beta[\tilde{h}] + \int \frac{|1 - \tilde{h}(z)|^2}{|z|^{2\beta}}[|z/\delta|^{2\beta}|\tilde{p}(z/\delta)|^2 - B_p^2]\,\mathrm{d}z\right\},$$

where the second integral tends to 0 by the dominated convergence theorem because of assumption (B.1). □

In the next remark, which is essentially due to Davis (1977), section 3, we consider a sufficient condition for a function $h \in L^1(\mathbb{R})$ to satisfy requirement (B.1).

Remark 4. If $h \in L^1(\mathbb{R})$, then $\int_1^\infty t^{-2\beta}|1 - \tilde{h}(t)|^2\,\mathrm{d}t < \infty$ for $\beta > 1/2$. Suppose further that there exists an integer $r \geq 2$ such that $\int x^m h(x)\,\mathrm{d}x = 0$, for $m = 1, \ldots, r-1$, and $\int x^r h(x)\,\mathrm{d}x \neq 0$. Then,

$$\frac{[1 - \tilde{h}(t)]}{t^r} = -t^{-r}\int\left[e^{itx} - \sum_{j=0}^{r-1}\frac{(itx)^j}{j!}h(x)\right]\mathrm{d}x = -\frac{i^r}{(r-1)!}\int x^r h(x)\int_0^1(1-u)^{r-1}e^{itxu}\,\mathrm{d}u\,\mathrm{d}x \to -\frac{i^r}{r!}\int x^r h(x)\,\mathrm{d}x,$$

as $t \to 0$. For $r \geq \beta$, the integral $\int_0^1 t^{-2\beta}|1 - \tilde{h}(t)|^2\,\mathrm{d}t < \infty$. Conversely, for $r < \beta$, the integral diverges. Therefore, for $1/2 < \beta \leq 2$, any symmetric probability density h with finite second moment is such that $I_\beta[\tilde{h}] < \infty$ and condition (B.1) is verified.

References

Barron, A., Schervish, M.J., Wasserman, L., 1999. The consistency of posterior distributions in nonparametric problems. The Annals of Statistics 27 (2), 536–561.

Castillo, I., 2014. On Bayesian supremum norm contraction rates. The Annals of Statistics, 42 (5), 2058–2091.

Dattner, I., Reiß, M., Trabs, M., 2013. Adaptive quantile estimation in deconvolution with unknown error distribution. Technical Report. URL <`http://arxiv.org/pdf/1303.1698.pdf`>. *Bernoulli, to appear*

Davis, K.B., 1977. Mean integrated square error properties of density estimates. The Annals of Statistics 5 (3), 530–535.

Gao, F., van der Vaart, A., 2015. Posterior contraction rates for deconvolution of Dirichlet-Laplace mixtures. Technical Report. URL <`http://arxiv.org/pdf/1507.07412v1.pdf`>.

Ghosal, S., Ghosh, J.K., van der Vaart, A.W., 2000. Convergence rates of posterior distributions. The Annals of Statistics 28 (2), 500–531.

Ghosal, S., van der Vaart, A.W., 2001. Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities. The Annals of Statistics 29 (5), 1233–1263.

Ghosal, S., van der Vaart, A., 2007. Posterior convergence rates of Dirichlet mixtures at smooth densities. The Annals of Statistics 35 (2), 697–723.

Giné, E., Nickl, R., 2011. Rates of contraction for posterior distributions in L^r-metrics, $1 \leq r \leq \infty$. The Annals of Statistics 39 (6), 2883–2911.

Giné, E., Nickl, R., 2015. Mathematical Foundations of Infinite-dimensional Statistical Models. Cambridge Series in Statistical and Probabilistic Mathematics.

McDiarmid, C., 1989. On the method of bounded differences. In: Surveys in Combinatorics, Cambridge University Press, Cambridge, pp. 148–188.

Parzen, E., 1962. On estimation of a probability density function and mode. Annals of Mathematical Statistics 33 (3), 1065–1076.

Scricciolo, C., 2007. On rates of convergence for Bayesian density estimation. Scandinavian Journal of Statistics 34 (3), 626–642.

Scricciolo, C., 2011. Posterior rates of convergence for Dirichlet mixtures of exponential power densities. Electronic Journal of Statistics 5, 270–308.

Scricciolo, C., 2014. Adaptive Bayesian density estimation in L^p-metrics with Pitman-Yor or normalized inverse-Gaussian process kernel mixtures. Bayesian Analysis 9 (2), 475–520.

Shen, W., Tokdar, S.T., Ghosal, S., 2013. Adaptive Bayesian multivariate density estimation with Dirichlet mixtures. Biometrika 100 (3), 623–640.

Shen, X., Wasserman, L., 2001. Rates of convergence of posterior distributions. The Annals of Statistics 29 (3), 687–714.

Watson, G.S., Leadbetter, M.R., 1963. On the estimation of the probability density, I. Annals of Mathematical Statistics 34 (2), 480–491.