# Rates of contraction for posterior distributions in $L^r$ metrics, $1 \leq r \leq \infty$

**Evarist Giné, University of Connecticut, and**

**Richard Nickl, Cambridge University**

Let $\{P_\theta : \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^d$ or perhaps more general, be a family of probability measures on $\mathbb{R}$, and let $\Pi$ be a probability measure on $\Theta$, the prior distribution of $\theta$. Let $X_i$ be i.i.d. with law $\theta_0 \in \Theta$, $\theta_0$ unknown. In finite dimensional problems, under fairly general assumptions, the posterior distribution of $\theta$ given the data $X_i$, $i = 1, \ldots, n$, concentrates about the true value $\theta_0$ at the rate $n^{-1/2}$, relative to any norm, for example as a consequence of the Bernstein-von Mises theorem. For infinite dimensional parameters, it is known that the posterior could even be inconsistent (e.g., Diaconis and Freedman (1986)). Ghosal, Gosh and van der Vaart (2000), among others, developed techniques that allow to obtain best rates of concentration of the posterior in Hellinger metric in *density estimation*, assuming the prior is chosen adequately (not too large a support, enough mass on Kullback-Leibler divergence neighborhoods of the true value). In particular, van der Vaart and Zanten (2008) applied this theory to Gaussian process priors.

With the aim of obtaining rates of concentration in the stronger supremum norm distance, we replaced the entropy properties of the support of the prior by approximation theoretic properties related to wavelets, that are satisfied by Hölder (and Besov) balls, while keeping the second condition about enough mass on Kullback-Leibler neighborhoods. This allows for relatively more transparent proofs that use concrete tests based on wavelets and to get rates in sup norm. However, our rates in sup norm may not be optimal, in fact we obtain a continuous transition of rates, from optimal in $L^r$-norms, $1 \leq r \leq 2$, to steadily lower for $L^r$ as $r$ grows from 2 to $\infty$. Question: is this is the best one can do?

On the other hand, in non-parametric Gaussian regression the minimax rate of contraction in the sup norm (in all the $L^r$ norms simultaneously) obtains for certain diagonal priors.

Our proofs use Borell and Talagrand's inequalities and small balls probabilities. Here is a more detailed description.

**1. Rate of posterior contraction in sup norm in non-parametric Gaussian regression with conjugate prior.** Consider the problem of estimating a function observed in Gaussian white noise: given a noise level $1/\sqrt{n}, n \in \mathbb{N}$, we observe

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

for $f = f_0 \in L_2([0,1])$, where $B$ is Brownian motion on $[0,1]$. What we observe is $\int_0^1 h(t)dY^{(n)}(t)$ for $h$ the elements of an orthonormal basis of $L_2([0,1])$, hence, for any $h \in L_2$. One wishes to estimate $f_0$ based on this observation. Assuming $f_0 \in C^\alpha([0,1])$ for some $\alpha > 0$, we take a Cohen-Daubechies-Vial wavelet basis of regularity $r > \alpha$ (a modification of a Daubechies wavelet basis to fit the interval). Then, observing $Y^{(n)}$ is equivalent to observing its action on the basis:

$$
\begin{aligned}
y_k &= \int_0^1 \phi_k(t)dY^{(n)}(t) = \langle f, \phi_k \rangle + \frac{1}{\sqrt{n}} \int_0^1 \phi_k(t)dB(t) \\
&:= \theta_k + \frac{1}{\sqrt{n}} g_k, \quad k = 1, \ldots, N,
\end{aligned} \tag{2}
$$

$$
\begin{aligned}
y_{\ell k} &= \int_0^1 \psi_{\ell k}(t)dY^{(n)}(t) = \langle f, \psi_{\ell k} \rangle + \frac{1}{\sqrt{n}} \int_0^1 \psi_{\ell k}(t)dB(t) \\
&:= \theta_{\ell k} + \frac{1}{\sqrt{n}} g_{\ell k}, \; k = 1, \ldots, 2^\ell, \ell \geq 0,
\end{aligned} \tag{3}
$$

with the variables $g_k$, $g_{\ell k}$ all i.i.d. $N(0,1)$. So,
$Y^{(n)} = (y_k, y_{\ell k}) \in \ell_2$, where $y_k$ is $N(\theta_k, 1/n)$ and $y_{\ell k}$ is $N(\theta_{\ell k}, 1/n)$, all independent. $f_0$ 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 $\Pi$ on $L_2$ maps onto a prior, still denoted by $\Pi$, on the parameter space $\{\theta : \theta = (\theta_k, \theta_{\ell k})^t \in \ell_2\}$.

Take the prior

$$\Pi = \mathcal{L}\left(\sum_{k=1}^{N} g_k\phi_k + \sum_{\ell=0}^{\infty}\sum_{k=1}^{2^\ell}\sqrt{\mu_\ell}g_{\ell k}\psi_{\ell k}\right) \tag{4}$$

in $L_2([0,1])$, with the $g$'s i.i.d. $N(0,1)$ and with

$$\mu_0 = 1, \mu_\ell = \ell^{-1}2^{-\ell(2\alpha+1)} \ \forall \ell > 0. \tag{5}$$

(or $\Pi = \mathcal{L}\left(g_k, \sqrt{\mu_\ell}g_{\ell k}\right)^t$ in $\ell_2$, for the wavelet coefficients of $f$). The posterior $\hat{\Pi}_n^Y = \Pi(\cdot|Y^{(n)})$ is then the law of $\theta$ given the observed process $Y^{(n)}$, and, just as in one dimension, it is also Gaussian, with mean and covariance

$$\hat{\theta}(Y) = E_\Pi(\theta|Y^{(n)}) = \Sigma(\Sigma + I/n)^{-1}Y^{(n)} = \Sigma(\Sigma+I/n)^{-1}(y_k; y_{\ell k})^t \tag{6}$$

$$\Sigma|Y^{(n)} = \Sigma(n\Sigma + I)^{-1}, \tag{7}$$

where $\Sigma$ is the covariance of $\Pi$, in our case, diagonal.

In one dimension, $Y^{(n)} = \theta + g/\sqrt{n}$ is $N(\theta, 1/n)$ and $\theta$ is $N(0, \mu)$ independent of $g$, and it is well known and easy to see that the posterior distribution of $\theta$ given $Y^{(n)} = y$ is

$$N\left(\frac{\mu y}{\mu + 1/n}, \frac{\mu}{n\mu + 1}\right).$$

The equations (6) are just this in Hilbert space.

**Theorem 1.** *Let $0 < \alpha < r$ and let $\Pi$ be the Gaussian prior on $L_2([0,1])$ defined by (4) based on a Cohen-Daubechies-Vial wavelet basis of $L^2([0,1])$ of smoothness at least $r$. Let $f_0 \in C^\alpha([0,1])$, let*

$$\varepsilon_n = \left(\frac{\log n}{n}\right)^{\alpha/(2\alpha+1)},$$

*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, $\alpha$ and $\|f_0\|_{\alpha,\infty}$ such that, for every $M_0 < M < \infty$, and for all $n \in \mathbb{N}$,*

$$E_{Y_0^{(n)}} \Pi\left(f : \|f - f_0\|_\infty > M\varepsilon_n \middle| Y_0^{(n)}\right) \leq \exp(-C^2(M - M_0)^2 \log n). \tag{8}$$

This is a result on the rate of posterior concentration about the true parameter $f_0$ in the sup norm. Some consequences follow:

The posterior mean given $Y_0^{(n)}$ is a formal Bayes estimator of $f_0$ and equals

$$E_\Pi(f|Y_0^{(n)}) = \sum_{k=1}^{N} \frac{1}{1+1/n} y_k \phi_k + \sum_{\ell=0}^{\infty} \sum_{k=1}^{2^\ell} \frac{\mu_\ell}{\mu_\ell + 1/n} y_{\ell k} \psi_{\ell k} \tag{9}$$

with $y_k = \langle f_0, \phi_k \rangle + g_k/\sqrt{n}$ and $y_{\ell k} = \langle f_0, \psi_{\ell k} \rangle + g_{\ell k}/\sqrt{n}$.

**Corollary 1.** *Under the same conditions and notation as in Theorem 1,*

$$\Pr\left( \|E_\Pi(f|Y_0^{(n)}) - f_0\|_\infty > 2M\varepsilon_n \right) = 0 \tag{10}$$

*from some $n$ onwards, in fact, at least for all $n$ such that* $\exp(-C^2(M - M_0)^2 \log n) < 1/2$.

Theorem 1 is thus best possible as its consequence for the Bayes estimator $E_\Pi(f|Y_0)$ is.

Not surprisingly, the proof of the theorem follows easily by the Borel-Sudakov-Tirelson exponential inequality and a simple estimate of the supremum of a Gaussian sequence, and that of the corollary uses also Anderson's inequality.

**2. Density estimation.** Let $\mathcal{P}$ be a class of bounded continuous probability densities on $[0,1]$, and let $X_1, \ldots, X_n$ be a random sample drawn from some unknown probability density $p_0$ with joint law $P_0^{\mathbb{N}}$. Suppose one is given a prior probability distribution $\Pi$ defined on some $\sigma$-algebra $\mathcal{B}$ of $\mathcal{P}$. The posterior is the random probability measure defined on $\mathcal{B}$ by

$$\Pi(B|X_1, \ldots, X_n) = \frac{\int_B \prod_{i=1}^n p(X_i) d\Pi(p)}{\int \prod_{i=1}^n p(X_i) d\Pi(p)}, \quad B \in \mathcal{B}.$$

We describe our results on two general examples.

**2.1. Uniform wavelet series.** Suppose an apriori upper bound on the Hölder norm $\|p_0\|_{\alpha,\infty}$ is available, so that the prior can be chosen to have bounded support in $\mathcal{C}^\alpha([0,1])$.

An example is obtained by uniformly distributing wavelet coefficients on a Hölder ball. Let $\{\phi_k, \psi_{\ell k}\}$ be a $N$-regular Cohen-Daubechies-Vial-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

$$U_\alpha(x) = \sum_k u_{0k}\phi_k(x) + \sum_{\ell=J_0}^{\infty}\sum_k 2^{-\ell(\alpha+1/2)} u_{\ell k}\psi_{\ell k}(x), \qquad (11)$$

which has trajectories in $\mathcal{C}^\alpha([0,1]) \subset L^r([0,1]), 1 \le r \le \infty$, almost surely since $C^\alpha([0,1]) = B^\alpha_{\infty,\infty}([0,1])$ has a definition in terms of the coefficients of wavelet expansions,

$$\|f\|_{\alpha,\infty,\infty} = \sup_k |\alpha_k(f)| + \sup_{\ell,k} 2^{\ell(\alpha+1/2)}|\beta_{\ell,k}(f)|.$$

Since $\|U_\alpha\|_\infty \leq C(B, \alpha, \psi)$ 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 $\Pi^\alpha$.

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

$$\Pi^\alpha \left\{ p \in \mathcal{P} : \|p - p_0\|_r \geq M n^{-\frac{\alpha - 1/2 + 1/\bar{r}}{2\alpha+1}} (\log n)^\eta \Big| X_1, \ldots, X_n \right\} \to^{P_0^{\mathbb{N}}} 0. \tag{12}$$

For $1 \leq r \leq 2$, the concentration rate $n^{-\frac{\alpha}{2\alpha+1}}(\log n)^{\eta}$ is best possible up to the log term, but not necessarily for $\alpha > 2$.

**2.2. Gaussian process priors.** A variety of Gaussian process priors have been considered in the nonparametric Bayes literature recently, mainly by van der Vaart and coauthors. For simplicity, we restrict consideration to integrated Brownian motions.

**Definition 1.** *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 $\alpha$, $B_\alpha$ is defined as the $[\alpha]$-fold integral*

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

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

We would like to define our prior on densities as the probability law of the random process

$$\frac{e^{B_\alpha}}{\int_0^1 e^{B_\alpha(t)}dt} \tag{13}$$

(Lenk, 1991) but we must make two corrections: First, since $B_\alpha^{(k)}(0) = 0$ a.s., $k \leq [\alpha]$, would impose unwanted conditions on the value at zero of the density and its derivatives, we should release $B_\alpha$ at zero, i.e., 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 $B_\alpha$ (see van der Vaart and Zanten (2008)). In order to deal with bounded densities, we introduce a second modification to (13), and define our prior (on the Borel sets of $C([0,1])$) as

$$\Pi = \mathcal{L}\left(\frac{e^{\bar{B}_\alpha}}{\int_0^1 e^{\bar{B}_\alpha(t)}dt}\middle| \|\bar{B}_\alpha\|_\infty \leq c\right) \tag{14}$$

where $c$ is a fixed arbitrary positive constant.

**Proposition 2.** *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* $\Pi$ *be the prior defined by (14) where* $\alpha$ *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.*

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

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

The result in Proposition 2 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.

The properties of these Gaussian processes that make them suitable for these results are the following (only stated for integrated Brownian motion).

One is that $\bar{B}_\alpha$ has its trajectories a.s. in $C^{\alpha,\infty,1/2}$, the space of bounded continuous functions with $[\alpha]$ bounded derivatives and with the $[\alpha]$-th derivative with modulus of continuity $t^{\alpha-[\alpha]}(\log t^{-1})^{1/2}$, and that Borell's concentration inequality applies for the norm in this space; we also need the more refined inequality about concentration in sup-norm-$\varepsilon$-neighborhoods of RKHS balls -in the case of released at zero integrated B.M., for each $\alpha$, the RKHS is precisely the Sobolev space $H^{\alpha+1/2}$-.

The second property we use is that the small ball estimate

$$\Pr\left\{\|\bar{B}_\alpha - w\|_\infty < \varepsilon\right\} \geq e^{-\phi_\alpha(\varepsilon)},\ \phi_\alpha(\varepsilon) = O(\varepsilon^{-1/\alpha}) \text{ as } \varepsilon \to 0,$$

holds for all $w \in C^\alpha$ (Li and Linde, 1999).

These properties, suitably modified, are inherited by the processes defining the priors.

**2.3. A general result.** The previous propositions are consequences of more general results, and the properties of Gaussian processes just mentioned are used for checking their conditions. Here is one of the two general results we have. Will only consider $T = [0, 1]$ ($T = \mathbb{R}$ is also possible). We need

**Condition 1.** *Let $T = [0, 1]$. The sequence of operators $K_j(x, y) = 2^j K(2^j x, 2^j y); x, y \in T, j \geq 0$ is called an admissible approximating sequence if $K(x, y) = K(x - y)$ is a standard convolution kernel of bounded p-variation for some $p \geq 1$ or if $K(x, y) = \sum_k \phi_k(x)\phi_k(y)$ is the projection kernel of a Cohen-Daubechies-Vial (CDV) wavelet basis (boundary corrected Daubechies wavelets) -or a Daubechies wavelet basis if $T = \mathbb{R}$-.*

**Theorem 2.** *Let $\mathcal{P}$ be a set of probability densities on $[0,1]$, and let $\Pi_n$ be priors defined on some $\sigma$-algebra of $\mathcal{P}$ for which the maps $p \mapsto p(x)$ are measurable for all $x \in [0,1]$. Let $X_i$ be i.i.d. with law $P_0$ and density $p_0 \in \mathcal{P}$. Let $1 \le r \le \infty$ 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*

$$\delta_n = \varepsilon_n (n\varepsilon_n^2)^{\frac{1}{2}-\frac{1}{2r}} \gamma_n \tag{15}$$

*for some $\gamma_n \ge 1$ $\forall n$. Let $J_n$ be any sequence satisfying $2^{J_n} \le cn\varepsilon_n^2$ 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*

$$\left\{p \in \mathcal{P} : \ \|K_{J_n}(p) - p\|_r \le C(K)\delta_n\right\}. \tag{16}$$

*where $C(K)$ is a constant that depends only on the operator kernel $K$. Assume there exists $C > 0$ such that, for every $n$ large enough,*

*1)* $\Pi_n(\mathcal{P} \setminus \mathcal{P}_n) \le e^{-(C+4)n\varepsilon_n^2}$ *and*

*2)* $\Pi_n \left\{ p \in \mathcal{P} : -P_0 \log \frac{p}{p_0} \le \varepsilon_n^2, \quad P_0 \left(\log \frac{p}{p_0}\right)^2 \le \varepsilon_n^2 \right\} \ge e^{-Cn\varepsilon_n^2}$.

*Let* $p_0 \in L^r([0,1])$ *s.t.* $\|K_{J_n}(p_0) - p_0\|_r = O(\delta_n)$*. If* $\delta_n \to 0$ *as* $n \to \infty$*, then there exists* $M < \infty$ *such that*

$$\Pi_n \left\{ p \in \mathcal{P} : \|p - p_0\|_r \ge M\delta_n \middle| X_1, \ldots, X_n \right\} \to 0 \quad \textit{as } n \to \infty \tag{17}$$

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

If we also impose that $p_0$ is bounded and that

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

as $n \to \infty$ in $P_0^{\mathbb{N}}$-probability for some $B < \infty$, then, for $1 < r < \infty$, we can improve the rate to $\delta_n = \varepsilon_n (n\varepsilon_n^2)^{\frac{1}{2} - \frac{1}{r \vee 2}} \gamma_n$, assuming $\varepsilon_n = O(\gamma_n (n\varepsilon_n^2)^{1/(r \wedge 2) - 1})$.

Note that the rates interpolate between $\delta_n = \varepsilon_n$ for $r = 1$ (or, in the bounded case, for $r = 2$) and $\delta_n = \sqrt{n}\varepsilon_n^2$ for $r = \infty$ (slower than $\varepsilon_n$ as $\sqrt{n}\varepsilon_n \to \infty$).

Condition 2) handles the denominator in

$$\Pi_n \left\{ p \in \mathcal{P} : \|p - p_0\|_r \geq M\delta_n | X_1, \ldots, X_n \right\}$$
$$= \frac{\int_{\|p-p_0\|_r \geq M\delta_n} \prod_{i=1}^n \frac{p}{p_0}(X_i) d\Pi_n(p)}{\int_{\mathcal{P}} \prod_{i=1}^n \frac{p}{p_0}(X_i) d\Pi_n(p)}$$

in the sense that, basically by Chebyshev's inequality (Ghosal, Ghosh and van der Vaart, 2000), if $\Pi$ is supported by the set in condition 2), then one has, for all $c > 0$,

$$P_0^n \left\{ \int \prod_{i=1}^n \frac{p}{p_0}(X_i) d\Pi(p) \leq e^{-(1+c)n\varepsilon_n^2} \right\} \leq \frac{1}{c^2 n \varepsilon_n^2}.$$

The numerator can be handled with the existence of 'tests' -indicator functions- $\phi_n = \phi_n(X_1, \ldots, X_n; p_0)$ such that

$$\lim_n P_0^n \phi_n = 0 \text{ and } \sup_{p \in \mathcal{P}_n : \|p - p_0\|_r \geq M\delta_n} P^n(1 - \phi_n) \leq 2e^{-(C+4)n\varepsilon_n^2}$$

(for $n$ large enough). Van der Vaart et al. used tests specific for the Hellinger distance derived from some abstract tests proved to exist by LeCam and by Birgé by means of entropy conditions. Our tests have the form

$$\phi_n = I(\|\hat{p}_n - p_0\|_r > M\delta_n\},$$

where $\hat{p}_n$ is a kernel or wavelet type density estimator. Then, the proofs of the above inequalities are obtained via Talagrand's inequality and moment bounds for $\|(P_n - P)K_j(x, \cdot)\|_r$, $1 \leq r \leq \infty$, together with control of bias term $\|P^n\hat{p}_n - p\|_r = \|K_{J_n}p - p\|_r$, $p = p_0$ and $p \in \mathcal{P}_n$, using (16).

In particular, $\mathcal{P}$ must satisfy some good theoretic approximation properties, like the $C^\alpha$ or the Besov spaces.

The different rates for different $r$ come from the estimation of the $p_0$-variance of $(K_j * f)(X)$ (or with the dominating convolution kernel $\Phi_j$ in the wavelet case), $f \in L^s$, $s$ conjugate of $r$, $K_j(x,y) = 2^j K(2^j x, 2^j y)$, using Young's inequalities: for $p_0 \in L^r$, we can only get

$$(E_{p_0}(K_j * f)^2)^{1/2} \le \|p_0\|_r^{1/2} \|K_j * f\|_{2s} \le C 2^{j(1/2-1/2r)},$$

but for $1 < r < 2$ and $p_0$ bounded, hence in $L^{s/(s-2)}$,

$$E_{p_0}(K_j * f)^2 \le \|p_0\|_{s/(s-2)} \|K_j * f\|_s \le \|p_0\|_{s/(s-2)} \|K_j\|_1^2 \|f\|_s^2 = const.$$

For $r = \infty$, $(E_{p_0} K_j(x-X)^2)^{1/2} \le C(K) \|p_0\|_\infty^{1/2} 2^{j/2}$.

Regarding the use of small balls probabilities in the verification of condition 2) of the theorem, first we observe that if $\|w_0\|_\infty \leq c/2$ and $\varepsilon \leq c/2$,

$$\Pr\left\{\|\bar{B}_\alpha - w_0\|_\infty| < \varepsilon \Big| \|\bar{B}_\alpha\|_\infty < c\right\} = \frac{e^{-\phi_{w_0}(\varepsilon)}}{\Pr\{\|\bar{B}_\alpha\|_\infty \leq c\}};$$

next, that setting $I(w) = e^w / \int_0^1 e^w(t)dt$, that with elementary computations, if $w_0 = \log p_0$ (so, $p_0 = I(w_0)$) and $p = I(w)$ with $\|w\|_\infty \leq c$, then

$$-P_0 \log \frac{p}{p_0} \leq R(c)\|w - w_0\|_\infty^2,\ P_0 \left(\log \frac{p}{p_0}\right)^2 \leq R(c)\|w - w_0\|_\infty^2$$

hence, for any $\varepsilon$ s.t. $R^{-1/2}\varepsilon < c/2$, if $w = \bar{B}_\alpha(\omega)$,

$$\Pi\left\{p \in \mathcal{P} : -P_0 \log \frac{p}{p_0} \leq \varepsilon^2,\ P_0 \left(\log \frac{p}{p_0}\right)^2 \leq \varepsilon^2\right\}$$

$$\geq \Pr\left\{\|\bar{B}_\alpha - w_0\|_\infty| < R^{-1/2}\varepsilon \Big| \|\bar{B}_\alpha\|_\infty < c\right\} = C(c)e^{-\phi_{w_0}(\varepsilon)}.$$

With $\varepsilon_n = 1/n^{\alpha/(2\alpha+1)}$, so that $\varepsilon_n^{-1/\alpha} = n\varepsilon_n^2$, we get the last quantity bounded by $e^{-Cn\varepsilon_n^2}$, which verifies Condition 2) in the general theorem, and modulo many details, yields Proposition 2.

Research to appear in Ann. Statist., [11].

**References**

BOUSQUET, O. (2003). Concentration inequalities for sub-additive functions using the entropy method. In: *Stochastic inequalities and applications., Progr. Probab.* **56**, E. Giné, C. Houdré, D. Nualart, eds., Birkhäuser, Boston, 213-247.

BORELL, C. (1975). The Brunn-Minkowski inequality in Gauss space. *Invent. Math.*, 30(2):207–216.

CIESIELSKI, Z. (1961). Hölder conditions for realizations of Gaussian processes. *Trans. Amer. Math. Soc.*, 99:403–413, 1961.

CIESIELSKI,Z., KERKYACHARIAN, G., and ROYNETTE, B. (1993). Quelques espaces fonctionnels associés à des processus gaussiens. *Studia Math.*, 107(2):171–204.

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

Donoho, D. L.; Johnstone, I. M.; Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24** 508-539.

Einmahl, U. and Mason, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.* **13** 1-37.

Ghosal, S., Ghosh, J.K., and van der Vaart, A.W. (2000). Convergence rates of posterior distributions. *Ann. Statist.*, 28(2):500–531.

Giné, E. and Guillou, A. (2001). On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals. *Ann. Inst. H. Poincaré Probab. Statist.* **37** 503-522.

Giné, E. and Nickl, R. (2009b). Uniform limit theorems for wavelet density estimators. *Ann. Probab.* **37** 1605-1646.

Giné, E. and Nickl, R. (2011). Rates of contraction for posterior distributions in $L^r$ metrics, $1 \leq r \leq \infty$.*Ann. Statist.*, to appear.

Härdle, W.; Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications.* Lecture Notes in Statistics **129**. Springer, New York.

Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33** 1060-1077.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes.* Springer, New York

Lenk, P.J. (1991). Towards a practicable Bayesian nonparametric density estimator. *Biometrika*, 78(3):531–543.

Li, W.V. and Linde, W. (1999). Approximations, metric entropy and small ball estimates for Gaussian measures. *Ann. Probab.* **27** 1556-1578.

Lounici, K. and Nickl, R. (2010). Global uniform risk bounds for wavelet deconvolution estimators. *Ann. Statist.*, to appear.

Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505-563.

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(3):1435–1463.

Zhao, L.H. (2000). Bayesian aspects of some nonparametric problems. *Ann. Statist.*, 28(2):532–552.