Probab. Theory Relat. Fields (2008) 141:333–387
DOI 10.1007/s00440-007-0087-9

Uniform central limit theorems for kernel density estimators

Evarist Giné · Richard Nickl

Received: 12 December 2006 / Revised: 5 April 2007 / Published online: 21 July 2007

Abstract Let $\mathbb{P}_n * K_{h_n}(x) = n^{-1}h_n^{-d} \sum_{i=1}^n K((x - X_i)/h_n)$ be the classical kernel density estimator based on a kernel K and n independent random vectors X_i each distributed according to an absolutely continuous law $\mathbb{P}$ on $\mathbb{R}^d$. It is shown that the processes $f \longmapsto \sqrt{n}\int f d(\mathbb{P}_n * K_{h_n} - \mathbb{P})$, $f \in \mathcal{F}$, converge in law in the Banach space $\ell^\infty(\mathcal{F})$, for many interesting classes $\mathcal{F}$ of functions or sets, some $\mathbb{P}$-Donsker, some just $\mathbb{P}$-pregaussian. The conditions allow for the classical bandwidths h_n that simultaneously ensure optimal rates of convergence of the kernel density estimator in mean integrated squared error, thus showing that, subject to some natural conditions, kernel density estimators are 'plug-in' estimators in the sense of Bickel and Ritov (Ann Statist 31:1033–1053, 2003). Some new results on the uniform central limit theorem for smoothed empirical processes, needed in the proofs, are also included.

Mathematics subject classification (2000) Primary: 62G07; Secondary: 60F05

Keywords Kernel density estimation · Uniform central limit theorem · Plug-in property · Smoothed empirical processes

1 Introduction

Let $X_1, \ldots, X_n$ be independent identically distributed random vectors with common law $\mathbb{P}$ on $\mathbb{R}^d$ and let $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ be the corresponding empirical measure. If nothing is known about the probability measure $\mathbb{P}$, one typically estimates $\mathbb{P}$ by $\mathbb{P}_n$,

E. Giné (✉) · R. Nickl
Department of Mathematics, University of Connecticut, Storrs, CT 06269-3009, USA
e-mail: gine@math.uconn.edu

R. Nickl
e-mail: nickl@math.uconn.edu

and this can be justified in many ways, in particular because the approximation error $\mathbb{P}_n - \mathbb{P}$ is asymptotically of the order $n^{-1/2}$ uniformly over many classes of functions $\mathcal{F}$, that is,

$$\sqrt{n}\left(\int f d\mathbb{P}_n - \int f d\mathbb{P}\right) = O_{\mathbb{P}}(1), \tag{1}$$

in fact, the processes $f \mapsto \sqrt{n}\int f d(\mathbb{P}_n - \mathbb{P})$, $f \in \mathcal{F}$, converge in law to a nice Gaussian process in $\ell^\infty(\mathcal{F})$ (the $\mathbb{P}$-Brownian bridge indexed by $\mathcal{F}$). Such classes of functions are known as $\mathbb{P}$-Donsker classes. If on the other hand $\mathbb{P}$ has a density p_0 with respect to Lebesgue measure, the empirical measure $\mathbb{P}_n$, which is a discrete random measure, is not adequate for estimating p_0. Rather, p_0 is then estimated in a variety of other ways, one of the oldest being by kernel smoothing of $\mathbb{P}_n$, that is, by

$$p_n(x) = \mathbb{P}_n * K_{h_n}(x) = \frac{1}{nh_n^d}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right) \tag{2}$$

where K is integrable and integrates to 1, $K_{h_n}(x) := h_n^{-d}K(x/h_n)$, and $h_n \to 0$, $h_n > 0$.

Under suitable conditions it is well known (see e.g. [29, Sect. 24.3]) that p_n is an optimal estimator of p_0 with respect to the mean integrated squared error (MISE) in the sense of achieving the minimax rate over all estimators for densities in certain classes. Similar results can be shown for rates of convergence of p_n to p_0 in the L^1-distance, cf. [5]. However these are not the only ways in which to measure performance of estimators for p_0, in particular given that the empirical measure performs already very good in the sense mentioned above (see (1)). The question arises as to whether $p_n(x)dx$, at least for suitably selected K and h_n, is not only best in the minimax sense for the MISE and the L^1 error, but is also as good as $\mathbb{P}_n$ in the sense that

$$\sqrt{n}\sup_{f\in\mathcal{F}}\left|\int f(x)p_n(x)dx - \int f(x)p_0(x)dx\right| = O_{\mathbb{P}}(1), \tag{3}$$

or, more specifically, in the sense that the processes $f \mapsto \sqrt{n}\int f(x)(p_n - p_0)(x)dx$, $f \in \mathcal{F}$, converge in law to a Gaussian process in the space $\ell^\infty(\mathcal{F})$ for many classes of functions $\mathcal{F}$. In fact the question as to whether (3) holds has given rise in the literature to several papers on 'smoothed empirical processes' (see [21,23,28,31], among others). Although these authors prove uniform central limit theorems (henceforth UCLTs) that apply to certain kernel density estimators, none of them obtains explicit results for the most interesting bandwidths h_n that are necessary to simultaneously obtain optimality in MISE and L^1 -error.

A first object of this article is to contribute to this literature in three ways: (1) by revisiting the very nice central limit theorem of van der Vaart [29], whose proof, we believe, requires clarification at a crucial point, and whose statement requires a (minor) additional condition to ensure correctness. (2) Van der Vaart's theorem concerns $\mathbb{P}$-Donsker classes only and, as indicated by Radulović and Wegkamp [21], the smoothed empirical process should converge to a Gaussian process even in situations when the regular empirical processes are not tight, as long as the limiting Gaussian process exists and is ‘nice’ (that is, for classes of functions $\mathcal{F}$ that are $\mathbb{P}$-pregaussian but not necessarily $\mathbb{P}$-Donsker). Proving such a result is very different from the Donsker case, in which case the UCLT is proved by showing that the smoothed empirical is close to the regular non-smoothed empirical, which is not a viable strategy of proof in the non-Donsker case. We provide a general uniform CLT for the smoothed empirical process indexed by $\mathbb{P}$-pregaussian classes of functions that is much more applicable than the Radulović–Wegkamp result. Then, (3), we give meaningful examples and applications of the previous results to particular (not necessarily Donsker-) classes of functions, including bounded-variation, Hölder and Lipschitz, Sobolev, and Besov classes of functions as well as many classes of sets in $\mathbb{R}^d$, of interest in statistics. In particular we obtain CLTs for kernel density estimators uniform over a continuous scale of Besov classes of functions that range from very irregular pregaussian (non-Donsker) to uniform Donsker.

To return to the question raised in (3) above, note that if p_0 has smoothness of order t (to be defined below), optimality in the MISE and L^1-error is achieved for very concrete h_n ($h_n \simeq n^{-1/(2t+d)}$) and for kernels satisfying certain properties (kernels of ‘order’ $r \geq t$, also to be defined below), and the question above was whether one has (3)—or even the UCLT—in precisely the situation when MISE and L^1-error optimality occurs. Estimators simultaneously satisfying these two kinds of optimality are called *plug-in* estimators by Bickel and Ritov [3], who introduced the notion. As a second object of this article—statistically more relevant than the first—we show that kernel density estimators can be made to satisfy this ‘plug-in property’ for all the classes of functions mentioned in the previous paragraph—even for some Besov classes that are $\mathbb{P}$-pregaussian but not $\mathbb{P}$-Donsker – by just increasing the order of the kernel by $d/2$. We thus provide concrete, meaningful examples of estimators satisfying the plug-in property for a large variety of different classes of functions.

We should point out that the main difficulty in proving the plug-in property, aside from the use of the right UCLTs for the ‘variance term’, resides in the treatment of the bias. The bias term in the MISE-case is treated by combining the order of the kernel K and the smoothness of p_0, but this is not so simple if one is interested in the UCLT, and we must find ways to use the order of K with the *combined* smoothness of p_0 and the elements of the class $\mathcal{F}$. This is already implicit in comments in Bickel and Ritov [3] regarding the estimation of the distribution function ($\mathcal{F} = \{\mathbf{1}_{(-\infty,t]} : t \in \mathbb{R}\}$): in this case one must essentially use that indicators of intervals are differentiable in the sense of distributions.

Nickl [18,19] studies UCLTs and the plug-in property for other density estimators, particularly, for nonparametric maximum likelihood estimators, sieved maximum likelihood estimators (for trigonometric sieves), as well as trigonometric series estimators. The UCLTs proved in this article constitute another step in the direction of a better understanding of the central limit theorem problem for density estimators.

This article is organized as follows: Sect. 2 introduces some notation and definitions, Sect. 3 contains the uniform central limit theorems for smoothed empirical processes, Sect. 4 deals with kernel density estimators over particular classes of functions and contains the more statistically relevant results of the article, and Sect. 5 collects the more technical proofs.

2 Notation and definitions

For an arbitrary (non-empty) set M, $\ell^\infty(M)$ will denote the Banach space of bounded real-valued functions H on M normed by

$$\|H\|_M := \sup_{m\in M} |H(m)|.$$

We denote by $\mathcal{B}_S$ the Borel-σ-algebra of a (non-empty) topological space S. For $h : \mathbb{R}^d \to \mathbb{R}$ a Borel-measurable function and μ a Borel measure on $\mathbb{R}^d$, we set $\mu h := \int_{\mathbb{R}^d} h d\mu$ and $\|h\|_{p,\mu} := (\int_{\mathbb{R}^d} |h|^p \, d\mu)^{1/p}$, $1 \le p \le \infty$ (where $\|h\|_{\infty,\mu}$ denotes the μ-essential supremum of $|h|$). We write $\mathcal{L}^p(\mathbb{R}^d, \mu)$ for the vector space of all Borel-measurable functions $h : \mathbb{R}^d \to \mathbb{R}$ that satisfy $\|h\|_{p,\mu} < \infty$, and $L^p(\mathbb{R}^d, \mu)$ for the corresponding Banach spaces of equivalence classes $[h]_\mu$ (modulo equality μ-almost everywhere), $h \in \mathcal{L}^p(\mathbb{R}^d, \mu)$. We shall sometimes omit the underlying space $\mathbb{R}^d$ and just write $\mathcal{L}^p(\mu)$. The symbol λ will always denote Lebesgue measure on $\mathbb{R}^d$. Also we will use obvious analogues of these spaces and norms for complex valued functions.

The symbol $\mathsf{C}(\mathbb{R}^d)$ denotes the Banach space of bounded real-valued continuous functions on $\mathbb{R}^d$ normed by the usual sup-norm $\|\cdot\|_\infty$. Let $\alpha = (\alpha_1, \dots, \alpha_d)$ be a multi-index of nonnegative integers α_i, set $|\alpha| = \sum_{i=1}^d \alpha_i$, and let

$$D^\alpha = \frac{\partial^{|\alpha|}}{(\partial x_1)^{\alpha_1} \cdots (\partial x_d)^{\alpha_d}}$$

denote the partial differential operator of order α. For $\alpha = 0$ set $D^\alpha = id$, and if $d = 1$ we set $D^1 = D$. For any nonnegative integer s, $\mathsf{C}^s(\mathbb{R}^d)$ denotes the Banach space of all bounded continuous real-valued functions that are s-times continuously differentiable on $\mathbb{R}^d$, equipped with the norm

$$\|f\|_{s,\infty} = \sum_{0\le|\alpha|\le s} \left\|D^\alpha f\right\|_\infty .$$

The Hölder spaces, for noninteger $s > 0$, are defined as ($[s]$ denotes the integer part of s)

$$\mathsf{C}^s(\mathbb{R}^d) = \left\{ f \in \mathsf{C}(\mathbb{R}^d) : \|f\|_{s,\infty} := \sum_{0\le|\alpha|\le[s]} \left\|D^\alpha f\right\|_\infty + \sum_{\alpha:|\alpha|=[s]} \sup_{x\ne y} \frac{|D^\alpha f(x) - D^\alpha f(y)|}{|x-y|^{s-[s]}} < \infty \right\}. \tag{4}$$

The symbol $\mathsf{C}_0(\mathbb{R}^d)$ denotes the closed subspace of $\mathsf{C}(\mathbb{R}^d)$ consisting of bounded continuous real-valued functions f that vanish at infinity. Denote by $\mathsf{C}_0(\mathbb{R}^d)'$ the (topological) dual space of $\mathsf{C}_0(\mathbb{R}^d)$ normed by the usual (operator) norm $\|\cdot\|'_{C_0}$. Then

$M(\mathbb{R}^d) = \mathsf{C}_0(\mathbb{R}^d)'$ is the space of signed Borel measures of finite variation on $\mathbb{R}^d$, and, as is well known, $\|\mu\|'_{C_0} = \|\mu\|$, where $\|\mu\| := |\mu|(\mathbb{R}^d)$ is the total variation norm of μ, $|\mu|$ being the total variation measure of $\mu \in M(\mathbb{R}^d)$.

The convolution of two signed Borel measures μ and ν on $\mathbb{R}^d$ is defined by $\mu * \nu(E) = \mu \times \nu(T^{-1}(E))$ where $E \in \mathcal{B}_{\mathbb{R}^d}$ and where $T : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is addition $T(x, y) = x + y$, cf. Sect. 8.6 in [9]. This gives the usual convolution $\int_{\mathbb{R}^d} f(x-y)g(y)dy$ of functions f, g (if it is defined) by setting $d\mu = f d\lambda$ and $d\nu = g d\lambda$. We recall here a few well-known facts on convolution of measures and functions on $\mathbb{R}^d$, that can be found, e.g., in Sect. 8.6 in [9] or in Sect. III.1.8 in [15]: If $f \in \mathcal{L}^p(\mathbb{R}^d, \lambda)$ $(1 \le p \le \infty)$ and $g \in \mathcal{L}^q(\mathbb{R}^d, \lambda)$ $(1 \le q \le \infty)$ with $1/p + 1/q = 1$, then $f * g(x)$ defines an element of $\mathsf{C}(\mathbb{R}^d)$ and (a special case of) Young's inequality gives

$$\|f * g\|_\infty \le \|f\|_{p,\lambda} \|g\|_{q,\lambda} . \tag{5}$$

Furthermore, if $g \in \mathcal{L}^p(\mathbb{R}^d, \lambda)$ $(1 \le p \le \infty)$ and $\mu \in M(\mathbb{R}^d)$, then the function $g * \mu$ is well defined λ-a.e. and satisfies

$$\|g * \mu\|_{p,\lambda} \le \|g\|_{p,\lambda} \|\mu\| . \tag{6}$$

In the case where $g \in \mathsf{C}(\mathbb{R}^d)$, $g * \mu$ is in fact defined everywhere and contained in $\mathsf{C}(\mathbb{R}^d)$.

Let now $(\Omega, \mathcal{A}, Pr)$ be a probability space and let $\mathbb{P}$ be a (Borel) probability measure on $\mathbb{R}^d$. Let $\emptyset \ne \mathcal{F} \subseteq \mathcal{L}^2(\mathbb{R}^d, \mathbb{P})$. A Gaussian process $\mathbb{G} : (\Omega, \mathcal{A}, Pr) \times \mathcal{F} \to \mathbb{R}$ with mean zero and covariance $E\mathbb{G}(f)\mathbb{G}(g) = \mathbb{P}[(f - \mathbb{P}f)(g - \mathbb{P}g)]$ for $f, g \in \mathcal{F}$ is called a (generalized) $\mathbb{P}$-*Brownian bridge* process indexed by $\mathcal{F}$. The covariance induces a semimetric $\rho^2_{\mathbb{P}}(f, g) = E[\mathbb{G}(f) - \mathbb{G}(g)]^2$ for $f, g \in \mathcal{F}$. A class of functions $\mathcal{F} \subseteq \mathcal{L}^2(\mathbb{R}^d, \mathbb{P})$ will be called $\mathbb{P}$-*pregaussian* if such a Gaussian process $\mathbb{G}$ can be defined such that for every $\omega \in \Omega$, the map $f \longmapsto \mathbb{G}(f, \omega)$ is bounded and uniformly continuous w.r.t. the semimetric $\rho_{\mathbb{P}}$ from $\mathcal{F}$ into $\mathbb{R}$. See also pp. 92–93 in [8].

Given n independent random vectors $X_1, \ldots, X_n$ identically distributed according to some law $\mathbb{P}$ on $\mathbb{R}^d$, we denote by $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ the usual empirical measure. Throughout the paper, E denotes expectation w.r.t. the law $\mathbb{P}$. [Also, we assume throughout that the X_i are the coordinate projections of the infinite product probability space $((\mathbb{R}^d)^{\mathbb{N}}, \mathcal{B}_{(\mathbb{R}^d)^{\mathbb{N}}}, \mathbb{P}^{\mathbb{N}})$.] For $\mathcal{F} \subseteq \mathcal{L}^2(\mathbb{R}^d, \mathbb{P})$, the $\mathcal{F}$-indexed *empirical process* is given by

$$f \longmapsto \sqrt{n}\,(\mathbb{P}_n - \mathbb{P})\, f = \frac{1}{\sqrt{n}} \sum_{i=1}^n (f(X_i) - \mathbb{P}f).$$

Convergence in law of random elements in $\ell^\infty(\mathcal{F})$ is defined in the usual way, see p. 94 in [8], and will be denoted by the symbol $\rightsquigarrow_{\ell^\infty(\mathcal{F})}$. The class $\mathcal{F}$ is said to be $\mathbb{P}$ -*Donsker* if it is $\mathbb{P}$-pregaussian and if $\sqrt{n}\,(\mathbb{P}_n - \mathbb{P}) \rightsquigarrow_{\ell^\infty(\mathcal{F})} \mathbb{G}$ where $\mathbb{G}$ is the (generalized) Brownian bridge process indexed by $\mathcal{F}$. If $\mathcal{F}$ is $\mathbb{P}$-Donsker for all probability measures $\mathbb{P}$ on $\mathbb{R}$ it is called *universal* Donsker. It is called *uniform* Donsker if convergence in law of $\sqrt{n}\,(\mathbb{P}_n - \mathbb{P})$ to $\mathbb{G}$ is uniform in a sense made precise in [13].

3 Uniform CLTs for smoothed empirical measures

Given a (pregaussian) class of functions $\mathcal{F}$ and the empirical measure $\mathbb{P}_n$, we want to study the limiting behavior in $\ell^\infty(\mathcal{F})$ of the random convolution product $\mathbb{P}_n * \mu_n$ where the measures $\mu_n \in M(\mathbb{R}^d)$ converge weakly to δ_0, the Dirac measure at zero. The leading special case is kernel density estimation, see Sect. 4, but in principle also other random measures $\mathbb{P}_n * \mu_n$ could be thought of (e.g., orthogonal series estimators). Sometimes the convolutions $\mathbb{P}_n * \mu_n$ are called 'smoothed' empirical measures, but we do not exclude discrete measures μ_n in our setup.

The fact that the signed measures μ_n converge weakly to δ_0, that is, that $\int_{\mathbb{R}^d} f d\mu_n \to f(0)$ for all real-valued bounded continuous functions on $\mathbb{R}^d$, implies, by the uniform boundedness principle, that $\sup_n \|\mu_n\| < \infty$ holds, and even that the sequence of the total variation measures $|\mu_n|$ is uniformly tight (see, e.g., [15, p. 98]; in this book as well as in some other references, what we call weak convergence is referred to as narrow convergence). In most situations in statistics (e.g., in kernel density estimation) the sequence μ_n enjoys an additional property, namely that $|\mu_n|(\mathbb{R}^d \backslash [-a, a]^d) \to 0$ for all $a > 0$, and this property plays a role in some proofs. This property is also natural in the sense that it does not allow for sequences such as $\mu_n = \delta_0 + \delta_{x_n} - \delta_{y_n}$ with $x_n \to x \neq 0$ and $x_n \neq y_n \to x$, which do not conform with the general intuition of an approximate identity. So, we make the following definition:

Definition 1 A sequence $\{\mu_n\}_{n=1}^\infty$ of finite signed Borel measures on $\mathbb{R}^d$ is an approximate convolution identity if it converges weakly to point mass δ_0 at 0. If, in addition, for every $a > 0$, $\lim_n |\mu_n|(\mathbb{R}^d \backslash [-a, a]^d) = 0$, then we call the sequence $\{\mu_n\}_{n=1}^\infty$ a proper approximate convolution identity.

For example, if $K \in \mathcal{L}^1(\mathbb{R}^d, \lambda)$ satisfies $\int_{\mathbb{R}^d} K(y)dy = 1$, and if $d\mu_n(y) = h_n^{-d} K(h_n^{-1} y)dy$ for $0 < h_n \to 0$, then the sequence μ_n is a proper approximate identity (we will often drop the word convolution): If f is bounded and continuous, then

$$\int_{\mathbb{R}^d} h_n^{-d} K(h_n^{-1} y) f(y) dy = \int_{\mathbb{R}^d} K(u) f(h_n u) du \to f(0)$$

by dominated convergence, and

$$\int_{\mathbb{R}^d \backslash [-a,a]^d} h_n^{-d} |K(h_n^{-1} y)| dy = \int_{\mathbb{R}^d \backslash [-a/h_n, a/h_n]^d} |K(u)|\, du \to 0$$

as n tends to infinity.

Now given an approximate identity μ_n, the *centered smoothed empirical process* is defined, for $f \in \mathcal{F}$, as

$$\sqrt{n}(\mathbb{P}_n - \mathbb{P}) * \mu_n(f) = \sqrt{n}\left(\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x+y)d\mathbb{P}_n(x)d\mu_n(y) - \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x+y)d\mathbb{P}(x)d\mu_n(y)\right)$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(f * \bar{\mu}_n(X_i) - \int_{\mathbb{R}^d} f * \bar{\mu}_n(x)d\mathbb{P}(x)\right)$$

where $\bar{\mu}_n(A) = \mu_n(-A)$ for $A \in \mathcal{B}_{\mathbb{R}^d}$ and where it is assumed that $f(x+\cdot) \in \mathcal{L}^1(|\mu_n|)$ for all n and $\mathbb{P}$-almost every $x \in \mathbb{R}^d$. Clearly, if μ_n arises from a symmetric kernel K, we have $\mu_n = \bar{\mu}_n$. Accordingly, the (uncentered) *smoothed empirical process*, is given by

$$\sqrt{n}(\mathbb{P}_n * \mu_n - \mathbb{P})(f) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (f * \bar{\mu}_n(X_i) - \mathbb{P}f), \quad f \in \mathcal{F}.$$

3.1 Donsker classes

Let $\mathcal{F}$ be a translation invariant $\mathbb{P}$-Donsker class of functions $f : \mathbb{R}^d \mapsto \mathbb{R}$, that is, for any $f \in \mathcal{F}$ also $f(\cdot + y)$ belongs to $\mathcal{F}$ whenever $y \in \mathbb{R}^d$. Given the usual theory of empirical processes, the proof of van der Vaart's [29] theorem mentioned in the introduction is based on the statement "the functions $x \mapsto \int f(x+y)d\mu(y)$, for signed measures [of finite variation], are weighted averages of elements of $\mathcal{F}$". However, this assertion is not formally correct, the correct assertion requires clarification and we can only verify it for Donsker classes of functions that satisfy an additional condition.

First it should be observed that Dudley's [7] theorem to the effect that if $\mathcal{F}$ is a $\mathbb{P}$-Donsker class then the sequential closure, in $\mathcal{L}^2(\mathbb{P})$ and pointwise simultaneously, of its symmetric convex hull is also a $\mathbb{P}$-Donsker class, can be slightly strengthened by deleting the word 'sequential'. Consider in $\mathcal{L}^2(\mathbb{P})$ (or in a subset of it) the following topologies: τ_1, the topology of pointwise convergence, defined by the neighborhood base $N(f; x_1, \ldots, x_r; \varepsilon) = \{g : |g(x_i) - f(x_i)| < \varepsilon,\ 1 \le i \le r\}$, $f : \mathbb{R}^d \mapsto \mathbb{R}$, $x_i \in \mathbb{R}$, $\varepsilon > 0$, $r \in \mathbb{N}$; and τ_2, defined by the (semi)metric $\|f - g\|_{2,\mathbb{P}}$. We recall that $\tau = \tau_1 \vee \tau_2$, the coarsest topology finer than τ_1 and τ_2, is defined as follows (e.g., [24, p. 5]): if $T_0 = \{A_1 \cap A_2 : A_i \in \tau_i, i = 1, 2\}$, then τ is the collection of arbitrary unions of sets in T_0. T_0 is a neighborhood base for τ, and if a map is continuous for either τ_1 or τ_2, then it is continuous for τ. Let us also recall that the symmetric convex hull $\mathcal{G}$ of a class of functions $\mathcal{F}$ is the collection of functions of the form $\sum_{i=1}^r \alpha_i f_i$, $r \in \mathbb{N}$, $f_i \in \mathcal{F}$, and $\sum_{i=1}^r |\alpha_i| \le 1$. With these definitions, we have the following:

Theorem 1 (Extension of Dudley [7, Theorem 5.3]) *Let $\mathcal{F}$ be a $\mathbb{P}$-Donsker class of functions. Then, the closure $\mathcal{H}$ in $\mathcal{L}^2(\mathbb{P})$ for the τ topology of the symmetric convex hull $\mathcal{G}$ of $\mathcal{F}$ is a $\mathbb{P}$-Donsker class.*

The proof is essentially the same as that of Dudley's theorem: One combines almost sure representations and the fact that, if L_i, $i \le n < \infty$, are linear functionals defined on the linear span of $\mathcal{F}$ that are continuous in either of the two topologies, then,

$$\left\| \sum_{i=1}^{n} L_i \right\|_{\mathcal{F}} = \left\| \sum_{i=1}^{n} L_i \right\|_{\mathcal{G}} = \left\| \sum_{i=1}^{n} L_i \right\|_{\mathcal{H}}$$

[these maps L_i consist of $\delta_{X_j(\omega)}$, $\mathbb{P}$ and (a suitable 'linear' version of) $\mathbb{G}$]. The first equality is clear and for the second equality we note the following. If $h \in \mathcal{H}$, then δ_x for every x, $\mathbb{P}$ and $\mathbb{G}$ are all τ-continuous at h (in the case of $\mathbb{G}$ because of Theorem 5.1(a) in [7]), hence, given $\varepsilon > 0$ arbitrary, there exists a neighborhood N of h such that $\left|\sum_{i=1}^{n} L_i(f-h)\right| < \varepsilon$ for all $f \in N$, and since $N \cap \mathcal{G} \neq \emptyset$ we have $\left|\sum_{i=1}^{n} L_i(h)\right| < \sup_{g\in\mathcal{G}} \left|\sum_{i=1}^{n} L_i(g)\right| + \varepsilon$, which gives the second identity by arbitrariness of ε.

Note that sequential closure coincides with topological closure in the case of τ_2, but not in the topology of pointwise convergence, so the above theorem is more general. Also, note that the above theorem applies to any dilation of $\mathcal{H}$, $c\mathcal{H} = \{\lambda h : |\lambda| \le c, h \in \mathcal{H}\}$, $0 < c \le \infty$.

The following two lemmas will make van der Vaart's observation precise and therefore, combined with the previous theorem, validate his theorem under an extra condition. The first lemma is well known if $\mathbb{Q} = \lambda$ and $f \in \mathcal{L}^2(\mathbb{R}^d, \lambda)$ (note that conditions (a)–(c) are then automatically satisfied), but in our setup we will need a result for not necessarily translation-invariant measures $\mathbb{Q}$, in particular, for finite measures; and also for functions f in more general spaces.

Lemma 1 *Let $\mathbb{Q}$ be a positive Borel measure on $\mathbb{R}^d$, let $\mu \in M(\mathbb{R}^d)$ be finite signed measure on $\mathbb{R}^d$, and let $f : \mathbb{R}^d \to \mathbb{R}$ be a Borel-measurable function. Assume that* (a) *$f(\cdot - y) \in \mathcal{L}^2(\mathbb{Q})$ for all $y \in \mathbb{R}^d$, b) $f(x - \cdot) \in \mathcal{L}^1(|\mu|)$ for $\mathbb{Q}$-almost every $x \in \mathbb{R}^d$, and (c) the function $y \mapsto \|f(\cdot - y)\|_{2,\mathbb{Q}}$ is in $\mathcal{L}^1(|\mu|)$. Then, the function*

$$h(x) := \int_{\mathbb{R}^d} f(x-y)d\mu(y)$$

is in the $\mathcal{L}^2(\mathbb{Q})$-closure of $\|\mu\|$ times the symmetric convex hull of $\mathcal{F}_f := \{f(\cdot - y) : y \in \mathbb{R}^d\}$.

Proof The space $\mathcal{L}^2(\mathbb{Q})$ is separable, and therefore so is $\mathcal{F}_f$. Hence there exists a countable set $\{y_k : k \in \mathbb{N}\}$ such that the set of functions $\{f(\cdot - y_k) : k \in \mathbb{N}\}$ is dense in $\mathcal{F}_f$ for the $\mathcal{L}^2(\mathbb{Q})$-norm. By standard arguments, for any g measurable, the function $y \mapsto \int_{\mathbb{R}^d}(f(x-y) - g(x))^2 d\mathbb{Q}(x)$ is measurable. Given $\varepsilon > 0$, set $\varepsilon' = \varepsilon/2\|\mu\|$ and define the following measurable partition $\{A_k\}_{k=1}^{\infty}$ of $\mathbb{R}^d$:

$$A_1 = \{y \in \mathbb{R}^d : \|f(\cdot - y) - f(\cdot - y_1)\|_{2,\mathbb{Q}} < \varepsilon'\},$$

and, recursively, for all $k \in \mathbb{N}$,

$$A_k = \Big(\bigcup_{j=1}^{k-1} A_j\Big)^c \cap \{y \in \mathbb{R}^d : \|f(\cdot - y) - f(\cdot - y_k)\|_{2,\mathbb{Q}} < \varepsilon'\}.$$

Now,

$$\int_{\mathbb{R}^d} f(x-y)d\mu(y) - \sum_{k=1}^r f(x-y_k)\mu(A_k)$$
$$= \sum_{k=1}^r \int_{\mathbb{R}^d} (f(x-y) - f(x-y_k))I_{A_k}(y)d\mu(y) + \int_{\bigcup_{k=r+1}^\infty A_k} f(x-y)d\mu(y)$$

for $\mathbb{Q}$-almost every $x \in \mathbb{R}^d$, so that, by Minkowski for integrals and the definition of A_k, ε',

$$\left\| \int_{\mathbb{R}^d} f(\cdot - y)d\mu(y) - \sum_{k=1}^r f(\cdot - y_k)\mu(A_k) \right\|_{2,\mathbb{Q}}$$
$$\leq \sum_{k=1}^r \int_{\mathbb{R}^d} \|f(\cdot - y) - f(\cdot - y_k)\|_{2,\mathbb{Q}} I_{A_k}(y)d|\mu|(y)$$
$$+ \int_{\bigcup_{k=r+1}^\infty A_k} \|f(\cdot - y)\|_{2,\mathbb{Q}} d|\mu|(y)$$
$$\leq \varepsilon/2 + \int_{\bigcup_{k=r+1}^\infty A_k} \|f(\cdot - y)\|_{2,\mathbb{Q}} d|\mu|(y).$$

Since the function $y \mapsto \|f(\cdot - y)\|_{2,\mathbb{Q}}$ is $|\mu|$ integrable and $\bigcup_{k=r}^\infty A_k \downarrow \emptyset$ as $r \to \infty$, it follows that there exists $r < \infty$ such that

$$\left\| \int_{\mathbb{R}^d} f(x-y)d\mu(y) - \sum_{k=1}^r f(x-y_k)\mu(A_k) \right\|_{2,\mathbb{Q}} < \varepsilon$$

holds, which completes the proof. □

Lemma 2 *Let $\mathcal{F}$ be a translation invariant $\mathbb{P}$-Donsker class of real-valued functions on $\mathbb{R}^d$ and let $\mathcal{M}$ be a collection of signed Borel measures of finite variation such that* $\sup_{\mu \in \mathcal{M}} \|\mu\| < \infty$. *Assume that for all $f \in \mathcal{F}$ and $\mu \in \mathcal{M}$, the functions*

$$y \mapsto f(x-y) \textit{ for all } x \in \mathbb{R}^d \quad \textit{and} \quad y \mapsto \|f(\cdot - y)\|_{2,\mathbb{P}}$$

are in $\mathcal{L}^1(|\mu|)$*. Then, the class of functions*

$$\tilde{\mathcal{F}} := \{f * \mu : f \in \mathcal{F}, \mu \in \mathcal{M}\}$$

is $\mathbb{P}$*-Donsker.*

Proof By Theorem 1 it suffices to show that for each $f \in \mathcal{F}$ and $\mu \in \mathcal{M}$, every neighborhood of $f * \mu$ for the τ -topology has a non-void intersection with $\|\mu\|$-times the symmetric convex hull of $\mathcal{F}_f$. By definition of the neighborhood base it suffices to prove this only for any set of the form

$$A_{x_1,\dots,x_r,\varepsilon} = \left\{g \in \mathcal{L}^2(\mathbb{P}) : \|f * \mu - g\|_{2,\mathbb{P}} < \varepsilon, |f * \mu(x_i) - g(x_i)| < \varepsilon, 1 \le i \le r\right\},$$

where $r < \infty$, $x_i \in \mathbb{R}^d$, and $\varepsilon > 0$. Define

$$\mathbb{Q} = \mathbb{P} + \delta_{x_1} + \cdots + \delta_{x_r}$$

and note that the hypotheses of Lemma 1 are satisfied by $\mathbb{Q}$, $\mu \in \mathcal{M}$ and $f \in \mathcal{F}$. The conclusion of that lemma is that $\|\mu\|$ times the symmetric convex hull of $\mathcal{F}_f$ intersects any neighborhood of $f * \mu$ for the $\mathcal{L}^2(\mathbb{Q})$-(semi)-norm,

$$B_\varepsilon = \{g \in \mathcal{L}^2(\mathbb{P}) : \|f * \mu - g\|_{2,\mathbb{Q}} < \varepsilon\},\ 0 < \varepsilon < \infty.$$

But obviously, $B_\varepsilon \subseteq A_{x_1,\dots,x_r,\varepsilon}$, which proves the lemma. □

Next we give the modified van der Vaart theorem for smoothed empirical measures over Donsker classes.

Theorem 2 (Modification of van der Vaart's [28] theorem) *Let* $\mathcal{F}$ *be a translation invariant* $\mathbb{P}$*-Donsker class of real-valued functions on* $\mathbb{R}^d$*, and let* $\{\mu_n\}_{n=1}^\infty$ *be an approximate convolution identity such that* $\mu_n(\mathbb{R}^d) = 1$ *for every n. Further assume that for every n,* $\mathcal{F} \subseteq \mathcal{L}^1(|\mu_n|)$ *and* $\int_{\mathbb{R}^d} \|f(\cdot - y)\|_{2,\mathbb{P}} d\,|\mu_n|\,(y) < \infty$ *for all* $f \in \mathcal{F}$*. Then,*

(a) *the condition*

$$\sup_{f \in \mathcal{F}} E \left(\int_{\mathbb{R}^d} (f(X+y) - f(X)) d\mu_n(y) \right)^2 \to_{n\to\infty} 0 \tag{7}$$

is necessary and sufficient for

$$\sup_{f \in \mathcal{F}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(f * \bar{\mu}_n(X_i) - f(X_i) - E[f * \bar{\mu}_n(X) - f(X)]\right) \right| \to_{n\to\infty} 0$$

in outer probability, hence, for the centered smoothed empirical measures

$$\left\{\sqrt{n}((\mathbb{P}_n - \mathbb{P}) * \mu_n(f) : f \in \mathcal{F}\right\}$$

to converge in law in $\ell^\infty(\mathcal{F})$ to the $\mathbb{P}$ -Brownian bridge $\mathbb{G}$ indexed by $\mathcal{F}$.

(b) *Consequently, Condition* (7) *and*

$$\sup_{f\in\mathcal{F}} \sqrt{n}\left|E\int_{\mathbb{R}^d}(f(X+y) - f(X))d\mu_n(y)\right| \to_{n\to\infty} 0 \tag{8}$$

are necessary and sufficient for $\sqrt{n}\,\|\mathbb{P}_n * \mu_n - \mathbb{P}_n\|_{\mathcal{F}}$ to converge to zero in outer probability and for

$$\sqrt{n}(\mathbb{P}_n * \mu_n - \mathbb{P}) \rightsquigarrow_{\ell^\infty(\mathcal{F})} \mathbb{G} \tag{9}$$

to hold, where $\mathbb{G}$ is the $\mathbb{P}$-Brownian bridge indexed by $\mathcal{F}$.

(c) *The sufficiency parts of* (a) *and* (b) *hold as well if:* (i) *the approximate identities are allowed to be random, i.e., $\mu_n : (\Omega, \mathcal{A}, Pr) \to M(\mathbb{R}^d)$, assuming that* $\sup_{n,\omega}\|\mu_n(\omega)\| < \infty$ *and $\mu_n \to \delta_0$ weakly in probability.* (ii) *$\mathcal{F}$ is not necessarily translation invariant but contained in a translation-invariant class of functions $\mathcal{G}$ so that $\mathcal{G} \subseteq \mathcal{L}^1(|\mu_n|)$, $\int_{\mathbb{R}^d}\|f(\cdot - y)\|_{2,\mathbb{P}}d\,|\mu_n|\,(y) < \infty$ for all $f \in \mathcal{G}$ and Condition* (7) *holds with the supremum extending over $\mathcal{G}$.*

Proof Combine Lemma 2 with the proof of the theorem in [29], who considers $\int f(x+y)d\mu(y) = f * \bar{\mu}(x)$ in our notation. [Note that, if $\mu_n \to \delta_0$ weakly, then also $\bar{\mu}_n \to \delta_0$ and the total variation norms $\|\bar{\mu}_n\|$ are uniformly bounded.] Since we shall make extensive use of the sufficiency part of Part (b), we give a short, simple proof of it. Consider the decomposition

$$\mathbb{P}_n * \mu_n - \mathbb{P}_n = (\mathbb{P}_n - \mathbb{P}) * \mu_n - (\mathbb{P}_n - \mathbb{P}) + \mathbb{P} * \mu_n - \mathbb{P}.$$

Since $(\mathbb{P} * \mu_n - \mathbb{P})f = E\int_{\mathbb{R}}(f(X+y) - f(X))\,d\mu_n(y)$, Condition (8) gives

$$\|\mathbb{P} * \mu_n - \mathbb{P}\|_{\mathcal{F}} = o(1/\sqrt{n}).$$

For the remaining part of the decomposition, note that

$$((\mathbb{P}_n - \mathbb{P}) * \mu_n - (\mathbb{P}_n - \mathbb{P}))\,f = (\mathbb{P}_n - \mathbb{P})\,(\bar{\mu}_n * f - f).$$

Now, $\bigcup_n\{\bar{\mu}_n * f - f : f \in \mathcal{F}\}$ is a $\mathbb{P}$-Donsker class by Lemma 2 (together with a simple permanence property, e.g., van der Vaart and Wellner [30], p. 192), so, since by Condition (7) $\sup_{f\in\mathcal{F}}\mathbb{P}(\bar{\mu}_n * f - f)^2 \to 0$, it follows that

$$\sup_{f\in\mathcal{F}}|(\mathbb{P}_n - \mathbb{P})(\bar{\mu}_n * f - f)| = o_{\mathbb{P}}(1/\sqrt{n}).$$

The last two estimates prove (9) since $\mathcal{F}$ is $\mathbb{P}$-Donsker. $\square$

Remark 1 Part (c) of the theorem is of practical interest. By (i)—as was already remarked by van der Vaart [29]—one may use data-driven choices of the bandwidth if μ_n comes from a kernel function. Also, in (ii), if one is interested in a (not necessarily translation-invariant) subset $\mathcal{F}$ of a translation-invariant Donsker class $\mathcal{G}$, then one may restrict the supremum in the bias condition (8) to this subclass. A leading example where this may be useful is where it is known in advance that $\mathbb{P}$ is supported by a proper subset of $\mathbb{R}^d$.

There are other ways (besides Remark 1) of dispensing with the translation invariance condition on the class $\mathcal{F}$. In particular, this can be done if instead of assuming that $\mathcal{F}$ is $\mathbb{P}$-Donsker we impose the considerably more restrictive condition that $\mathcal{F}$ satisfy

$$\int_0^\infty \sup_{\mathbb{Q}} \sqrt{H(\mathcal{F}, L^2(\mathbb{Q}), \varepsilon \|F\|_{2,\mathbb{Q}})} d\varepsilon < \infty, \tag{10}$$

where the supremum is extended over all finitely supported probability measures $\mathbb{Q}$ on $\mathbb{R}^d$, F is an measurable envelope of $\mathcal{F}$, and

$$H(\mathcal{F}, L^2(\mathbb{Q}), \varepsilon) := \log N(\mathcal{F}, L^2(\mathbb{Q}), \varepsilon)$$

denotes the usual $L^2(\mathbb{Q})$-metric entropy of $\mathcal{F}$. In fact Rost [23], based on [32], who generalized results from [1], proved that if $\mathcal{F}$ is uniformly bounded, countable (or suitably measurable) and satisfies conditions (10), (7) and (8) then the central limit theorem (9) holds. We will not use this result because the entropy condition is either not satisfied or has not been proved for many classes of functions in this article, whereas translation invariance does hold (at least in the sense of Remark 1).

3.2 Pregaussian classes

Radulović and Wegkamp [21] make the interesting observation that the smoothed empirical process in (9) may in some cases converge in law in $\ell^\infty(\mathcal{F})$ to $\mathbb{G}$ even if $\mathcal{F}$ is *not* $\mathbb{P}$-Donsker. This situation is entirely different from the situation considered in the last section: whereas in the Donsker case the smoothed process gets closer and closer to the un-smoothed empirical process, in the present situation the smoothed and the un-smoothed empirical processes should drift away from each other as one will converge and the other one will not. This means that the amount of smoothing allowed in the non-Donsker case should have a lower bound: if we do not smooth enough, the smoothed process might be too close to the original for it to converge.

Radulović and Wegkamp [21] prove their theorem by adapting the proof of Theorem 3.2 in [11] on the relation between the pregaussian and the Donsker properties for uniformly bounded classes of functions. Their result has a limited scope since they have to impose stringent conditions on $\mathcal{F}$ and $\mathbb{P}$. In particular, their method seems only to work if $\mathcal{F}$ and $\mathbb{P}$ have compact support, and if the density of $\mathbb{P}$ is twice differentiable and bounded from above and below on the support of $\mathcal{F}$, see also Remarks 3 and 9. In the following theorem, we will also adapt the Giné and Zinn

[11] result. In the proof, we will use a dominating Gaussian process different from the one Radulović and Wegkamp [21] use. The general theorem—which also allows for unbounded classes of functions—will imply that classical kernel density estimators can converge in law in $\ell^\infty(\mathcal{F})$ for pregaussian classes of functions $\mathcal{F}$ that are *not* Donsker under only the assumption that $\mathbb{P}$ has a bounded density, see Theorem 9. We refer to Sect. 4.2 for examples and more discussion.

For a given class of measurable functions $\mathcal{F}$, we write

$$\mathcal{F}'_\delta = \{f - g : f, g \in \mathcal{F}, \|f - g\|_{2,\mathbb{P}} \le \delta\}.$$

We will randomize the point masses δ_{X_i} with a Rademacher sequence $\{\varepsilon_i\}_{i=1}^\infty$, that is, a sequence of independent symmetric random variables taking only the values $+1$ and -1, independent of the sequence $\{X_i\}$. In fact, we take all the variables, X_i, ε_j, to be coordinate projections of an infinite product probability space. We will also randomize by an orthogaussian sequence g_i, which is taken to be independent from the X_j and the ε_j in the same product space sense. The product probability measure in this large product space is denoted by Pr.

Theorem 3 *Let $\mathcal{F}$ be a $\mathbb{P}$-pregaussian class of real-valued functions on $\mathbb{R}^d$ that is invariant by translations and such that $\|\mathbb{P}f\|_{\mathcal{F}} < \infty$ holds. Let $\{\mu_n\}_{n=1}^\infty$ be an approximate convolution identity such that $\mu_n(\mathbb{R}^d) = 1$ for every n. Assume that $\mathcal{F} \subseteq \mathcal{L}^1(|\mu_n|)$ holds for every n and, in addition,*

(a) *for each n, the class of functions $\tilde{\mathcal{F}}_n := \{f * \bar{\mu}_n : f \in \mathcal{F}\}$ consists of functions whose absolute values are bounded by a constant M_n;*
(b) $\sup_{f\in\mathcal{F}} E(f * \bar{\mu}_n(X) - f(X))^2 \to 0$ *as* $n \to \infty$ (*that is,* (7) *holds*);
(c)
$$\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{(\tilde{\mathcal{F}}_n)'_{1/n^{1/4}}} \to 0 \tag{11}$$
as $n \to \infty$ in outer probability;
(d) *For all $0 < \varepsilon < 1$, $H(\tilde{\mathcal{F}}_n, L^2(\mathbb{P}), \varepsilon) \le \lambda_n(\varepsilon)/\varepsilon^2$ for functions $\lambda_n(\varepsilon)$ such that $\lambda_n(\varepsilon) \to 0$ and $\lambda_n(\varepsilon)/\varepsilon^2 \to \infty$ as $\varepsilon \to 0$, uniformly in n, and then, the bounds M_n of part* (a) *satisfy*
$$M_n \le \left(5\sqrt{\lambda_n(1/n^{1/4})}\right)^{-1} \tag{12}$$
for all n large enough.
Then,
$$\sqrt{n}(\mathbb{P}_n - \mathbb{P}) * \mu_n \rightsquigarrow_{\ell^\infty(\mathcal{F})} \mathbb{G}, \tag{13}$$
where $\mathbb{G}$ is the $\mathbb{P}$-Brownian bridge indexed by $\mathcal{F}$. If, in addition, the bias condition (8) *is satisfied, then we also have*
$$\sqrt{n}(\mathbb{P}_n * \mu_n - \mathbb{P}) \rightsquigarrow_{\ell^\infty(\mathcal{F})} \mathbb{G}. \tag{14}$$

Proof In the proof, we set $\|f - g\|_{2,\mathbb{P}} = e_{\mathbb{P}}(f, g)$ for the sake of brevity. Let $\mathbb{Z}_{\mathbb{P}}(f) = \mathbb{G}(f) + (\mathbb{P}f)g$, $f \in \mathcal{F}$, where g is $N(0, 1)$ independent of $\mathbb{G}$. Then

$E(\mathbb{Z}_{\mathbb{P}}(f)-\mathbb{Z}_{\mathbb{P}}(h))^2 = e_{\mathbb{P}}^2(f,h)$. Also, since $|\mathbb{P}(f-h)| \le e_{\mathbb{P}}(f,h)$ and $\|\mathbb{P}f\|_{\mathcal{F}} < \infty$, this process has bounded and $e_{\mathbb{P}}$-uniformly continuous sample paths (sample continuous for short). This implies in particular, by Sudakov's theorem (e.g., [14, p. 81]), that $\sup_{f\in\mathcal{F}} \mathbb{P}f^2 = C_1^2$ for some $C_1 < \infty$. By the uniform boundedness principle, $\sup_n \|\mu_n\| = \sup_n \|\bar{\mu}_n\| < C_2$ for some $C_2 < \infty$. Then, $\mathcal{F}$ being invariant by translations, we have for all $f \in \mathcal{F}$ and $\mu \in \mathcal{M}$ that $\int \|f(\cdot - y)\|_{2,\mathbb{P}} d|\mu|(y) \le C_1C_2$, where $\mathcal{M}$ is the collection of all signed Borel measures whose total variation is bounded by C_2 and which integrate all the functions in $\mathcal{F}$ (note that $\{\bar{\mu}_n\}_{n=1}^{\infty} \subseteq \mathcal{M}$). Then Lemma 1 gives that the class of functions $\tilde{\mathcal{F}} := \{f * \bar{\mu} : f \in \mathcal{F}, \mu \in \mathcal{M}\}$ is contained in the $e_{\mathbb{P}}$-closure of C_2 times the symmetric convex hull of $\mathcal{F}$. Then, by Theorem 0.3 in [6], the process $\mathbb{Z}_{\mathbb{P}}$ extends to the whole class $\tilde{\mathcal{F}}$ as a centered Gaussian process with bounded and $e_{\mathbb{P}}$-uniformly continuous sample paths. This implies, again by Sudakov's theorem, that the class $\tilde{\mathcal{F}}$ and (therefore also) the classes $\tilde{\mathcal{F}}_n$ for every n, are $e_{\mathbb{P}}$ totally bounded, in fact,

$$\varepsilon^2 H(\tilde{\mathcal{F}}, e_{\mathbb{P}}, \varepsilon) \to 0 \quad \text{as } \varepsilon \to 0. \tag{15}$$

This shows that a function $\lambda(\varepsilon) \to 0$ such that $\lambda(\varepsilon)/\varepsilon^2 \to \infty$ as $\varepsilon \to 0$ as specified in condition (d) always exists if $\mathcal{F}$ is $\mathbb{P}$-pregaussian. We will use these observations in the proof of tightness, but first we must deal with the finite dimensional distributions and with randomization.

(1) *Convergence of finite dimensional distributions.* By the conditions on M_n [in (d)], $M_n/n^{1/2} \le (5n^{1/2})^{-1/2}$ for all n large enough, and, by condition (b), $E(g * \bar{\mu}_n(X))^2 \to Eg^2(X)$ for any linear combination g of functions in $\mathcal{F}$. Then, by Lindeberg's theorem, $\frac{1}{\sqrt{n}}\sum_{i=1}^n (g * \bar{\mu}_n(X_i) - Eg * \bar{\mu}_n(X))$ converges in law to $\mathbb{G}_{\mathbb{P}}(g)$, hence, by Cramér-Wold, there is convergence of the finite dimensional distributions in (13).

(2) *Randomization in the asymptotic equicontinuity condition.* By (1) and by the asymptotic equicontinuity theorem on convergence in law of bounded processes (e.g., Theorem 5.1.2 in [4]), the limit (13) will follow if we show

$$\lim_{\delta\to 0} \limsup_n Pr^* \left\{ \sup_{f\in\mathcal{F}'_\delta} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^n (f * \bar{\mu}_n(X_i) - \mathbb{P}(f * \bar{\mu}_n)) \right| > \gamma \right\} = 0 \tag{16}$$

for all $\gamma > 0$ (notice that, since $\mathbb{Z}_{\mathbb{P}}$ is sample continuous on $\mathcal{F}$, Sudakov's theorem implies that $(\mathcal{F}, e_{\mathbb{P}})$ is totally bounded). In order to apply the proof of Theorem 3.2 in [11], which is the model for the proof of this theorem, we should both randomize the random variables involved in (16) and modify the set on which the sup is taken, from $\mathcal{F}'_\delta$ to $(\tilde{\mathcal{F}}_n)'_\delta$, as will be seen in the development of the proof. To randomize, we use Lemma 2.5 in [11], to the effect that if $X(t)$, $X'(t)$, $t \in T$, are two independent stochastic processes (defined in different components of a product probability space),

$$Pr^*\{\|X\|_T > s\} \le \frac{1}{1 - \sup_T Pr\{|X'(t)| \ge u\}} Pr^*\{\|X - X'\|_T > s - u\}.$$

In our case, $T = \mathcal{F}'_\delta$, X is the smoothed empirical process and X' an independent copy. If we choose $n_\delta < \infty$ such that for $n \geq n_\delta$, $\sup_{f \in \mathcal{F}'_\delta} E(f * \bar{\mu}_n(X) - f(X))^2 < \delta^2$, which we can by condition b), we have $\sup_{f \in \mathcal{F}'_\delta} E(f * \bar{\mu}_n(X))^2 \leq 4\delta^2$, and the above inequality and Chebyshev, together with Rademacher randomization, give

$$Pr^* \left\{ \sup_{f \in \mathcal{F}'_\delta} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} (f * \bar{\mu}_n(X_i) - \mathbb{P}(f * \bar{\mu}_n)) \right| > \gamma \right\}$$
$$\leq 4Pr^* \left\{ \sup_{f \in \mathcal{F}'_\delta} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} \varepsilon_i f * \bar{\mu}_n(X_i) \right| > \frac{\gamma - 2\sqrt{2}\delta}{2} \right\}.$$

So, by choosing $\delta < \gamma/4$, we conclude that in order to prove (16) it is sufficient to show that for all $\gamma > 0$,

$$\lim_{\delta \to 0} \limsup_{n} Pr^* \left\{ \sup_{f \in \mathcal{F}'_\delta} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} \varepsilon_i (f * \bar{\mu}_n)(X_i) \right| > \gamma \right\} = 0. \tag{17}$$

Since for $n \geq n_\delta$ we also have that $f \in \mathcal{F}'_\delta$ implies $f * \bar{\mu}_n \in (\tilde{\mathcal{F}}_n)'_{2\delta}$, we conclude that, in order to prove (17), it is sufficient to prove

$$\lim_{\delta \to 0} \limsup_{n} Pr^* \left\{ \sup_{f \in (\tilde{\mathcal{F}}_n)'_\delta} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right| > \gamma \right\} = 0. \tag{18}$$

(3) *The main arguments of the proof, following the proof of Theorem 3.2 in [11].* We take $\mathcal{H} = \mathcal{H}_n$ to be a maximal collection $h_1, \ldots, h_m$ of functions from $\tilde{\mathcal{F}}_n$ such that $\mathbb{P}(h_i - h_j)^2 > 1/n^{1/2}$ if $i \neq j$ and, noting that the $e_{\mathbb{P}}$ closed balls with these centers and radius $1/n^{1/4}$ cover $\tilde{\mathcal{F}}_n$, we have, for n large enough so that $1/n^{1/4} \leq \delta/2$,

$$Pr^* \left\{ \sup_{f \in (\tilde{\mathcal{F}}_n)'_\delta} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right| > 3\gamma \right\}$$
$$\leq 2Pr^* \left\{ \sup_{f \in (\tilde{\mathcal{F}}_n)'_{1/n^{1/4}}} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} \varepsilon_i f(X_i) \right| > \gamma \right\}$$
$$+ Pr \left\{ \max_{h \in \mathcal{H}'_{2\delta}} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{n} \varepsilon_i h(X_i) \right| > \gamma \right\}.$$

The $\lim_{\delta\to 0}\limsup_n$ of the first probability is zero by condition (c), and we are left with checking that the same is true for the second. Letting

$$A_n := \left\{ \max_{f\in\mathcal{H}'_{2\delta}\setminus\{0\}} \frac{\sum_{i=1}^n h^2(X_i)}{n\mathbb{P}h^2} < 2 \right\}$$

we get, for the second term,

$$\begin{aligned} Pr\left\{ \max_{h\in\mathcal{H}'_{2\delta}} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^n \varepsilon_i h(X_i) \right| > \gamma \right\} \;\le\; & Pr\{A_n^c\} \\ & + \gamma^{-1} E_X E_\varepsilon \left(\left\| \frac{\sum_{i=1}^n \varepsilon_i h(X_i)}{n^{1/2}} \right\|_{\mathcal{H}'_{2\delta}} I(A_n) \right) \\ := & \text{ (I)} + \text{(II)}. \end{aligned}$$

For (I) we notice that, by condition (d),

$$\#\mathcal{H}'_n \le \exp\left(2\log N(\tilde{\mathcal{F}}_n, e_{\mathbb{P}}, 1/n^{1/4})\right) \le \exp\left(2\lambda_n(1/n^{1/4})n^{1/2}\right).$$

On the other hand, by Bernstein's inequality, for any $h \in \mathcal{H}'_n \setminus \{0\}$,

$$\begin{aligned} Pr\left\{ \sum_{i=1}^n (h^2(X_i) - \mathbb{P}h^2) > n\mathbb{P}h^2 \right\} &\le \exp\left(-\frac{n^2(\mathbb{P}h^2)^2}{2n\mathbb{P}h^4 + 8M_n^2 n\mathbb{P}h^2/3} \right) \\ &\le \exp\left(-\frac{n\mathbb{P}h^2}{11M_n^2} \right) \le \exp\left(-\frac{n^{1/2}}{11M_n^2} \right). \end{aligned}$$

Hence, the hypotheses on M_n and $\lambda_n(\varepsilon)$ give

$$Pr\{A_n^c\} \le \exp\left(2\lambda_n(1/n^{1/4})n^{1/2} - \frac{n^{1/2}}{11M_n^2} \right) \le \exp\left(-\frac{3\lambda_n(1/n^{1/4})n^{1/2}}{11} \right) \to 0,$$

as $n \to \infty$.

For (II), for each $n \in \mathbb{N}$ and ω in A_n, consider the Gaussian process

$$Z_{\omega,n}(h) = \frac{1}{\sqrt{n}} \sum_{i=1}^n g_i h(X_i(\omega)), \quad h \in \mathcal{H}_n,$$

where g_i are i.i.d. $N(0, 1)$ random variables independent of the variables X_j (in the sense described above). Its increments obviously satisfy, by the definition of A_n, that

$$
\begin{aligned}
E_g \left(Z_{\omega,n}(h) - Z_{\omega,n}(h')\right)^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(h(X_i(\omega)) - h'(X_i(\omega))\right)^2 \\
&\le 2\mathbb{P}(h-h')^2 \\
&= E\left(\sqrt{2}\mathbb{Z}_{\mathbb{P}}(h) - \sqrt{2}\mathbb{Z}_{\mathbb{P}}(h')\right)^2,
\end{aligned}
$$

where $\mathbb{Z}_{\mathbb{P}}$ is the Gaussian process defined on $\tilde{\mathcal{F}}$ at the beginning of the proof, which is sample continuous, and where E_g denotes integration only with respect to the variables g_i (with the $X_j(\omega)$ fixed). So, we can apply a comparison theorem for Gaussian process due to Fernique (e.g., Theorem 2.17(b) in [11]), to the effect that, for all n and $\omega \in A_n$,

$$
\begin{aligned}
E_g \left\| \frac{\sum_{i=1}^{n} g_i h(X_i(\omega))}{\sqrt{n}} \right\|_{(\mathcal{H}_n)'_{2\delta}} &\le 4\sqrt{2}E \sup_{\mathbb{P}(h-h')^2 \le 4\delta^2 h, h' \in \tilde{\mathcal{F}}} |\mathbb{Z}_{\mathbb{P}}(h) - \mathbb{Z}_{\mathbb{P}}(h')| \\
&\quad + 26\sqrt{2}\left(\delta H^{1/2}(\tilde{\mathcal{F}}, e_{\mathbb{P}}, \delta)\right).
\end{aligned}
$$

The second term tends to zero as $\delta \to 0$ by (15) (Sudakov), and the first term by sample path uniform continuity of $Z_{\mathbb{P}}$ and integrability of suprema of sample continuous Gaussian processes. Now, $\lim_{\delta\to 0} \limsup_n$ of (II) is zero by a simple comparison principle (e.g., the first inequality in Lemma 2.9 in [11]). This proves (18) and therefore concludes the proof of (13). (14) immediately follows from (13) and (8). □

Remark 2 If the class $\mathcal{F}$ is uniformly bounded, then Conditions (a) and (d) are automatically satisfied. [Note that $\tilde{\mathcal{F}}_n$ is then also uniformly bounded by (6). Also, as already mentioned in the proof, a function $\lambda(\varepsilon)$ as specified in condition (d) always exists, so (12) is then satisfied for $M \equiv M_n$.] If one has additional information on $\lambda(\varepsilon)$, this can be used to treat unbounded classes, see the proof of Theorem 10.

Remark 3 Invariance of $\mathcal{F}$ by translations is only used, in the previous proof, in order to ensure that the processes $\{\mathbb{Z}_{\mathbb{P}}(f * \bar{\mu}_n) : f \in \mathcal{F}\} = \{\mathbb{Z}_{\mathbb{P}}(h) : h \in \tilde{\mathcal{F}}_n\}$ have the $L^2(\mathbb{P})$ norms of their increments dominated by the $L^2(\mathbb{P})$ norms of the increments of a *single* well behaved Gaussian process, in our case $\{\mathbb{Z}_{\mathbb{P}}(h) : h \in \bigcup_{n=1}^{\infty} \tilde{\mathcal{F}}_n\}$, so that we can apply Fernique's comparison inequality (at the end of the proof). The same effect is achieved by Radulović and Wegkamp [21] by imposing, instead of invariance by translation, the condition

$$
E\left[(f-g) * \bar{\mu}_n(X)\right]^2 \le C E[(f-g)(X)]^2, \quad f, g \in \mathcal{F} \cup \{0\}, \tag{19}
$$

for some $C < \infty$: then $\mathbb{Z}_{\mathbb{P}}(f * \bar{\mu}_n)$, $f \in \mathcal{F}$, is dominated in the stated sense by the nice process $\mathbb{Z}_{\mathbb{P}}(f)$, $f \in \mathcal{F}$ (actually they impose a slightly stronger condition). Notice that condition (19) also allows for control of the probability of the sets A_n^c. The problem with this condition is that in practice it only applies to $\mathbb{P}$ with differentiable density p_0 bounded away from zero, restricting the results, in particular, to p_0 with support of finite Lebesgue measure. If we impose condition (19), then it is easy to see that the above proof does not require the full force of condition (7), but only

$$E[(f * \bar{\mu}_n(X)]^2 \to E[f(X)]^2, \quad E[(f * \bar{\mu}_n(X)] \to E[f(X)], \quad f \in \mathcal{F} \tag{20}$$

(which Radulović and Wegkamp [21] also need in their Theorem 2.1, but omit to mention). With these two changes, that is (19) and (20) replacing the hypothesis of translation invariance and condition (7), Theorem 3 still holds true. The above result then contains (and slightly corrects) Theorem 2.1 in [21].

Remark 4 (*Condition (c) in Theorem 3*) Condition (c) is difficult to verify in general. The typical tools for this are either uniform entropy bounds or bracketing entropy bounds for $\tilde{\mathcal{F}}_n$. We state here the two most useful ones for further reference. See Theorems 8 and 10 for an application of the inequalities below.

(1) It follows from the second maximal inequality in Theorem 2.14.2, p. 240 in [30], a simple computation on bracketing numbers, and symmetrization, that

$$E^* \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{(\tilde{\mathcal{F}}_n)'_{n^{-1/4}}} \leq L \left(\int_0^{n^{-1/4}} \sqrt{1 \vee \log N_{[]}(\tilde{\mathcal{F}}_n, L^2(\mathbb{P}), \varepsilon)}\, d\varepsilon + \sqrt{n} M_n I(M_n > \sqrt{n} a_n) \right) \tag{21}$$

where $a_n = n^{-1/4}/\sqrt{1 + 2\log N_{[]}(\tilde{\mathcal{F}}_n, L^2(\mathbb{P}), 2^{-1}n^{-1/4})}$, $L < \infty$ is a universal constant, and $\log N_{[]}(\mathcal{F}, L^2(\mathbb{P}), \varepsilon)$ is the usual $L^2(\mathbb{P})$-bracketing metric entropy of $\mathcal{F}$, cf., e.g., p. 83 in [30].

(2) It follows from Theorem 3.1 in [10] that if $\tilde{\mathcal{F}}_n$ satisfies

$$\log N(\tilde{\mathcal{F}}_n, L^2(Q), \varepsilon) \leq H_n(M_n/\varepsilon), \; 0 < \varepsilon \leq M_n,$$

for all finitely supported probability measures Q, where H_n is a non-decreasing regularly varying function of exponent $\alpha \in [0, 2)$ with $H_n(1) > 0$, then

$$E^* \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{(\tilde{\mathcal{F}}_n)'_{1/n^{1/4}}} \leq L \max \left[C_{H_n} \frac{1}{n^{1/4}} \sqrt{H_n(2M_n n^{1/4})},\; C_{H_n}^2 \frac{M_n}{\sqrt{n}} H_n(2M_n n^{1/4}) \right] \tag{22}$$

where $L < \infty$ is a universal constant and

$$C_{H_n} := \sup_{x \geq 1} \frac{\int_x^\infty u^{-2} \sqrt{H_n(u)}\, du}{x^{-1}\sqrt{H_n(x)}}.$$

[The theorem in [10] simplifies if the envelope is taken to be a constant, as we do here: in this case, the bound (3.9) there is zero.]

3.3 Verification of Condition (7) in Theorems 2 and 3

We now discuss how to verify Condition (7)—that appears in both Theorems 2 and 3—for general classes of functions and sets.

3.3.1 Classes of functions on $\mathbb{R}$

We first treat the case of functions on $\mathbb{R}$. The symbol $\mathsf{BV}(\mathbb{R})$ will denote the space of measurable functions $\mathbb{R} \mapsto \mathbb{R}$ of bounded variation, equipped with the total variation norm

$$\|f\|_{TV} = \sup\left\{\sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| : n \in \mathbb{N}, -\infty < x_1 < \cdots < x_n < +\infty\right\}. \tag{23}$$

Furthermore, consider a class of functions $\mathcal{F} \subseteq \mathcal{L}^2(\mathbb{R}, \lambda)$ that satisfies

$$\sup_{f\in\mathcal{F}}\left(\|f\|_{2,\lambda} + \sup_{0\neq z\in\mathbb{R}} |z|^{-s}\left(\int_{\mathbb{R}} |f(x+z) - f(x)|^2\, dx\right)^{1/2}\right) < \infty \tag{24}$$

for some $s > 0$, that is, $\mathcal{F}$ is a bounded subset of the Besov space $\mathcal{B}^s_{2\infty}(\mathbb{R})$ for some $s > 0$; see also Remark 11ii. Note also that, if p_0 is a bounded function, then the conditions in Part (d) of the following proposition are automatically satisfied for any precompact subset $\mathcal{F}$ of $\mathcal{L}^2(\mathbb{R}, \lambda)$.

Proposition 1 *Let $\mathbb{P}$ be an absolutely continuous probability measure, $d\mathbb{P}(x) = p_0(x)d\lambda(x)$, and let $\mathcal{F} \subseteq \mathcal{L}^2(\mathbb{R}, \mathbb{P})$. Let further $\{\mu_n\}_{n=1}^\infty$ be a proper approximate convolution identity. Assume one of the following four conditions*:

(a) *$\mathcal{F}$ is a bounded subset of $\mathsf{BV}(\mathbb{R})$, or*
(b) *$\mathcal{F}$ is a bounded subset of $\mathsf{C}^s(\mathbb{R})$ for some $s > 0$, or*
(c) *$\mathcal{F}$ satisfies (24) for some $s > 0$ and p_0 is a bounded function, or*
(d) $\sup_{f\in\mathcal{F}} E(f(X+y)-f(X))^2 \to 0$ *as* $y \to 0$ *and* $\sup_{y\in\mathbb{R}} \sup_{f\in\mathcal{F}} \mathbb{P}(f(\cdot+y))^2 < \infty$.
Then Condition (7) holds, that is,

$$\sup_{f\in\mathcal{F}} E\left(\int_{\mathbb{R}} (f(X+y) - f(X))d\mu_n(y)\right)^2 \to_{n\to\infty} 0$$

Proof Note first that in Part (a) we may assume without loss of generality that $\mathcal{F}$ is uniformly bounded. [Otherwise, we define $\mathcal{F}' = \{f - f(-\infty+) : f \in \mathcal{F}\}$ which is again a bounded set in $\mathsf{BV}(\mathbb{R})$, see Sect. 3.5 in [9]. Then $\mathcal{F}'$ is uniformly supnorm-bounded by $\sup_{f\in\mathcal{F}}\|f\|_{TV}$, and the expression in (7) stays the same if the sup is extended only over $\mathcal{F}'$, as $\int_{\mathbb{R}}(f(X+y)-f(X))d\mu_n(y)$ is zero for f constant.]

Now to prove the proposition, by Minkowski for integrals we have

$$\begin{aligned}
&\sup_{f\in\mathcal{F}}\left(E\left(\int_{\mathbb{R}}(f(X+y)-f(X))d\mu_n(y)\right)^2\right)^{1/2}\\
&\le \sup_{f\in\mathcal{F}}\int_{|y|\le\delta}\left(E(f(X+y)-f(X))^2\right)^{1/2}d|\mu_n|(y)\\
&\quad+\sup_{f\in\mathcal{F}}\int_{|y|>\delta}\left(E(f(X+y)-f(X))^2\right)^{1/2}d|\mu_n|(y)\\
&:=(\mathrm{I})_{n,\delta}+(\mathrm{II})_{n,\delta}.
\end{aligned}\tag{25}$$

In all four cases, we have $\sup_{y\in\mathbb{R}}\sup_{f\in\mathcal{F}}\mathbb{P}(f(\cdot+y))^2\le D^2<\infty$: For (a) and (b) this follows from uniform boundedness of $\mathcal{F}$, for (d) this holds by assumption, and for (c) we have $\sup_{f\in\mathcal{F}}\mathbb{P}(f(\cdot+y))^2\le\sup_{f\in\mathcal{F}}\|f\|_{2,\lambda}^2\|p_0\|_\infty<\infty$ since $\mathcal{F}$ is a bounded subset of $\mathcal{L}^2(\mathbb{R},\lambda)$. Now, $\{\mu_n\}_{n=1}^\infty$ being a proper approximate identity, $|\mu_n|\{|y|>\delta\}\to 0$ holds for all $\delta>0$, and hence

$$\lim_n(\mathrm{II})_{n,\delta}\le 2D|\mu_n|\{|y|>\delta\}=0$$

for all $\delta>0$. We now treat $(\mathrm{I})_{n,\delta}$, where we recall that $\sup_n\|\mu_n\|<\infty$ by the uniform boundedness principle. If (a) holds, for $f\in\mathcal{F}$, let ν_f be the measure of bounded variation defined by $\nu_f(a,b]=f(b+)-f(a+)$, and note that except for a countable set, $f(x)=f(x+)$ (see after (26)). Also, in this case, the functions in $\mathcal{F}$ are uniformly bounded, say by D. Then, for $0<y<\delta$,

$$\begin{aligned}
E(f(X+y)-f(X))^2&\le 2DE|f(X+y)-f(X)|\\
&\le 2D\int_{-\infty}^{\infty}\left(\int_x^{x+y}d|\nu_f|(u)\right)p_0(x)dx\\
&=2D\int_{-\infty}^{\infty}\int_{u-y}^{u}p_0(x)dxd|\nu_f|(u)\\
&\le 2D|\nu_f|(\mathbb{R})\sup_{\lambda(A)\le\delta}\int_A p_0(x)dx,
\end{aligned}$$

and likewise if $-\delta < y < 0$. Hence, by the absolute continuity of the integral we have

$$\lim_{\delta\to 0}\sup_n (\mathrm{I})_{n,\delta} \le \lim_{\delta\to 0}\left(2D|\nu_f|(\mathbb{R}) \sup_{\lambda(A)\le\delta}\int_A p_0(x)dx\right)^{1/2} \sup_n \|\mu_n\| = 0.$$

If (b) holds, there is a constant c such that, for $\alpha = \min(s, 1)$,

$$(\mathrm{I})_{n,\delta} \le c \int_{|y|\le\delta} |y|^\alpha d|\mu_n|(y) \le c\delta^\alpha \sup_n \|\mu_n\| \to 0$$

as $\delta \to 0$, uniformly in n. If (c) holds, by (24) and boundedness of p_0, there is $c' < \infty$ such that

$$(E(f(X+y) - f(X))^2)^{1/2} \le c' \left(\int_{\mathbb{R}} (f(x+y) - f(x))^2 dx\right)^{1/2} \le c'|y|^s$$

holds uniformly in $\mathcal{F}$, and hence $(\mathrm{I})_{n,\delta} \to 0$ as $\delta \to 0$, uniformly in n, as in case (b). Finally, if (d) holds, $(\mathrm{I})_{n,\delta} \to 0$ uniformly in n as $\delta \to 0$ since $\sup_{|y|\le\delta}\sup_{f\in\mathcal{F}} E(f(X+y) - f(X))^2 \to 0$ as $\delta \to 0$ and, by assumption, $\sup_n \|\mu_n\| < \infty$. □

3.3.2 Classes of sets and functions in $\mathbb{R}^d$

Parts (b), (c) and (d) of the previous proposition could be considered in higher dimensions with only formal changes. However, if one is interested in applying Theorems 2 or 3 to classes of sets (or non-smooth functions) in $\mathbb{R}^d$, then the smoothness requirements of (b) and (c) will not be appropriate. In this section we show how Condition (7) can be verified for such classes, by using the notion of functions of bounded variation in higher dimensions. Indicators of many relevant classes of sets will be shown to be functions of bounded variation on $\mathbb{R}^d$.

We need some notation: As usual, if $f : \mathbb{R}^d \to \mathbb{R}$ is a locally integrable function, it gives rise to a distribution T_f acting on the space $\mathcal{D}(\mathbb{R}^d)$ of all infinitely differentiable real-valued functions on $\mathbb{R}^d$ with compact support via integration. We define the partial distributional derivative $D_w^\alpha f$ of a locally integrable function f as usual by the relation $D_w^\alpha T_f(\phi) = (-1)^{|\alpha|}\int_{\mathbb{R}^d} f(x)(D^\alpha\phi)(x)dx$, where $\phi \in \mathcal{D}(\mathbb{R}^d)$. If $D_w^\alpha T_f(\cdot)$ is, for every α with $|\alpha| = 1$, also a 'regular' distribution given by another locally integrable function g (or, alternatively, by a signed Borel measure μ), then we say that g (resp. μ) is the *weak derivative* of f, and we write $g = D_w^\alpha f$ ($\mu = D_w^\alpha f$); cf., e.g., p.42 in Ziemer [33]. If f is differentiable in the classical sense, then of course $D^\alpha f = D_w^\alpha f$ holds λ-a.e. (as can be easily checked by integration by parts). Also, we set $D_w f = D_w^\alpha f$ in case $\alpha = d = 1$.

The definition of bounded variation in (23) above does not immediately generalize to higher dimensions. On the other hand, the following definition is possible:

$$\mathcal{BV}(\mathbb{R}^d) = \{f : \mathbb{R}^d \to \mathbb{R} \text{ locally integrable, } D_w^\alpha f \in M(\mathbb{R}^d) \text{ for every } \alpha \text{ with } |\alpha| = 1\},$$

that is, $\mathcal{BV}(\mathbb{R}^d)$ is the space of all locally integrable functions that have weak partial derivatives of order one which are finite signed measures. The space $\mathcal{BV}(\mathbb{R}^d)$ can be equipped with the seminorm

$$\|f\|_{BV} = \max_{\alpha:|\alpha|=1} \left\| D_w^\alpha f \right\|$$

where we recall that $\left\| D_w^\alpha f \right\| = \left\| D_w^\alpha f \right\|'_{C_0}$ is the total variation of the measure $D_w^\alpha f$.

If $d = 1$, then $f \in \mathcal{BV}(\mathbb{R})$ if and only if there exists $g \in \mathsf{BV}(\mathbb{R})$ such that $f = g$ holds λ-a.e. To see this, for $f \in \mathcal{BV}(\mathbb{R})$ consider $\tilde{f}(x) = D_w f(-\infty, x]$. Clearly $\tilde{f} \in \mathsf{BV}(\mathbb{R})$ as it is the cumulative distribution function of $D_w f \in M(\mathbb{R})$. Then $D_w \tilde{f} = D_w f$ implies that $f - \tilde{f}$ equals a constant c almost everywhere, cf. p.51 in Schwartz [25], so $f \in [\tilde{f} + c]_\lambda$ where $\tilde{f} + c \in \mathsf{BV}(\mathbb{R})$. Conversely, if $f \in \mathsf{BV}(\mathbb{R})$, then $\tilde{f}(x) = f(x+) - f(-\infty+)$ defines a function in $\mathsf{BV}(\mathbb{R})$, right-continuous and with left limits, equal to zero at $-\infty$, which coincides with $f(x) - f(-\infty+)$ except at most for a countable number of points (see, e.g., Sect. 3.5 in Folland [9]). The finite signed measure $\nu_f(a, b] = \tilde{f}(b) - \tilde{f}(a)$ for $-\infty < a < b < \infty$ is precisely $D_w f$, so $f \in \mathcal{BV}(\mathbb{R})$.

As a consequence, if $f \in \mathcal{BV}(\mathbb{R})$ and $\tilde{f}(x) = D_w f(-\infty, x]$, then for all x, y in a set Ω_f such that $\lambda(\Omega_f^c) = 0$, $x < y$, we have

$$|f(y) - f(x)| = \left|\tilde{f}(y) - \tilde{f}(x)\right| = |D_w f(x, y]| \leq |D_w f|(x, y] \tag{26}$$

and Ω_f^c is countable if $f \in \mathsf{BV}(\mathbb{R})$. We shall use this fact repeatedly in the rest of this paper, where we will usually write ν_f for the measure $D_w f$. Also, by right-continuity of $\tilde{f}$ and the definitions, we have

$$\begin{aligned} \left\|\nu_f\right\| &= \|f\|_{BV} = \left\|\tilde{f}\right\|_{BV} = \left\|\tilde{f}\right\|_{TV} \\ &= \sup\left\{\sum_{i=1}^n |f(x_i) - f(x_{i-1})| : n \in \mathbb{N}, -\infty < x_1 < \cdots < x_n < +\infty,\ x_i \in \Omega_f\right\} \\ &\leq \|f\|_{TV}, \end{aligned} \tag{27}$$

where $\|f\|_{TV}$ may be infinite (if f is in $\mathcal{BV}(\mathbb{R})$ but not in $\mathsf{BV}(\mathbb{R})$).

In case $d > 1$, one has a characterization of $\mathcal{BV}(\mathbb{R}^d)$ which is in the same spirit. First, some notation: For $f \in \mathcal{BV}(\mathbb{R}^d)$, $1 \leq i \leq d$, $\tilde{x}_i := (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d) \in \mathbb{R}^{d-1}$ and $t \in \mathbb{R}$ define

$$f_{i,\tilde{x}_i}(t) = f(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_d),$$

the functions of one variable obtained from f by freezing all but the i-th coordinate. We will use the notation $x = (\tilde{x}_i, x_i)$ for $x = (x_1, \ldots, x_d)$.

Proposition 2 *Let $f \in \mathcal{L}^1(\mathbb{R}^d, \lambda)$. Then $f \in \mathcal{BV}(\mathbb{R}^d)$ if and only if $f_{i,\tilde{x}_i} \in \mathcal{BV}(\mathbb{R})$ for almost every $\tilde{x}_i \in \mathbb{R}^{d-1}$ and*

$$\int\limits_{\mathbb{R}^{d-1}} \left\| \nu_{f_{i,\tilde{x}_i}} \right\| d\tilde{x}_i < \infty$$

for every $1 \le i \le d$, and then, this integral is dominated by $\|f\|_{BV}$.

Proof The proposition can be deduced from the proof of Theorem 5.3.5 in Ziemer [33]. One only has to note that $\left\| \nu_{f_{i,\tilde{x}_i}} \right\|$ coincides—for $f \in \mathcal{BV}(\mathbb{R})$—with the essential variation defined in that theorem, and that Theorem 5.3.1 in Ziemer [33] in fact holds with $U = \mathbb{R}^d$. □

Using Proposition 2, one can prove the following result.

Proposition 3 *Let $\mathbb{P}$ be an absolutely continuous probability measure, $d\mathbb{P}(x) = p_0(x)d\lambda(x)$, and let $\{\mu_n\}_{n=0}^{\infty}$ be a proper approximate convolution identity. Let $\mathcal{F}$ be a uniformly bounded class of functions and assume either that*

(a) *the density p_0 is a bounded function, $\mathcal{F} \subseteq \mathcal{L}^1(\mathbb{R}^d, \lambda) \cap \mathcal{BV}(\mathbb{R}^d)$ and $\sup_{f\in\mathcal{F}} \|f\|_{BV} < \infty$, or*

(b) *the functions $f_{i,\tilde{x}_i}$ exist for all $1 \le i \le d$, $\tilde{x}_i \in \mathbb{R}^{d-1}$, $f \in \mathcal{F}$ and $\sup\{\| f_{i,\tilde{x}_i} \|_{BV} : 1 \le i \le d,\ \tilde{x}_i \in \mathbb{R}^{d-1},\ f \in \mathcal{F}\} < \infty$ holds.*

Then Condition (7) holds, that is

$$\sup_{f\in\mathcal{F}} E\left(\int\limits_{\mathbb{R}^d} (f(X+y) - f(X))d\mu_n(y)\right)^2 \to 0$$

as n tends to infinity.

Proof Using the same decomposition as in (25) above, we have

$$\begin{aligned}
&\sup_{f\in\mathcal{F}} \left(E\left(\int\limits_{\mathbb{R}^d} (f(X+y) - f(X))d\mu_n(y)\right)^2\right)^{1/2} \\
&\le \sup_{f\in\mathcal{F}} \int\limits_{|y|\le\delta} \left(E(f(X+y) - f(X))^2\right)^{1/2} d|\mu_n|(y) \\
&\quad + \sup_{f\in\mathcal{F}} \int\limits_{|y|>\delta} \left(E(f(X+y) - f(X))^2\right)^{1/2} d|\mu_n|(y) \\
&:= (\mathrm{I})_{n,\delta} + (\mathrm{II})_{n,\delta}
\end{aligned}$$

and, just as in the proof of Proposition 1, $\lim_n (\mathrm{II})_{n,\delta} = 0$ for all $\delta > 0$, noting that uniform boundedness of $\mathcal{F}$ implies $\sup_{y\in\mathbb{R}^d}\sup_{f\in\mathcal{F}} \mathbb{P}(f(\cdot+y))^2 < \infty$.

About the term $(\mathrm{I})_{n,\delta}$, we have the following: Let $u_i = (x_1+y_1,\dots,x_i+y_i, x_{i+1},\dots,x_d)$ for $i=1,\dots,d$ and let $u_0 = x = (x_1,\dots,x_d)$. Then, if D is the uniform bound for $\mathcal{F}$, we have

$$\int_{\mathbb{R}^d} (f(x+y)-f(x))^2 p_0(x)dx \le 2D \int_{\mathbb{R}^d} |f(x+y)-f(x)|\, p_0(x)dx$$
$$\le 2D \sum_{i=1}^{d} \int_{\mathbb{R}^d} |f(u_i)-f(u_{i-1})|\, p_0(x)dx.$$

We now consider the i-th summand. Set $\tilde{u}_i = (x_1+y_1,\dots,x_{i-1}+y_{i-1},x_{i+1},\dots,x_d)$ so that in the notation from above $u_i = (\tilde{u}_i, x_i+y_i)$ and $u_{i-1} = (\tilde{u}_i, x_i)$. Also, recall the measure $\nu_{f_{i,\tilde{u}_i}}$ from Proposition 2 and set $\nu_{i,\tilde{u}_i} := \nu_{f_{i,\tilde{u}_i}}$ for brevity. Then we have for (in case (a) almost) every $\tilde{u}_i$ that $|f(u_i)-f(u_{i-1})| \le \left|\nu_{i,\tilde{u}_i}\right|(x_i, x_i+y_i]$ for $y_i \ge 0$ and $|f(u_i)-f(u_{i-1})| \le \left|\nu_{i,\tilde{u}_i}\right|(x_i+y_i, x_i]$ for $y_i < 0$ except for x_i, x_i+y_i in a set of measure zero (depending on $\tilde{u}_i$), cf. (26). We only consider $y_i \ge 0$, the case $y_i < 0$ is similar. Since $0 \le y_i \le \delta$ we have—recalling $x = (\tilde{x}_i, x_i)$—that

$$\begin{aligned}\int_{\mathbb{R}^d} |f(u_i)-f(u_{i-1})|\, p_0(x)dx &\le \int_{\mathbb{R}^{d-1}}\int_{\mathbb{R}} \left|\nu_{i,\tilde{u}_i}\right|(x_i, x_i+y_i] p_0(\tilde{x}_i, x_i)dx_i d\tilde{x}_i \\ &\le \int_{\mathbb{R}^{d-1}}\int_{\mathbb{R}} \left|\nu_{i,\tilde{u}_i}\right|(x_i, x_i+\delta] p_0(\tilde{x}_i, x_i)dx_i d\tilde{x}_i \\ &= \int_{\mathbb{R}^{d-1}}\int_{\mathbb{R}} \left[\int_{u-\delta}^{u} p_0(\tilde{x}_i, x_i) d\left|\nu_{i,\tilde{u}_i}\right|(u)\right] dx_i d\tilde{x}_i \\ &= \int_{\mathbb{R}^{d-1}}\int_{\mathbb{R}} \left[\int_{\mathbb{R}} \mathbf{1}_{[u-\delta,u]} p_0(\tilde{x}_i, x_i)dx_i\right] d\left|\nu_{i,\tilde{u}_i}\right|(u) d\tilde{x}_i.\end{aligned} \tag{28}$$

Define $\lambda_\delta(\tilde{x}_i, u) = \int_{\mathbb{R}} \mathbf{1}_{[u-\delta,u]} p_0(\tilde{x}_i, x_i)dx_i$. Then

$$\sup_{u\in\mathbb{R}} \lambda_\delta(\tilde{x}_i, u) = \int_{\mathbb{R}} \mathbf{1}_{[u-\delta,u]} p_0(\tilde{x}_i, x_i)dx_i \to 0$$

as $\delta \to 0$ for each $\tilde{x}_i$ by absolute continuity of the integral of the function $p_0(\tilde{x}_i,\cdot) \in \mathcal{L}^1(\mathbb{R},\lambda)$. Now distinguish cases (a) and (b):

In case (a), set $\sup_{x\in\mathbb{R}^d} p_0(x) = C$. Then—using a change of variables $\tilde{x}_i \to \tilde{u}_i$ and Proposition 2—the last integral in (28) is bounded by

$$C\delta \int_{\mathbb{R}^{d-1}} \int_{\mathbb{R}} d\left|\nu_{i,\tilde{u}_i}\right|(u)d\tilde{x}_i \leq C\delta \int_{\mathbb{R}^{d-1}} \left\|\nu_{i,\tilde{u}_i}\right\| d\tilde{u}_i \leq C\delta \left\|f\right\|_{BV},$$

and this immediately gives $\lim_{\delta\to 0}\limsup_n ((\mathrm{I})_{n,\delta}) = 0$ since $\sup_n \|\mu_n\| < \infty$.

In case (b), the last integral in (28) is dominated by

$$\sup_{\tilde{x}_i\in\mathbb{R}^{d-1}} \left\|\nu_{i,\tilde{u}_i}\right\| \int_{\mathbb{R}^{d-1}} \sup_{u\in\mathbb{R}} \lambda_\delta(\tilde{x}_i, u)d\tilde{x}_i$$

and—since $\sup_{u\in\mathbb{R}} \lambda_\delta(\tilde{x}_i, u) \leq \int_{\mathbb{R}} p_0(\tilde{x}_i, x_i)dx_i$ and $\int_{\mathbb{R}^{d-1}}\int_{\mathbb{R}} p_0(\tilde{x}_i, x_i)dx_i d\tilde{x}_i = 1$—we have by dominated convergence that

$$\begin{aligned}&\lim_{\delta\to 0}\limsup_n (\mathrm{I})_{n,\delta}\\ &\quad\leq 2Dd \lim_{\delta\to 0}\sup_{f\in\mathcal{F}}\left(\sup_{\tilde{x}\in\mathbb{R}^{d-1}}\left\|\nu_{i,\tilde{u}_i}\right\|\sup_n \|\mu_n\|\right)\int_{\mathbb{R}^{d-1}}\sup_{u\in\mathbb{R}}\lambda_\delta(\tilde{x}_i, u)d\tilde{x}_i = 0,\end{aligned}$$

which completes the proof of Part (b). □

We note that Part (a) can also be proved by using results on Besov spaces, see Lemma 7 in Sect. 5. [This lemma implies that any $f \in \mathcal{BV}(\mathbb{R}^d)$ is contained in the Besov space $\mathcal{B}^1_{1\infty}(\mathbb{R}^d)$, which in turn yields that $\sup_{z\neq 0} |z|^{-r} \left(\int_{\mathbb{R}^d} |f(x+z) - f(x)|\, dx\right) < \infty$ holds for every $r < 1$. This L^1-Hölder condition—together with uniform boundedness of $\mathcal{F}$—could then be applied just as in the proof of Part (c) of Proposition 1.]

Now any class of sets $\mathcal{C} \subseteq \mathcal{B}_{\mathbb{R}^d}$ gives rise to a class of locally integrable functions $\{\mathbf{1}_C : C \in \mathcal{C}\}$. Furthermore, results in geometric measure theory imply that many sets $C \in \mathcal{B}_{\mathbb{R}^d}$ correspond to functions $\mathbf{1}_C$ that are contained in $\mathcal{BV}(\mathbb{R}^d)$. The class of all sets whose indicators are in $\mathcal{BV}(\mathbb{R}^d)$ is the class of all sets of finite perimeter, see p. 299 in [33]. The following corollary gives two simple examples (convex sets and sets with smooth differentiable boundaries) to which Proposition 3 applies. An open set $C \subseteq \mathbb{R}^d$ is said to be a C^α-domain ($\alpha \geq 2$) if its boundary ∂C is a $(d-1)$-dimensional compact Riemannian submanifold of $\mathbb{R}^d$ of smoothness of order α. The $(d-1)$-dimensional Hausdorff measure $H^{d-1}(\cdot)$ on ∂C defined in [33] then coincides with the usual Riemannian 'surface area' on ∂C.

Corollary 1 *Let $\mathbb{P}$ be an absolutely continuous probability measure on $\mathbb{R}^d$, $d\mathbb{P}(x) = p_0(x)d\lambda(x)$, and let $\{\mu_n\}_{n=0}^\infty$ be a proper approximate convolution identity. Let $\mathcal{C}$ be one of the following classes*:

(a) *the class $\mathcal{C}$ of all convex sets in $\mathbb{R}^d$*
(b) *the class $\mathcal{C}$ of bounded C^2-domains with $H^{d-1}(\partial C)$ bounded by a fixed constant.*

Then in both cases, but assuming in Part (b) *that* p_0 *is also bounded, we have*

$$\sup_{C\in\mathcal{C}} E\left(\int_{\mathbb{R}^d} (\mathbf{1}_C(X+y)-\mathbf{1}_C(X))d\mu_n(y)\right)^2 \to_{n\to\infty} 0.$$

Proof Part (a) follows from Proposition 3b, since the intersection of a convex set C with any line parallel to the coordinate axes is either empty, a point, or an interval. Hence $\sup\{\|\mathbf{1}_{C_{i,\tilde{x}_i}}\|_{BV} : 1 \le i \le d,\ \tilde{x}_i \in \mathbb{R}^{d-1},\ C \in \mathcal{C}\}$ is at most 2. For Part (b), if C is a C^2-domain, then it follows from the Gauß–Green theorem that $\|\mathbf{1}_C\|_{BV}$ can be bounded by a constant times $H^{d-1}(\partial C)$, see, e.g., Remark 5.4.2 in [33], so the result follows from Proposition 3a. □

Remark 5 We note that the indicator of an arbitrary convex set C is not necessarily in $\mathcal{BV}(\mathbb{R}^d)$, but it is if C has finite diameter: In this case, the orthogonal projection $\pi_i(C)$ of C on the subspace $x_i = 0$ $(1 \le i \le d)$ satisfies $H^{d-1}(\pi_i(C)) < \infty$, and it is easy to see that with $\alpha = (0, \dots, 1, \dots 0)$, where 1 is in the i-th place, one has $\left\|D_w^\alpha \mathbf{1}_C\right\| \le 2H^{d-1}(\pi_i(C))$, so $\|\mathbf{1}_C\|_{BV} \le 2\max_{1\le i\le d} H^{d-1}(\pi_i(C))$. In particular, if $\mathcal{C}$ is a collection of convex sets of diameter bounded by a fixed constant D, then we have in fact $\sup_{C\in\mathcal{C}} \|\mathbf{1}_C\|_{BV} \le 2v_{d-1}D^{d-1}$ where v_{d-1} is the volume of the unit sphere in $\mathbb{R}^{d-1}$. We will use this observation in Proposition 4.

4 Application of Theorems 2 and 3 to kernel estimators

In this section we illustrate how Theorems 2 and 3 can be applied—for several classes $\mathcal{F}$—to kernel density estimators. In most of this section, we discuss the case where $\mathcal{F}$ is a class of functions on the real line, i.e. $d = 1$. In Sect. 4.1.3 we treat the case of higher dimensions with a focus on classes of sets. Also, for simplicity, we will restrict ourselves to *symmetric* kernel functions.

Definition 2 A kernel $K : \mathbb{R} \to \mathbb{R}$ of real order $r > 0$ is a Lebesgue integrable function, symmetric around the origin, such that

$$\int_{\mathbb{R}} K(y)dy = 1, \quad \int_{\mathbb{R}} y^j K(y)dy = 0 \quad \text{for } j = 1, \dots, \{r\}, \quad \text{and}$$

$$\int_{\mathbb{R}} |y|^r |K(y)|dy < \infty$$

where $\{r\}$ is the largest integer strictly smaller than r.

If $h_n > 0$ is a sequence of real numbers converging to zero, then the classical kernel density estimator is given by

$$\mathbb{P}_n * K_{h_n}(x) = \frac{1}{nh_n}\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right). \tag{29}$$

We recall that $d\mu_n(y) = h_n^{-1}K(h_n^{-1}y)dy$ is a proper approximate identity (if $h_n \to 0$). In density estimation, the following assumption is natural:

Condition 1 *The random variables $X_1, \ldots, X_n$ are i.i.d. according to the law $\mathbb{P}$ on $\mathbb{R}$, and $d\mathbb{P}(x) = p_0(x)d\lambda(x)$. In fact, we take the variables X_i to be the coordinate projections of the infinite product probability space* $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}_{\mathbb{R}^{\mathbb{N}}}, \mathbb{P}^{\mathbb{N}})$.

We will give sufficient conditions on h_n and the order r of the kernel that imply the UCLT

$$\sqrt{n}(\mathbb{P}_n * K_{h_n} - \mathbb{P}) \rightsquigarrow_{\ell^\infty(\mathcal{F})} \mathbb{G} \tag{30}$$

for many concrete classes $\mathcal{F}$.

4.1 Donsker classes

In this subsection, we discuss UCLTs for kernel density estimators when the class of functions $\mathcal{F}$ is Donsker. The main additional effort will be to give sharp bounds for the bias term (8),

$$\sqrt{n}h_n^{-1}\left|E\int_{\mathbb{R}^d}(f(X+y)-f(X))K(y/h_n)dy\right| = \sqrt{n}|E(f * K_{h_n}(X) - f(X))| \tag{31}$$

(noting the symmetry of K), in Theorem 2. We first treat the case where nothing else is known about the Donsker class $\mathcal{F}$ other than that it is translation-invariant. Then, for kernel density estimators it was already noticed in van der Vaart [29] that the bias term can be controlled by straightforward methods that use smoothness assumptions on the true density p_0.

For integer $t > 0$, we denote by $\mathcal{W}_1^t(\mathbb{R})$ the space of functions f whose derivatives up to order $t-1$ are in $\mathcal{L}^1(\mathbb{R}, \lambda)$, $D^{t-1}f$ is contained in $\mathcal{BV}(\mathbb{R})$ and $D_w^t f$ is absolutely continuous (i.e. $D_w^t f \in \mathcal{L}^1(\mathbb{R}, \lambda)$, also see Sect. 3.3.2). By convention we set $\mathcal{W}_1^0(\mathbb{R}) = \mathcal{L}^1(\mathbb{R}, \lambda)$.

Lemma 3 *Let $d\mu_h(x) = h^{-1}K(x/h)d\lambda(x)$ where K is a kernel of order t and let $d\mathbb{P}(x) = p_0(x)dx$ be a probability measure with a bounded density $p_0 \in \mathcal{W}_1^t(\mathbb{R})$. Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded function. Then*

$$|E(f * \mu_h(X) - f(X))| \le 2\|f\|_\infty \|D_w^t p_0\|_{1,\lambda} \left(\int_{\mathbb{R}} |y|^t K(y)|dy\right) h^t.$$

Proof Routine application of Taylor's formula and the definition of $\mathcal{W}_1^t(\mathbb{R})$. □

Given the preceding lemma, the following theorem—which is essentially van der Vaart's [28] observation for kernel smoothing—is then a straightforward application of Theorem 2. [Note that conditions for (7) to hold were already given in Proposition 1.]

Theorem 4 *Let Condition 1 hold with $p_0 \in \mathcal{W}_1^t(\mathbb{R})$ for some $t > 0$. Let K be a kernel of order t and let $h_n > 0$ be such that $nh_n^{2t} \to_{n\to\infty} 0$. Let $\mathcal{F}$ be a translation invariant uniformly bounded $\mathbb{P}$-Donsker class of functions satisfying Condition (7). Then*

$$\sqrt{n}(\mathbb{P}_n * K_{h_n} - \mathbb{P}) \rightsquigarrow_{\ell^\infty(\mathcal{F})} \mathbb{G},$$

where $\mathbb{G}$ is the $\mathbb{P}$-Brownian bridge indexed by $\mathcal{F}$.

Proof This follows easily from Theorem 2 and Lemma 3. □

4.1.1 The bias term revisited

Theorem 4 is restrictive in various ways. First, the requirement that the true density p_0 possesses some smoothness seems unnatural. One would rather expect that—as soon as p_0 is known to exist—a little smoothing should not make the density estimator to deviate much from the empirical process in $\ell^\infty(\mathcal{F})$; more concretely, one would expect the smoothness parameter $t = 0$ to be admissible if one is only interested in a UCLT in $\ell^\infty(\mathcal{F})$. Secondly, Theorem 4 rules out classical kernel density estimators with 'textbook' bandwidth $h_n \simeq n^{-1/(2t+1)}$, since then $\lim_n nh_n^{2t} = \lim_n n^{1-2t/(2t+1)} = \infty$ holds for every $t \in \mathbb{N}$. Consequently one would have to take faster bandwidths than $h_n \simeq n^{-1/(2t+1)}$ which would subsequently slow down the rate of convergence of the kernel density estimator in mean integrated squared error. In particular, Theorem 4 does not allow for the construction of a 'plug-in estimator' in the sense of Bickel and Ritov [3], see the introduction.

Both deficiencies have their origin in the debiasing Lemma 3. Maybe not surprisingly, it turns out that this bias bound is too crude for many interesting cases. Note that, while Lemma 3 uses the fact that p_0 is smooth, it does *not* use the fact that also f may possess some 'regularity'. We begin with a simple observation in this direction, where the idea will be to use smoothness of f and p_0 *simultaneously*.

Lemma 4 *Let $d\mu_h(x) = h^{-1}K(x/h)d\lambda(x)$ where K is a kernel of order $m \geq 0$ and let $d\mathbb{P}(x) = p_0(x)dx$ be a probability measure with a density $p_0 \in \mathcal{L}^q(\mathbb{R}, \lambda)$. Let further $f \in \mathcal{L}^p(\mathbb{R}, \lambda)$ where $1/p + 1/q = 1$ and set $\bar{f}(x) = f(-x)$ for $x \in \mathbb{R}$. Then $\bar{f} * p_0 \in \mathsf{C}(\mathbb{R})$ and we have for the bias term that*

$$|E(f * \mu_h(X) - f(X))| = \left| \int_{\mathbb{R}} K(t)[\bar{f} * p_0(ht) - \bar{f} * p_0(0)]dt \right|.$$

Proof Note that $f * K \in \mathcal{L}^p(\mathbb{R}, \lambda)$ by (6) which implies (by Hölder's inequality) that $(f * K) \cdot p_0$ is Lebesgue-integrable for every $f \in \mathcal{F}$ since $p_0 \in \mathcal{L}^q(\mathbb{R}, \lambda)$. Now, by change of variables and Fubini, we have that

$$
\begin{aligned}
E(f * \mu_h(X) - f(X)) &= h^{-1} \int_{\mathbb{R}} \int_{\mathbb{R}} (f(x-y) - f(x)) K(y/h) dy p_0(x) dx \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} (f(x-th) - f(x)) K(t) dt p_0(x) dx \\
&= \int_{\mathbb{R}} K(t) \left[\int_{\mathbb{R}} f(x-th) p_0(x) dx - \int_{\mathbb{R}} f(x) p_0(x) dx \right] dt \\
&= \int_{\mathbb{R}} K(t) \left[\int_{\mathbb{R}} \bar{f}(th-x) p_0(x) dx - \int_{\mathbb{R}} \bar{f}(0-x) p_0(x) dx \right] dt \\
&= \int_{\mathbb{R}} K(t) [\bar{f} * p_0(ht) - \bar{f} * p_0(0)] dt.
\end{aligned}
$$

The expression on the right hand side is well defined since $f \in \mathcal{L}^p(\mathbb{R}, \lambda)$ implies $\bar{f} \in \mathcal{L}^p(\mathbb{R}, \lambda)$ and hence $\bar{f} * p_0 \in \mathsf{C}(\mathbb{R})$ by (5). □

So in the analysis of the bias term one really looks at smoothness of $\bar{f} * p_0$ instead of at smoothness of only one of the factors of the convolution product. If nothing at all is known about $\mathcal{F}$ except that it is Donsker, then $\bar{f} * p_0$ will inherit the smoothness of p_0 only and we are back at Theorem 4. But the advantage of Lemma 4 is that one can combine the information on p_0 and $\mathcal{F}$. This can be done, e.g., by using the following lemma. Recall from (26) that, if $f \in \mathcal{BV}(\mathbb{R})$, then there exists $\tilde{f} \in [f]_\lambda$ such that $\nu_f((a,b]) = \tilde{f}(b) - \tilde{f}(a)$ holds, where $\nu_f = D_w f \in M(\mathbb{R})$ is a finite signed measure.

Lemma 5 (a) *Let $f \in \mathsf{C}(\mathbb{R})$ be such that Df exists and is bounded, and let $\nu \in M(\mathbb{R})$ be a finite signed measure. Then, for every $x \in \mathbb{R}$, $D(f * \nu)(x)$ exists and $D(f * \nu)(x) = (Df * \nu)(x)$ holds.*

(b) *Let $g \in \mathsf{C}(\mathbb{R})$, let $f \in \mathcal{BV}(\mathbb{R})$, and suppose that $g * f(x)$ is defined for every $x \in \mathbb{R}$. Then, for every $x \in \mathbb{R}$, $D(g * f)(x)$ exists and $D(g * f)(x) = (g * \nu_f)(x)$ holds, where ν_f is the finite signed measure defined by $\nu_f((a,b]) = \tilde{f}(b) - \tilde{f}(a)$.*

Proof Part (a) By the mean value theorem and boundedness of Df, $h^{-1}[f(x-y+h) - f(x-y)]$ is uniformly bounded, hence, by dominated convergence, we have

$$
\begin{aligned}
D(f * \nu)(x) &= \lim_{h \to 0} h^{-1} \int_{-\infty}^{\infty} (f(x-y+h) - f(x-y)) \, d\nu(y) dy \\
&= \int_{-\infty}^{\infty} \lim_{h \to 0} h^{-1} [f(x-y+h) - f(x-y)] d\nu(y) dy \\
&= \int_{-\infty}^{\infty} Df(x-y) d\nu(y) dy = (Df * \nu)(x),
\end{aligned}
$$

the last integral being convergent for every $x \in \mathbb{R}$ since Df is bounded.

Part (b) Clearly, $(g * f)(x) = (g * \tilde{f})(x)$ holds for every $x \in \mathbb{R}$, so we have

$$\begin{aligned} D(g * f)(x) &= \lim_{h \to 0} h^{-1} \int_{-\infty}^{\infty} \left(\tilde{f}(x - y + h) - \tilde{f}(x - y) \right) g(y) dy \\ &= \lim_{h \to 0} h^{-1} \int_{-\infty}^{\infty} \int_{x-y}^{x-y+h} d\nu_f(t) g(y) dy \\ &= \lim_{h \to 0} h^{-1} \int_{-\infty}^{\infty} \int_{x-t}^{x-t+h} g(y) dy d\nu_f(t) \\ &= \int_{-\infty}^{\infty} \lim_{h \to 0} h^{-1} \int_{x-t}^{x-t+h} g(y) dy d\nu_f(t) \\ &= \int_{-\infty}^{\infty} g(x - t) d\nu_f(t) = g * \nu_f(x) \end{aligned}$$

for every x. The first two equalities follow from the definition of convolution, and from definition of the measure ν_f. The third is Fubini and the fourth equality follows from $g \in \mathsf{C}(\mathbb{R})$ and Lebesgue's dominated convergence theorem. The fifth equality follows from the fundamental theorem of calculus. The integral in the last line converges for every $x \in \mathbb{R}$ by boundedness of g. □

If now, for example, $f \in \mathsf{C}^1(\mathbb{R})$ and $p_0 \in \mathsf{BV}(\mathbb{R})$, then, by applying the above lemma twice, we have

$$D^2(f * p_0) = D(Df * p_0) = (Df * \nu_{p_0}) \tag{32}$$

and hence

$$\left\| D^2(f * p_0) \right\|_\infty \leq \|Df\|_\infty \left\| \nu_{p_0} \right\| < \infty \tag{33}$$

by (6) and since $\left\| \nu_{p_0} \right\| \leq \|p_0\|_{TV} < \infty$ holds (see (27)), so it follows that the convolution product $p_0 * f$ is twice differentiable whereas p_0 is only once (and this only in the weak sense). The above ideas (and variations thereof) are exploited in the next sections to improve upon Theorem 4 in many interesting special cases. Note that a general version of Lemma 5 (including the multivariate case) is proved in Lemma 12 (by using Fourier-analytical methods).

Remark 6 It is instructive to consider the special case where $f_t(x) = 1_{(-\infty,t]}(x)$ in Lemma 4. Then $p_0 * \bar{f_t} = p_0 * 1_{[-t,\infty)}(x)$ just equals the distribution function $\int_{-\infty}^{t+x} p_0 d\lambda$, which—as was already noted by Bickel and Ritov [3]—can be seen directly to have one more derivative than $p_0 \in \mathsf{C}(\mathbb{R})$. [Note that then $\left\| \nu_{1_{(-\infty,t]}} \right\| = \|\delta_t\| = 1$ for every t in (33).]

4.1.2 Examples of Donsker classes

1. *Bounded Variation Classes* Recall that any bounded subset $\mathcal{U}$ of $\mathsf{BV}(\mathbb{R})$ is a uniform Donsker class, (see, e.g., Dudley [8, p. 329]). We introduce the parameter k in order to gain some flexibility in the choice of the order of the kernel, see the remark after the theorem.

Theorem 5 *Let Condition 1 hold and suppose that p_0 is a bounded function, in which case we set $t = 0$ in what follows, or assume $p_0 \in \mathsf{C}^t(\mathbb{R})$ for some real $t > 0$. Let $\mathcal{U}$ be a bounded subset of $\mathsf{BV}(\mathbb{R})$. Let K be a kernel of order $r = t + 1 - k$ for some k, $0 \leq k < t + 1$. If $h_n > 0$ is such that $h_n^{t+1-k} n^{1/2} \to_{n\to\infty} 0$, then*

$$\sqrt{n}(\mathbb{P}_n * K_{h_n} - \mathbb{P}) \rightsquigarrow_{\ell^\infty(\mathcal{U})} \mathbb{G},$$

where $\mathbb{G}$ is the $\mathbb{P}$-Brownian bridge indexed by $\mathcal{U}$.

Proof Since $(\mathbb{P}_n * K_{h_n} - \mathbb{P})(f + c) = (\mathbb{P}_n * K_{h_n} - \mathbb{P})(f)$ for any constant c, we may assume without loss of generality that the class $\mathcal{U}$ is uniformly bounded. [Otherwise, consider $\mathcal{U}' = \{f - f(-\infty+) : f \in \mathcal{U}\}$, which is uniformly bounded by $\sup_{f\in\mathcal{U}} \|f\|_{TV}$.] Now, $\mathcal{U}$ being uniformly bounded, we have $\mathcal{U} \subseteq \mathcal{L}^1(|\mu_n|)$ as $K \in \mathcal{L}^1(\mathbb{R}, \lambda)$ and also $\sup_{f\in\mathcal{U}} \int_{\mathbb{R}} \|f(\cdot - y)\|_{2,\mathbb{P}} d\,|\mu_n|\,(y) < \infty$. Moreover, Condition (7) of Theorem 2 is verified in Part (a) of Proposition 1.

Hence it remains to verify the bias condition (8). If $t = 0$ it follows as in the proof of Lemma 5 (without limits) that $\{\bar{f} * p_0 : f \in \mathcal{U}\}$ is a bounded subset of the space of bounded Lipschitz functions on $\mathbb{R}$; therefore, using Lemma 4, we obtain that the bias is dominated by Ch_n^{1-k} with C depending only on K and $\sup_{f\in\mathcal{U}} \|f\|_{TV}$, which completes the proof of the theorem for $t = 0$.

Finally, we consider the bias for $t > 0$. By using Lemma 5 iteratively, we have, for $f \in \mathcal{U}$,

$$D^{[t]+1}(p_0 * f) = D^{[t]}(p_0 * \nu_f) = D^{[t]} p_0 * \nu_f \tag{34}$$

recalling that $[t]$ denotes the integer part of t. In particular, $D^\alpha(p_0 * f)$ exists and is a bounded and continuous function for every $0 \leq \alpha \leq [t] + 1$. We distinguish two cases:

(a) In case $t = [t]$ is an integer, we have from (34) and (6) that

$$\sup_{f\in\mathcal{U}} \left\| D^{t+1}(p_0 * f) \right\|_\infty \leq \left\| D^t p_0 \right\|_\infty \sup_{f\in\mathcal{U}} \left\| \nu_f \right\| < \infty, \tag{35}$$

recalling $\|\nu_f\| \leq \|f\|_{TV}$ (see (27) above).

(b) In the noninteger case $t > [t]$ we show that $D^{[t]+1}(p_0 * f)$ is Hölder-continuous of order $t - [t]$. We have

$$\begin{aligned}
&|h|^{[t]-t}\left|D^{[t]+1}(p_0 * f)(x+h) - D^{[t]+1}(p_0 * f)(x)\right| \\
&= \left|\int_{\mathbb{R}} |h|^{[t]-t}\left(D^{[t]}p_0(x+y+h) - D^{[t]}p_0(x+y)\right) d\nu_f(y)\right| \\
&\leq \left\|D^{[t]}p_0\right\|_{t-[t],\infty} \left\|\nu_f\right\|,
\end{aligned} \tag{36}$$

which is bounded uniformly in $f \in \mathcal{U}$, since $p_0 \in \mathsf{C}^t(\mathbb{R})$ implies $D^{[t]}p_0 \in \mathsf{C}^{t-[t]}(\mathbb{R})$ and since the variation of ν_f is uniformly bounded (using again $\|\nu_f\| \leq \|f\|_{TV}$).

Now, to bound the bias term, we use Lemma 4 (and the identity (31)): Note that, if $f \in \mathcal{U}$, then also $\bar{f} \in \mathcal{U}$. Consider first $t + 1 - k$ noninteger. Then by a Taylor expansion,

$$\begin{aligned}
&\left|E(f * K_{h_n}(X) - f(X))\right| \\
&= \Bigg|\int_{\mathbb{R}} K(t)[D(p_0 * \bar{f})(0)th_n \\
&\quad + \cdots + \frac{1}{[t+1-k]!} D^{[t+1-k]}(p_0 * \bar{f})(0)t^{[t+1-k]}h_n^{[t+1-k]}]dt \\
&\quad + \frac{1}{[t+1-k]!} h_n^{[t+1-k]} \int_{\mathbb{R}} K(t)t^{[t+1-k]}[D^{[t+1-k]}(p_0 * \bar{f})(\zeta h_n t) \\
&\quad - D^{[t+1-k]}(p_0 * \bar{f})(0)]dt\Bigg|
\end{aligned} \tag{37}$$

for some $0 < \zeta < 1$. The first $[t + 1 - k]$ terms in the above display are all equal to zero by the choice of the kernel (cf. Definition 2). Note next that by either (35) or (36) above, $D^{[t+1-k]}(p_0 * \bar{f})$ is contained in $\mathsf{C}^{\alpha}(\mathbb{R})$ where $\alpha = t+1-k-[t+1-k] > 0$, and—using again the assumption on the kernel—we have that the last term in (37) is bounded in absolute value by

$$C h_n^{t+1-k} \int_{\mathbb{R}} |K(t)|\, |t|^{t+1-k},$$

the constant C being equal to

$$C = \frac{1}{[t+1-k]!} \sup_{f \in \mathcal{U}} \left\|D^{[t+1-k]}(p_0 * f)\right\|_{\infty,\alpha} < \infty.$$

Consequently, by the choice of h_n, we obtain the bound

$$\sup_{f\in\mathcal{U}} \sqrt{n}\left|E(f * K_{h_n}(X) - f(X))\right| = O(\sqrt{n}h_n^{t+1-k}) = o(1) \tag{38}$$

for the bias term. In case $t+1-k$ integer,

$$\begin{aligned}
&\left|E(f * K_{h_n}(X) - f(X))\right| \\
&= \Bigg| \int_{\mathbb{R}} K(t)[D(p_0 * \bar{f})(0)th_n \\
&\quad + \cdots + \frac{1}{\{t+1-k\}!} D^{\{t+1-k\}}(p_0 * \bar{f})(0)t^{\{t+1-k\}}h_n^{\{t+1-k\}}]dt \\
&\quad + \frac{1}{(t+1-k)!} h_n^{t+1-k} \int_{\mathbb{R}} K(t)t^{t+1-k}[D^{t+1-k}(p_0 * \bar{f})(\zeta h_n t)]dt \Bigg|
\end{aligned} \tag{39}$$

for some $0 < \zeta < 1$ yields (38) by choice of the kernel and since $D^{t+1-k}(p_0 * \bar{f})$ is sup-norm bounded uniformly in $f \in \mathcal{U}$ by either (35) or (36) above. □

Remark 7 (Order of the kernel and MISE-optimal rates) We note that the parameter $k \geq 0$ allows for some flexibility in the choice of the order of the kernel. In the most interesting case $h_n \simeq n^{-1/(2t+1)}$, where the kernel estimator simultaneously achieves optimal rates of convergence in squared error loss, we have $h_n^{t+1-k}n^{1/2} \to_{n\to\infty} 0$ if $0 \leq k < 1/2$, so the kernel has to be of order

$$r > t + 1/2. \tag{40}$$

In other words, if one wants classical kernel density estimators to satisfy a UCLT over bounded variation classes, the order of the kernel has to be chosen larger by $1/2$ than usual. See also Remark 8 below for more discussion.

We also state the following immediate corollary for the cumulative distribution function of the kernel estimator, which is essentially due to Bickel and Ritov [3].

Corollary 2 *Let Condition 1 hold and suppose that p_0 is a bounded function (case $t = 0$) or assume that $p_0 \in \mathsf{C}^t(\mathbb{R})$ for some real $t > 0$. Let K be a kernel of order $r > t + 1/2$. Choose $h_n > 0$ of order $h_n \simeq n^{-1/(2t+1)}$. Define the cumulative distribution functions $\tilde{F}_n(t) = \int_{-\infty}^t (\mathbb{P}_n * K_{h_n})(x)dx$ as well as $F(t) = \int_{-\infty}^t p_0(x)dx$. Then*

$$\sqrt{n}(\tilde{F}_n - F) \rightsquigarrow_{\mathsf{C}(\mathbb{R})} \mathbb{G},$$

where $\mathbb{G}$ is the $\mathbb{P}$-Brownian bridge in $\mathsf{C}(\mathbb{R})$.

2. *Hölder and Lipschitz classes* We now deal with the Hölder classes

$$\mathcal{F}_{s,\infty} = \{f \in \mathsf{C}(\mathbb{R}) : \|f\|_{s,\infty} \leq 1\}, \tag{41}$$

Note that $\mathcal{F}_{s,\infty}$ for $s < 1$ contains the usual Lipschitz classes (considered, e.g., in [12]). If $s > 1/2$ and $\int_{\mathbb{R}} |x|^{1+\eta} \, d\mathbb{P}(x) < \infty$ holds for some $\eta > 0$, then $\mathcal{F}_{s,\infty}$ is a $\mathbb{P}$-Donsker class (see, e.g., Corollary 5 and Sect. 3.3.1 in Nickl and Pötscher [20]), and these conditions can be shown to be (essentially) sharp (cf. Theorem 2 in [12] for the Lipschitz case and also Theorems 4 and 6 in Nickl [17]). Recall the spaces $\mathcal{W}_1^t(\mathbb{R})$ defined before Lemma 3.

Theorem 6 *Let Condition 1 hold and suppose that* $\int_{\mathbb{R}} |x|^{1+\eta} \, d\mathbb{P}(x) < \infty$ *is satisfied for some* $\eta > 0$ *and that* $p_0 \in \mathcal{W}_1^t(\mathbb{R})$ *for some* $t \geq 0$*. Let* $\mathcal{F}_{s,\infty}$ *with* $s > 1/2$ *be given as in* (41)*. Let* K *be a kernel of order* $r = t + s - k$ *for some* k*,* $0 \leq k < t + s$*. If* $h_n > 0$ *is such that* $h_n^{t+s-k} n^{1/2} \to_{n\to\infty} 0$*, then*

$$\sqrt{n}(\mathbb{P}_n * K_{h_n} - \mathbb{P}) \rightsquigarrow_{\ell^\infty(\mathcal{F}_{s,\infty})} \mathbb{G},$$

where $\mathbb{G}$ *is the* $\mathbb{P}$*-Brownian bridge indexed by* $\mathcal{F}_{s,\infty}$*.*

Proof Note first that, since $\mathcal{F}_{s,\infty}$ is $\mathbb{P}$-Donsker and closed under translations, we can apply Theorem 2b. Since $\mathcal{F}_{s,\infty}$ is uniformly bounded, we have $\mathcal{F}_{s,\infty} \subseteq \mathcal{L}^1(|\mu_n|)$, $\sup_{f \in \mathcal{F}_{s,\infty}} \int_{\mathbb{R}} \|f(\cdot - y)\|_{2,\mathbb{P}} d\,|\mu_n|\,(y) < \infty$, and Part 2 of Proposition 1 verifies Condition (7). It remains to treat the bias term from Condition (8). First, observe the following. For $f \in \mathcal{W}_1^t(\mathbb{R})$, we have, for every $0 \leq \alpha \leq t - 1$, that $D^\alpha f \in \mathcal{BV}(\mathbb{R})$, and the corresponding finite signed measures $\nu_{D^\alpha f}$ defined in Part (b) of Lemma 5 are in fact absolutely continuous. Using the convention $d\nu_{D^{t-1}p_0} = p_0 d\lambda$ in case $t = 0$, we have, for $f \in \mathcal{F}_{s,\infty}$,

$$D^{t+[s]}(p_0 * f) = D^t(p_0 * D^{[s]} f) = \nu_{D^{t-1}p_0} * D^{[s]} f$$

by Lemma 5. In particular, $D^\alpha(p_0 * f)$ exists and is a bounded and continuous function for every $0 \leq \alpha \leq t + [s]$. If $s = [s]$ integer, we conclude

$$\sup_{f \in \mathcal{F}_{s,\infty}} \left\| D^{t+s}(p_0 * f) \right\|_\infty \leq \left\| \nu_{D^{t-1}p_0} \right\| \sup_{f \in \mathcal{F}_{s,\infty}} \left\| D^s f \right\|_\infty < \infty \tag{42}$$

by (6), and if s is noninteger, it follows as in (36) that

$$\sup_{f \in \mathcal{F}_{s,\infty}} \left\| D^{t+[s]}(p_0 * f) \right\|_{s-[s],\infty} < \infty \tag{43}$$

since $D^{[s]} f \in \mathsf{C}^{s-[s]}(\mathbb{R})$ by assumption.

Now, to finish the proof, we proceed as in Theorem 5 and apply Lemma 4 (and the identity (31)). Consider first $t + s - k$ noninteger. Then we have from a Taylor expansion that

$$
\begin{aligned}
&\left|E(f * K_{h_n}(X) - f(X))\right| \\
&= \Bigg| \int_{\mathbb{R}} K(t)[D(p_0 * \bar{f})(0) t h_n \\
&\quad + \cdots + \frac{1}{[t+s-k]!} D^{[t+s-k]}(p_0 * \bar{f})(0) t^{[t+s-k]} h_n^{[t+s-k]}] dt \\
&\quad + \frac{1}{[t+s-k]!} h_n^{[t+s-k]} \int_{\mathbb{R}} K(t) t^{[t+s-k]} [D^{[t+s-k]}(p_0 * \bar{f})(\zeta h_n t) \\
&\quad - D^{[t+s-k]}(p_0 * \bar{f})(0)] dt \Bigg|
\end{aligned}
$$

holds for some $0 < \zeta < 1$. The first $[t+s-k]$ terms in the above display are all equal to zero by choice of the kernel. By either (42) or (43), $D^{[t+s-k]}(p_0 * \bar{f})$ is contained in $\mathsf{C}^{\alpha}(\mathbb{R})$ for $\alpha = t+s-k-[t+s-k] > 0$, and, by the same arguments as below (37) in the proof of Theorem 5, we conclude that the last term in is bounded in absolute value by

$$
C h_n^{t+s-k} \int_{\mathbb{R}} |K(t)|\,|t|^{t+s-k},
$$

for some fixed constant C. This gives

$$
\sup_{f \in \mathcal{U}} \sqrt{n}\left|E(f * K_{h_n}(X) - f(X))\right| = O(\sqrt{n} h_n^{t+s-k}) = o(1)
$$

for the bias term. The case $t+s-k$ integer is also similar as in the proof of Theorem 5, and we omit it. □

Remark 8 (Order of the kernel and MISE-optimal rates II) Again (cf. Remark 7), the parameter $k \geq 0$ allows for some flexibility in the choice of the order of the kernel. In the most interesting case $h_n \simeq n^{-1/(2t+1)}$, one now has that $h_n^{t+s-k} n^{1/2} \to_{n\to\infty} 0$ if $0 \leq k < s - 1/2$, so—*even as s varies*—the order r of the kernel only has to satisfy

$$
r > t + 1/2
$$

as in (40). So the 'rule of thumb' to obtain UCLTs—and hence Bickel and Ritov's [3] plug-in property—for classical kernel density estimators on the real line with bandwidth $h_n \simeq n^{-1/(2t+1)}$ is to choose the kernel of an order higher by 1/2 than usual.

3. *Sobolev classes* Denote by F the Fourier–Plancherel transform on $\mathcal{L}^2(\mathbb{R}, \lambda)$ (see Sect. 5). The Sobolev space of order $s > 0$ is defined as

$$
\mathcal{W}_2^s(\mathbb{R}) = \{f \in \mathcal{L}^2(\mathbb{R}, \lambda) : \|f\|_{s,2} := \left\| F^{-1}[(1+|\cdot|^2)^{s/2} F f(\cdot)] \right\|_{2,\lambda} < \infty\}. \quad (44)
$$

As is well known (e.g., Theorem 3.5.1 in Malliavin [15]), $f \in \mathcal{W}_2^s(\mathbb{R})$ is equivalent to $D_w^\alpha f \in \mathcal{L}^2(\mathbb{R}, \lambda)$ for $0 \le \alpha \le s$ and $s > 0$ integer, where D_w^α denotes the weak differential operator of order α, see Sect. 3.3.2 above. Note that—if $s > 1/2$—the Sobolev space can be viewed as consisting of bounded continuous functions (in the sense that, if $f \in \mathcal{W}_2^s(\mathbb{R})$, then $[f]_\lambda$ contains one element which is in $\mathsf{C}(\mathbb{R})$). It is known that norm-balls in $\mathcal{W}_2^s(\mathbb{R})$ are universal Donsker classes if and only if $s > 1/2$ holds, see Marcus [16].

Theorem 7 *Let Condition 1 hold and suppose that $p_0 \in \mathcal{W}_2^t(\mathbb{R})$ for some $t \ge 0$. Let $\mathcal{F}_{s,2} = \{f \in \mathsf{C}(\mathbb{R}) : \|f\|_{s,2} \le 1\}$ with $s > 1/2$. Let K be a kernel of order $r = t + s - k$ for some k, $0 \le k < t + s$. If $h_n > 0$ is such that $h_n^{t+s-k} n^{1/2} \to_{n\to\infty} 0$, then*

$$\sqrt{n}(\mathbb{P}_n * K_{h_n} - \mathbb{P}) \rightsquigarrow_{\ell^\infty(\mathcal{F}_{s,2})} \mathbb{G},$$

where $\mathbb{G}$ is the $\mathbb{P}$-Brownian bridge indexed by $\mathcal{F}_{s,2}$.

Proof We set shorthand $\langle u\rangle^\alpha = (1 + |u|^2)^{\alpha/2}$ in what follows. Since $\mathcal{F}_{s,2}$ is universal Donsker and closed under translations, we can apply Theorem 2b. Note that $\mathcal{F}_{s,2}$ is uniformly bounded in view of

$$\sup_{f\in\mathcal{F}_{s,2}} \|f\|_\infty \le \sup_{f\in\mathcal{F}_{s,2}} \|Ff\|_{1,\lambda} \le \sup_{f\in\mathcal{F}_{s,2}} \|f\|_{s,2} \left\|\langle u\rangle^{-s}\right\|_{2,\lambda} < \infty$$

by Fourier-inversion and the Plancherel theorem (see, e.g., Sect. III.2.4 in Malliavin [15]), hence $\mathcal{F}_{s,2} \subseteq \mathcal{L}^1(|\mu_n|)$ and

$$\sup_{f\in\mathcal{F}_{s,2}} \int_{\mathbb{R}} \|f(\cdot - y)\|_{2,\mathbb{P}} d\,|\mu_n|\,(y) < \infty$$

for every n. Even more, one can show that $\mathcal{F}_{s,2}$ is a bounded subset of $\mathsf{C}^{s-1/2}(\mathbb{R})$ by using embeddings in 2.7.1 and Theorem 2.5.6/2 in Triebel [26], hence Part 2 of Proposition 1 verifies Condition (7). We next verify the bias Condition (8). In the case $t + s$ integer, one has by well known facts in Fourier analysis (the relationships $F(Df)(u) = iuFf(u)$, $F(p_0 * f) = \sqrt{2\pi}\,Fp_0Ff$, Fourier inversion as well as Plancherel's theorem) that

$$\begin{aligned}
&\|p_0 * f\|_\infty + \|D(p_0 * f)\|_\infty + \cdots + \left\|D^{t+s}(p_0 * f)\right\|_\infty \\
&\le \sqrt{2\pi}\left[\int_{\mathbb{R}} (1 + |u| + \cdots + |u|^{t+s})\,|Fp_0(u)Ff(u)|\,du\right] \\
&\le C\left\|\langle u\rangle^{t+s} Fp_0Ff\right\|_{1,\lambda} = C\left\|\langle u\rangle^{t} Fp_0 \langle u\rangle^{s} Ff\right\|_{1,\lambda} \\
&\le C\,\|p_0\|_{t,2}\,\|f\|_{s,2}
\end{aligned}$$

for some $0 < C < \infty$, in particular, $p_0 * f$ is contained in $\mathsf{C}^{t+s}(\mathbb{R})$ and

$$\sup_{f\in\mathcal{F}_{s,2}} \|p_0 * f\|_{t+s,\infty} < \infty$$

is satisfied. [To be precise, the above argument establishes the inequality in the last display only for an element of the equivalence class of $p_0 * f$. This element cannot be different from $p_0 * f$ as the latter is itself a continuous function (see before (5) above).] The case of noninteger $t + s$ follows from standard generalizations of the above Fourier-analytical arguments, and we omit it. [It also follows from Lemma 12 below since $\mathcal{W}_2^s = \mathcal{B}_{22}^s \subset \mathcal{B}_{2\infty}^s$.] Now the bound

$$\sup_{f \in \mathcal{U}} \sqrt{n} \left| E(f * K_{h_n}(X) - f(X)) \right| = O(\sqrt{n} h_n^{t+s-k}) = o(1)$$

for the bias term follows from the same arguments as in Theorem 6 above, noting again that $f \in \mathcal{F}_{s,2}$ implies $\bar{f} \in \mathcal{F}_{s,2}$. □

Again, a remark similar to Remark 8 applies for the bandwidth choice $h_n \simeq n^{-1/(2t+1)}$. We also note that the above result can be generalized in a simple way to the Sobolev spaces $\mathcal{W}_p^s(\mathbb{R})$ with $1 < p < \infty$ if $p_0 \in \mathcal{W}_q^t(\mathbb{R})$ and $1/p + 1/q = 1$.

4.1.3 Extensions to higher dimensions

On the one hand, Theorems 6 and 7 could be obtained in higher dimensions (i.e., in $\mathbb{R}^d$) with only formal changes, if the smoothness index s satisfies $s > d/2$—which is necessary for these classes to be Donsker (and pregaussian)—and if the kernel is of order larger than $t + d/2$. We note here that, under the conditions of this multivariate extension of Theorem 7, Bickel and Ritov [3, p. 1036f]. remarked that one can construct a plug-in kernel estimator for the respective Sobolev classes for certain kernels K satisfying $|FK(u) - 1| \le B(1 \wedge |u|^{t+s})$ for some constant B. Theorem 7 shows that this condition is not necessary, and that one can use classical kernels without any problems to obtain even a UCLT, which implies the plug-in property, by only choosing the order of the kernel larger by $d/2$. We also note here that, when dealing with classes of functions in higher dimensions, it may be more convenient to treat the bias term by Fourier-analytical methods (as in Theorem 7, cf. also Lemma 12 in Sect. 5) rather than as in Lemma 5.

On the other hand, Theorem 5 does not easily generalize to higher dimensions. Note that, whereas the spaces $\mathsf{BV}(\mathbb{R})$ can be generalized to higher dimensions, cf. Sect. 3.3.2 above, balls in $\mathcal{BV}(\mathbb{R}^d)$ are neither Donsker nor pregaussian if $d > 1$. [This can be shown, e.g., by using results in Nickl [17].] But it is certainly of interest to prove central limit theorems for the kernel density estimator that are uniform over classes of suitable subsets of $\mathbb{R}^d$. We give some results in this direction. In analogy to Definition 2, define a kernel $K : \mathbb{R}^d \to \mathbb{R}$ of real order $r > 0$ to be a Lebesgue integrable function, symmetric around the origin, such that

$$\int_{\mathbb{R}^d} K(y)dy = 1, \int_{\mathbb{R}^d} |y|^r |K(y)| dy < \infty \text{ as well as}$$

$$\int_{\mathbb{R}} y_l^j K(y) dy_l = 0 \text{ for } l = 1, \dots, d \text{ and } j = 1, \dots, \{r\}$$

where $y \in \mathbb{R}^d$ is denoted by $(y_1, \ldots, y_d)$. If h_n is a sequence of positive real numbers converging to zero, then the sequence $\mu_n \in M(\mathbb{R}^d)$ given by $K_{h_n}(y)d\lambda(y) := h_n^{-d}K(y/h_n)d\lambda(y)$ is a proper approximate identity in the sense of Definition 1 above. The usual kernel density estimator is then $\mathbb{P}_n * K_{h_n}(x)$, and this notation is kept throughout the rest of this section.

Note that the classes in Part (a) of the following proposition include all parallelepipeds and ellipsoids of uniformly bounded diameter, (by using the fact that these classes are VC- and hence uniform Donsker classes, cf. Theorems 4.2.1 and 4.5.4 in Dudley [8]), so, in particular, Euclidean balls and boxes in $\mathbb{R}^d$. For Part (b), recall the definition of a C^α domain from before Corollary 1. We set $d > 1$, as $d = 1$ follows from Theorem 5.

Proposition 4 *Let $X_1, \ldots, X_n$ be i.i.d. according to the law $\mathbb{P}$ on $\mathbb{R}^d$, $d > 1$, where $d\mathbb{P}(x) = p_0(x)d\lambda(x)$ with density p_0. Suppose that p_0 is a bounded function—in which case we set $t = 0$ in what follows—or suppose $p_0 \in \mathsf{C}^t(\mathbb{R}^d)$ for some $t > 0$. Let K be a kernel of order $r = t + 1 - k$ for some k, $0 < k < t + 1$. Let $\mathcal{C}$ be one of the following classes:*

(a) *any translation-invariant $\mathbb{P}$-Donsker class $\mathcal{C}$ of convex sets of diameter bounded by a fixed constant, or*
(b) *all C^α-domains with $a \geq 2$, $\alpha > (d-1)/2$ and with both diameter and $H^{d-1}(\partial C)$ bounded by a fixed constant.*
If $h_n > 0$ is such that $h_n^{t+1-k}n^{1/2} \to_{n\to\infty} 0$, and if, in Part (b), p_0 is supported in $[-M, M]^d$ for some finite positive M, then

$$\sqrt{n}(\mathbb{P}_n * K_{h_n} - \mathbb{P}) \rightsquigarrow_{\ell^\infty(\mathcal{C})} \mathbb{G},$$

where $\mathbb{G}$ is the $\mathbb{P}$-Brownian bridge indexed by $\mathcal{C}$.

Proof We first show that the classes in (b) are $\mathbb{P}$-Donsker: Since the diameters of the domains $C \in \mathcal{C}$ are uniformly bounded, by the constant D say, and since p_0 is concentrated on $[-M, M]^d$, we have $\|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{C}} = \|\mathbb{P}_n - \mathbb{P}\|_{\mathcal{C}_{MD}}$ where $\mathcal{C}_{MD}$ is the class of all elements in $\mathcal{C}$ that are 'properly' contained in the cube $[-M-D, M+D]^d$, i.e., the intersection of each $C \in \mathcal{C}_{MD}$ with the complement of $[-M-D, M+D]^d$ is empty. Then, w.l.o.g., $[-M-D, M+D]^d$ is equal to the unit cube, and Theorem 8.2.15 in Dudley [8] together with Ossiander's CLT give the Donsker property of $\mathcal{C}_{MD}$, and hence of $\mathcal{C}$, for $\alpha > (d-1)/2$.

Now, since all classes considered are uniformly bounded translation-invariant $\mathbb{P}$-Donsker classes for the laws in question, we only have to verify Conditions (7) and (8) to apply Theorem 2. Corollary 1 verifies Condition (7) in both cases. Furthermore, the proof of this corollary, the remark below it and uniform boundedness of the volume of elements of $\mathcal{C}$, imply

$$\sup_{C\in\mathcal{C}}(\|\mathbf{1}_C\|_{1,\lambda} + \|\mathbf{1}_C\|_{BV}) < \infty,$$

which we will use together with some results on Besov spaces in Sect. 5 to bound (8): By Lemma 8 (and Remark 11)ii) below, the last display implies that $\{\mathbf{1}_C : C \in \mathcal{C}\}$ is a bounded subset of the Besov space $\mathcal{B}^1_{1\infty}(\mathbb{R}^d)$. Also, $p_0 \in \mathsf{C}^t(\mathbb{R}^d) \subseteq \mathcal{B}^t_{\infty\infty}(\mathbb{R}^d)$ by (55). Hence, we can apply Lemma 12 below to obtain that the set $\{\mathbf{1}_C * p_0 : C \in \mathcal{C}\}$ is bounded in $\mathcal{B}^{t+1}_{\infty\infty}(\mathbb{R}^d)$ and, by Remark 11, also in $\mathsf{C}^{t+1-\eta}(\mathbb{R}^d)$ for every $\eta > 0$, in particular for $\eta = k$. Then, as in Lemma 4, the bias term equals

$$|E(\mathbf{1}_C * \mu_{h_n}(X) - \mathbf{1}_C(X))| = \left| \int\limits_{\mathbb{R}^d} K(t)[\mathbf{1}_{\bar{C}} * p_0(h_n t) - \mathbf{1}_{\bar{C}} * p_0(0)]dt \right|,$$

where $\bar{C} = -C \in \mathcal{C}$ in all three cases. Consequently, by obvious generalizations to dimension d of the Taylor expansion arguments in Theorem 5, we conclude

$$\sup_{C\in\mathcal{C}} \sqrt{n} \left| E(\mathbf{1}_C * K_{h_n}(X) - \mathbf{1}_C(X)) \right| = O(\sqrt{n} h_n^{t+1-k}) = o(1).$$

□

We note that the bias in the above proposition can also be bounded by adapting arguments from the proof of Proposition 3.

On the one hand, the above proposition shows that—only under the assumption of a bounded density—one can construct smoothed empirical measures that satisfy the CLT uniformly over many Donsker classes of sets, and hence allows to improve on the 'naive' Theorem 4 above. Also, the bandwidth choice $n^{-1/(2t+d)}$—which gives optimal results in mean squared error—is 'almost' admissible in case $d = 2$, as then $h_n \simeq n^{-1/(2t+2)-\delta}$ for some arbitrary $\delta > 0$ implies $h_n^{t+1-k} n^{1/2} \to_{n\to\infty} 0$ for some $k > 0$. On the other hand, in dimensions higher than 2, the bandwidths allowed for in Proposition 4 have to be significantly faster than those that would be necessary to obtain the plug-in property. Already Bickel and Ritov [3] noted this problem for the case of distribution functions. The bias bound used in the proof above tries to use 'generalized' smoothness of the functions $\mathbf{1}_C$, i.e., it uses that $\mathbf{1}_C$ has first order distributional derivatives, regardless of the dimension, for a very large class of sets C, cf. also Sect. 3.3.2.

In special cases, one may be able to improve our general bound. Note that the bias for a given set C is in fact equal to

$$\begin{aligned}|E(\mathbf{1}_C * \mu_{h_n}(X) - \mathbf{1}_C(X))| &= \left| \int\limits_{\mathbb{R}^d} K(t)[\mathbf{1}_{\bar{C}} * p_0(h_n t) - \mathbf{1}_{\bar{C}} * p_0(0)]dt \right| \\ &= \left| \int\limits_{\mathbb{R}^d} K(t)[\mathbb{P}(C + h_n t) - \mathbb{P}(C)]dt \right|,\end{aligned}$$

where $C + h_n t$ denotes the translate of the set C by the vector $h_n t$. Now consider, for example, the case where $\mathbb{P} = \lambda\,|I$ is uniform on some cube $I = [a, b]^d$, and C belongs to a class $\mathcal{C}' \subseteq \mathcal{C} \cap [a+\delta, b-\delta]^d$ where $\delta > 0$ and where $\mathcal{C}$ is one of the classes from the above proposition. [For example, $\mathcal{C}'$ could consist of all balls contained in $[a+\delta, b-\delta]^d$.] If K has compact support, then both $C \in \mathcal{C}'$ and its translate $C + h_n t$ are contained in I for n large enough (independent of C) and every t in the support of K. Consequently $\lambda((C + h_n t) \cap I) = \lambda(C \cap I)$ from some n onwards by translation-invariance of Lebesgue measure, so the bias is eventually exactly zero. Then the UCLT follows directly from Theorem 2c (see also Remark 1) and Corollary 1, with the only condition that $h_n \to 0$ as $n \to \infty$. For more general classes $\mathcal{C}$ and measures $\mathbb{P}$, similar ideas might yield improvements on Proposition 4 above, but this is not the focus of the present paper.

4.2 Pregaussian classes

Having verified that Theorem 2 applies to a large variety of interesting Donsker classes, we now wish to discuss the case of pregaussian classes that are *not* Donsker, in which case one has to apply Theorem 3. In particular, it is interesting to see to which extent the general conditions imposed in that theorem work. We will be mostly interested in classes of functions, so we restrict ourselves to the case $d = 1$ for simplicity.

We first consider arbitrary pregaussian classes of functions as in Radulović and Wegkamp [21]. We will need the following well-known fact to verify Condition (11) from Theorem 3, where we recall that $\log N_{[]}(\cdot)$ denotes entropy with bracketing, see, e.g., p. 83 in van der Vaart and Wellner [30].

Lemma 6 *Define*

$$\mathcal{V}_M := \{f : \mathbb{R} \mapsto \mathbb{R} : \|f\|_{TV} \le M\}$$

and let $P(\mathbb{R})$ denote the set of all (Borel-) probability measures on $\mathbb{R}$. Then, for $0 < \varepsilon \le M$, we have

$$\sup_{Q \in P(\mathbb{R})} \log N(\mathcal{V}_M, L^2(Q), \varepsilon) \le \sup_{Q \in P(\mathbb{R})} \log N_{[]}(\mathcal{V}_M, L^2(Q), \varepsilon) \le \frac{KM}{\varepsilon}.$$

Proof The first inequality follows from the definition of bracketing numbers. The second inequality follows directly from Theorem 2.7.5 in van der Vaart and Wellner [30] together with the Hahn-Jordan decomposition of $f - f(-\infty+)$ in $\mathcal{V}_M$ into the difference of two non-decreasing functions f_P and f_N that are zero at $-\infty$ and such that $f_P(+\infty-) + f_N(+\infty-) \le M$, see Sect. 3.5 in Folland [9]. □

Using the lemma above, one can prove the following theorem, where we recall the spaces $\mathcal{W}_1^t(\mathbb{R})$ from before Lemma 3. [Also note that conditions for (7) to hold were already given in Proposition 1.]

Theorem 8 *Let Condition 1 hold with $p_0 \in \mathcal{W}_1^t(\mathbb{R})$ for some $t > 2$. Let $K \in \mathsf{BV}(\mathbb{R})$ be a kernel of order t and let $h_n > 0$ be such that $nh_n^4 \to \infty$ and $nh_n^{2t} \to 0$ as n tends to infinity. Let $\mathcal{F}$ be a translation invariant, uniformly bounded and $\mathcal{L}^1(\mathbb{R}, \lambda)$-bounded $\mathbb{P}$-pregaussian class of functions satisfying Condition (7). Then,*

$$\sqrt{n}(\mathbb{P}_n * K_{h_n} - \mathbb{P}) \rightsquigarrow_{\ell^\infty(\mathcal{F})} \mathbb{G},$$

where $\mathbb{G}$ *is the* $\mathbb{P}$*-Brownian bridge indexed by* $\mathcal{F}$.

Proof The proof consists in checking the hypotheses of Theorem 3. Condition (b) in Theorem 3 holds by assumption. Also, the bias condition (8) holds because, by Lemma 3,

$$\sqrt{n}\sup_{f\in\mathcal{F}}|E(f*K_{h_n}(X)-f(X))| = O(n^{1/2}h_n^t) = o(1)$$

is satisfied. Note further that $\mathcal{F}\subseteq\mathcal{L}_1(|\mu_n|)$ for each n because $\mathcal{F}$ is uniformly bounded. Also, by (6), $\|f*\mu_n\|_\infty \le \|f\|_\infty\|K\|_{1,\lambda}$ and we can take $M_n = M = \sup_{f\in\mathcal{F}}\|f\|_\infty\|K\|_{1,\lambda}$, which gives condition a) in Theorem 3 with M_n independent of n. Now we prove Condition (c). It is easy to see that, for all $f\in\mathcal{F}$, $\|f*\mu_n\|_{TV}\le \sup_{f\in\mathcal{F}}\|f\|_{1,\lambda}\|K\|_{TV}/h_n := D/h_n$. Therefore, we have $\tilde{\mathcal{F}}_n = \{f*\mu_n : f\in\mathcal{F}\}\subset \mathcal{V}_{D/h_n}$, and Lemma 6 gives that there exists a constant $B<\infty$ such that for every probability measure Q on $\mathbb{R}$ and $\varepsilon\le M$,

$$\log N(\tilde{\mathcal{F}}_n, L^2(Q),\varepsilon) \le \frac{B}{\varepsilon h_n} = H_n\left(\frac{M}{\varepsilon}\right)$$

where $H_n(x) = \frac{x}{MBh_n}$. Since $C_{H_n} = 2/3$ for all n, inequality (22) above then gives

$$E^*\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i f(X_i)\right\|_{(\tilde{\mathcal{F}}_n)'_{1/n^{-1/4}}} \le L\max\left[\frac{2}{3}\sqrt{\frac{2}{Bn^{1/4}h_n}}, \frac{8}{9Bn^{1/4}h_n}\right]\to 0$$

if $n^{1/4}h_n\to\infty$, which is Condition (c) in Theorem 3. Finally, we check Condition (d) in Theorem 3, for $M_n = M$. By the first remark after Theorem 3, there exists $\lambda(\varepsilon)$ such that $H(\tilde{\mathcal{F}}, L^2(\mathbb{P}),\varepsilon)\le\lambda(\varepsilon)/\varepsilon^2$, $\lambda(\varepsilon)/\varepsilon^2\to\infty$ and $\lambda(\varepsilon)\to 0$ as $\varepsilon\to 0$. In particular, $M\le 1/5\sqrt{\lambda(1/n^{1/4})}$ is eventually satisfied, thus completing the proof of condition (d). □

We see that—if no special knowledge is available about the pregaussian class—rather restrictive conditions have to be imposed on the true density p_0 in order to obtain the CLT. This is the price to be paid for the fact that closeness of the smoothed empirical measure to $\mathbb{P}_n$ cannot be used to prove the CLT. However, we will show in the next sections that one can improve on the above result for many concrete pregaussian classes of functions that are not Donsker.

Remark 9 If instead of (a) and invariance by translations in the above theorem, one imposes $Ef^j(X+y)\to Ef^j(X)$ as $y\to 0$ for $j=1,2$, (19), and

$$\int_{\mathbb{R}}(f-g)^2(x)dx \le CE((f-g)^2(X)) \tag{45}$$

for some $0<C<\infty$ and for all $f,g\in\mathcal{F}\cup\{0\}$, then the bandwidth condition $nh_n^4\to\infty$ can be weakened to $nh_n^2\to\infty$, which is in the spirit of the Radulović and Wegkamp [21] result. However, it seems that Conditions (19) and (45) can only be verified if p_0 is bounded from above and below on the support of $\mathcal{F}$ (as they do in their Theorem 2.2), so this result is restricted to compact support of p_0 and $\mathcal{F}$. Since one simultaneously needs that p_0 is at least absolutely continuous on the *whole* real line (note that Radulović and Wegkamp [21] forget to require this in the statement of their Theorem 2.2), the additional restriction that the support of $\mathcal{F}$ is a proper subset of the support of p_0 seems to be required.

4.2.1 The Besov class $\mathcal{U}^1_{1\infty}$

We now treat the class of functions $U^1_{1\infty}$ given by

$$U^1_{1\infty} = \Bigg\{ [f]_\lambda \in L^1(\mathbb{R}, \lambda) : \|f\|_{1,\lambda} + \sup_{0 \neq z \in \mathbb{R}} |z|^{-1} \left(\int_{\mathbb{R}} \Big| f(x+z) + f(x-z) - 2f(x) \Big| dx \right) \le 1 \Bigg\}. \quad (46)$$

[It can be shown that $U^1_{1\infty}$ is a bounded subset of the Besov space $B^1_{1\infty}(\mathbb{R})$ defined in Definition 3 below, see Remark 11ii. Note that any Lebesgue-integrable function of bounded variation is contained in $cU^1_{1\infty}$ for some $c > 0$, see Lemma 8 below.]

We wish to show that $\mathbb{P}_n * K_{h_n}$ *does* satisfy a UCLT over $U^1_{1\infty}$, whereas $\mathbb{P}_n$ does *not*. First note that $U^1_{1\infty}$ does not consist of functions, so we have to give a rule that selects elements out of each equivalence class $[f]_\lambda \in U^1_{1\infty}$. Choosing all functions $f \in [f]_\lambda$ with $[f]_\lambda \in U^1_{1\infty}$ would not give a fair comparison between $\mathbb{P}_n * K_{h_n}$ and $\mathbb{P}_n$: On the one hand, the absolutely continuous measure $\mathbb{P}_n * K_{h_n}$ is constant on each $[f]_\lambda$. On the other hand, the set $[f]_\lambda$ contains any modification of f at a set of Lebesgue-measure zero, so $\{f : [f]_\lambda \in U^1_{1\infty}\}$ is by far too large to be Donsker. But the following result is meaningful. We recall the shorthand notation $\langle x \rangle = (1 + |x|^2)^{1/2}$, and note that the condition

$$\|p_0 \langle x \rangle\|_\infty = \sup_{x \in \mathbb{R}} |p_0(x) \langle x \rangle| < \infty$$

in the following proposition is satisfied, e.g., by any bounded eventually monotone density p_0.

Proposition 5 *Let $\mathcal{U}$ be be any set constructed by selection of one arbitrary representative out of every $[f]_\lambda \in U^1_{1\infty}$, where $U^1_{1\infty}$ is given in (46), and assume that the law $\mathbb{P}$ possesses a Lebesgue density p_0 satisfying $\|p_0 \langle x \rangle\|_\infty < \infty$. Then $\mathcal{U}$ is $\mathbb{P}$-pregaussian but not $\mathbb{P}$-Donsker.*

Proof Follows immediately as a special case of Theorems 5 and 7 in Nickl [17], upon noting that we used a different but equivalent definition of Besov spaces, see Remark 11ii below. □

The proof of Theorem 7 in Nickl [17] in fact implies that $\mathcal{U}^1_{1\infty}$ is not even $\mathbb{P}$-Glivenko-Cantelli for the laws considered in the above theorem. Consequently, the empirical measure will be an *inconsistent* estimator for $\mathbb{P}$ in $\ell^\infty(\mathcal{U}^1_{1\infty})$. We now show that smoothed empirical measures—in particular kernel density estimators with classical bandwidths and kernels as in Definition 2—satisfy the CLT not only in $\ell^\infty(\mathcal{U})$ for any $\mathcal{U}$ as in the above proposition, but even in $\ell^\infty(\mathcal{U}^1_{1\infty})$ where $\mathcal{U}^1_{1\infty}$ consists of all functions contained in the equivalence classes from $U^1_{1\infty}$. Clearly, if any selection of representatives of $U^1_{1\infty}$ is $\mathbb{P}$-pregaussian, and if $\mathbb{P}$ is absolutely continuous, then also $\mathcal{U}^1_{1\infty}$ is $\mathbb{P}$-pregaussian, where $\mathbb{G}$ is constant on all elements of any given equivalence class.

Theorem 9 *Let Condition 1 hold and suppose that* p_0 *satisfies* $\|p_0\langle x\rangle\|_\infty < \infty$. *Set* $t = 0$ *in what follows or assume, in addition, that* $p_0 \in \mathsf{C}^t(\mathbb{R})$ *for some real* $t > 0$. *Let* $\mathcal{U}^1_{1\infty} = \{f : [f]_\lambda \in U^1_{1\infty}\}$. *Let* $K \in \mathsf{BV}(\mathbb{R})$ *be a kernel of order* $r = t + 1 - k$ *for some* k, $0 < k < t + 1$. *If* $h_n > 0$ *is such that* $h_n^{t+1-k} n^{1/2} \to_{n\to\infty} 0$ *as well as* $h_n n^\alpha \to_{n\to\infty} \infty$ *for some* $\alpha > 0$, *then*

$$\sqrt{n}(\mathbb{P}_n * K_{h_n} - \mathbb{P}) \rightsquigarrow_{\ell^\infty(\mathcal{U}^1_{1\infty})} \mathbb{G},$$

where $\mathbb{G}$ *is the* $\mathbb{P}$*-Brownian bridge indexed by* $\mathcal{U}^1_{1\infty}$.

Proof By (53) below, $\mathcal{U}^1_{1\infty}$ is, for every $\epsilon > 0$, a subset of the set $\mathcal{U}^{1-\epsilon}_{11}$, which allows us to apply Theorem 10 below with $s = 1 - \epsilon$. The bandwidth condition $h_n n^{(2s-1)/[8s(1-s)]} \to_{n\to\infty} \infty$ from Theorem 10 then simplifies to $h_n n^\alpha \to_{n\to\infty} \infty$ for some positive α by straightforward calculations. □

Next to some conditions on the kernel, the only price that one has to pay here for the fact that $\mathcal{U}^1_{1\infty}$ is not Donsker but only pregaussian (for p_0 as in the theorem) is that h_n is not allowed to decay exponentially fast. In particular, the choice $h_n \simeq n^{-1/(2t+1)}$ is admissible if the kernel is of order $r > t + 1/2$, so the theorem implies that classical kernel density estimators can possess Bickel and Ritov's [3] 'plug-in property' for pregaussian classes of functions that are not Donsker. We refer to after Theorem 10 below for more discussion.

4.2.2 $\mathcal{L}^1$-Hölder classes

For $0 < s < 1$ and $f \in \mathcal{L}^1(\mathbb{R}, \lambda)$, define the functional

$$\|f\|^*_{s,1,1} := \|f\|_{1,\lambda} + \int_{\mathbb{R}} |z|^{-s-1} \int_{\mathbb{R}} |f(x+z) - f(x)|\, dx dz. \tag{47}$$

For $s = 1$ and $f \in \mathcal{L}^1(\mathbb{R}, \lambda)$ define

$$\|f\|^*_{1,1,1} = \|f\|_{1,\lambda} + \int_{\mathbb{R}} |z|^{-2} \int_{\mathbb{R}} |f(x+z) + f(x-z) - 2f(x)|\, dx dz. \tag{48}$$

For general s, let $\{s\}$ be the largest integer strictly smaller than $s > 0$, and define for $f \in \mathcal{L}^1(\mathbb{R}, \lambda)$, the (semi)norm

$$\|f\|^*_{s,1,1} = \sum_{0 \le \alpha \le \{s\}} \left\| D^\alpha_w f \right\|_{1,\lambda} + \left\| D^{\{s\}}_w f \right\|^*_{s-\{s\},1,1}, \tag{49}$$

where D^α_w denotes the weak differential operator, see Sect. 3.3.2 above. For $s > 0$, this norm is an equivalent norm on the Besov space $B^s_{11}(\mathbb{R})$ defined in Definition 3 below, see Remark 11ii.

Let now $U^s_{11} = \{[f]_\lambda \in L^1(\mathbb{R}, \lambda) : \|f\|^*_{s,1,1} \le 1\}$. The following results were proved in Nickl and Pötscher [20] and Nickl [17], where we recall $\langle x \rangle = (1 + |x|^2)^{1/2}$ and where we note that in case $s \ge 1$, each $[f]_\lambda \in U^s_{11}$ contains a bounded continuous function.

Proposition 6 (a) *Let $s \ge 1$ and let $\mathcal{U}^s_{11} = \{f \in \mathsf{C}(\mathbb{R}) \cap \mathcal{L}^1(\mathbb{R}, \lambda) : \|f\|^*_{s,1,1} \le 1\}$. Then $\mathcal{U}^s_{11}$ is a uniform Donsker class.*

(b) *Let $1/2 < s < 1$ and assume that the law $\mathbb{P}$ has a Lebesgue density p_0 satisfying $\|p_0 \langle x \rangle\|_\infty < \infty$. Let $\mathcal{U}$ be any set constructed by selection of one arbitrary representative out of every $[f]_\lambda \in U^s_{1,1}$. Then $\mathcal{U}$ is $\mathbb{P}$-pregaussian but not $\mathbb{P}$-Donsker.*

Proof Part (a) follows immediately from Theorem 2 in Nickl and Pötscher [20] and Part (b) follows from Theorems 5 and 7 in Nickl [17] upon noting that we used a different but equivalent definition of Besov spaces, see Remark 11)ii below. □

We now apply Theorem 3 to these classes.

Theorem 10 *Let Condition 1 hold, and suppose that p_0 is a bounded function—in which case we set $t = 0$ in what follows—or assume that $p_0 \in \mathsf{C}^t(\mathbb{R})$ for some real $t > 0$. Let $\mathcal{U}^s_{11}$ be as in Proposition 6a if $s \ge 1$ and set $\mathcal{U}^s_{11} = \{f : [f]_\lambda \in U^s_{11}\}$ in case $s < 1$. Let $K \in \mathsf{BV}(\mathbb{R})$ be a kernel of order $r = t+s-k$ for some k, $0 < k < t+s$. Let $h_n > 0$ be such that $h_n^{t+s-k} n^{1/2} \to_{n\to\infty} 0$. Assume further that one of the following conditions holds*

(a) *$s \ge 1$, or*

(b) *$1/2 < s < 1$, p_0 satisfies $\|p_0 \langle x \rangle\|_\infty < \infty$, and $h_n n^{(2s-1)/8s(1-s)} \to_{n\to\infty} \infty$. Then we have*

$$\sqrt{n}(\mathbb{P}_n * K_{h_n} - \mathbb{P}) \rightsquigarrow_{\ell^\infty(\mathcal{U}^s_{11})} \mathbb{G},$$

where $\mathbb{G}$ is the $\mathbb{P}$-Brownian bridge indexed by $\mathcal{U}^s_{11}$.

Proof The proof is given in the next section. □

Remark 10 (i) Since the classes $\mathcal{U}^s_{11}$ are not $\mathbb{P}$-Donsker but only $\mathbb{P}$-pregaussian for $s < 1$, one needs additional conditions in Part (b) of the above theorem. First, one needs $\|p_0 \langle x \rangle\|_\infty < \infty$ to ensure that $\mathcal{U}^s_{11}$ is $\mathbb{P}$-pregaussian (cf. Proposition 6). More importantly, one needs the additional bandwidth condition

$$h_n n^{(2s-1)/8s(1-s)} \to_{n\to\infty} \infty \tag{50}$$

(in addition to the bias-related condition $h_n^{t+s-k} n^{1/2} \to_{n\to\infty} 0$). This additional condition prevents too fast rates of convergence of h_n to zero, since otherwise $\mathbb{P}_n * K_{h_n}$ would be too close to the empirical measure $\mathbb{P}_n$ which behaves badly in $\ell^\infty(\mathcal{F})$ for classes $\mathcal{F}$ that are not Donsker. Note that the classes $\mathcal{U}_{11}^s$ with $s < 1$ are not in any $\mathcal{L}^p(\mathbb{R}, \lambda)$ for p larger than $p = 1/(1-s)$, and, for any $x \in \mathbb{R}$, the set $\mathcal{U}_{11}^s$ contains functions with a pole at x. So it is not surprising that relatively stringent conditions have to be imposed on h_n so that $\mathbb{P}_n * K_{h_n}$ stays away from the discrete measure $\mathbb{P}_n$. In contrast, if $s = 1$, then $\mathcal{U}_{11}^1$ is uniformly bounded (and also uniformly Donsker) and one needs no additional conditions. The case $\mathcal{U}_{1\infty}^1$ considered in Theorem 9 above lies exactly between the uniform Donsker case and the classes treated in Theorem 10b. [Note that, although $\mathcal{U}_{1\infty}^1$ also contains unbounded functions, it has 'good integrability' in the sense that it is contained in every $\mathcal{L}^p(\mathbb{R}, \lambda)$, $p < \infty$.]

(ii) In the proof of Theorem 10 we shall apply Theorem 3 even under Condition (a), where $s \geq 1$. Clearly, in case $s \geq 1$, one could also apply Theorem 2 since then $\mathcal{U}_{11}^s$ are uniform Donsker classes (in view of Proposition 6). It is interesting to note that application of Theorem 2 in this case would give exactly the same results. Similarly, Theorems 5 to 7 could have been proved by using Theorem 3. In this sense our Theorem 3 is sharp.

5 Remaining proofs

In this section, we provide the proof of Theorem 10 as well as the promised generalization of Lemma 5. We will also establish some relations between Besov spaces, the space of finite signed measures, and functions of bounded variation that were used in the proof of Proposition 4 and will also be needed to prove Theorem 10.

5.1 Some definitions

We start by reviewing some facts on tempered distributions. Let $\mathcal{S}(\mathbb{R}^d)$ denote the Schwartz space of rapidly decreasing infinitely differentiable complex-valued functions and let $\mathcal{S}'(\mathbb{R}^d)$ denote the (dual) space of complex tempered distributions on $\mathbb{R}^d$. We shall restrict attention to real-valued tempered distributions T (i.e., $T = \bar{T}$, where $\bar{T}$ is defined via $\bar{T}(\phi) = \overline{T(\bar{\phi})}$ for $\phi \in \mathcal{S}(\mathbb{R})$), but we shall view real-valued distributions as elements of $\mathcal{S}'(\mathbb{R}^d)$. Let now F denote the Fourier transform acting on $\mathcal{S}(\mathbb{R}^d)$, i.e., for $\phi \in \mathcal{S}(\mathbb{R}^d)$,

$$F\varphi(u) = (2\pi)^{-d/2} \int\limits_{\mathbb{R}} e^{-ixu} \varphi(x) dx,$$

with inverse $F^{-1}\phi(u) = F\phi(-u)$. The operator F (as well as F^{-1}) is a bijection of $\mathcal{S}(\mathbb{R}^d)$ and extends—by duality—to a continuous bijection of $\mathcal{S}'(\mathbb{R}^d)$ (again denoted by F and F^{-1}), and this extension coincides with the usual Fourier–Plancherel transform when restricted to $\mathcal{L}^2(\mathbb{R}^d, \lambda)$. [See, e.g., Theorems III.4.2 and III.4.3.4 in Malliavin [15] or p. 295 in Folland [9].]

We will use the Fourier-analytical definition of Besov spaces, see Triebel [26, 2.3.1] or Chap. 6 in Bergh and Löfström [2]. There exists $\psi \in \mathcal{S}(\mathbb{R}^d)$ such that supp $\psi \subseteq \{x : 1/2 \le |x| \le 2\}$, $\psi(x) > 0$ if $2^{-1} < |x| < 2$ as well as $\sum_{k=-\infty}^{\infty} \psi(2^{-k}x) = 1$ for every $x \ne 0$, see, e.g., Lemma 6.1.7 in Bergh and Löfström [2]. Define the functions $\varphi_k = \psi(2^{-k}x)$ for $k > 0$ and $\varphi_0(x) = 1 - \sum_{k=1}^{\infty} \varphi_k(x)$. Then the functions $\{\varphi_k\}_{k=0}^{\infty}$ form a dyadic partition (resolution) of unity as in Definition 2.3.1/1 in Triebel [26]. Note that $F^{-1}(\varphi_k FT)$ is an entire analytic function on $\mathbb{R}^d$ for any $T \in \mathcal{S}'(\mathbb{R}^d)$ and any k by the Paley–Wiener–Schwartz theorem (e.g., Theorem 1.2.1/2 in Triebel [26]).

Definition 3 Let $-\infty < s < \infty$, $1 \le p \le \infty$, and $1 \le q \le \infty$. For $T \in \mathcal{S}'(\mathbb{R}^d)$ define

$$\|T\|_{s,p,q} := \left(\sum_{k=0}^{\infty} 2^{ksq} \left\| F^{-1}(\varphi_k FT) \right\|_{p,\lambda}^{q} \right)^{1/q} \tag{51}$$

(usual modification if $q = \infty$). The (real) Besov spaces are defined as

$$B^s_{pq}(\mathbb{R}^d) := \{T \in \mathcal{S}'(\mathbb{R}) : T = \bar{T},\ \ \|T\|_{s,p,q} < \infty\}.$$

$B^s_{pq}(\mathbb{R}^d)$ is a Banach space of distributions. The definition is independent of the choice of ψ (in fact, any system $\{\varphi_k\}_{k=0}^{\infty}$ as in Definition 2.3.1/1 in Triebel [26] may be used) and any ψ gives rise to an equivalent norm on $B^s_{pq}(\mathbb{R}^d)$, cf. Triebel [26, Sect. 2.3.2]. We should recall that the more classical definition of Besov spaces in terms of L_p-Hölder conditions coincides with the one just given (Triebel [26, Sect. 2.5.7]).

Remark 11 We summarize here some properties of Besov spaces, which can be found, e.g., in Sects. 2.3.2, 2.5.7, 2.5.12, 2.7 of Triebel [26].

(i) We say that a normed vector space X is continuously embedded in Y, or $X \hookrightarrow Y$, if $X \subseteq Y$ (containment possibly in the sense of a linear map, e.g., $f \to [f]_\lambda$, or $[f]_\lambda \to \tilde{f}$ with $\tilde{f} \in [f]_\lambda$) and $\|x\|_Y \le C\|x\|_X$ for all $x \in X$ and some constant C independent of x. For $1 \le p \le \infty$, the following continuous embeddings

$$B^0_{p1}(\mathbb{R}^d) \hookrightarrow L^p(\mathbb{R}^d, \lambda) \hookrightarrow B^0_{p\infty}(\mathbb{R}^d) \tag{52}$$

hold. Also

$$B^{s_1}_{pq_1}(\mathbb{R}^d) \hookrightarrow B^{s_2}_{pq_2}(\mathbb{R}^d) \tag{53}$$

for $s_1 > s_2$ and q_1, q_2 arbitrary, as well as

$$B^{s_1}_{p_1 q}(\mathbb{R}^d) \hookrightarrow B^{s_2}_{p_2 q}(\mathbb{R}^d) \tag{54}$$

for $p_1 \le p_2$ and $s_1 - 1/p_1 \ge s_2 - 1/p_2$. Finally, for $0 \le s < \infty$

$$B^s_{\infty 1}(\mathbb{R}^d) \hookrightarrow \mathsf{C}^s(\mathbb{R}^d) \hookrightarrow B^s_{\infty\infty}(\mathbb{R}^d), \tag{55}$$

where the second imbedding in the last display is an identity if s is not an integer.

(ii) From the above it follows that in case $s > 0$ or $s = 0$ and $q = 1$, the space $B^s_{pq}(\mathbb{R}^d)$ consists of elements of $L^p(\mathbb{R}^d, \lambda)$. In particular, one can define the seminormed vector spaces $\mathcal{B}^s_{pq}(\mathbb{R}^d) = \{f \in \mathcal{L}^p(\mathbb{R}^d, \lambda) : [f]_\lambda \in B^s_{pq}(\mathbb{R}^d)\}$, which coincide with $B^s_{pq}(\mathbb{R}^d)$ if one takes the usual quotient modulo equality almost everywhere. Furthermore, for $s > 0$, the seminorms $\|\cdot\|^*_{s,1,1}$ introduced in (47), (48) and (49) above are equivalent seminorms on $\mathcal{B}^s_{11}(\mathbb{R})$ (in fact, the norms characterize the Besov space in the sense that finiteness of $\|T\|^*_{s,1,1}$ implies containment in $B^s_{11}(\mathbb{R})$ for arbitrary $T \in \mathcal{S}'(\mathbb{R})$). Similarly so for the seminorm occurring in (24) in the case of $\mathcal{B}^s_{2\infty}(\mathbb{R})$, and the seminorm occurring in (46) in the case of $\mathcal{B}^1_{1\infty}(\mathbb{R})$.

5.2 Relationships of Besov spaces to $M(\mathbb{R}^d)$ and $\mathcal{BV}(\mathbb{R}^d)$

On the one hand, by (52), we know that the space of finite signed measures $M(\mathbb{R}^d)$ contains $B^0_{11}(\mathbb{R}^d)$. On the other hand, we have the following result.

Lemma 7 *Let $\mu \in M(\mathbb{R}^d)$. Then μ, interpreted as a tempered distribution, belongs to $B^0_{1\infty}(\mathbb{R}^d)$ and the inequality $\|\mu\|_{0,1,\infty} \le C\,\|\mu\|$ is satisfied for some constant C independent of μ.*

Proof Note that $B^0_{1\infty}(\mathbb{R}^d)$ is the dual space of the Banach space $\overline{B}^0_{\infty 1}(\mathbb{R}^d)$, where $\overline{B}^0_{\infty 1}(\mathbb{R}^d)$ is the completion of $\mathcal{S}(\mathbb{R}^d) \cap \{f : \mathbb{R}^d \to \mathbb{R}\}$ in the norm $\|\cdot\|_{0,\infty,1}$ (see Triebel [26], Remark 2 on page 180). Since $\|\cdot\|_\infty \le C\,\|\cdot\|_{0,\infty,1}$ holds on $\mathcal{S}(\mathbb{R}^d)$ (see expression (13) on p.131 in Triebel [26]), and since the completion of $\mathcal{S}(\mathbb{R}^d) \cap \{f : \mathbb{R}^d \to \mathbb{R}\}$ in $\|\cdot\|_\infty$ is $\mathsf{C}_0(\mathbb{R}^d)$ (elementary proof), it follows that the continuous embedding $\overline{B}^0_{\infty 1}(\mathbb{R}^d) \hookrightarrow \mathsf{C}_0(\mathbb{R}^d)$ holds, which in turn, by duality, implies the continuous embedding $M(\mathbb{R}^d) = \mathsf{C}_0(\mathbb{R}^d)' \hookrightarrow (\overline{B}^0_{\infty 1}(\mathbb{R}^d))' = B^0_{1\infty}(\mathbb{R}^d)$. □

The following lemma relates the spaces $\mathsf{BV}(\mathbb{R})$ (defined before (23)) and $\mathcal{BV}(\mathbb{R}^d)$ (defined in Sect. 3.3.2) to Besov spaces. We note in advance that a result similar to Part (a) of the following lemma could be proved in higher dimensions, but we do not need it in the present paper.

Lemma 8 (a) *If $[f]_\lambda \in B^1_{11}(\mathbb{R})$, then $[f]_\lambda$ contains a (unique) continuous function $g \in \mathsf{BV}(\mathbb{R})$ and the inequality $\|g\|_{TV} \le D\,\|f\|_{1,1,1}$ holds for some constant D independent of f.*
(b) *If $f \in \mathcal{BV}(\mathbb{R}^d) \cap \mathcal{L}^1(\mathbb{R}^d, \lambda)$, then $[f]_\lambda \in B^1_{1\infty}(\mathbb{R}^d)$ and the inequality $\|f\|_{1,1,\infty} \le D(\|f\|_{1,\lambda} + \|f\|_{BV})$ holds for some constant D independent of f.*

Proof The embedding in Part (a) is known and is proved, e.g., in the second paragraph of the proof of Theorem 2 in Nickl and Pötscher [20]. We believe the second part to be known also, but have no reference for it. To prove it, observe the following: We know that $f \in \mathcal{BV}(\mathbb{R}^d)$ implies $D^\alpha_w f \in M(\mathbb{R}^d)$ for every α with $|\alpha| = 1$, hence by (52) and Lemma 7 we have

$$\sum_{0 \le |\alpha| \le 1} \left\| D^\alpha_w f \right\|_{0,1,\infty} \le C[\|f\|_{1,\lambda} + \|f\|_{BV}]$$

for some finite constant C. But $\sum_{0\le|\alpha|\le 1}\left\|D_w^\alpha(\cdot)\right\|_{0,1,\infty}$ is an equivalent norm that characterizes the Besov space $B^1_{1\infty}(\mathbb{R}^d)$, see the proof of Theorem 2.3.8 in Triebel [26]. □

5.3 A generalization of Lemma 5

We first need some auxiliary lemmas, which will also be used in the proof of Theorem 10.

Lemma 9 *Consider T and S in $\mathcal{S}'(\mathbb{R}^d)$ such that $FT \in \mathcal{L}^p(\mathbb{R}^d,\lambda)$ and $FS \in \mathcal{L}^q(\mathbb{R}^d,\lambda)$ where $1/p+1/q \ge 1$. Then the multiplication TS is defined as an element of $\mathcal{S}'(\mathbb{R}^d)$ and*

$$F(TS) = \sqrt{2\pi}(FT * FS) \tag{56}$$

holds true (in $\mathcal{S}'(\mathbb{R}^d)$). The same holds if F is replaced by F^{-1}.

Proof This follows from Theorem 7.6 on p.123 as well as from the proposition on p.128 in Richards and Youn [22]. □

For $f \in \mathcal{S}'(\mathbb{R}^d)$, the expression $F^{-1}(\psi F f)$ always makes sense if ψ is infinitely differentiable and slowly increasing, since then $\psi F f \in \mathcal{S}'(\mathbb{R}^d)$, see p.246 in Schwartz [25], and then also $F^{-1}(\psi F f) \in \mathcal{S}'(\mathbb{R}^d)$. Let $\langle u\rangle^\alpha = (1+|u|^2)^{\alpha/2}$, which is slowly increasing for every $\alpha \in \mathbb{R}$.

Lemma 10 *Let $\alpha, r \in \mathbb{R}$ and let $1 \le p, q \le \infty$. The linear mapping $f \longmapsto F^{-1}(\langle u\rangle^\alpha\, Ff)$ is a norm-continuous isomorphism from $B^r_{pq}(\mathbb{R}^d)$ to $B^{r-\alpha}_{pq}(\mathbb{R}^d)$, and $\left\|F^{-1}(\langle u\rangle^\alpha\, F(\cdot))\right\|_{r-\alpha,p,q}$ is an equivalent norm on $B^r_{pq}(\mathbb{R}^d)$.*

Proof Theorem 2.3.8 in Triebel [26]. □

Lemma 11 *Let $f \in \mathcal{L}^p(\mathbb{R}^d,\lambda)$ and $r \in \mathbb{R}$, and let φ_k be defined as before Definition 3. Then*

$$\left\|F^{-1}(\varphi_k\,\langle u\rangle^r\, Ff)\right\|_{p,\lambda} \le C2^{rk}\left\|F^{-1}(\varphi_k Ff)\right\|_{p,\lambda}$$

for $k \ge 0$.

Proof See Lemma 6.2.1/1,3 in Bergh and Löfström [2] (with the only difference that we have applied $F^{-1}F$ to the expression in the norm of the r.h.s in their lemma, and noting that their φ_k equals our $F^{-1}\varphi_k$). □

The following lemma parallels and generalizes the 'debiasing' Lemma 5. Recall that by Lemma 4 an efficient treatment of the bias term requires bounds on the Hölder norm of $\bar{f} * p_0$, which follow from bounds on the Besov norm $\|p_0 * f\|_{s+t,\infty,\infty}$ by Remark 11i.

Lemma 12 *Let $1 \le p, q \le \infty$ with $1/p + 1/q = 1$ and let $s, t \ge 0$ be such that $s+t > 0$. Let $p_0 \in \mathcal{B}^t_{p\infty}(\mathbb{R}^d) \cap \mathcal{L}^p(\mathbb{R}^d,\lambda)$ for some $t \ge 0$, and let $\mathcal{U} \subseteq \mathcal{B}^s_{q\infty}(\mathbb{R}^d) \cap \mathcal{L}^q(\mathbb{R}^d,\lambda)$ such that $\sup_{f\in\mathcal{U}}\|f\|_{s,q,\infty} < \infty$. Then $p_0 * f \in \mathsf{C}(\mathbb{R}^d)$ for every $f \in \mathcal{U}$ and $\sup_{f\in\mathcal{U}}\|p_0 * f\|_{s+t,\infty,\infty} < \infty$.*

Proof Note that $p_0 * f \in \mathsf{C}(\mathbb{R}^d)$ by (6). We start with

$$\begin{aligned}
\|p_0 * f\|_{s+t,\infty,\infty} &\le c \left\| F^{-1}(\langle u\rangle^{s+t} Fp_0 Ff)\right\|_{0,\infty,\infty} \\
&= c \sup_{k\ge 0} \left\| F^{-1}(\varphi_k \langle u\rangle^{s+t} Fp_0 Ff)\right\|_{\infty,\lambda} \\
&\le c' \sup_{k\ge 0} \left\| F^{-1}(\varphi_k^2 \langle u\rangle^{s+t} Fp_0 Ff)\right\|_{\infty,\lambda} \\
&= c'' \sup_{k\ge 0} \left\| F^{-1}(\varphi_k \langle u\rangle^{s} Ff) * F^{-1}(\varphi_k \langle u\rangle^{t} Fp_0)\right\|_{\infty,\lambda} \\
&\le c'' \sup_{k\ge 0} \left[\left\| F^{-1}(\varphi_k \langle u\rangle^{s} Ff)\right\|_{q,\lambda} \left\| F^{-1}(\varphi_k \langle u\rangle^{t} Fp_0)\right\|_{p,\lambda}\right] \\
&\le c''' \sup_{k\ge 0} \left[2^{ks} \left\| F^{-1}(\varphi_k Ff)\right\|_{q,\lambda} 2^{kt} \left\| F^{-1}(\varphi_k Fp_0)\right\|_{p,\lambda}\right] \\
&\le c''' \|f\|_{s,q,\infty} \|p_0\|_{t,p,\infty} \qquad (57)
\end{aligned}$$

where the constant c''' does not depend on f. The first inequality follows from Lemma 10. The identity in the second line is just the definition of the seminorm (51). The inequality in the third line follows from the observation that if φ_k is replaced by φ_k^2 in (51) then one obtains an equivalent norm for $B^s_{\infty\infty}$ if $s > 0$. [This follows directly from Theorem 2.3.2, p. 172, Triebel [27] since both systems $\{\varphi_k\}$ and $\{\varphi_k^2\}$ satisfy the conditions of this theorem.] For the equality in the fourth line, we apply Lemma 9 with $T = \varphi_k \langle u\rangle^s Ff$ and $S = \varphi_k \langle u\rangle^t Fp_0$. To do this, we have to verify the conditions of the lemma by showing that $F^{-1}T \in \mathcal{L}^q(\mathbb{R}^d, \lambda)$, $F^{-1}S \in \mathcal{L}^p(\mathbb{R}^d, \lambda)$ as well as that $T, S \in \mathcal{S}'(\mathbb{R}^d)$. For T we have the following: Clearly, $T \in \mathcal{S}'(\mathbb{R}^d)$ since, $Ff \in \mathcal{S}'(\mathbb{R}^d)$ and since $\varphi_k \langle u\rangle^s \in \mathcal{S}(\mathbb{R}^d)$ (noting that $\varphi_k \langle u\rangle^s$ is infinitely differentiable and compactly supported). Then Lemma 9 and (6) imply

$$F^{-1}T = F^{-1}(\varphi_k \langle u\rangle^s Ff) = \sqrt{2\pi} F^{-1}(\varphi_k \langle u\rangle^s) * F^{-1}Ff \in \mathcal{L}^q(\mathbb{R}^d, \lambda)$$

since $f \in \mathcal{L}^q(\mathbb{R}^d, \lambda)$ and since $\varphi_k \langle u\rangle^s \in \mathcal{S}(\mathbb{R}^d)$ implies $F^{-1}\varphi_k \langle u\rangle^s \in \mathcal{S}(\mathbb{R}^d) \subseteq \mathcal{L}^1(\mathbb{R}^d, \lambda)$. The same arguments with q replaced by p also show that $S \in \mathcal{S}'(\mathbb{R}^d)$ and $F^{-1}S \in \mathcal{L}^p(\mathbb{R}^d, \lambda)$ The inequality in the fifth line follows from (5). The inequality in the sixth line follows from Lemma 11. The last inequality follows from the definition of the Besov norm. □

For instance, this lemma implies the following fact (using equivalent characterizations of Besov spaces). If p and q are conjugate, $0 < s, t \le 1$, $s + t$ non-integer, and

$$|f(x+h) - f(x)| \le g_1(x)|h|^s, \quad |p_0(x+h) - p_0(x)| \le g_2(x)|h|^t, \quad |h| \le 1, \ x \in \mathbb{R},$$

with $f,\ g_1 \in \mathcal{L}^p(\mathbb{R}, \lambda)$ and $p_0,\ g_2 \in \mathcal{L}^q(\mathbb{R}, \lambda)$, then $f * p_0 \in \mathsf{C}^{s+t}(\mathbb{R})$. If $s + t \in \mathbb{N}$, one still has the same conclusion with $s + t$ replaced by $s + t - \delta$ for any $\delta > 0$.

5.4 Proof of Theorem 10

The proof of Theorem 10 is based on verifying the conditions of Theorem 3. First, we need the following proposition.

Proposition 7 *Let* $d\mu_n(y) = h_n^{-1}K(h_n^{-1}y)dy$ *where* $K \in \mathsf{BV}(\mathbb{R})$ *is a kernel of order* $t \geq 0$ *and where* $h_n > 0$ *satisfies* $h_n \to_{n\to\infty} 0$. *Let* $0 < s < \infty$ *and let* $\mathcal{U}_{11}^s$ *be given as in Theorem 10. Define* $\mathcal{G}_n^s = \{g * \mu_n : g \in \mathcal{U}_{11}^s\}$.

(a) *Let* $0 < s < 1$. *If* $d\mathbb{P} = p_0 d\lambda$ *with* $\|p_0 \langle x\rangle\|_\infty < \infty$ *is satisfied, then we have*

$$H(\mathcal{G}_n^s, L^2(\mathbb{P}), \varepsilon) \leq D\varepsilon^{-1/s} \tag{58}$$

for some $0 < D < \infty$ *independent of* n.
Let further

$$C_n = \begin{cases} h_n^{s-1} & \text{if } s < 1 \\ const. & \text{if } s \geq 1 \end{cases}.$$

Then we have
(b) $\sup_{g\in\mathcal{U}_{11}^s} \|g * \mu_n\|_\infty \leq D'C_n$ *for some* $0 < D' < \infty$ *independent of* n *as well as*
(c)

$$\sup_{Q\in P(\mathbb{R})} H(\mathcal{G}_n^s, L^2(Q), \varepsilon) \leq \sup_{Q\in P(\mathbb{R})} H_{[]}(\mathcal{G}_n^s, L^2(Q), \varepsilon) \leq \frac{D''C_n}{\varepsilon}$$

where $P(\mathbb{R})$ *denotes the set of all p.m.'s on* $\mathbb{R}$ *and where* $0 < D'' < \infty$ *is independent of* n.

Proof Throughout the proof, we use the fact that the set $\mathcal{U}_{11}^s$ is a bounded subset of the Besov space $\mathcal{B}_{11}^s(\mathbb{R})$, see Remark 11ii.

Part (a) Expression (5) on p.127 in Triebel [26] gives the convolution inequality

$$\|f * g\|_{s,1,q} \leq c\,\|f\|_{0,1,\infty}\,\|g\|_{s,1,q}$$

for $f \in B_{1\infty}^0(\mathbb{R})$ and $g \in B_{1q}^s(\mathbb{R})$. But this immediately implies $\|\mu * g\|_{s,1,q} \leq cC\,\|\mu\|\,\|g\|_{s,1,q}$ for any $\mu \in M(\mathbb{R})$ by Lemma 7 above. So $\sup_n \sup_{g\in\mathcal{U}_{11}^s} \|\mu_n * g\|_{s,1,1} < \infty$ follows from $\sup_n \|\mu_n\| < \infty$. But bounded subsets $\mathcal{U}$ of $\mathcal{B}_{11}^s(\mathbb{R})$ with $0 < s < 1$ satisfy the entropy bound $H(\mathcal{U}, L^2(\mathbb{P}), \varepsilon) \leq K\varepsilon^{-1/s}$ for some $0 < K < \infty$, see the proof of Theorem 5 in Nickl [17].

Parts (b), (c) Recall the notation $\langle u\rangle^r$ as shorthand for $(1 + |u|^2)^{r/2}$. As the main step, we bound the quantity $\|g * \mu_n\|_{1,1,1}$. We will verify the following relations:

$$\begin{aligned}
\|g * \mu_n\|_{1,1,1} &= \sqrt{2\pi} \sum_{k=0}^{\infty} 2^k \left\| F^{-1}(\varphi_k F g F \mu_n) \right\|_{1,\lambda} \\
&= \sqrt{2\pi} \sum_{k=0}^{\infty} 2^k \left\| F^{-1}(\varphi_k F g \langle u \rangle^{s-1} \langle u \rangle^{1-s} F \mu_n) \right\|_{1,\lambda} \\
&= 2\pi \sum_{k=0}^{\infty} 2^k \left\| F^{-1}(\varphi_k F g \langle u \rangle^{s-1}) * F^{-1}(\langle u \rangle^{1-s} F \mu_n) \right\|_{1,\lambda} \\
&\leq \pi \left\| F^{-1}(\langle u \rangle^{1-s} F \mu_n) \right\|_{1,\lambda} \cdot \sum_{k=0}^{\infty} 2^k \left\| F^{-1}(\varphi_k F g \langle u \rangle^{s-1}) \right\|_{1,\lambda} \\
&:= \pi A_n \cdot B(g)
\end{aligned} \tag{59}$$

The first equality follows from Definition 3 and the second equality is trivial. The last inequality follows immediately from (6), so it remains to show that the third equality holds true: We accomplish this by using Lemma 9 above with $T = \varphi_k F g \langle u \rangle^{s-1}$, $S = \langle u \rangle^{1-s} F \mu_n$ and $p = q = 1$. To do this, we have to verify the conditions of the lemma by showing that $F^{-1}T, F^{-1}S \in \mathcal{L}^1(\mathbb{R}, \lambda)$ as well as that $T, S \in \mathcal{S}'(\mathbb{R})$. The facts that $T \in \mathcal{S}'(\mathbb{R})$ and that $F^{-1}T \in \mathcal{L}^1(\mathbb{R}, \lambda)$ follow by similar arguments as in the proof of Lemma 12 above. That $S \in \mathcal{S}'(\mathbb{R})$ follows from $F\mu_n \in \mathcal{S}'(\mathbb{R})$ and since $\langle u \rangle^{1-s}$ is slowly increasing. To show $F^{-1}S \in \mathcal{L}^1(\mathbb{R}, \lambda)$ we have

$$K_{h_n} \in \mathcal{L}^1(\mathbb{R}, \lambda) \cap \mathsf{BV}(\mathbb{R}) \subseteq \mathcal{B}^1_{1\infty}(\mathbb{R})$$

by assumption and Lemma 8. But this implies that

$$F^{-1}(\langle u \rangle^{1-s} F \mu_n) \in \mathcal{B}^s_{1\infty}(\mathbb{R}) \subseteq \mathcal{L}^1(\mathbb{R}, \lambda) \tag{60}$$

by using Lemma 10, (53) and (52) above. This completes the verification of (59).

We now treat the expressions A_n and $B(g)$ in (59). We first bound $B(g)$. Using Lemma 11 and $g \in \mathcal{L}^1(\mathbb{R}, \lambda)$ we have

$$\begin{aligned}
B(g) &= \sum_{k=0}^{\infty} 2^k \left\| F^{-1}(\varphi_k \langle k \rangle^{s-1} F g) \right\|_{1,\lambda} \leq C \sum_{k=0}^{\infty} 2^k 2^{k(s-1)} \left\| F^{-1}(\varphi_k F g) \right\|_{1,\lambda} \\
&= C \|g\|_{s,1,1} \leq C'
\end{aligned}$$

for some $0 < C' < \infty$ independent of g.

To bound A_n, consider first the case $s = 1$: Then

$$\sup_n A_n = \sup_n \left\| F^{-1}(\langle u \rangle^{1-s} F \mu_n) \right\|_{1,\lambda} = \sup_n \|\mu_n\| < \infty$$

holds, which also gives the case $s > 1$ (by Lemma 10, (53) and (52)). So let $0 < s < 1$. Then

$$A_n \leq C \left\| F^{-1}(\langle u\rangle^{1-s} F\mu_n) \right\|_{0,1,1} \leq C' \left\| K_{h_n} \right\|_{1-s,1,1} \leq C'' \left\| K_{h_n} \right\|^*_{1-s,1,1}$$

by (52), Lemma 10 and the fact the the norm $\|\cdot\|^*_{1-s,1,1}$ from (47) is an equivalent norm, see Remark 11)ii. Using the definition of $\|\cdot\|^*_{1-s,1,1}$, we now bound the last expression in the above display. Observe that

$$\begin{aligned}
&\int_{\mathbb{R}} \left(|z|^{s-2} \int_{\mathbb{R}} \left| K_{h_n}(x+z) - K_{h_n}(x) \right| dx \right) dz \\
&= h_n^{-1} \int_{\mathbb{R}} \left(|z|^{s-2} \int_{\mathbb{R}} |K((x+z)/h_n) - K(x/h_n)|\, dx \right) dz \\
&= h_n^{s-2} \int_{\mathbb{R}} \left(|t|^{s-2} \int_{\mathbb{R}} |K((x/h_n)+t) - K(x/h_n)|\, dx \right) dt \\
&= h_n^{s-1} \int_{\mathbb{R}} \left(|t|^{s-2} \int_{\mathbb{R}} |K(u+t) - K(u)|\, du \right) dt.
\end{aligned}$$

For $|t| \geq \xi > 0$, the r.h.s. is bounded by Ch_n^{s-1} since $K \in \mathcal{L}^1(\mathbb{R}, \lambda)$. For small $|t|$, since $K \in \mathsf{BV}(\mathbb{R})$, we have from (26), (27) and Fubini

$$\begin{aligned}
h_n^{s-1} \int_{|t|\leq\xi} |t|^{s-2} \int_{\mathbb{R}} |K(u+t) - K(u)|\, du dt &= h_n^{s-1} \int_{|t|\leq\xi} |t|^{s-2} \int_{\mathbb{R}} \left| \int_u^{u+t} d\nu_K \right| du dt \\
&\leq h_n^{s-1} \|K\|_{TV} \int_{|t|\leq\xi} |t|^{s-1}\, dt < \infty.
\end{aligned}$$

Since also $\sup_n \left\| K_{h_n} \right\|_{1,\lambda} = \sup_n \|K\|_{1,\lambda} < \infty$, we obtain that $A_n \leq d h_n^{s-1}$ holds for some constant d independent of n.

Summarizing this shows

$$\|g * \mu_n\|_{1,1,1} \leq \begin{cases} K' h_n^{s-1} & \text{if } s < 1 \\ const. & \text{if } s \geq 1, \end{cases} \tag{61}$$

where the constants on the r.h.s do not depend on g or n. Since $g * \mu_n$ is continuous (see before (6)), we have from Part (a) of Lemma 8 that $\|g * \mu_n\|_{TV} \leq K'' \|g * \mu_n\|_{1,1,1}$ and by (54) and (55) we also have $\|g * \mu_n\|_\infty \leq K''' \|g * \mu_n\|_{1,1,1}$ which give Part (b) of the proposition directly and Part (c) by Lemma 6. □

Proof (**Theorem 10**)**:** Since $\mathcal{U}_{11}^s$ is $\mathbb{P}$-pregaussian (by Theorem 6 and an obvious remark as before Theorem 9) and closed under translations, we can apply Theorem 3. Note that $\mathcal{U}_{11}^s \subseteq \mathcal{L}^1(|\mu_n|)$ for every n since K_{h_n} is bounded and since $\mathcal{U}_{11}^s \subseteq \mathcal{L}^1(\mathbb{R}, \lambda)$. Also

$$\|\mathbb{P}f\|_{\mathcal{U}_{11}^s} = \sup_{f \in \mathcal{U}_{11}^s} |\mathbb{P}f| \leq \sup_{f \in \mathcal{U}_{11}^s} \|f\|_{1,\lambda} \|p_0\|_\infty < \infty.$$

We now proceed to verify Conditions (a) to (d) as well as the bias condition (8) from Theorem 3 in order to prove Theorem 10.

Condition (a) By Part (b) of Proposition 7, we have that M_n can be chosen such that $M_n = D'C_n$.

Condition (b) is satisfied by Proposition 1c, since $\mathcal{U}_{11}^s$ is bounded in $\mathcal{B}_{2\infty}^{s-1/2}(\mathbb{R})$ (by (54) and (53)), where $s - 1/2 > 0$.

Condition (c) Recall $\mathcal{G}_n^s := \{f * K_{h_n} : f \in \mathcal{U}_{11}^s\}$. Using the inequality (21) we have

$$\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i f(X_i) \right\|_{(\mathcal{G}_n^s)'_{1/n^{1/4}}} \leq L \int_0^{n^{-1/4}} \sqrt{1 \vee \log N_{[]}(\mathcal{G}_n^s, L^2(\mathbb{P}), \varepsilon)}\, d\varepsilon$$
$$+ L\sqrt{n} M_n I(M_n > \sqrt{n} a_n) := I + II \qquad (62)$$

for some constant $0 < L < \infty$. We show that both I and II tend to zero under the conditions of the theorem. For I we have

$$\int_0^{n^{-1/4}} \sqrt{1 \vee \log N_{[]}(\mathcal{G}_n^s, L^2(\mathbb{P}), \varepsilon)}\, d\varepsilon \leq \int_0^{n^{-1/4}} \sqrt{1 \vee \varepsilon^{-1} D'' C_n}\, d\varepsilon \leq L' C_n^{1/2} n^{-1/8}$$

where L' is a fixed constant and where C_n is given in Part (c) of Proposition 7 above. If $s \geq 1$ then $C_n^{1/2}$ is constant and the integral above converges to zero. For $s < 1$ the integral converges to zero if the condition

$$h_n^{s-1} n^{-1/4} \to_{n\to\infty} 0 \qquad (63)$$

holds. Since $h_n^{s-1} n^{(1-2s)/8s} \to_{n\to\infty} 0$ by the assumptions of the theorem, (63) follows. We now bound II by showing that the indicator $I(M_n > \sqrt{n} a_n)$ equals zero eventually. Here recall from (21) that

$$a_n = n^{-1/4} / \sqrt{1 + 2 \log N_{[]}(\mathcal{G}_n^s, L^2(\mathbb{P}), 2^{-1} n^{-1/4})}$$
$$\geq n^{-1/4} / \sqrt{1 + D'' C_n n^{-1/4}}$$

where C_n is given in Part (c) of Proposition 7. If $s \geq 1$, both C_n and the envelope $M_n = M$ are constant and the result follows immediately since $\sqrt{n} a_n \simeq n^{1/4}(1 +$

$n^{-1/4})^{-1/2} \to_{n\to\infty} \infty$ so $I(M > \sqrt{n}a_n)$ equals zero eventually. If $s < 1$, the condition $M_n \le \sqrt{n}a_n$ is implied by

$$D' h_n^{s-1} \le n^{1/4}/\sqrt{1 + D'' C_n n^{-1/4}},$$

which, by definition of C_n, is equivalent to

$$D' n^{-1/4} h_n^{s-1} \sqrt{1 + D'' h_n^{s-1} n^{-1/4}} \le 1. \tag{64}$$

But this inequality is implied by (63) above. This completes verification of convergence to zero in (62).

Condition (d) Note that this condition is automatically satisfied in case $s \ge 1$ since then $\mathcal{U}_{11}^s$ is uniform Donsker and hence uniform pregaussian by Proposition 6 as well as uniformly bounded (by (54) and (55)), so $M_n = const.$ and $\lambda(\varepsilon)$ exists by Sudakov's inequality (e.g., Ledoux and Talagrand [14, p. 81]). If $0 < s < 1$, we have by Part (a) of Proposition 7 that $H(\varepsilon, \mathcal{G}_n^s, \|\cdot\|_{2,\mathbb{P}}) \le D\varepsilon^{-1/s}$ holds. Hence we can choose $\lambda(\varepsilon)$ in (d) equal to

$$\lambda(\varepsilon) = D\varepsilon^{2-(1/s)} = D\varepsilon^{(2s-1)/s} \to_{\varepsilon\to 0} 0,$$

since $s > 1/2$, and obviously $\lambda(\varepsilon)\varepsilon^{-2} \to_{\varepsilon\to 0} \infty$. Consequently, Condition (12) in Part (d) becomes

$$h_n^{s-1} \le c/\sqrt{\lambda(n^{-1/4})} = c' n^{(2s-1)/8s}, \tag{65}$$

for some constants $0 < c, c' < \infty$. Since $h_n^{s-1} n^{(1-2s)/8s} \to_{n\to\infty} 0$ by assumption of the theorem, this inequality is satisfied for n large enough.

Bias Condition (8). We have $p_0 \in \mathsf{C}^t(\mathbb{R}) \subseteq \mathcal{B}_{\infty\infty}^t(\mathbb{R})$ (see (55)). Also, $\mathcal{U}_{11}^s$ is a bounded subset of $\mathcal{B}_{11}^s(\mathbb{R}) \subset \mathcal{B}_{1\infty}^s(\mathbb{R})$, cf. Remark 11 above. Consequently we conclude from Lemma 12 (with $p = \infty$ and $q = 1$) that $\sup_{f\in\mathcal{U}_{11}^s} \|p_0 * f\|_{s+t-\delta,\infty} < \infty$ holds for every $\delta > 0$, so in particular for $\delta = k$. Now we apply Lemma 4 and the fact that $f \in \mathcal{U}_{11}^s$ implies $\bar{f} \in \mathcal{U}_{11}^s$ to obtain from the same Taylor series arguments as in the proof of Theorem 6 above that

$$\sup_{f\in\mathcal{U}_{11}^s} \sqrt{n}\left|E(f * K_{h_n}(X) - f(X))\right| = O(\sqrt{n}h_n^{s+t-k}) = o(1),$$

which completes the proof. □

Acknowledgements The authors would like to thank V. Losert, W. R. Madych and H. Triebel for helpful comments on subjects related to this article.

References

1. Alexander, K.S.: Central limit theorems for stochastic processes under random entropy conditions. Probab. Theory Relat. Fields **75**, 351–378 (1987)
2. Bergh, J., Löfström, J.: Interpolation Spaces. Springer, Berlin (1976)

3. Bickel, J.P., Ritov, Y.: Nonparametric estimators which can be ‘plugged-in’. Ann. Statist. **31**, 1033–1053 (2003)
4. de la Peña, V., Giné, E.: Decoupling. From Dependence to Independence. Springer, New York (1999)
5. Devroye, L., Györfi, L.: Nonparametric Density Estimation. The L_1 View. Wiley, New York (1985)
6. Dudley, R.M.: Sample functions of the Gaussian process. Ann. Probab. **1**, 66–103 (1973)
7. Dudley, R.M.: An extended Wichura theorem, definition of Donsker class, and weighted empirical distributions. Lecture Notes in Mathematics, vol. 1153, pp. 141–178. Springer, New York (1985)
8. Dudley, R.M.: Uniform Central Limit Theorems. Cambridge University Press, Cambridge, England (1999)
9. Folland, G.B.: Real Analysis. Wiley, New York (1999)
10. Giné, E., Koltchinskii, V.: Concentration inequalities and asymptotic results for ratio type empirical processes. Ann. Probab. **34**, 1143–1216 (2006)
11. Giné, E., Zinn, J.: Some limit theorems for empirical processes. Ann. Probab. **12**, 929–989 (1984)
12. Giné, E., Zinn, J.: Empirical processes indexed by Lipschitz functions. Ann. Probab. **14**, 1329–1338 (1986)
13. Giné, E., Zinn, J.: Gaussian characterization of uniform Donsker classes. Ann. Probab. **19**, 758–782 (1991)
14. Ledoux, M., Talagrand, M.: Probability in Banach Spaces. Springer, Berlin (1991)
15. Malliavin, P.: Integration and Probability. Springer, New York (1995)
16. Marcus, D.J.: Relationships between Donsker classes and Sobolev spaces. Z. Wahrsch. Verw. Gebiete **69**, 323–330 (1985)
17. Nickl, R.: Empirical and Gaussian processes on Besov classes. In: Giné, E., Koltchinskii, V., Li, W., Zinn, J. (eds.): High dimensional probability IV, IMS Lecture Notes, vol. 51, pp. 185–195 (2006a)
18. Nickl, R.: Uniform central limit theorems for density estimators. preprint (2006b)
19. Nickl, R.: Donsker-type theorems for nonparametric maximum likelihood estimators. Probab. Theory Relat. Fields **138**, 411–449 (2007)
20. Nickl, R., Pötscher, B.M.: Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov- and Sobolev-type. J. Theoret. Probab. **20**, 177–199 (2007)
21. Radulović, D., Wegkamp, M.: Weak convergence of smoothed empirical processes. Beyond Donsker classes. In: Giné, E., Mason, D.M., Wellner, J.A. (eds.) High dimensional probability II, Progr. Probab. **47**, pp. 89–105. Birkhäuser, Boston (2000)
22. Richards, J.I., Youn, H.K.: Theory of Distributions. Cambridge University Press, Cambridge England (1990)
23. Rost, D.: Limit theorems for smoothed empirical processes. In: Giné, E., Mason, D.M., Wellner, J.A. (eds.) High Dimensional Probability II, Progr. Probab. **47**, pp. 107–113. Birkhäuser, Boston (2000)
24. Schaefer, H.H.: Topological Vector Spaces. Springer, New York (1971)
25. Schwartz, L.: Théorie des Distributions. Hermann, Paris (1966)
26. Triebel, H.: Theory of Function Spaces. Birkhäuser, Basel (1983)
27. Triebel, H.: Interpolation Theory, Function Spaces, Differential Operators. Barth, Heidelberg (1995)
28. van der Vaart, A.W.: Weak convergence of smoothed empirical processes. Scand. J. Statist. **21**, 501–504 (1994)
29. van der Vaart, A.W.: Asymptotic Statistics. Cambridge University Press, Cambridge (1998)
30. van der Vaart, A.W., Wellner, J.A.: Weak Convergence and Empirical Processes. Springer, New York (1996)
31. Yukich, J.E.: Weak convergence of smoothed empirical processes. Scand. J. Statist. **19**, 271–279 (1992)
32. Ziegler, K.: Functional central limit theorems for triangular arrays of function-indexed processes under uniformly integrable entropy conditions. J. Multivar. Anal. **62**, 233–272 (1997)
33. Ziemer, W.P.: Weakly Differentiable Functions. Springer, New York (1989)