J Theor Probab (2009) 22: 38–56
DOI 10.1007/s10959-008-0177-3

On Convergence and Convolutions of Random Signed Measures

Richard Nickl

Received: 20 September 2006 / Revised: 5 February 2008 / Published online: 1 July 2008

Abstract Let μ_n be a sequence of random finite signed measures on the locally compact group G equal to either $\mathbb{T}^d$ or $\mathbb{R}^d$. We give weak conditions on the sequence μ_n and on functions K such that the convolution product $\mu_n * K$, and its derivatives, converge in law, in probability, or almost surely in the Banach spaces $\mathsf{C}_0(G)$ or $L^p(G)$. Examples for sequences μ_n covered are the empirical process (possibly arising from dependent data) and also random signed measures $\sqrt{n}(\tilde{\mathbb{P}}_n - \mathbb{P})$ where $\tilde{\mathbb{P}}_n$ is some (nonparametric) estimator for the measure $\mathbb{P}$, including the usual kernel and wavelet based density estimators with MISE-optimal bandwidths. As a statistical application, we apply the results to study convolutions of density estimators.

Keywords Convolutions · Limit theorems · Banach space · Density estimator

Mathematics Subject Classification (2000) Primary 60F17 · 28A33 · Secondary 60B15 · 46E27

1 Introduction

Consider two independent random variables Y and Z that take values in the locally compact group G equal to the d-dimensional Torus $\mathbb{T}^d$ or Euclidean space $\mathbb{R}^d$. The law of $Y + Z$ is equal to the convolution $\mathbb{P}^Y * \mathbb{P}^Z$ of the laws of Y and Z. Recently, in [17, 18] as well as [7], results of the following type were obtained: If the densities of Y and Z exist and satisfy mild smoothness assumptions and if $\tilde{\mathbb{P}}_n^Y$ and $\tilde{\mathbb{P}}_n^Z$ are the random measures obtained from (nonparametric) kernel density estimators for the densities of Y and Z, then the (centered) convolution products $\tilde{\mathbb{P}}_n^Y * \tilde{\mathbb{P}}_n^Z - \mathbb{P}^Y * \mathbb{P}^Z$

R. Nickl (✉)
Department of Mathematics, University of Connecticut, 196 Auditorium Road—U 3009, Storrs, CT 06269, USA
e-mail: nickl@math.uconn.edu

were shown to converge in law at rate $\sqrt{n}$ in the space $\mathsf{C}_0(\mathbb{R}^d)$ of bounded continuous functions on $\mathbb{R}^d$ that vanish at infinity; and also in $L^p(\mathbb{R}^d)$-spaces.

Motivated by these results, the aim of the present paper is to shed some light on the general question of convergence behavior of convolutions of density estimators. From a mathematical point of view, this question can be phrased in general terms of convergence properties of convolution products of sequences of random finite signed measures. As is well known, regularity (e.g., smoothness or integrability) properties of a function K are typically inherited by the convolution product of this function with another integrable function or finite measure. For example, the mapping $\mu \mapsto \mu * K$ from the space $M(G)$ of finite signed measures on G into $\mathsf{C}_0(G)$ (or $L^p(G)$) is continuous for arbitrary $K \in \mathsf{C}_0(G)$ (or $L^p(G)$) if both spaces are equipped with the respective norm topologies. Consequently, limiting properties of a (possibly random) sequence μ_n of finite signed measures in the total variation norm carry over to limiting properties of convolution products $\mu_n * K$ in $\mathsf{C}_0(G)$ (or $L^p(G)$). Hence a first attempt to reproduce the results by the abovementioned authors might be to set $\mu_n = \sqrt{n}(\tilde{\mathbb{P}}_n^Y - \mathbb{P}^Y)$ and K equal to the Lebesgue density of $\mathbb{P}^Z$. (We consider here for the moment the even simpler case where $\mathbb{P}^Z$ is known.) The continuous mapping argument just given would then work if we could establish that $\sqrt{n}(\tilde{\mathbb{P}}_n^Y - \mathbb{P}^Y)$ is uniformly tight in (the norm topology of) $M(G)$. But this is impossible if the postulated statistical model for $\mathbb{P}^Y$ is genuinely infinite-dimensional: In this case, the total variation norm of $\sqrt{n}(\tilde{\mathbb{P}}_n^Y - \mathbb{P}^Y)$ will become unbounded with positive probability as n increases, since it is well known that the best rate of convergence in the total variation norm of a nonparametric estimator for $\mathbb{P}^Y$ is, in general, strictly slower than $1/\sqrt{n}$. We refer to Chap. 15 in [2] for a detailed discussion of this fact.

A second attempt might be to use the fact that convolution with an arbitrary function K in $\mathsf{C}_0(G)$ (or $L^p(G)$) is still a continuous mapping from $M(G)$ into $\mathsf{C}_0(G)$ (or $L^p(G)$) if $M(G)$ is equipped with its *weak-star* topology, cf., e.g., Theorem III.1.9 in [12]. But the proof of this result uses the fact that a (nonrandom) weak-star convergent sequence μ_n in $M(G)$ stays uniformly bounded in the total variation norm by the Banach–Steinhaus theorem. As mentioned above, $\sqrt{n}(\tilde{\mathbb{P}}_n^Y - \mathbb{P}^Y)$ becomes unbounded with positive probability in the total variation norm as n increases, so this approach does not work either.

Both attempts fail because the chosen topologies on $M(G)$ are too strong in the sense that $\sqrt{n}(\tilde{\mathbb{P}}_n^Y - \mathbb{P}^Y)$ is not uniformly tight in these topologies. Consequently, one could try to choose an even weaker topology on $M(G)$ (so that $\sqrt{n}(\tilde{\mathbb{P}}_n^Y - \mathbb{P}^Y)$ is uniformly tight in this topology) and investigate whether the operation of convolution with certain functions still possesses suitable continuity properties on $M(G)$ with respect to this topology. This can be done by imbedding $M(G)$ into certain spaces of Schwartz distributions, which is the approach taken in the present paper, see Sects. 2 and 3: We first study convergence of random signed measures in a general class of Banach spaces of distributions and we establish a simple connection between norm-convergence in these spaces to limiting results for random measure processes (such as density estimators or the empirical process) that are available in the literature. We then use the fact that some of these distribution spaces have simple Fourier-analytical characterizations to show that convolution with a function that possesses some regularity properties is still a continuous mapping into $\mathsf{C}_0(G)$ and

$L^p(G)$. The last observation has already been used in the mathematical physics and interacting particle system literature (e.g., [16]), but it seems that this was not widely appreciated by probabilists and statisticians.

We then discuss in Sect. 4 how our general results can be applied to the statistical problems considered by Frees [5], Schick and Wefelmeyer [17, 18], and Giné and Mason [7]. Our general results rely on the techniques developed in Sects. 2 and 3 and on recent results on *uniform central limit theorems* for density estimators, see [13, 14] and [8, 9]. Next to reproducing the results Schick, Wefelmeyer, Giné and Mason in the kernel density estimator case, our results also apply to other density estimators, in particular, to wavelet, maximum likelihood, and trigonometric series estimators, possibly in a setting with dependent data and adaptive choice of bandwidth or tuning parameters. In particular for estimators that are nonlinear in the empirical measure, our results can still be used, in contrast to the U-process approach used by Giné and Mason [7].

2 Convergence of Random Signed Measures

2.1 Some Function and Distribution Spaces on G

We denote by $\mathcal{B}_S$ the Borel-σ-algebra of a (nonempty) topological space S. To unify the presentation, we shall deal with the locally compact group G equal either to the d-dimensional Torus $\mathbb{T}^d$, Euclidean space $\mathbb{R}^d$, or the integer vectors $\mathbb{Z}^d$. We define the following function spaces on $G \in \{\mathbb{T}^d, \mathbb{R}^d, \mathbb{Z}^d\}$: The symbol $\mathsf{C}(G)$ denotes the Banach space of bounded real-valued continuous functions on G normed by the usual sup-norm $\|\cdot\|_\infty$. The symbol $\mathsf{C}_0(G)$ denotes the closed subspace of $\mathsf{C}(G)$ consisting of bounded continuous real-valued functions f such that, for every $\varepsilon > 0$, there exists a compact set $K \subseteq G$ such that $f(x) < \varepsilon$ holds for $x \notin K$. (Clearly, $\mathsf{C}_0(\mathbb{T}^d) = \mathsf{C}(\mathbb{T}^d)$, and $\mathsf{C}_0(\mathbb{R}^d)$ is the subspace of $\mathsf{C}(\mathbb{R}^d)$ consisting of functions that vanish at infinity.) We denote by $\mathcal{L}^0(G)$ the set of real-valued $\mathcal{B}_G$-measurable functions on G. For G equal to $\mathbb{T}^d$ or $\mathbb{R}^d$, the symbol λ will always denote the (product-)Lebesgue measure on G, and for $G = \mathbb{Z}^d$, the symbol λ stands for the (product-)counting measure. For $h \in \mathcal{L}^0(G)$, we set $\|h\|_{p,\lambda} := (\int_G |h|^p \, d\lambda)^{1/p}$ for $1 \le p \le \infty$ (where $\|h\|_{\infty,\lambda}$ denotes the λ-essential supremum of $|h|$). We denote by $\mathcal{L}^p(G)$ the vector space of all $h \in \mathcal{L}^0(G)$ that satisfy $\|h\|_{p,\lambda} < \infty$. In accordance, $L^p(G)$ denotes the corresponding Banach space of equivalence classes $[h]_\lambda$ (modulo equality λ-almost everywhere), $h \in \mathcal{L}^p(G)$. We shall also use obvious analogues of these spaces and norms for complex-valued functions.

Let $\mathcal{D}(\mathbb{T}^d)$ denote the space of complex-valued infinitely differentiable functions defined on $\mathbb{T}^d$, and let $\mathcal{S}(\mathbb{R}^d)$ denote the space of rapidly decreasing infinitely differentiable complex-valued functions on $\mathbb{R}^d$. We equip both spaces with their usual locally convex topology and obtain the topological dual spaces $\mathcal{D}'(\mathbb{T}^d)$—the space of (complex) Schwartz-distributions on $\mathbb{T}^d$—as well as $\mathcal{S}'(\mathbb{R}^d)$, the space of (complex) tempered distributions on $\mathbb{R}^d$. We refer to [20] for details. We will, with some abuse of notation, set $\mathcal{D}(G)$ and $\mathcal{D}'(G)$ equal to $\mathcal{D}(\mathbb{T}^d)$ and $\mathcal{D}'(\mathbb{T}^d)$, respectively, for $G = \mathbb{T}^d$ and equal to $\mathcal{S}(\mathbb{R}^d)$ and $\mathcal{S}'(\mathbb{R}^d)$, respectively, for $G = \mathbb{R}^d$.

We denote by F the (distributional) Fourier transform acting on $\mathcal{D}'(G)$ (see, e.g., p. 225 and p. 249 in [20]), whose restriction to $L^1(G)$ coincides with the classical Fourier transform and whose restriction to $L^2(G)$ coincides with the Fourier–Plancherel transform. For a (tempered) distribution T on G, its complex conjugate $\bar{T}$ is defined via $\bar{T}(\phi) = \overline{T(\bar{\phi})}$ for $\phi \in \mathcal{D}(G)$. Let $\langle k \rangle^s = (1 + |k|^2)^{s/2}$, where k is an element of the dual group $\hat{G} \in \{\mathbb{Z}^d, \mathbb{R}^d\}$ of $G \in \{\mathbb{T}^d, \mathbb{R}^d\}$ and where $|\cdot|$ is the Euclidean norm. We define, for $s \in \mathbb{R}$ and $G \in \{\mathbb{T}^d, \mathbb{R}^d\}$, the Sobolev(–Hilbert) spaces

$$W_2^s(G) = \left\{T \in \mathcal{D}'(G) : T = \bar{T},\ \|T\|_{s,2,\lambda} := \left\| \langle k \rangle^s FT \right\|_{2,\lambda} < \infty \right\}. \tag{1}$$

If $s \geq 0$, every distribution in $W_2^s(G)$ can be identified with an element of $L^2(G)$. In particular, $W_2^0(G) = L^2(G)$ holds with isometric norms by Plancherel's theorem. Furthermore, for integer $m > 0$, an equivalent norm on $W_2^m(G)$ is given by $\|f\| = \sum_{0 \leq |\alpha| \leq m} \|D^\alpha f\|_{2,\lambda}$, where $D^\alpha = \frac{\partial^{|\alpha|}}{(\partial x_1)^{\alpha_1} \cdots (\partial x_d)^{\alpha_d}}$ denotes the partial differential operator of order $|\alpha|$ in the sense of distributions. (Here, as usual, $\alpha = (\alpha_1, \ldots, \alpha_d)$ is a multi-index of nonnegative integers α_i, and $|\alpha| = \sum_{i=1}^d \alpha_i$.) Note that $[Ff]_\lambda \in L^1(\hat{G})$ follows from the definition of $W_2^s(G)$ if $s > d/2$, and hence, by Fourier inversion, $[F^{-1}Ff]_\lambda = [f]_\lambda$ contains a unique element of $\mathsf{C}_0(G)$. For such s, we can define the Hilbert spaces

$$\mathsf{W}_2^s(G) = \left\{ f \in \mathsf{C}_0(G) : [f]_\lambda \in L^2(G, \lambda),\ \|f\|_{s,2,\lambda} < \infty \right\} \tag{2}$$

consisting of bounded continuous functions.

We will also need Sobolev spaces with $p = 1$: For $G \in \{\mathbb{T}^d, \mathbb{R}^d\}$ and integer $m \geq 0$, we denote by $W_1^m(G)$ the space of equivalence classes of functions $[f]_\lambda \in L^1(G)$ such that $\sum_{0 \leq |\alpha| \leq m} \|D^\alpha f\|_{1,\lambda} < \infty$ holds. For $0 < s < 1$, we define

$$W_1^s(G) = \left\{ [f]_\lambda \in L^1(G) : \int_G \left\| f(\cdot - y) \right\|_{1,\lambda} |y|^{-(d+s)}\, dy < \infty \right\},$$

the space of equivalence classes of functions in $L^1(G)$ that are L^1-Hölder of order s. For noninteger $s > 1$, let m be the largest integer smaller than s. We then define $W_1^s(G)$ to be the space of $[f]_\lambda \in L^1(G)$ whose distributional partial derivatives up to order m are all contained in $L^1(G)$ and $[D^m f]_\lambda$ is contained in $W^{s-m}(G)$.

Finally, on $\mathbb{R}^d$, we will need certain Besov spaces. Let φ_0 be a complex-valued C^∞-function on $\mathbb{R}^d$ with $\varphi_0(k) = 1$ if $|k| \leq 1$ and $\varphi_0(k) = 0$ if $|k| \geq 3/2$. Define $\varphi_1(k) = \varphi_0(k/2) - \varphi_0(k)$ and $\varphi_u(k) = \varphi_1(2^{-u+1}k)$ for $u \in \mathbb{N}$. Then the system φ_u forms a (locally finite) dyadic partition of unity. Let now $-\infty < s < \infty$, $1 \leq p \leq \infty$, and $1 \leq q \leq \infty$. For $T \in \mathcal{S}'(\mathbb{R}^d)$, define the Besov norms

$$\|T\|_{s,p,q,\lambda} := \left(\sum_{u=0}^{\infty} 2^{usq} \left\| F^{-1}(\varphi_u FT) \right\|_{p,\lambda}^q \right)^{1/q}$$

with modification

$$\|T\|_{s,p,\infty,\lambda} := \sup_{0 \leq u < \infty} 2^{us} \left\| F^{-1}(\varphi_u FT) \right\|_{p,\lambda}.$$

We define the Banach spaces

$$B^s_{pq}(\mathbb{R}^d) = \{T \in \mathcal{S}'(\mathbb{R}^d) : T = \bar{T},\ \|T\|_{s,p,q,\lambda} < \infty\}. \tag{3}$$

The Besov-norms are independent of the choice of φ_0, and different φ_0 result in equivalent norms, cf. Sect. 2.3.2 in [21]. For $s > 0$, the Besov (or Hölder–Zygmund) space $\mathring{\mathsf{B}}^s_{\infty\infty}(\mathbb{R}^d)$ is obtained by taking the closure of $\mathcal{S}(\mathbb{R}^d) \cap \{f : f \text{ real-valued}\}$ in the norm-topology of $\|\cdot\|_{s,\infty,\infty,\lambda}$.

2.2 Random Measures in Distribution Spaces

Denote by $M(G)$ the Banach space of (real) finite signed (Borel) measures on $G \in \{\mathbb{T}^d, \mathbb{R}^d\}$ normed by the total variation norm $\|\cdot\|_{TV}$. Let $(Z, \mathcal{Z}, \zeta)$ be a probability space, and let $\mu_n : (Z, \mathcal{Z}, \zeta) \to M(G)$ be a sequence of random finite signed measures on G. Throughout the paper, we do *not* require any measurability of the mapping μ_n. Convergence of random elements (in law, in (outer) probability, or almost surely) in a metric space is defined in the usual way, cf. Sects. 3.1–3.3 in [4].

For $\mathcal{F}$ a nonempty uniformly bounded class of Borel-measurable functions $f : G \to \mathbb{R}$, denote by $\ell^\infty(\mathcal{F})$ the Banach space of real-valued bounded functions on $\mathcal{F}$ equipped with the supnorm $\|\cdot\|_{\infty,\mathcal{F}}$. For $f \in \mathcal{F}$ and $\mu \in M(G)$, we set $\mu f := \int_G f\, d\mu$. As usual, μ_n gives rise to a sequence of random elements in $\ell^\infty(\mathcal{F})$. An alternative to imbedding $M(G)$ into $\ell^\infty(\mathcal{F})$ is to view elements of $M(G)$ as distributions, that is, as elements of $\mathcal{D}'(G)$. To obtain sharp results, we will focus on certain *subspaces* of $\mathcal{D}'(G)$ that still contain $M(G)$. A topological vector space X is imbedded into another topological vector space Y if X is a linear subspace of Y and if the identity map is continuous. We write $X \hookrightarrow Y$ for such an imbedding.

Definition 1 Let G be either $\mathbb{R}^d$ or $\mathbb{T}^d$. We say that a Banach space $\mathsf{X}(G)$ of real-valued Borel-measurable functions on G (with norm $\|\cdot\|_X$) satisfies property D if the following conditions are satisfied:

1. $\mathcal{D}(G) \cap \{f : f \text{ real-valued}\}$ is dense in $\mathsf{X}(G)$, and $\mathcal{D}(G) \hookrightarrow \mathsf{X}(G)$ holds.
2. $\|\varphi\|_\infty \le C\|\varphi\|_X$ holds for some $0 < C < \infty$ and all $\varphi \in \mathcal{D}(G)$.

Clearly, every element of the dual space $(\mathsf{X}(G))'$ of a space $\mathsf{X}(G)$ satisfying property D can be uniquely identified with a real-valued (tempered) distribution, that is, $(\mathsf{X}(G))'$ is contained in $\mathcal{D}'(G)$. Equipped with the operator norm $\|\cdot\|'_X$, the space $(\mathsf{X}(G))'$ becomes a Banach space. Also, property D implies that $\mathsf{X}(G)$ is dense and continuously imbedded into $\mathsf{C}_0(G)$ since the closure of $\mathcal{D}(G)$ w.r.t. $\|\cdot\|_\infty$ is $\mathsf{C}_0(G)$. Consequently, since $M(G)$ is isometrically isomorphic to $\mathsf{C}_0(G)'$ (see, e.g., Theorem 6.6 in [12]), we have

$$M(G) = \mathsf{C}_0(G)' \hookrightarrow (\mathsf{X}(G))'. \tag{4}$$

That is, every dual space of a Banach space $\mathsf{X}(G)$ *that satisfies property D is a space of distributions that contains all finite signed measures*. Furthermore, convergence of a sequence of (possibly random) signed measures μ_n in $\ell^\infty(\mathcal{U})$ for $\mathcal{U}$ the unit ball $\mathsf{X}(G)$ can be linked to convergence in $(\mathsf{X}(G))'$ in a simple way:

Lemma 1 *Let G be either $\mathbb{R}^d$ or $\mathbb{T}^d$, and let $\mathsf{X}(G)$ satisfy property D. Let $(Z, \mathcal{Z}, \zeta)$ be a probability space, and let $\mu_n : (Z, \mathcal{Z}, \zeta) \to M(G)$ be a sequence of random signed measures. Let $\mathcal{U}$ be the unit ball of $\mathsf{X}(G)$. We then have*

$$\|\mu\|_{\infty,\mathcal{U}} = \|\mu\|'_X$$

for every $\mu \in M(G)$. In particular, the sequence μ_n converges in law, in probability, or almost surely in the Banach space $\ell^\infty(\mathcal{U})$ if and only if μ_n converges in the respective sense in the Banach space $(\mathsf{X}(G))'$.

Proof Denote by $\mathrm{lin}^\infty(\mathcal{U})$ the space of all real-valued functionals L defined on $\mathcal{U}$ which satisfy

$$L\big(\alpha f + (1-\alpha)g\big) = \alpha L(f) + (1-\alpha)L(g)$$

for all $f, g \in \mathcal{U}$ and $0 \le \alpha \le 1$; $L(0) = 0$ as well as $\sup_{f\in\mathcal{U}} |L(f)| < \infty$. Then $\mu_n \in \mathrm{lin}^\infty(\mathcal{U})$ for every n. It is clear that $\mathrm{lin}^\infty(\mathcal{U})$ (with the induced norm) is a closed subspace of $\ell^\infty(\mathcal{U})$. Consequently, μ_n converges in law in $l^\infty(\mathcal{U})$ if and only if it converges in law in $\mathrm{lin}^\infty(\mathcal{U})$.

Let now L be any element of $\mathrm{lin}^\infty(\mathcal{U})$. Then it is easy to see that there exists a unique linear functional $T_L : \mathsf{X}(G) \to \mathbb{R}$ satisfying $T_L(f) = L(f)$ for every $f \in \mathcal{U}$ (this follows, e.g., from Lemma 2.5.3 in [4]). Also, every $T_L \in (\mathsf{X}(G))'$ gives rise to an element of $\mathrm{lin}^\infty(\mathcal{U})$. Finally, $\|T_L\|'_X = \sup_{f\in\mathcal{U}} |L(f)|$ holds, and hence the linear mapping

$$T : \mathrm{lin}^\infty(\mathcal{U}) \to \big(\mathsf{X}(G)\big)'$$

associating to each $L \in \mathrm{lin}^\infty(\mathcal{U})$ the respective continuous linear functional $T_L \in (\mathsf{X}(G))'$ is an isometric isomorphism. □

The following lemma gives examples of spaces satisfying property D that will be used in this paper.

Lemma 2 *The spaces $\mathsf{W}_2^s(G)$ with $s > d/2$ and $\mathring{\mathsf{B}}^s_{\infty\infty}(\mathbb{R}^d)$ with $s > 0$ satisfy property D.*

Proof It is easy to see that the space $\mathsf{W}_2^s(G)$ contains $\mathcal{D}(G)$ as a dense subset, and this imbedding is continuous. (For a reference, see, e.g., 3.5.1/6 in [19] and Theorem 2.3.3 in [21].) Furthermore, $\mathsf{W}_2^s(G)$ is continuously imbedded into $\mathsf{C}_0(G)$, see the reasoning before (2) above. In the second case, note that $\mathcal{S}(\mathbb{R}^d)$ is dense in $\mathring{\mathsf{B}}^s_{\infty\infty}(\mathbb{R}^d)$ by definition. Furthermore, $\mathcal{S}(\mathbb{R}^d) \hookrightarrow \mathring{\mathsf{B}}^s_{\infty\infty}(\mathbb{R}^d)$ holds since convergence of elements of $\mathcal{S}(\mathbb{R}^d)$ in the norm $\|\cdot\|_{s,\infty,\infty,\lambda}$ is implied by convergence in the norm $\sum_{0\le|\alpha|\le r} \|D^\alpha(\cdot)\|_\infty$ for some integer $r > s$ (cf. 2.5.7/6, 11 in [21]), which is in turn implied by convergence in $\mathcal{S}(\mathbb{R}^d)$. Furthermore, $\|\varphi\|_\infty \le C\|\varphi\|_{s,\infty,\infty,\lambda}$ holds for $\varphi \in \mathcal{S}(\mathbb{R}^d)$ by Triebel ([21], 2.7.1/12, 13). □

The following observation will be a key to the main results of the paper. Together with Lemmas 1 and 2, it implies that sup-norm on the space $\ell^\infty(\mathcal{F})$ for $\mathcal{F}$

the unit balls of $\mathsf{W}_2^s(G)$ and $\mathring{\mathsf{B}}^s_{\infty\infty}(\mathbb{R}^d)$ is equivalent to a norm that has a much simpler Fourier-analytical structure. Recall that two norms $\|\cdot\|_{X,1}$, $\|\cdot\|_{X,2}$ on a vector space X are equivalent if $\|\cdot\|_{X,2} \le c\|\cdot\|_{X,1} \le C\|\cdot\|_{X,2}$ holds on X for some $0 < c \le C < \infty$.

Lemma 3 *Let $s > 0$.*

1. $(\mathsf{W}_2^s(G))' = W_2^{-s}(G)$, *and the norms $\|\cdot\|'_{s,2,\lambda}$ and $\|\cdot\|_{-s,2,\lambda}$ are equivalent.*
2. $(\mathring{\mathsf{B}}^s_{\infty\infty}(\mathbb{R}^d))' = B_{11}^{-s}(\mathbb{R}^d)$, *and the norms $\|\cdot\|'_{s,\infty,\infty,\lambda}$ and $\|\cdot\|_{-s,1,1,\lambda}$ are equivalent.*

Proof Part 1 is a simple duality argument in weighted L^2-type spaces, proved, e.g., in Theorem 2.11.2/2 of [21] and Theorem 3.5.6 in [19]. (Note that the Triebel spaces $F^s_{22}(G)$ coincide with $W_2^s(G)$ (with equivalent norms).) For Part 2, cf. Remark 2.11.2/2 in [21]. □

Before we proceed to the main results of the paper, we discuss a number of important examples of random signed measures that converge in the distribution spaces $W_2^{-s}(G)$ and $B_{11}^{-s}(\mathbb{R}^d)$.

2.2.1 Empirical Processes

Let $\{X_j\}_{j=1}^n$ be i.i.d. according to the (Borel) law $\mathbb{P}$ on the locally compact group $G \in \{\mathbb{T}^d, \mathbb{R}^d\}$. Define the empirical measure $\mathbb{P}_n = n^{-1}\sum_{j=1}^n \delta_{X_j}$. The *empirical process* is the sequence of random finite signed measures $\sqrt{n}(\mathbb{P}_n - \mathbb{P})$.

Proposition 1 *Let $s > d/2$, and let $\mathcal{U}_2^s$ and $\mathcal{U}_\infty^s$ denote the unit balls of $\mathsf{W}_2^s(G)$ and $\mathring{\mathsf{B}}^s_{\infty\infty}(\mathbb{R}^d)$, respectively.*

1. $\sqrt{n}(\mathbb{P}_n - \mathbb{P})$ *converges in law in $\ell^\infty(\mathcal{U}_2^s)$ and in $W_2^{-s}(G)$.*
2. *If $\int_{\mathbb{R}^d} |x|^\gamma \, d\mathbb{P}(x) < \infty$ for some $\gamma > d$, then $\sqrt{n}(\mathbb{P}_n - \mathbb{P})$ converges in law in $\ell^\infty(\mathcal{U}_\infty^s)$ and in $B_{11}^{-s}(\mathbb{R}^d)$.*

Proof In view of Lemmata 1, 2, and 3 above, it is sufficient to prove that $\mathcal{U}_2^s$ and $\mathcal{U}_\infty^s$ are $\mathbb{P}$-Donsker classes (cf. Chap. 3 in [4]). The universal Donsker property of $\mathcal{U}_2^s$ is proved in [6] for $G = \mathbb{T}^d$ and follows from Proposition 1 in [15] in the case $G = \mathbb{R}^d$. (Both cases also follow from the CLT in Hilbert spaces, Theorem 10.5 in [11].) The $\mathbb{P}$-Donsker property of $\mathcal{U}_\infty^s$ follows from Corollary 5 in [15]. □

Similarly, one can obtain strong invariance principles or laws of the iterated logarithm for the empirical process in $\ell^\infty(\mathcal{U}_2^s)$ or $\ell^\infty(\mathcal{U}_\infty^s)$ by using Proposition 1 and, e.g., the results in Sects. 9.4 and 9.5 in [4].

If the $\{X_j\}_{j=1}^n$ are not i.i.d. but form a strictly stationary β-mixing sequence of random variables with marginal probability distribution $\mathbb{P}$, it is known that the central limit theorem still holds in $\ell^\infty(\mathcal{F})$ under summability conditions on the mixing coefficients and under bracketing metric entropy conditions on the class $\mathcal{F}$, see Theorem 1 in [3]. Bracketing metric entropy bounds for $\mathcal{U}_2^s$ and $\mathcal{U}_\infty^s$ the unit balls of $\mathsf{W}_2^s(G)$

and $\mathring{\mathsf{B}}^s_{\infty\infty}(\mathbb{R}^d)$ are derived in Theorem 1 and Corollary 4 in [15]. Consequently, a result similar to Proposition 1 can be proved for $n^{1/2}(\mathbb{P}_n - \mathbb{P})$ under certain mixing conditions on the process $\{X_j\}_{j=1}^n$, cf., in particular, p. 405 in [3].

2.2.2 Density Estimators

If $\tilde{p}_n \in \mathcal{L}^1(G)$ is some estimator for some probability density p_0 on G, then $\mu_n = \sqrt{n}(\tilde{p}_n - p_0)\,d\lambda$ is a sequence of random finite signed measures on G. Convergence results for μ_n in $\ell^\infty(\mathcal{U}_2^s)$ and $\ell^\infty(\mathcal{U}_\infty^s)$ have been recently established in the literature for kernel, wavelet, trigonometric series, and maximum likelihood estimators, see [13, 14] and [8, 9]. For instance, if $\{X_j\}_{j=1}^n$ are i.i.d. with law $\mathbb{P}$ on $\mathbb{R}$, and if $K : \mathbb{R} \to \mathbb{R}$ is a kernel of order r (e.g., Definition 2 in [9]), then the usual Rosenblatt–Parzen kernel estimator is the convolution

$$\mathbb{P}_n * K_{h_n}(y) = \frac{1}{nh_n}\sum_{j=1}^n K\left(\frac{y - X_j}{h_n}\right),$$

and one can prove the following:

Proposition 2 *Let $s > 1/2$, and let $\mathcal{U}_2^s$ and $\mathcal{U}_\infty^s$ denote the unit balls of $\mathsf{W}_2^s(\mathbb{R})$ and $\mathring{\mathsf{B}}^s_{\infty\infty}(\mathbb{R})$, respectively. Let K be a kernel of order r, and suppose that the density p_0 of $\mathbb{P}$ exists.*

1. *If $[p_0]_\lambda \in W_2^t(\mathbb{R})$ for some $t \geq 0$, if $r \in (0, t+s]$, and if $\sqrt{n}h_n^r \to 0$, then $\sqrt{n}(\mathbb{P}_n * K_{h_n} - \mathbb{P})$ converges in law in $\ell^\infty(\mathcal{U}_2^s)$ and in $W_2^{-s}(\mathbb{R})$.*
2. *If $[p_0]_\lambda \in W_1^t(\mathbb{R})$ for some $t \geq 0$, if $r \in (0, t+s)$, if $\sqrt{n}h_n^r \to 0$, and if $\int_\mathbb{R} |x|^\gamma\, d\mathbb{P}(x) < \infty$ is satisfied for some $\gamma > 1$, then $\sqrt{n}(\mathbb{P}_n * K_{h_n} - \mathbb{P})$ converges in law in $\ell^\infty(\mathcal{U}_\infty^s)$ and in $B_{11}^{-s}(\mathbb{R})$.*

Proof The first part is Theorem 7 in [9]. The second part follows from Theorem 6 in [9], upon observing (cf., e.g., Remark 11 in the last reference) that bounded subsets of $\mathring{\mathsf{B}}^s_{\infty\infty}(\mathbb{R})$ are, for every $\delta > 0$, bounded in the Hölder space $\mathsf{C}^{s-\delta}(\mathbb{R})$ considered there (whence $r = s + t$ has to be excluded in the second part of the proposition). □

Note that the MISE-optimal choice $h_n \simeq n^{-1/(2t+1)}$ is admissible if one chooses $r > t + 1/2$, which is always possible in the above proposition.

A similar result holds for wavelet density estimators. For details about wavelets in statistics we refer to [10]. If ϕ is a compactly supported father wavelet with associated projection kernel $K_j(x, y) = \sum_{k\in\mathbb{Z}} 2^j\phi(2^j x - k)\phi(2^j y - k)$ and if $\hat{\alpha}_{jk} = \int_\mathbb{R} 2^{j/2}\phi(2^j x - k)\, d\mathbb{P}_n(x)$, then the linear wavelet density estimator can be written as

$$\mathbb{P}_n\big(K_{j_n}(\cdot, y)\big) = \sum_k 2^{j_n/2}\hat{\alpha}_{j_n k}\phi\big(2^{j_n}y - k\big),$$

and we have, with $\mathbb{P}_n(K_j)$ denoting the finite signed measure induced by $\mathbb{P}_n(K_j(\cdot, y))\, d\lambda(y)$:

Proposition 3 *Let $s > 1/2$, and let $\mathcal{U}_2^s$ and $\mathcal{U}_\infty^s$ denote the unit balls of* $\mathsf{W}_2^s(\mathbb{R})$ *and* $\mathring{\mathsf{B}}_{\infty\infty}^s(\mathbb{R})$, *respectively. Let ϕ be a compactly supported father wavelet that has $r \geq 1$ bounded derivatives, $s < r + 1$.*

1. *If $[p_0]_\lambda \in W_2^t(\mathbb{R})$ for some $0 \leq t < r+1$ and if $\sqrt{n}2^{-j_n(t+s)} \to 0$, then $\sqrt{n}(\mathbb{P}_n(K_{j_n}) - \mathbb{P})$ converges in law in $\ell^\infty(\mathcal{U}_2^s)$ and in $W_2^{-s}(\mathbb{R})$.*
2. *If $[p_0]_\lambda \in W_1^t(\mathbb{R})$ for some $0 \leq t < r+1$, if $\sqrt{n}2^{-j_n(t+s)} \to 0$, and if $\int_{\mathbb{R}} |x|^\gamma \, d\mathbb{P}(x) < \infty$ is satisfied for some $\gamma > 1$, then $\sqrt{n}(\mathbb{P}_n(K_{j_n}) - \mathbb{P})$ converges in law in $\ell^\infty(\mathcal{U}_\infty^s)$ and in $B_{11}^{-s}(\mathbb{R})$.*

Proof Both cases follow from Theorem 7 in [8]. (In the first case, observe that $W_2^s(\mathbb{R})$ equals the Besov space $B_{22}^s(\mathbb{R})$ for all s.) □

Again, the MISE-optimal choice $2^{j_n} \simeq n^{1/(2t+1)}$ is always admissible in the above proposition (for sufficiently smooth ϕ).

If $G = \mathbb{T}$, such results also hold for nonparametric maximum likelihood and trigonometric series estimators, see [13, 14]. The essential advantage of $\mathbb{P}_n * K_h$, $\mathbb{P}_n(K_{j_n})$ (and other density estimators) over $\mathbb{P}_n$—which is related to the fact that $\mathbb{P}_n$ is inconsistent in stronger loss functions (such as L^p-loss)—will be used in Sect. 4.

3 Convolutions of Random Signed Measures

We recall the definition of the convolution product of functions and measures on $G \in \{\mathbb{T}^d, \mathbb{R}^d\}$. For two real-valued Borel measurable functions h and g on G, we set

$$(h * g)(x) = \int_G h(x-y)g(y)\, d\lambda(y),$$

provided that the integral exists and is finite for λ-a.e. $x \in G$. For a Borel measurable function $h : G \to \mathbb{R}$ and a finite signed measure μ on G, we set

$$(h * \mu)(x) = \int_G h(x-y)\, d\mu(y),$$

provided that the integral exists and is finite for λ-a.e. $x \in G$. For two finite signed measures μ and ν on G, their convolution defines another finite signed measure $\mu * \nu$ by

$$(\mu * \nu)(f) = \int_G \int_G f(x+y)\, d\mu(x)\, d\nu(y), \qquad f \in \mathsf{C}_0(G).$$

For $h, g \in \mathcal{L}^1(G)$, all definitions coincide by setting $d\mu = h\, d\lambda$ and $d\nu = g\, d\lambda$. It is easily seen that convolution is commutative and associative. Furthermore, for $1 \leq p \leq \infty$, $\mu \in M(G)$, and $K \in \mathcal{L}^p(G)$, we have the inequality

$$\|\mu * K\|_{p,\lambda} \leq \|\mu\|_{TV} \|K\|_{p,\lambda}, \tag{5}$$

which also holds if K is replaced by another element of $M(G)$ and if $\|\cdot\|_{p,\lambda}$ is replaced by $\|\cdot\|_{TV}$. We refer to Chap. 3 in [12] for these well-known facts.

Let $\mathcal{U}_2^s$ be the unit ball of $\mathsf{W}_2^s(G)$. The following theorem shows that a sequence of random signed measures μ_n that converges in $\ell^\infty(\mathcal{U}_2^s)$ (or in $W_2^{-s}(G)$) will also converge in $\mathsf{C}_0(G)$ and in $L^p(G)$-spaces when convolved with a (nonrandom) function K that possesses some regularity properties. Results for convolutions with *random* functions are discussed in Sect. 4.1 below.

Theorem 1 *Let G be either $\mathbb{R}^d$ or $\mathbb{T}^d$. Let $(Z, \mathcal{Z}, \zeta)$ be a probability space, and let $\mu_n : (Z, \mathcal{Z}, \zeta) \to M(G)$ be a random sequence of signed measures. Let $\mathcal{U}_2^s$ be the unit ball of $\mathsf{W}_2^s(G)$ with $s > d/2$. Suppose that μ_n converges in the space $\ell^\infty(\mathcal{U}_2^s)$ in law, in probability, or almost surely.*

1. *If $K \in \mathsf{W}_2^s(G)$, then $\mu_n * K$ converges in the space $\mathsf{C}_0(G)$ in law, in probability, or almost surely, respectively.*
2. *If $[K]_\lambda \in W_1^s(G)$, then $[\mu_n * K]_\lambda$ converges in the space $L^2(G)$, in law, in probability, or almost surely, respectively.*

Proof We recall here some basic properties of the (distributional) Fourier transform that are used in all subsequent proofs. For $G \in \{\mathbb{T}^d, \mathbb{R}^d, \mathbb{Z}^d\}$, the operator F (and its inverse F^{-1}) acts as a linear and continuous, injective mapping from $M(G)$ (or $L^1(G)$) into $\mathsf{C}(\hat{G})$ (or $\mathsf{C}_0(\hat{G})$). We consistently denote elements of G by x and elements of $\hat{G}$ by k. Furthermore F is a bijection of both $\mathcal{S}(\mathbb{R}^d)$ and $\mathcal{S}'(\mathbb{R}^d)$, and $T = F^{-1}FT$ holds for every $T \in \mathcal{D}'(G)$. As is well known, convolution is transformed into pointwise multiplication under the Fourier transform, that is, $F(\mu * \nu) = (2\pi)^{d/2} F\mu F\nu$ holds, with respective modifications if μ and/or ν are replaced by suitable functions. We refer, e.g., to Chap. 3 in [12] and Chap. 7 in [20] for these (and further) results from Fourier analysis of the groups $G \in \{\mathbb{T}^d, \mathbb{R}^d\}$.

We now prove Part 1: The following preliminary observation is necessary: For $\mu \in M(G)$, note that FK and $F\mu FK \in \mathcal{L}^1(\hat{G})$ by definition of $\mathsf{W}_2^s(G)$ and since $F\mu \in \mathsf{C}(\hat{G})$. Consequently $\mu * K = F^{-1}F(\mu * K) \in \mathsf{C}_0(G)$ holds, and, in particular, $F^{-1}F(\mu_n * K)$ is a random element in $\mathsf{C}_0(G)$ for every n.

We can view $M(G)$ as a linear subspace of the Banach space $W_2^{-s}(G)$ by (4) and Lemmata 2 and 3 above. We show that the mapping

$$\kappa : \mu \to \mu * K$$

is a continuous linear map between the normed spaces $(M(G), \|\cdot\|_{-s,2,\lambda})$ and $\mathsf{C}_0(G)$. To see this, observe that

$$\|\mu * K\|_\infty = \left\|F^{-1}F(\mu * K)\right\|_\infty \le (2\pi)^{d/2}\left\|F\mu\langle x\rangle^{-s}\langle x\rangle^s FK\right\|_{1,\lambda} \le C\|\mu\|_{-s,2,\lambda}$$

holds for μ in $M(G)$ and $C = (2\pi)^{d/2}\|K\|_{s,2,\lambda}$. Since $M(G)$ contains the dense subset $\mathcal{D}(G)$ of $W_2^{-s}(G)$ (cf. Theorem 2.3.3 in [21] and 3.5.1/6 in [19]), there exists a unique continuous linear mapping

$$\bar{\kappa} : W_2^{-s}(G) \to \mathsf{C}_0(G)$$

such that $\bar{\kappa} = \kappa$ on $M(G)$. We conclude that, since the random elements μ_n, when viewed as (tempered) distributions, converge in law, in probability, or almost surely in the Banach space $W_2^{-s}(G)$ by Lemmas 1, 2, and 3 above, the random variables

$$\bar{\kappa}(\mu_n) = \kappa(\mu_n) = \mu_n * K$$

converge in law, in probability, or almost surely in $\mathsf{C}_0(G)$ by the continuous mapping theorem.

We now prove Part 2: Note that $\mu * K \in L^2(G) \cap L^1(G)$ by (5), since $K \in W_1^s(G) \subseteq L^1(G) \cap L^2(G)$ holds for $s > d/2$ (see 2.7.1 in [21]). As in the proof of Part 1 above, we show that

$$\kappa : \mu \to \mu * K$$

is a continuous linear map between the normed spaces $(M(G), \|\cdot\|_{-s,2,\lambda})$ and $L^2(G)$. By Lemma 4 below, Plancherel's theorem, and Hölder's inequality we have

$$\begin{aligned}\|\mu * K\|_{2,\lambda} &= (2\pi)^{d/2}\|F\mu F K\|_{2,\lambda} \\ &\le (2\pi)^{d/2}\big\|F\mu\langle k\rangle^{-s}\big\|_{2,\lambda}\big\|\langle k\rangle^{s} F K\big\|_{\infty} \\ &\le C\|\mu\|_{-s,2,\lambda}\end{aligned}$$

for $\mu \in M(G)$ and $C < \infty$. Given the following lemma, the rest of the proof is identical to the one given for Part 1 above upon replacing $\mathsf{C}_0(G)$ by $L^2(G)$.

Lemma 4 *Let G be either $\mathbb{R}^d$ or $\mathbb{T}^d$, and let $0 \le s < \infty$. If $f \in W_1^s(G)$, then $\|\langle k\rangle^s F f\|_\infty < \infty$.*

Proof We only proof the case of $\mathbb{R}^d$ with $d = 1$; the case of $\mathbb{T}$ is the same, and the cases $d > 1$ follow from an obvious generalization of the argument. Recall the basic properties of the Fourier transform mentioned at the beginning of the proof of Theorem 1 above. We shall use repeatedly the inequality $\langle k\rangle^s \le 2^{s-1}(1 + |k|^s)$ for real $s > 0$. For $m \in \mathbb{N} \cup \{0\}$, we have that $[f]_\lambda \in W_1^m(G)$ implies $[D^\alpha f]_\lambda \in L^1(\mathbb{R})$ for $0 \le \alpha \le m$ by the definition of $W_1^m(G)$, and hence

$$\big\|\langle k\rangle^m F f\big\|_\infty \le 2^{m-1}\big(\|Ff\|_\infty + \big\||k|^m F f\big\|_\infty\big) \le 2^{m-1}\big(\|f\|_1 + \big\|D^m f\big\|_1\big) < \infty$$

holds by using the well-known relationship $|F(D^m f)| \le |k|^m |Ff|$. We next prove the case $0 < s < 1$: For $[f]_\lambda \in W_1^s(\mathbb{R})$, define the function

$$h(x) = \int_{\mathbb{R}} \big(f(x-y) - f(x)\big)|y|^{-(1+s)}\,dy,$$

which is contained in $\mathcal{L}^1(\mathbb{R})$ by the definition of $W_1^s(\mathbb{R})$ and Fubini's theorem. By applying the Fourier transform and Fubini's theorem, we obtain

$$(Fh)(k) = \int_{\mathbb{R}} \big(e^{-iyk} - 1\big) F f(k)|y|^{-(1+s)}\,dy.$$

It can be shown by elementary calculus that

$$\begin{aligned}\big|(Fh)(k)\big| &= \big|Ff(k)\big| \int_{\mathbb{R}} |y|^{-(1+s)} \sqrt{2-2\cos(yk)}\, dy \\ &= \big|Ff(k)k^s\big| \int_{\mathbb{R}} |z|^{-(1+s)} \sqrt{2-2\cos(z)}\, dz \\ &= C\big|Ff(k)k^s\big|\end{aligned}$$

holds for some $0 < C < \infty$. Consequently

$$\big\| Ff|k|^s \big\|_\infty = (1/C)\|Fh\|_\infty \le (1/C)\|h\|_{1,\lambda} < \infty.$$

Since also $f \in \mathcal{L}^1(\mathbb{R})$ holds, we have that

$$\big\| \langle k\rangle^s Ff \big\|_\infty \le 2^{s-1}\big(\|Ff\|_\infty + \big\||k|^s Ff\big\|_\infty\big) < \infty,$$

which proves this case. For $s > 1$ noninteger, one applies the last argument to the mth derivative, where m is the largest integer smaller than s. □

□

In the second part of the theorem, it is allowed to convolve with discontinuous and/or unbounded functions. For example, one may convolve with the indicator function of some interval in G or with the Gamma density on $\mathbb{R}$ with shape parameter $1/2 < \alpha < 1$.

Note that convergence in $\mathsf{C}(\mathbb{T}^d)$ implies convergence in $L^p(\mathbb{T}^d)$ for every $1 \le p \le \infty$. In the case $G = \mathbb{R}^d$, there are no inclusion relationships for the spaces $L^p(\mathbb{R}^d)$, but given the results for $L^2(\mathbb{R}^d)$ and $\mathsf{C}_0(\mathbb{R}^d)$ in the theorem above, one can use interpolation properties of these spaces to obtain results for $L^p(\mathbb{R}^d)$ with $p \in [2, \infty]$; and, by the subsequent theorem, also for $p \in [1, 2]$.

Convergence in $L^1(\mathbb{R}^d)$ is not covered by Theorem 1. Here, things are somewhat different. In particular, convergence of μ_n in the space $\ell^\infty(\mathcal{U}_2^s)$ is not the appropriate requirement. Instead of convergence in $\ell^\infty(\mathcal{U}_2^s)$, we have to require convergence in $B_{11}^{-s}(\mathbb{R}^d)$, or, equivalently, in $\ell^\infty(\mathcal{U}_\infty^s)$, where $\mathcal{U}_\infty^s$ is the unit ball of $\mathring{\mathsf{B}}_{\infty\infty}^s(\mathbb{R}^d)$. As a consequence, also the convolution kernel is required to lie in the Besov space $B_{11}^s(\mathbb{R}^d)$. We note that $W_1^s(\mathbb{R}^d) = B_{11}^s(\mathbb{R}^d)$ holds for noninteger s (e.g., [21], 2.5.7/1), so this condition is similar to the one in Part 2 of Theorem 1. (In particular, for $G = \mathbb{T}^d$, a result similar to Theorem 2 (with $s > d/2$) can be directly deduced from Part 2 of Theorem 1 by the continuous imbedding $L^2(\mathbb{T}^d) \hookrightarrow L^1(\mathbb{T}^d)$.)

Theorem 2 *Let $(Z, \mathcal{Z}, \zeta)$ be a probability space, and let $\mu_n : (Z, \mathcal{Z}, \zeta) \to M(\mathbb{R}^d)$ be a sequence of random signed measures. Let $\mathcal{U}_\infty^s$ be the unit ball of $\mathring{\mathsf{B}}_{\infty\infty}^s(\mathbb{R}^d)$ with $s > 0$. Suppose that μ_n converges in the space $\ell^\infty(\mathcal{U}_\infty^s)$ in law, in probability, or almost surely. If $[K]_\lambda \in B_{11}^s(\mathbb{R}^d)$, then $[\mu_n * K]_\lambda$ converges in the space $L^1(\mathbb{R}^d)$ in law, in probability, or almost surely, respectively.*

Proof Throughout the proof, we shall use the basic properties of the Fourier transform mentioned at the beginning of the proof of Theorem 1 above.

The following three preliminary observations are necessary: First, let φ_u be a dyadic resolution of unity as defined before (3), and set $h(k) = \varphi_u(k)\langle k\rangle^{-s} F\mu(k)$, where $\mu \in M(\mathbb{R}^d)$ and $u \in \mathbb{N}$ are arbitrary. We show that $F^{-1}h \in \mathcal{L}^1(\mathbb{R}^d)$. Note that $\varphi_u\langle k\rangle^{-s} \in \mathcal{S}(\mathbb{R}^d)$ holds (since φ_u has compact support and since both φ_u and $\langle k\rangle^{-s}$ are infinitely differentiable). Now

$$F^{-1}h = (2\pi)^{-d/2}\big(F^{-1}\varphi_u\langle k\rangle^{-s}\big) * F^{-1}F\mu$$

holds by using Théorème 15 on p. 268 in [20]. Using (5), this proves $F^{-1}h \in \mathcal{L}^1(\mathbb{R}^d)$ upon noting that $F^{-1}\varphi_u\langle k\rangle^{-s} \in \mathcal{S}(\mathbb{R}^d) \subseteq \mathcal{L}^1(\mathbb{R}^d)$ and that $F^{-1}F\mu = \mu \in M(\mathbb{R}^d)$.

Second, $[F^{-1}\langle k\rangle^s FK]_\lambda \in L^1(\mathbb{R}^d)$ holds since $[K]_\lambda \in B^s_{11}(\mathbb{R}^d)$ and since

$$[f]_\lambda \to \big[F^{-1}\langle k\rangle^s Ff\big]_\lambda$$

is a norm-continuous mapping from $B^s_{11}(\mathbb{R}^d)$ to $B^0_{11}(\mathbb{R}^d)$ by Theorem 2.3.8 in [21] and since $B^0_{11}(\mathbb{R}^d) \subseteq L^1(\mathbb{R}^d)$ holds by Triebel ([21], 2.5.7/1).

Third, consider two functions h and g on $\mathbb{R}^d$ such that Fh and Fg are in $\mathcal{L}^1(\mathbb{R}^d)$. Then $hg \in \mathcal{S}'(\mathbb{R}^d)$, and we have $F(hg) = (2\pi)^{d/2} Fh * Fg$ by, e.g., Lemma 9 in [9], and an analogous result can be proved if F^{-1} and F are interchanged.

We now proceed to prove the theorem. We view $M(\mathbb{R}^d)$ as a linear subspace of $B^{-s}_{11}(\mathbb{R}^d)$ by Lemmata 2 and 3 and by (4) above. We show that the mapping

$$\kappa : \mu \to \mu * K$$

is a continuous linear map between the normed spaces $(M(\mathbb{R}^d), \|\cdot\|_{-s,1,1,\lambda})$ and $L^1(\mathbb{R}^d)$. Note first that $M(\mathbb{R}^d)$ is contained in the Besov space $B^0_{1\infty}(\mathbb{R}^d)$, see, e.g., Lemma 7 in [9], and that $K \in B^s_{11}(\mathbb{R}^d) \subset B^0_{11}(\mathbb{R}^d)$. Consequently the continuous imbedding $B^0_{11}(\mathbb{R}^d) * B^0_{1\infty}(\mathbb{R}^d) \hookrightarrow B^0_{11}(\mathbb{R}^d)$ (2.6.5/5 in [21]) implies $\mu * K \in B^0_{11}(\mathbb{R}^d)$. Recalling $B^0_{11}(\mathbb{R}^d) \hookrightarrow L^1(\mathbb{R}^d)$ and using the three observations from above (the third observation with $h = \varphi_u\langle k\rangle^{-s}F\mu$ and $g = \langle k\rangle^s FK$), (5) as well as Lemma 11 in [9], we obtain

$$\begin{aligned}
\|\mu * K\|_{1,\lambda} &\le c\|\mu * K\|_{0,1,1,\lambda} \\
&= c' \sum_{u=0}^{\infty} \big\|F^{-1}\big(\varphi_u F\mu\langle k\rangle^{-s}\langle k\rangle^s FK\big)\big\|_{1,\lambda} \\
&= c'' \sum_{u=0}^{\infty} \big\|F^{-1}\big(\varphi_u\langle k\rangle^{-s}F\mu\big) * F^{-1}\langle k\rangle^s FK\big\|_{1,\lambda} \\
&\le c'' \sum_{u=0}^{\infty} \big\|F^{-1}\big(\varphi_u\langle k\rangle^{-s}F\mu\big)\big\|_{1,\lambda}\big\|F^{-1}\langle k\rangle^s FK\big\|_{1,\lambda} \\
&\le c''' \sum_{u=0}^{\infty} 2^{-su}\big\|F^{-1}(\varphi_u F\mu)\big\|_{1,\lambda} \\
&= c'''\|\mu\|_{-s,1,1,\lambda}
\end{aligned} \tag{6}$$

for some $0 < c''' < \infty$.

To finish the proof, we proceed as in the proof of Part 1 of Theorem 1. Since $M(\mathbb{R}^d)$ is dense in the Banach space $B^{-s}_{11}(\mathbb{R}^d)$ w.r.t. $\|\cdot\|_{-s,1,1,\lambda}$ (note that $M(\mathbb{R}^d)$ contains $\mathcal{S}(\mathbb{R}^d)$ which is dense in $B^{-s}_{11}(\mathbb{R}^d)$ by Theorem 2.3.3 in [21]), there exists a unique continuous linear mapping

$$\bar{\kappa} : B^{-s}_{11}\big(\mathbb{R}^d\big) \to L^1\big(\mathbb{R}^d\big)$$

such that $\bar{\kappa} = \kappa$ on $M(\mathbb{R}^d)$. We conclude that, since the random variables μ_n, when viewed as random elements of $B^{-s}_{11}(\mathbb{R}^d)$, converge in law, in probability, or almost surely in that space by Lemmas 1, 2, and 3 above, the random variables

$$\bar{\kappa}(\mu_n) = \kappa(\mu_n) = \mu_n * K$$

converge in law, in probability, or a.s. in $L^1(\mathbb{R}^d)$ by the continuous mapping theorem. □

Before we turn to applications, we collect the following two remarks.

Remark 1 (Derivatives of $\mu_n * K$) Theorems 1 and 2 can be used to derive convergence results in $\mathsf{C}_0(G)$ or $L^p(G)$ for (distributional) derivatives of $\mu_n * K$. To keep the notation simple, we set here $d = 1$ without loss of generality.

Case $G = \mathbb{T}$: Let δ' be the distribution on $\mathbb{T}$ obtained by taking the distributional derivative of Dirac measure δ at zero. As is well known,

$$\delta' * (\mu_n * K) = D(\mu_n * K)$$

holds, that is, convolution (in the sense of distributions) of a distribution T with δ' gives the distributional derivative DT of T, see, e.g., [20, p. 159]. Since furthermore convolution of distributions on $\mathbb{T}$ is commutative and associative, we have $\delta' * (\mu_n * K) = (\delta' * K) * \mu_n$, and one can apply Theorems 1 and 2 if the convolution kernel $\delta' * K = DK$ satisfies the conditions of these theorems. For example, if μ_n converges in the space $\ell^\infty(\mathcal{U}^s_2)$ in law (or in probability, or almost surely), and if $K \in W^{1+s}_1(\mathbb{T})$ for some $s > 1/2$, then $\delta' * K = DK \in W^s_1(\mathbb{T})$, and hence both $[\mu_n * K]_\lambda$ and $[D(\mu_n * K)]_\lambda$ converge in $L^2(\mathbb{T})$ in the respective sense. (In fact, $[\mu_n * K]_\lambda$ converges in $W^1_2(\mathbb{T})$.) Higher derivatives follow from a successive application of this argument.

Case $G = \mathbb{R}$: If one is interested in limit theorems for the derivatives of $\mu_n * K$ in the case $G = \mathbb{R}$, one may try to proceed as in the preceding paragraph. However, here the complication arises that the convolution product of arbitrary distributions on $\mathbb{R}$ is not necessarily well defined (and even if it is, it is not necessarily associative). So one has to establish first that $\delta' * (\mu_n * K) = (\delta' * K) * \mu_n$, which basically amounts to showing that differentiation and integration can be interchanged. Since δ' has compact support and since every $\mu_n \in M(\mathbb{R})$ gives rise to a tempered distribution, a sufficient condition for this to hold is that K gives rise to a rapidly decreasing *distribution* (not a function), cf. p. 244 and Théorème 11 on p. 247 in [20].

Remark 2 (Limiting Random Variable) In the generality of Theorems 1 and 2, we cannot say anything about the nature of the limiting random variable of $\mu_n * K$ in the corresponding Banach space B. Since we can infer uniform tightness of $\mu_n * K$ in B from convergence in law (and hence also from convergence in probability, or almost surely), the limiting variable can be determined by calculating the limits of $L(\mu_n * K)$ for every L contained in the dual space B' (or every L in a dense subset of B'), see, e.g., Sect. 1.4 in [1] or Sect. 2.1 in [11].

4 Some Applications

4.1 Estimating Sums of Independent Random Variables

We start with the following simple example.

Example 1 (Estimation of the sum of a known and an unknown random variable) Let $X = Y + Z$, where Y and Z are independent random variables with values in G. Assume that Z possesses a Lebesgue density p^Z, which is known, whereas nothing is known about the distribution $\mathbb{P}^Y$ of Y. Then the density p^X of X exists and is given by the convolution $\mathbb{P}^Y * p^Z$. Given a sample from Y, we want to estimate p^X. If $\mathbb{P}_n^Y$ denotes the empirical measure and if p^Z satisfies the conditions on K from Part 1 of Theorem 1, we obtain from this theorem and Proposition 1 that $\sqrt{n}(\mathbb{P}_n^Y * p^Z - \mathbb{P}^Y * p^Z)$ converges in law in the space $\mathsf{C}_0(G)$. For example, we obtain the rate of convergence $\|\mathbb{P}_n^Y * p^Z - \mathbb{P}^Y * p^Z\|_\infty = O_{\mathbb{P}}(n^{-1/2})$ for p^Z equal to a Beta-density (which is easily seen to satisfy the conditions on K in Theorem 1).

If one just wanted to obtain results as in the above example, one can always work with the empirical measure, and then in fact more powerful techniques could be used, see Sect. 4.2 below. However, if both the laws of Y and Z are unknown, the empirical measures of samples from Y and Z cannot be used to obtain rates of convergence in L^p-spaces (the convolution $\mathbb{P}_n^Y * \mathbb{P}_n^Z$ is discrete for every n, hence not contained in any L^p-space). Random measures arising from density estimators are the natural alternative, and in the remainder of this section we show how the results of this paper can be applied to this problem.

Let $X = Y + Z$, where Y and Z are independent random variables with values in $G \in \{\mathbb{T}, \mathbb{R}\}$ with unknown Lebesgue-densities p^Y and p^Z. The density p^X of X is then the convolution $p^Y * p^Z$. The goal is to estimate p^X if one has i.i.d. samples from Y and Z. (Everything that will be said below easily generalizes to the case $d > 1$ and also to sums $X = \sum_{i=1}^m Y_i$ of independent random variables Y_i by setting $Y = Y_1$ and $Z = \sum_{i=2}^m Y_i$.) Given a sequence of estimators p_n^Y and p_n^Z for the densities of Y and Z, respectively, an obvious plug-in estimate of $p^X = p^Y * p^Z$ is the convolution product $p_n^X = p_n^Y * p_n^Z$. We have the following simple decomposition

$$\begin{aligned}\sqrt{n}\big(p_n^X - p^X\big) &= \sqrt{n}\big(p_n^Y * p_n^Z - p^Y * p^Z\big)\\ &= \sqrt{n}\big(p_n^Y - p^Y\big) * p^Z + \sqrt{n}\big(p_n^Z - p^Z\big) * p^Y\\ &\quad + \sqrt{n}\big(p_n^Y - p^Y\big) * \big(p_n^Z - p^Z\big).\end{aligned} \tag{7}$$

To the first two terms in the last line of (7) one can apply Theorem 1 (or Theorem 2) above as follows: Suppose, for instance, that p_n^Y and p_n^Z are density estimators such that $\sqrt{n}(p_n^Y - p^Y)$ and $\sqrt{n}(p_n^Z - p^Z)$ converge in law in $\ell^\infty(\mathcal{U}_2^s)$. (See Sect. 2.2.2 for examples.) Suppose furthermore that the densities p^Y and p^Z belong to $\mathsf{W}_2^t(G)$ (or $W_1^t(G)$) for some $t > 1/2$. Then, by Theorem 1, the corresponding convolution products $\sqrt{n}(p_n^Y - p^Y) * p^Z$ and $\sqrt{n}(p_n^Z - p^Z) * p^Y$ converge in law in the space $\mathsf{C}_0(G)$ (or $L^2(G)$). (Using Theorem 2, a similar result can be proved in $L^1(\mathbb{R})$.) Hence, to obtain that $\sqrt{n}(p_n^X - p^X)$ converges in law in the spaces $\mathsf{C}_0(G)$ (or $L^2(G)$), we only have to show that the last term in (7) satisfies

$$\sqrt{n}\big\|\big(p_n^Y - p^Y\big) * \big(p_n^Z - p^Z\big)\big\|_{p,\lambda} = o_{\mathbb{P}}(1) \tag{8}$$

for $p = \infty$ (or $p = 2$). For example, for the kernel density estimator, this can be achieved as follows: It is well known that $p_0 \in \mathsf{W}_2^t(\mathbb{R})$ implies that the usual kernel density estimator $\tilde{p}_n(y) = \mathbb{P}_n * K_h(y)$ (from Proposition 2) based on an i.i.d. sample of size n from the law p_0 achieves the rate of convergence $\|\tilde{p}_n - p_0\|_{2,\lambda} = O_{\mathbb{P}}(n^{-t/(2t+1)})$ if $h_n \simeq n^{-1/(2t+1)}$. If such kernel estimators are constructed both for Y and Z, then by Young's inequality

$$\begin{aligned}\big\|\big(p_n^Y - p^Y\big) * \big(p_n^Z - p^Z\big)\big\|_\infty &\le \big\|p_n^Y - p^Y\big\|_2\big\|p_n^Z - p^Z\big\|_2 \\ &= O_{\mathbb{P}}\big(n^{-2t/(2t+1)}\big) = o_{\mathbb{P}}\big(n^{-1/2}\big)\end{aligned}$$

since $t > 1/2$, and we have, for example, the following:

Proposition 4 *Let $X = Y + Z$, where Y and Z are independent and have densities p^Y and p^Z contained in $\mathsf{W}_2^t(\mathbb{R})$ for some $t > 1/2$. Suppose that $Y_1, \ldots, Y_n$ and $Z_1, \ldots, Z_n$ are i.i.d. with densities p^Y and p^Z, respectively, and denote by $\mathbb{P}_n^Y$ and $\mathbb{P}_n^Z$ the associated empirical measures. If K is a kernel of order $r > t + 1/2$ and if $h_n \simeq n^{-1/(2t+1)}$, then $\sqrt{n}((\mathbb{P}_n^Y * K_{h_n}) * (\mathbb{P}_n^Z * K_{h_n}) - p^X)$ converges in law in the space $\mathsf{C}_0(\mathbb{R})$.*

In practice, we need not assume more than $t = 1/2 + \delta$ for δ arbitrary, so that any compactly supported symmetric kernel with the bandwidth $h_n \simeq n^{-1/(2(1+\delta))}$ will do. We also note that the limiting variable is mean zero Gaussian and can be computed similarly as in [7, p. 1118].

For the wavelet estimator from Sect. 2.2.2, one can prove a similar result: Using the additional fact that the wavelet estimator with resolution level $2^{j_n} \simeq n^{1/(2t+1)}$ achieves the optimal rate of convergence $n^{-t/(2t+1)}$ in the $L^2(\mathbb{R})$-norm (in probability), see, e.g., Theorem 10.1 in [10], we have:

Proposition 5 *Let the conditions of Proposition 4 be satisfied. If ϕ satisfies the conditions of Proposition 2 with $t < r + 1$ and if $2^{j_n} \simeq n^{1/(2t+1)}$, then $\sqrt{n}(\mathbb{P}_n^Y(K_{j_n}) * \mathbb{P}_n^Z(K_{j_n}) - p^X)$ converges in law in the space $\mathsf{C}_0(\mathbb{R})$.*

We can make remarks similar to those after Proposition 4. It should further be noted that, if the sample size for Y is n but the one for Z is m, then the same results hold at least if $m := m(n)$ is such that $m(n)/n \to c < \infty$. Analogous results can be proved in $L^p(G)$-spaces, including derivatives of p^X. Also, if $G = \mathbb{T}$, one can obtain such results for MLEs and trigonometric series estimators, see [13], Corollary 6 and [14].

It should be noted that the above derivations do require $t > 1/2$, but nothing more, which parallels what Giné and Mason [7] obtain for the kernel estimator, although they do not explicitly state their smoothness conditions. In contrast, in terms of smoothness conditions on the underlying density, Theorem 1 in [17] is suboptimal, where they require that p^Y and p^Z are continuously differentiable with derivative satisfying a Hölder condition of arbitrary order. In the $L^1(\mathbb{R})$-setting, Schick and Wefelmeyer [18] impose sharp smoothness conditions ($p_0^Y, p_0^Z \in W_1^t(\mathbb{R})$ for $t > 1/2$).

4.2 Comparison with the CLT in Banach Spaces

One can immediately apply Theorems 1 and 2 to the empirical process by using Proposition 1 as in Example 1. Alternatively, one can use the central limit theorem (CLT) in Banach spaces. It is interesting to compare the conditions arising via the latter approach to those required by the former. One would expect that our general approach has to pay a price when applied to the empirical process, since our proofs cannot use the special structure of the latter. Clearly, sums of i.i.d. random variables in spaces with "nice geometry" concentrate well around their mean, whereas this concentration phenomenon cannot be used in our case of arbitrary finite-signed measures. Interestingly, our results are (essentially) sharp in certain spaces while not in others, which we believe to reflect the role of the geometry of the space.

For K in $\mathsf{C}_0(G)$ or $\mathcal{L}^p(G)$, the i.i.d. random variables $\{X_j\}_{j=1}^n$ induce the sequence of random functions $K * \delta_{X_j}(x) = K(x - X_j)$ and the sums of i.i.d. random variables

$$\sqrt{n}(\mathbb{P}_n * K - \mathbb{P} * K)(x) = n^{-1/2} \sum_{j=1}^{n} \left(K(x - X_j) - \int_G K(x-y)\, d\mathbb{P}(y) \right). \tag{9}$$

For $K \in \mathcal{L}^p(G)$ or $K \in \mathsf{C}_0(G)$, these are in fact $L^p(G)$- or $\mathsf{C}_0(G)$-valued sums of i.i.d. random variables. (Note that the mapping $x \mapsto K(\cdot - x)$ from G to $L^p(G)$ or $\mathsf{C}_0(G)$ is continuous, hence Borel-measurable, so that the $K * \delta_{X_j}$ are indeed Borel random variables in these Banach spaces.) The CLT in Banach spaces is described by the geometry of the space [1, 11], so we have to distinguish several cases. We summarize in advance that the generalization from empirical processes to arbitrary sequences of random signed measures seems to have a price in the Hilbert space case ($L^2(G)$), but is possible without serious additional restrictions if one is "far away" from the Hilbert space setting, that is, in the (nonreflexive) Banach spaces $\mathsf{C}_0(G)$ and $L^1(\mathbb{R})$ that do not possess "nice" geometric properties.

1. *The Case of* $\mathsf{C}_0(G)$ Part 1 of Theorem 1 and Proposition 1 above imply that the expression in (9) converges in law in $\mathsf{C}_0(G)$ if $K \in \mathsf{W}_2^s(G)$ holds for some $s > d/2$. Alternatively, one can apply the CLT in the space of bounded continuous functions

on some compact metric space S, cf. Corollary 3.7.17 in [1]. In the case $S = \mathbb{T}^d$, this requires verification of the condition

$$\left|K(s-X) - K(t-X)\right| \le M(X)d(s,t)$$

for every $s,t \in S$, some M with $E(M^2) < \infty$, and some metric d on $\mathbb{T}^d$ that generates the usual topology. This is essentially equivalent to assuming some Hölder or Lipschitz condition on the function K. An arbitrary function $K \in \mathsf{W}_2^s(\mathbb{T}^d)$ satisfies a Hölder condition at best of order $s - d/2 > 0$, hence our condition $s > d/2$ is comparable. In $\mathsf{C}_0(\mathbb{R}^d)$, application of the general CLT is more involved. (One may use the one-point (Alexandrov) compactification $\bar{\mathbb{R}}^d$ of $\mathbb{R}^d$ and verify the Lipschitz condition in the space of continuous functions on $\bar{\mathbb{R}}^d$.) In contrast, our Theorem 1 can be used in $\mathsf{C}_0(\mathbb{R}^d)$ without any complications.

2. *The Case of $L^2(G)$ (and $L^1(\mathbb{T}^d)$)* Part 2 of Theorem 1 and Proposition 1 imply that the expression in (9) converges in law in $L^2(G)$ (and hence also in $L^1(\mathbb{T}^d)$) if $K \in \mathsf{W}_1^s(G)$ holds for some $s > d/2$. Alternatively, one can apply the CLT in the Hilbert space $L^2(G)$. It follows from Theorem 10.5 in [11] that

$$E\left\|K(\cdot - X) - \mathbb{P} * K\right\|_{2,\lambda}^2 < \infty \tag{10}$$

is sufficient for (9) to satisfy the CLT in $L^2(G)$. Since the $L^2(G)$-norm is translation invariant, condition (10) is satisfied for *arbitrary* $[K]_\lambda \in L^2(G)$. (Note that $[\mathbb{P} * K]_\lambda \in L^2(G)$ by (5).) Clearly, this result also carries over to $L^1(\mathbb{T}^d)$ by the continuous injection $L^2(\mathbb{T}^d) \hookrightarrow L^1(\mathbb{T}^d)$. Consequently our results, when applied to the empirical process, are suboptimal in this particular case.

3. *The Case of $L^1(\mathbb{R}^d)$* Theorem 2 and Proposition 1 imply that the expression in (9) converges in law in $L^1(\mathbb{R}^d)$ if $[K]_\lambda \in B_{11}^s(\mathbb{R}^d)$ holds for some $s > d/2$ and if $\mathbb{P}$ possesses a moment of order $\gamma > d$. Alternatively, one can use the CLT in the cotype 2 Banach space $L^1(\mathbb{R}^d)$. It is known that an $L^1(\mathbb{R}^d)$-valued random variable Z satisfies the CLT if and only if $\int_{\mathbb{R}^d} (EZ^2)^{1/2}\, d\lambda < \infty$, cf. p. 205 in [1]. In our setting, this is equivalent to

$$\int_{\mathbb{R}^d} \left(K^2 * \mathbb{P}\right)^{1/2} d\lambda = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} K^2(x-y)\, d\mathbb{P}(y)\right)^{1/2} d\lambda(x) < \infty. \tag{11}$$

One way to verify (11) is by a moment condition on K^2 and $\mathbb{P}$ of order $\gamma > d$, see Lemma 1 in [7]. Alternatively, one can impose some smoothness on K together with moment conditions on $\mathbb{P}$ to verify (11). The latter approach is related to the result obtained in this paper.

Acknowledgements The author wishes to thank Benedikt M. Pötscher for helpful discussions on a preliminary version of the paper and Evarist Giné, Viktor Losert, and Harald Rindler for comments on the subject of the paper.

References

1. Araujo, A., Giné, E.: The Central Limit Theorem for Real and Banach Valued Random Variables. Wiley, New York (1980)
2. Devroye, L., Lugosi, G.: Combinatorial Methods in Density Estimation. Springer, New York (2001)
3. Doukhan, P., Massart, P., Rio, E.: Invariance principles for absolutely regular processes. Ann. Inst. Henri Poincaré Probab. Stat. **31**, 393–427 (1995)
4. Dudley, R.M.: Uniform Central Limit Theorems. Cambridge University Press, Cambridge (1999)
5. Frees, E.W.: Estimating densities of functions of observations. J. Am. Stat. Assoc. **89**, 517–525 (1994)
6. Giné, E.: Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev-norms. Ann. Stat. **3**, 1243–1266 (1975)
7. Giné, E., Mason, D.M.: On local U-statistic processes and the estimation of densities of functions of several variables. Ann. Stat. **35**, 1105–1145 (2007)
8. Giné, E., Nickl, R.: Uniform limit theorems for wavelet density estimators. Preprint (2007)
9. Giné, E., Nickl, R.: Uniform central limit theorems for kernel density estimators. Probab. Theory Relat. Fields **141**, 333–387 (2008)
10. Härdle, W., Kerkyacharian, G., Picard, D., Tsybakov, A.: Wavelets, Approximation, and Statistical Applications. Lecture Notes in Statistics, vol. 129. Springer, New York (1998)
11. Ledoux, M., Talagrand, M.: Probability in Banach spaces. Springer, Berlin (1991)
12. Malliavin, P.: Integration and Probability. Springer, New York (1995)
13. Nickl, R.: Donsker type theorems for nonparametric maximum likelihood estimators. Probab. Theory Relat. Fields **138**, 411–449 (2007)
14. Nickl, R.: Uniform central limit theorems for density estimators. Preprint (2008)
15. Nickl, R., Pötscher, B.M.: Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov- and Sobolev-type. J. Theor. Probab. **20**, 177–199 (2007)
16. Reed, M., Simon, B.: Methods of Modern Mathematical Physics. I: Functional Analysis. Academic Press, San Diego (1972)
17. Schick, A., Wefelmeyer, W.: Root n consistent density estimators for sums of independent random variables. J. Nonparametr. Stat. **16**, 925–935 (2004)
18. Schick, A., Wefelmeyer, W.: Root n consistent density estimators of convolutions in weighted L_1-norms. J. Stat. Plan. Inference **137**, 1765–1774 (2007)
19. Schmeisser, H.-J., Triebel, H.: Topics in Fourier Analysis and Function Spaces. Wiley, New York (1987)
20. Schwartz, L.: Théorie des Distributions. Hermann, Paris (1966)
21. Triebel, H.: Theory of Function Spaces. Birkhäuser, Basel (1983)