IMS Collections
High Dimensional Probability V: The Luminy Volume
Vol. 5 (2009) 338–356

DOI: 10.1214/09-IMSCOLL522

Uniform central limit theorems for sieved maximum likelihood and trigonometric series estimators on the unit circle*

Richard Nickl[1]

University of Cambridge

Abstract: Given an i.i.d. sample from the law $\mathbb{P}$ on the unit circle, we obtain uniform central limit theorems for the random measures induced by trigonometric series and sieved maximum likelihood density estimators. The limit theorems are uniform over balls in Sobolev-Hilbert spaces of order $s > 1/2$.

1. Introduction

Let $X_1, \ldots, X_n$ be independent identically distributed random variables with common law $\mathbb{P}$. The simplest way to estimate the probability measure $\mathbb{P}$ is by the empirical measure $\mathbb{P}_n = n^{-1} \sum_{j=1}^n \delta_{X_j}$. The *uniform central limit theorem* (UCLT) for empirical measures is an important tool for studying the asymptotic properties of $\mathbb{P}_n$ as an estimator for $\mathbb{P}$. It states that

$$\sqrt{n}\left(\frac{1}{n}\sum_{j=1}^{n} f(X_j) - \int f d\mathbb{P}\right)$$

converges in law to a Gaussian limit not only for a *given* function f (with $\int f^2 d\mathbb{P} < \infty$), but *uniformly* so over certain function classes $\mathcal{F}$. Classes $\mathcal{F}$ for which this holds (also called Donsker classes) have been extensively studied in empirical process theory, cf. van der Vaart and Wellner [17] and Dudley [5].

The usefulness of $\mathbb{P}_n$ as an estimator for $\mathbb{P}$, however, also has its limitations. For example, if one considers the strong (instead of the weak) topology on the set of all probability measures, then $\|\mathbb{P}_n - \mathbb{P}\|_{TV}$ – where $\|\cdot\|_{TV}$ denotes the total variation norm on the set of finite signed measures – will *not* converge to zero in general. However, if $\mathbb{P}$ possesses a density p_0 w.r.t. Lebesgue measure λ that belongs to a probability model $\mathcal{P}$ contained in some Hölder- or Sobolev ball, then it is well-known that classical nonparametric density estimators $\tilde{p}_n$ for p_0 satisfy $\|\tilde{\mathbb{P}}_n - \mathbb{P}\|_{TV} \to_{n\to\infty} 0$ (where $d\tilde{\mathbb{P}}_n = \tilde{p}_n d\lambda$), with explicit rates of convergence to zero available - cf., e.g., Devroye and Lugosi [4]. A natural question now to ask is whether such a density estimator $\tilde{p}_n$ – that achieves the optimal rate of convergence

*An earlier version of this article was circulated in August 2006 under the title 'Uniform Central Limit Theorems for Density Estimators'.

[1]Statistical Laboratory, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, UK, e-mail: nickl@statslab.cam.ac.uk

Keywords and phrases: density estimation, orthogonal series estimator, sieved maximum likelihood estimator, uniform central limit theorem.

AMS 2000 subject classifications: Primary 60F17; secondary 46E35.

in the strong $\|\cdot\|_{TV}$-norm – *simultaneously* satisfies a *uniform central limit theorem*, that is, whether the stochastic process

$$(1) \qquad \sqrt{n}\left(\int f(\tilde{p}_n - p_0)d\lambda\right)_{(f\in\mathcal{F})}$$

converges in law, with respect to uniform convergence over some class of functions $\mathcal{F}$, to some Gaussian process indexed by $\mathcal{F}$. Such results imply in particular that the estimator $\tilde{p}_n$ possesses the uniform 'plug-in property' introduced recently by Bickel and Ritov [1]. As argued already in Bickel and Ritov [1], there are many potential statistical applications of such results. For example, there exist many parameters $\theta(\mathbb{P})$ of statistical interest for which $\theta(\mathbb{P}_n)$ is an inconsistent (or not even a well-defined) estimator. By using UCLTs for density estimators, these parameters can be shown to be estimable at rate $\sqrt{n}$ by certain plug-in density-estimators. See also Section 3 in Nickl [12] for further examples.

In the recent articles Nickl [12] and Giné and Nickl [8, 9, 10] it was shown that certain nonparametric maximum likelihood estimators (MLEs), classical kernel density estimators and wavelet density estimators satisfy the central limit theorem uniformly over a many Donsker classes $\mathcal{F}$. Since the estimators considered there also achieve the minimax rate of convergence in 'strong' metrics, this gives a positive answer to the question raised before (1).

In this article we shall restrict our attention to the special case where the sample takes values in the unit circle $\mathbb{T}$, and where $\mathcal{F}$ is a ball $\mathcal{U}_s$ in a Sobolev space of order $s > 1/2$ on $\mathbb{T}$. This has the advantage that one can explicitly use Hilbert-space structure in the proofs, which is useful to lay out the main mechanisms behind UCLTs for certain density estimators without having to become too technical. The estimators we consider in this article are the classical *trigonometric series estimator* (TSE), as well as the *sieved maximum likelihood estimator* based on trigonometric sieves. No UCLTs are available in the literature for both estimators, and we will close this gap for the Sobolev-case. Whereas the proofs for the trigonometric series estimator is simple, the sieved MLE is much harder and requires several nontrivial adaptations of the proof in Nickl [12]. The interest in *sieved* MLEs (in contrast to just MLEs) stems often from computational or practicaly issues, see van de Geer [15, 16], Wong and Shen [18] and Birgé and Massart [3]. We should note that our results for the particular Sobolev class $\mathcal{U}_s$ directly apply to many other classes of functions: Imbedding theorems for function spaces (see, e.g., Sections 3.5.4 and 3.5.5 in Schmeisser and Triebel [14]) imply that balls in Besov, Lipschitz, Hölder spaces (with smoothness index $s > 1/2$) are bounded subsets of the Sobolev spaces considered here.

2. Notations 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

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

Let $\mathbb{T}$ be the unit circle equipped with its Borel sigma-algebra. Then $\mathsf{L}^\infty(\mathbb{T})$ will denote the Banach space of bounded measurable real-valued functions on $\mathbb{T}$, normed by $\|\cdot\|_\infty := \|\cdot\|_{\mathbb{T}}$. For measurable functions $h : \mathbb{T} \to \mathbb{R}$ and measures μ on $\mathbb{T}$, we set

$\mu h := \int_{\mathbb{T}} h d\mu$ and $\|h\|_{p,\mu} := (\int_{\mathbb{T}} |h|^p d\mu)^{1/p}$, $1 \le p \le \infty$ (where $\|h\|_{\infty,\mu}$ denotes the μ-essential supremum of $|h|$). We write $\mathcal{L}^p(\mathbb{T}, \mu)$ for the vector space of all measurable functions $h : \mathbb{T} \to \mathbb{R}$ that satisfy $\|h\|_{p,\mu} < \infty$. The symbol $L^p(\mathbb{T}, \mu)$ denotes the usual quotient spaces of $\mathcal{L}^p(\mathbb{T}, \mu)$ modulo equality μ-almost everywhere. The symbol $d\lambda$ will denote the usual Lebesgue-measure normalized by $(2\pi)^{-1}$ so that $\int_{\mathbb{T}} d\lambda = 1$, and we then set shorthand $\mathcal{L}^p(\mathbb{T}) := \mathcal{L}^p(\mathbb{T}, \lambda)$. The convolution $f * g(x)$ of two measurable real-valued functions f, g on $\mathbb{T}$ is defined by $\int g(x-y)f(y)d\lambda(y)$ if the integral converges. Similarly, if μ is a finite signed measure and f is a measurable function, we define the convolution $\mu * f(x) = \int f(x-y)d\mu(y)$ if the integral exists.

Let $\langle k \rangle^s = (1 + |k|^2)^{s/2}$ and define, for real $s \ge 0$, the Sobolev space

$$\mathcal{W}_2^s(\mathbb{T}) = \left\{ f \in \mathcal{L}^2(\mathbb{T}) : \ \|f\|_{s,2,\lambda}^2 = \sum_{k \in \mathbb{Z}} \langle k \rangle^{2s} |Ff(k)|^2 < \infty \right\} \tag{2}$$

where $Ff(k) = \int e^{-ixk} f(x) d\lambda(x)$, $k \in \mathbb{Z}$. Clearly $\mathcal{W}_2^s(\mathbb{T})$ is a vector space semi-normed by $\|\cdot\|_{s,2,\lambda}$. With each element $f \in \mathcal{W}_2^s(\mathbb{T})$, any element of $[f]_\lambda$ also belongs to $\mathcal{W}_2^s(\mathbb{T})$, and by taking the quotient w.r.t. the set $\{f : \|f\|_{s,2,\lambda} = 0\}$, one obtains the Hilbert space $W_2^s(\mathbb{T})$. For $s > 1/2$, each equivalence class $[f]_\lambda \in W_2^s(\mathbb{T})$, contains a (unique) continuous function (Sobolev's lemma), and one defines the Hilbert spaces $\mathsf{W}_2^s(\mathbb{T}) = \{f \text{ continuous } : [f]_\lambda \in W_2^s(\mathbb{T})\}$.

Given n independent random variables $X_1, \dots, X_n$ identically distributed according to some Borel law $\mathbb{P}$ on $\mathbb{T}$, we denote by $\mathbb{P}_n = n^{-1} \sum_{j=1}^n \delta_{X_j}$ the empirical measure. We assume throughout that the variables X_j are the coordinate projections of $\mathbb{T}^{\mathbb{N}}$ with product probability $\mathbb{P}^{\mathbb{N}}$. The empirical process indexed by $\mathcal{F} \subseteq \mathcal{L}^2(\mathbb{T}, \mathbb{P})$ is given by

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

Convergence in law of random elements in $\ell^\infty(\mathcal{F})$ is defined in the usual way, see, e.g., Section 3 in Dudley [5], 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 $\mathcal{F}$ with covariance $E\mathbb{G}(f)\mathbb{G}(g) = \mathbb{P}[(f - \mathbb{P}f)(g - \mathbb{P}g)]$.

3. Main Results

3.1. Trigonometric Series Estimator

Estimating a density on the cirlce $\mathbb{T}$ is equivalent to estimating a density p_0 supported by $[0, 2\pi)$. Note that if p_0 is continuous on $\mathbb{T}$, it is necessarily periodic when viewed as a function on $[0, 2\pi)$ (i.e., $p_0(0) = \lim_{x \to 2\pi} p_0(x)$ has to hold). A natural density estimator can be obtained using the trigonometric polynomials

$$\{e_k(x) \equiv e^{ixk} : k \in \mathbb{Z}\}.$$

Given the empirical measure $\mathbb{P}_n$ obtained from the sample, its (random) Fourier series coefficients are given by $F\mathbb{P}_n(k) = n^{-1} \sum_{j=1}^n e^{-iX_j k}$, $k \in \mathbb{Z}$. We define the trigonometric series estimator by

$$p_n^{TSE} = \sum_{k \in \mathbb{Z} \cap [-H(n), H(n)]} F\mathbb{P}_n(k) e_k \tag{3}$$

where $H(n) \in \mathbb{N}$ is the truncation point of the series expansion increasing with sample size n, i.e., $H(n) < \infty$, $H(n) \to \infty$ as $n \to \infty$. [It is not difficult to see that the mapping $[p_n^{TSE}(\cdot, X_1, \ldots, X_n)]_\lambda : \mathbb{T}^n \to L^2(\mathbb{T}, \lambda)$ is Borel measurable.] The following result is folklore, and we include it for completeness and since elements of the proof will be used later on.

Proposition 1. *Let $X_1, \ldots, X_n$ be i.i.d. with law $\mathbb{P}$ on $\mathbb{T}$, $d\mathbb{P}(x) = p_0(x)d\lambda(x)$, where $p_0 \in \mathcal{W}_2^t(\mathbb{T})$ for some $t \geq 0$. Then*

$$\left\|p_n^{TSE} - p_0\right\|_{2,\lambda} = o(H(n)^{-t}) + O_{\mathbb{P}}(n^{-1/2}H(n)^{1/2}). \tag{4}$$

Proof. Here and later we shall use the simple fact that

$$\begin{aligned} \mathbb{E}\left|F\mathbb{P}_n(k) - Fp_0(k)\right|^2 &= \mathbb{E}\left|n^{-1}\sum_{j=1}^{n} e^{-iX_jk} - \int_{\mathbb{T}} e^{-ixk}d\mathbb{P}(x)\right|^2 \\ &\leq n^{-2}\sum_{j=1}^{n}\mathbb{E}\left|e^{-iX_jk} - \mathbb{E}e^{-iX_jk}\right|^2 \leq n^{-1} \end{aligned} \tag{5}$$

holds by independence of the X_j and since the e^{-ixk} are uniformly bounded by 1.

Now define the truncated Fourier series

$$u_n(p_0) = \sum_{k\in\mathbb{Z}\cap[-H(n),H(n)]} Fp_0(k)e_k \tag{6}$$

so that $\|p_n^{TSE} - p_0\|_{2,\lambda} \leq \|p_n^{TSE} - u_n(p_0)\|_{2,\lambda} + \|u_n(p_0) - p_0\|_{2,\lambda}$. We treat the bias term first:

$$\begin{aligned} \|u_n(p_0) - p_0\|_{2,\lambda}^2 &= \sum_{k\in\mathbb{Z}\setminus[-H(n),H(n)]} |Fp_0(k)|^2 \\ &\leq \sup_{k\in\mathbb{Z}\setminus[-H(n),H(n)]} \langle k\rangle^{-2t} \sum_{k\in\mathbb{Z}\setminus[-H(n),H(n)]} \langle k\rangle^{2t}|Fp_0(k)|^2 \\ &= O(H(n)^{-2t})o(1) = o(H(n)^{-2t}) \end{aligned} \tag{7}$$

holds by assumption, Parseval's identity, Hölder's inequality and definition of $\|\cdot\|_{t,2,\lambda}$. By using (5) and again Parseval's identity, we obtain for the variance term

$$\begin{aligned} &\mathbb{E}\left\|p_n^{TSE} - u_n(p_0)\right\|_{2,\lambda}^2 \\ &= \mathbb{E}\sum_{k\in\mathbb{Z}\cap[-H(n),H(n)]} |F\mathbb{P}_n(k) - Fp_0(k)|^2 \\ &\leq \sum_{k\in\mathbb{Z}\cap[-H(n),H(n)]} n^{-1} = n^{-1}(2H(n)+1), \end{aligned}$$

which – after collecting terms – completes the proof. □

Consequently, for the choice $H(n) \sim n^{1/(2t+1)}$, we obtain the rate bound $O_{\mathbb{P}}(n^{-t/(2t+1)})$, which is the usual minimax rate of convergence over Sobolev balls of densities in $\|\cdot\|_{2,\lambda}$-loss for *any* estimator of p_0.

We will denote by $\mathbb{P}_n^{TSE}$ the random measure defined by

$$\mathbb{P}_n^{TSE}(A) = \int_A p_n^{TSE}(x)d\lambda(x)$$

for every Borel set $A \subseteq \mathbb{T}$. The following result has a simple proof. Note that balls in $\mathsf{W}_2^s(\mathbb{T})$ are $\mathbb{P}$-Donsker for any $\mathbb{P}$ if $s > 1/2$, see, e.g., Giné [7].

Theorem 1. *Let $X_1, \ldots, X_n$ be i.i.d. with law $\mathbb{P}$ on $\mathbb{T}$, $d\mathbb{P}(x) = p_0(x)d\lambda(x)$, where $p_0 \in \mathcal{W}_2^t(\mathbb{T})$ for some $t \geq 0$. Let $\mathcal{F}$ be a (non-empty) bounded subset of the space $\mathsf{W}_2^s(\mathbb{T})$ with $s > 1/2$. Then*

$$\mathbb{E}\sqrt{n}\left\|\mathbb{P}_n^{TSE} - \mathbb{P}_n\right\|_{\infty,\mathcal{F}} = o(n^{1/2}H(n)^{-s-t}) + O(H(n)^{1/2-s}). \tag{8}$$

If in addition $H(n)^{-s-t} = O(n^{-1/2})$ holds, we have

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

Proof. To prove Theorem 1, we use negative order Sobolev (distribution-) spaces $W_2^{-s}(\mathbb{T})$ with norm $\|f\|_{-s,2,\lambda}^2 \simeq \sum_{k\in\mathbb{Z}} |Ff(k)|^2 \langle k \rangle^{-s}$, which are the dual spaces of $\mathsf{W}_2^s(\mathbb{T})$. (The proof of this duality is elementary, using Fourier transforms and duality arguments for weighted ℓ^2-spaces.) We set w.l.o.g. $\mathcal{F}$ equal to the unit ball $\mathcal{U}_{s,1}$ of $\mathsf{W}_2^s(\mathbb{T})$. Define the finite signed measure $\mathbb{P}_n^{rem}$ by

$$d\mathbb{P}_n^{rem} = (u_n(p_0) - p_0)d\lambda$$

where $u_n(p_0)$ was defined in (6) above. Note further that

$$\mathbb{P}_n^{TSE} - \mathbb{P}_n = -\sum_{k\in\mathbb{Z}\setminus[-H(n),H(n)]} F\mathbb{P}_n(k)e_k$$

holds (in $W_2^{-s}(\mathbb{T})$ for $s > 1/2$). We thus have

$$\begin{aligned}\left\|\mathbb{P}_n^{TSE} - \mathbb{P}_n\right\|_{\infty,\mathcal{U}_{s,1}} &= \left\|\mathbb{P}_n^{TSE} - \mathbb{P}_n + \mathbb{P}_n^{rem} - \mathbb{P}_n^{rem}\right\|_{\infty,\mathcal{U}_{s,1}} \\ &\leq \left\|\mathbb{P}_n^{TSE} - \mathbb{P}_n + \mathbb{P}_n^{rem}\right\|_{\infty,\mathcal{U}_{s,1}} + \left\|\mathbb{P}_n^{rem}\right\|_{\infty,\mathcal{U}_{s,1}}.\end{aligned}$$

We first handle the second term. Using the duality between $W_2^{-s}(\mathbb{T})$ and $\mathsf{W}_2^s(\mathbb{T})$ together with Hölder's inequality and the definition of $\|\cdot\|_{t,2,\lambda}$, we obtain

$$\begin{aligned}\|\mathbb{P}_n^{rem}\|_{\infty,\mathcal{U}_{s,1}} &\sim \|\mathbb{P}_n^{rem}\|_{-s,2,\lambda} \\ &= \left(\sum_{k\in\mathbb{Z}\setminus[-H(n),H(n)]} |Fp_0(k))|^2 \langle k\rangle^{-2s}\right)^{1/2} \\ &= \left(\sum_{k\in\mathbb{Z}\setminus[-H(n),H(n)]} \langle k\rangle^{2t} |Fp_0(k))|^2 \langle k\rangle^{-2(s+t)}\right)^{1/2} \\ &\leq \left(\sup_{k\in\mathbb{Z}\setminus[-H(n),H(n)]} \langle k\rangle^{-2(s+t)}\right)^{1/2} \\ &\quad \times \left(\sum_{k\in\mathbb{Z}\setminus[-H(n),H(n)]} \langle k\rangle^{2t} |Fp_0(k)|^2\right)^{1/2} \\ &= O(H(n)^{-s-t})o(1) = o(H(n)^{-s-t})\end{aligned}$$

for some $0 < C < \infty$. For the first term, we use the same duality argument, Fubini and (5) to obtain

$$\begin{aligned}\mathbb{E}\left\|\mathbb{P}_n^{TSE}-\mathbb{P}_n+\mathbb{P}_n^{rem}\right\|_{\infty,\mathcal{U}_{s,1}}^2 &\simeq \mathbb{E}\left\|\mathbb{P}_n^{TSE}-\mathbb{P}_n+\mathbb{P}_n^{rem}\right\|_{-s,2,\lambda}^2\\ &= \mathbb{E}\sum_{k\in\mathbb{Z}\setminus[-H(n),H(n)]}\langle k\rangle^{-2s}\left|F\mathbb{P}_n(k)-Fp_0(k)\right|^2\\ &= n^{-1}\sum_{k\in\mathbb{Z}\setminus[-H(n),H(n)]}\langle k\rangle^{-2s} = O(n^{-1}H(n)^{1-2s})\end{aligned}$$

for some $0<C<\infty$. This completes the proof of (8). The second claim of the theorem follows immediately from (8) since $\|\mathbb{P}_n^{TSE}-\mathbb{P}_n\|_{\infty,\mathcal{F}}$ is then of order $o_{\mathbb{P}}(n^{-1/2})$, and the fact that $\mathcal{F}$ is a universal Donsker class. ☐

We discuss some interesting choices of $H(n)$. First, we obtain a corollary for the case $H(n)\sim n^{1/(2t+1)}$.

Corollary 1. *Let the conditions of Theorem 1 be satisfied. Choose $H(n)\sim n^{1/(2t+1)}$. We then have that*

$$\sqrt{n}\left\|\mathbb{P}_n^{TSE}-\mathbb{P}_n\right\|_{\infty,\mathcal{F}} = O_{\mathbb{P}}(n^{(1/2-s)/(2t+1)}) = o_{\mathbb{P}}(1) \tag{10}$$

holds. In particular, we have

$$\sqrt{n}(\mathbb{P}_n^{TSE}-\mathbb{P}) \rightsquigarrow_{\ell^\infty(\mathcal{F})}\mathbb{G}$$

where $\mathbb{G}$ is defined as in Theorem 1 above.

Hence, if $p_0\in\mathcal{W}_2^t(\mathbb{T})$ for some $t>0$, our results imply hat the trigonometric series estimator with $H(n)\sim n^{1/(2t+1)}$ achieves *both* the optimal rate of convergence (to p_0) in $\|\cdot\|_{2,\lambda}$-loss, *and* satisfies a UCLT.

3.2. Sieved Maximum Likelihood Estimator

Consider the probability model

$$\mathcal{P}(t,\zeta,D)=\left\{p\in \mathsf{W}_2^t(\mathbb{T}): p(x)\geq\zeta \text{ for all } x\in\mathbb{T},\quad \int_{\mathbb{T}} p\,d\lambda=1,\quad \|p\|_{t,2,\lambda}\leq D\right\} \tag{11}$$

with real constants $t>1/2$, $\zeta>0$, $0<D<\infty$ satisfying $\zeta\leq 1\leq D$; the latter condition ensuring that $\mathcal{P}(t,\zeta,D)$ is non-empty. Given the model $\mathcal{P}(t,\zeta,D)$, the (log)likelihood function is given by

$$L_n(p) := \mathbb{P}_n\log p = n^{-1}\sum_{j=1}^n \log p(X_j) \tag{12}$$

with $p\in\mathcal{P}(t,\zeta,D)$. A maximizer of the function $L_n(p)$ over $\mathcal{P}(t,\zeta,D)$ (if it exists) is a maximum likelihood estimator.

Instead of maximizing L_n over the whole parameter space $\mathcal{P}(t,\zeta,D)$, it is often convenient to rather maximize L_n over some *sieve* (approximating model) $\mathcal{P}_{H(n)}$, see, e.g., van de Geer [15], Wong and Shen [18] and Birgé and Massart [3]. We will consider the following simple trigonometric sieve

$$\mathcal{P}_{H(n)}(t,\zeta,D)=\left\{p\in\mathcal{P}(t,\zeta,D):\ p\in\langle e_k\rangle_{H(n)}\right\}$$

where $\langle e_k \rangle_H$ denotes the set of real-valued functions contained in the linear span of $\{e_k : k \in \mathbb{Z} \cap [-H, H]\}$, and where $H(n) \in \mathbb{N}$ for every n is the dimension of the sieve. [Note that $\langle e_k \rangle_H \subseteq \mathcal{W}_2^s(\mathbb{T})$ for every $H \in \mathbb{N}$, $s \in \mathbb{R}$. In particular, $\mathcal{P}_{H(n)}(t, \zeta, D)$ is non-empty.] The sieved maximum likelihood estimator is defined as an element $p_n^{MLE} \in \mathcal{P}_{H(n)}(t, \zeta, D)$ which satisfies

$$L_n\left(p_n^{MLE}\right) = \sup_{p \in \mathcal{P}_{H(n)}(t,\zeta,D)} L_n(p). \tag{13}$$

We will use the convention that $\langle e_k \rangle_\infty$ is equal to $\mathcal{L}^2(\mathbb{T}, \lambda)$, so that the unsieved MLE corresponds to the case $H(n) = \infty$.

Since $t > 1/2$, the set $\mathcal{P}_H(t, \zeta, D)$ is, for every $H \in \mathbb{N} \cup \{\infty\}$, a compact subset of $\mathsf{C}(\mathbb{T})$, see Lemma 2. Furthermore, the function $L_n(\cdot)$ is continuous on $\mathcal{P}_H(t, \zeta, D)$ w.r.t. the sup-norm topology. Consequently, the supremum in (13) is attained. [Furthermore, using Proposition 5 in Nickl [12], a Borel (e.g., for the sup-norm-topology) measurable element $p_n^{MLE} \in \mathcal{P}_{H(n)}(t, \zeta, D)$ satisfying (13) exists. We note that all results in the paper hold for *every* measurable selection. In fact, they hold for *any* selection (measurable or not) if one formulates all results in terms of outer probability $\mathbb{P}^*$.]

We will impose the following condition:

Condition 1. *Let $X_1, \ldots, X_n$ be i.i.d. with law $\mathbb{P}$ on $\mathbb{T}$, $d\mathbb{P}(x) = p_0(x) d\lambda(x)$, and p_0 is contained in $\mathcal{P}(t, \zeta, D)$ where $t > 1/2$ (and $0 < \zeta \le 1 \le D$). Furthermore, p_0 satisfies the strict inequalities $p_0(x) > \zeta$ for all $x \in \mathbb{T}$, as well as $\|p_0\|_{t,2,\lambda} < D$.*

The second part of the condition is an 'internality condition' – which is discussed in detail in Nickl [12]. With an eye on Bickel and Ritov's [1] plug-in property', we first wish to show that the sieved MLE is optimal in strong' metrics, by using a result in van de Geer [16]. The proof can be found in the next section.

Proposition 2. *Let $0 < H(n) < \infty$, $H(n) \to \infty$ as $n \to \infty$. Suppose Condition 1 is satisfied. Then*

$$\left\| p_n^{MLE} - p_0 \right\|_{2,\lambda} = o(H(n)^{-t}) + O_{\mathbb{P}}(n^{-t/(2t+1)}). \tag{14}$$

For the choice $H(n) \sim n^{1/(2t+1)}$, we obtain the rate bound $O_{\mathbb{P}}(n^{-t/(2t+1)})$. The conditions on the MLE in the above proposition are somewhat stronger than those needed for the kernel and series estimator, which is related to the well known fact that otherwise the bracketing integral of $\mathcal{P}_t$ is not convergent and MLEs are then known to be suboptimal (cf. Birgé and Massart [2]). Also, with a view on what follows, whether (14) can be proved under less restrictive assumptions is not our primary focus here.

Denote by $\mathbb{P}_n^{MLE}$ the random measure induced by p_n^{MLE}. While one can obtain results for general sequences $H(n)$ (see Theorem 2 below), in most cases only specific choices of $H(n)$ are of interest, which are given in the following corollary (to Theorem 2 below). Recall that balls in $\mathsf{W}_2^s(\mathbb{T})$ are $\mathbb{P}$-Donsker for any $\mathbb{P}$ if $s > 1/2$.

Corollary 2. *Suppose that Condition 1 is satisfied. Let $\mathcal{F}$ be a (non-empty) bounded subset of the space $\mathsf{W}_2^s(\mathbb{T})$ where $s > 1/2$. Let either $H(n) \sim n^{1/(2t+1)}$ or $H(n) = \infty$ for every n. We then have for every $1/2 < k < \min(t, s)$ that*

$$\sqrt{n} \left\| \mathbb{P}_n^{MLE} - \mathbb{P}_n \right\|_{\infty,\mathcal{F}} = \begin{cases} O_{\mathbb{P}}(n^{(k-s)/(2t+1)}) = o_{\mathbb{P}}(1) & \text{if } s < t \\ O_{\mathbb{P}}(n^{(k-t)/(2t+1)}) = o_{\mathbb{P}}(1) & \text{if } s \ge t \end{cases} \tag{15}$$

holds. In particular, we have

$$\sqrt{n}(\mathbb{P}_n^{MLE} - \mathbb{P}) \rightsquigarrow_{\ell^\infty(\mathcal{F})} \mathbb{G}.$$

Proof. The result follows from Theorem 2 below: If $H(n) \sim n^{1/(2t+1)}$, choose $K(n) = H(n)$. If $H(n) = \infty$, choose $K(n) \sim n^{1/(2t+1)}$. □

The corollary shows that – under the conditions of Theorem 2 – the sieved MLE (with $H(n) \sim n^{1/(2t+1)}$) as well as the unsieved MLE have a similar asymptotic behavior in $\ell^\infty(\mathcal{F})$ as the trigonometric series estimator with MISE-optimal bandwidths. In particular, also the sieved MLE satisfies Bickel and Ritov's [1] plug-in property'. Note also that for $H(n) = \infty$, the corollary gives a result analogous to Theorem 3 in Nickl [12] for the sample space $\mathbb{T}$. Inspection of Theorem 2 also shows that, in contrast to the series estimator, *any* approximating space growing faster than $H(n) \sim n^{1/(2t+1)}$ delivers a UCLT for the maximum likelihood estimator.

4. Proofs for Sieved MLEs

The purpose of this section is to prove the following result.

Theorem 2. *Suppose that Condition 1 is satisfied. Let $\mathcal{F}$ be a (non-empty) bounded subset of the space $\mathsf{W}_2^s(\mathbb{T})$ where $s > 1/2$. Let either $H(n) = \infty$ for every n or $H(n) < \infty$, $H(n) \to \infty$ as $n \to \infty$. Let $K(n)$ be any sequence of positive integers such that $K(n) \leq H(n)$ holds for every n, and let k be arbitrary subject to $1/2 < k < \min(t, s)$. Define the random sequence of real numbers*

$$C_n(s,t) = \sqrt{n} \left\| \mathbb{P}_n^{MLE} - \mathbb{P}_n \right\|_{\infty,\mathcal{F}}. \tag{16}$$

Then $C_n(s,t)$ can be stochastically bounded as follows: If $s < t$, then

$$\begin{aligned} C_n(s,t) = {} & o_{\mathbb{P}}(n^{1/2-t/(2t+1)} K(n)^{-s}) + o(n^{1/2} H(n)^{-t} K(n)^{-s}) \\ & + O_{\mathbb{P}^*}(n^{-(t-k)/(2t+1)} K(n)^{t-s}) + O(H(n)^{-(t-k)} K(n)^{t-s}) \\ & + o(n^{1/2} H(n)^{-2t} K(n)^{t-s}) + o_{\mathbb{P}}(K(n)^{-(s-k)}). \end{aligned}$$

If $s \geq t$, then

$$\begin{aligned} C_n(s,t) = {} & o_{\mathbb{P}}(n^{1/2-t/(2t+1)} K(n)^{-s}) + o(n^{1/2} H(n)^{-t} K(n)^{-s}) \\ & + O_{\mathbb{P}^*}(n^{-(t-k)/(2t+1)}) + O_{\mathbb{P}}(H(n)^{-(t-k)}) \\ & + o(n^{1/2} H(n)^{-2t}) + o_{\mathbb{P}}(K(n)^{-(s-k)}). \end{aligned}$$

In both displays, if $H(n) = \infty$ for every n, then any term that involves $H(n)$ can be set to zero.

Furthermore, if $C_n(s,t) = o_{\mathbb{P}^*}(1)$ holds, then we have

$$\sqrt{n}(\mathbb{P}_n^{MLE} - \mathbb{P}) \rightsquigarrow_{\ell^\infty(\mathcal{F})} \mathbb{G}. \tag{17}$$

The basic proof idea for Theorem 2 is inspired by the proof of Theorems 1-3 in Nickl [12], where unsieved MLEs are treated. The case of the sieved MLE needs considerable (and nontrivial) adaptations. In particular, since the true parameter p_0 is not generally contained in $\mathcal{P}_{H(n)}(t, \zeta, D)$ for given n, the proof of the central Lemma 4 in Nickl [12] cannot be directly used. On the other hand, some preliminary results can be taken from Nickl [12] without special efforts. Note also that the proof of Proposition 2 follows as a special case of Proposition 5 below.

The proof will be given in several steps. We first recall some simple facts on Sobolev spaces.

Proposition 3. *Let $s > 1/2$.*

1. *Every $[f]_\lambda \in W_2^s(\mathbb{T})$ contains an element $f \in \mathsf{C}(\mathbb{T})$. In particular,*
$$\|f\|_\infty \le C_s \|f\|_{s,2,\lambda}$$
holds for every $[f]_\lambda \in W_2^s(\mathbb{T})$ and imbedding constant $0 < C_s < \infty$.
2. *The set $\mathcal{P}(s,\zeta,D)$ defined in (11) is contained in $\{f \in \mathsf{C}(\mathbb{T}) : \zeta \le f(x) \le C_s D$ for all $x \in \mathbb{T}\}$.*
3. *Let $\mathcal{F}$ be a bounded subset of $\mathsf{W}_2^s(\mathbb{T})$. Then $\mathcal{F}$ is equicontinuous on $\mathbb{T}$.*

Proof. Part 1 (Sobolev's lemma) follows easily from Fourier inversion, and Part 2 follows from Part 1 and the definitions. Equicontinuity in Part 3 follows, e.g., from 3.5.4/19 and 3.5.5/4 in Schmeisser and Triebel [14]. □

4.1. Preliminary Results

We will need the Fréchet derivatives of the likelihood function $p \longmapsto L_n(p)$ as well as of its limiting function $p \longmapsto \mathbb{P}L(p) = \int_\mathbb{T} \log p(x) d\mathbb{P}(x)$ both viewed as mappings defined on a suitable open subset $\mathcal{V}$ of the Banach space $\mathsf{L}^\infty(\mathbb{T})$.

Proposition 4. *Let $\mathcal{V} = \{d \in \mathsf{L}^\infty(\mathbb{T}) : d(x) > \zeta/2$ for all $x \in \mathbb{T}\}$ where $0 < \zeta < \infty$. For $f_1, \ldots, f_\alpha \in \mathsf{L}^\infty(\mathbb{T})$, $\alpha \ge 1$, the multilinear mapping representing the α-th Fréchet-derivative of $L_n : \mathcal{V} \to \mathbb{R}$ at the point $d \in \mathcal{V}$ is given by*

$$D^\alpha L_n(d)(f_1, \ldots, f_\alpha) = n^{-1}(\alpha-1)!(-1)^{\alpha-1} \sum_{j=1}^n d^{-\alpha}(X_j) f_1(X_j) \cdot \cdots \cdot f_\alpha(X_j).$$

Furthermore, the multilinear mapping representing the α-th Fréchet-derivative of $\mathbb{P}L(\cdot)$ at the point $d \in \mathcal{V}$ is given by

$$\begin{aligned} D^\alpha \mathbb{P}L(d)(f_1, \ldots, f_\alpha) &= \mathbb{P}D^\alpha L(d)(f_1, \ldots, f_\alpha) \\ &= (\alpha-1)!(-1)^{\alpha-1} \int_\mathbb{T} d^{-\alpha} f_1 \cdot \cdots \cdot f_\alpha d\mathbb{P}. \end{aligned}$$

Proof. The result follows from Proposition 3 in Nickl [12] upon setting $\Omega = [0,1)$ in that proposition (and upon identifying $\mathsf{L}^\infty(\mathbb{T})$ with $\mathsf{L}^\infty([0,1))$). [Note that $[0,1)$ in Nickl [12] carries a different topology but the same σ-field.] □

By Part 2 of Proposition 2, the set $\mathcal{P}(t,\zeta,D)$ is contained in the $\mathsf{L}^\infty(\mathbb{T})$-open set $\mathcal{V}$. Thus the above result shows that the likelihood function and its limiting counterpart are Fréchet-differentiable in $\mathsf{L}^\infty(\mathbb{T})$ at each $p \in \mathcal{P}(t,\zeta,D)$ (the former for all $(X_1,\ldots,X_n)^T \in \mathbb{T}^n$).

In what follows, for a (possibly random) symmetric bilinear functional Ψ defined on $\mathcal{L}^0(\mathbb{T})$, we shall use the following notation where $\mathcal{H}$ and $\mathcal{G}$ are subsets of $\mathcal{L}^0(\mathbb{T})$:

$$\|\Psi\|_{\infty,\mathcal{H},\mathcal{G}} := \sup_{h\in\mathcal{H}} \sup_{g\in\mathcal{G}} |\Psi(h,g)|. \tag{18}$$

We use the same notation for multilinear functionals.

Lemma 1. *Let Condition 1 hold with $t > 1/2$ and assume $p_0(x) \ge \zeta > 0$ for every $x \in \mathbb{T}$. Let $1 \le \alpha < \infty$ and let $\mathcal{H}_j$, $j = 1, \ldots, \alpha$, be bounded subsets of $\mathsf{L}^\infty(\mathbb{T})$ that are $\mathbb{P}$-Donsker. We then have*

$$\sup_{p\in\mathcal{P}(t,\zeta,D)} \|D^\alpha L_n(p) - \mathbb{P}D^\alpha L(p)\|_{\infty,\mathcal{H}_1,\ldots,\mathcal{H}_\alpha} = O_{\mathbb{P}^*}(n^{-1/2}). \tag{19}$$

Proof. The proof is exactly the same as the one of Lemma 2 in Nickl [12]. □

We now turn to compactness of the (approximating) parameter space.

Lemma 2. *If $t > 1/2$, then $\mathcal{P}_{H(n)}(t, \zeta, D)$ is a compact subset of $\mathsf{C}(\mathbb{T})$.*

Proof. The proof is again a simple adaptation of a result in Nickl [12]. The set $\mathcal{U}_{H(n),t,D} = \{f : \mathbb{T} \to \mathbb{R}, f \in \langle e_k \rangle_{H(n)} : \|f\|_{t,2,\lambda} \leq D\}$ is, by Proposition 3, bounded in $\mathsf{C}(\mathbb{T})$ and uniformly equicontinuous on $\mathbb{T}$. Hence $\mathcal{U}_{H(n),t,D}$ is relatively compact in $\mathsf{C}(\mathbb{T})$ by the Arzelà-Ascoli theorem. That $\mathcal{U}_{H(n),t,D}$ is in fact compact in $\mathsf{C}(\mathbb{T})$ follows from similar arguments as in the proof of Lemma 3 in Nickl [12]. [Note that $\langle e_k \rangle_{H(n)}$ is a closed subspace of $\mathsf{W}_2^t(\mathbb{T})$, which implies that $\mathcal{U}_{H(n),t,D}$ is norm-closed and -bounded - and hence weakly compact - in $\mathsf{W}_2^t(\mathbb{T})$.]

Again similar as in Lemma 3 in Nickl [12], it follows that the sets

$$\mathcal{P}^{\zeta} = \{f \in \mathsf{C}(\mathbb{T}) : \zeta \leq f(x) < \infty \ \forall x \in \mathbb{T}\}$$

as well as

$$\mathcal{P}^{(1)} = \{g \in \mathsf{C}(\mathbb{T}) : \ \|g\|_{1,\lambda} = 1\}$$

are $\|\cdot\|_\infty$-closed. Then $\mathcal{P}_{H(n)}(t, \zeta, D) = \mathcal{P}^{(1)} \cap \mathcal{P}^{\zeta} \cap \mathcal{U}_{H(n),t,D}$ is compact in $\mathsf{C}(\mathbb{T})$ since it is the intersection of a compact and two closed subsets of $\mathsf{C}(\mathbb{T})$. □

4.2. Approximating Sequence, Rates of (Strong) Convergence

To develop the relevant asymptotic theory for the sieved MLE (i.e., $H(n) < \infty$ for given n), we have to take into account that the true density p_0 is not necessarily contained in $\mathcal{P}_{H(n)}(t, \zeta, D)$ for any given n. The main idea is based on the construction of a suitable approximating element $p_n^* \in \mathcal{P}_{H(n)}(t, \zeta, D)$ of p_0. This approach is also often used to obtain rates of convergence of the sieved MLE in the L^2-norm (or some closely related distance), see, e.g., Theorem 10.13 in van de Geer [16] or Section 4 in Wong and Shen [18]. In our case the approximating element p_n^* has to match some geometric properties of $\mathcal{P}_{H(n)}(t, \zeta, D)$. In particular, p_n^* has to satisfy the internality condition' in Condition 1. We construct such a sequence in the following lemma.

Lemma 3. *Let p_0 satisfy Condition 1. Let $H(n) < \infty$ tend to infinity as $n \to \infty$. Then there exists a sequence of functions $p_n^* \in \mathcal{P}_{H(n)}(t, \zeta, D)$ satisfying the following properties:*

1. *$\|p_n^*\|_{t,2,\lambda} < D$ for every n.*
2. *$p_n^*(x) > \zeta$ for every $x \in \mathbb{T}$ and every n.*
3. *$\|p_n^* - p_0\|_{s,2,\lambda} = o(H(n)^{s-t})$ for $s \leq t$ as $n \to \infty$.*

Proof. For $p_0 \in \mathsf{W}_2^t(\mathbb{T})$, $u_n(p_0)$ is the truncated Fourier series expansion of p_0 (see (6) above). Clearly, $u_n(p_0)$ is, for every n, a continuous and hence bounded and integrable function on $\mathbb{T}$. Observe first that for $s \leq t$

$$\begin{aligned} (20) \quad \|u_n(p_0) - p_0\|_{s,2,\lambda}^2 &= \sum_{k \in \mathbb{Z} \setminus [-H(n), H(n)]} |Fp_0(k)|^2 \langle k \rangle^{2t} \langle k \rangle^{2(s-t)} \\ &\leq \sup_{k \in \mathbb{Z} \setminus [-H(n), H(n)]} \langle k \rangle^{2(s-t)} \sum_{k \in \mathbb{Z} \setminus [-H(n), H(n)]} |Fp_0(k)|^2 \langle k \rangle^{2t} \\ &= O(H(n)^{2(s-t)}) o(1) = o(H(n)^{2(s-t)}) \end{aligned}$$

by Hölder's inequality and definition of $\|\cdot\|_{t,2,\lambda}$. In particular, we have the chain of inequalities

$$\begin{aligned}(21)\qquad \|u_n(p_0)-p_0\|_{1,\lambda} &\le \|u_n(p_0)-p_0\|_{2,\lambda}\\ &\le \|u_n(p_0)-p_0\|_\infty \le C_t\,\|u_n(p_0)-p_0\|_{t,2,\lambda} \to_{n\to\infty} 0\end{aligned}$$

where we have used Part 1 of Proposition 3 in the last inequality.

Now (21) implies that $\|u_n(p_0)\|_{1,\lambda} \to_{n\to\infty} \|p_0\|_{1,\lambda} = 1$ holds, and hence there exists some positive integer N such that the sequence $u_n(p_0)/\|u_n(p_0)\|_{1,\lambda}$ is well defined for every $n \ge N$. In particular we conclude that

$$(22)\qquad \left\|u_n(p_0)/\,\|u_n(p_0)\|_{1,\lambda} - p_0\right\|_{t,2,\lambda} \to 0$$

holds as n tends to infinity. Consequently, since $\|p_0\|_{t,2,\lambda} < D$ holds by assumption, we infer that

$$\left\|u_n(p_0)/\,\|u_n(p_0)\|_{1,\lambda}\right\|_{t,2,\lambda} < D$$

holds for every $n \ge N'$, where N' is a sufficiently large positive integer. Observe next that also

$$\left\|u_n(p_0)/\,\|u_n(p_0)\|_{1,\lambda} - p_0\right\|_{\infty} \to 0$$

holds as n tends to infinity by (22) and Part 1 of Proposition 3. Furthermore, $\inf_{x\in\mathbb{T}} p_0(x) > \zeta$ holds by continuity of p_0 on the compact metric space $\mathbb{T}$. Hence also $u_n(p_0)(x)/\|u_n(p_0)\|_{1,\lambda} > \zeta$ has to hold for every $x \in \mathbb{T}$ and for every $n \ge N''$, where N'' is a sufficiently large positive integer. This proves Parts 1 and 2 of the lemma upon setting

$$p_n^* = u_n(p_0)/\,\|u_n(p_0)\|_{1,\lambda}$$

for $n \ge N''' = \max(N, N', N'')$ and $p_n^* = p_{N'''}^*$ for $n < N'''$.

Finally, to prove the third part of the lemma, observe that

$$\begin{aligned}\|p_n^* - p_0\|_{s,2,\lambda} &= \left\|u_n(p_0)\,\|u_n(p_0)\|_{1,\lambda}^{-1} - u_n(p_0) + u_n(p_0) - p_0\right\|_{s,2,\lambda}\\ &\le \left|\|u_n(p_0)\|_{1,\lambda}^{-1} - 1\right| \|u_n(p_0)\|_{s,2,\lambda} + \|u_n(p_0) - p_0\|_{s,2,\lambda}\\ &= o(H(n)^{-t}) + o(H(n)^{s-t})\end{aligned}$$

holds for $s \le t$ by using (20) twice as well as the first inequality in (21), and noting $\|\cdot\|_{0,2,\lambda} \simeq \|\cdot\|_{2,\lambda}$. □

In the following proposition, we derive the rate of convergence of the sieved MLE in the L^2-norm by using results in van de Geer [16]. This provides, among others, the proof for Part 2 of Proposition 2. By a suitable interpolation inequality, these results also imply convergence rates in certain Sobolev norms, which will be of importance in the proof of Theorem 2 below.

Proposition 5. *Let the conditions of Proposition 2 hold. Then*

$$(23)\qquad \left\|p_n^{MLE} - p_0\right\|_{s,2,\lambda} = O_{\mathbb{P}^*}(n^{-(t-s)/(2t+1)}) + o(H(n)^{-(t-s)})$$

holds for $0 \le s \le t$. If $H(n) = \infty$ for every n, then the second term on the r.h.s. of (23) can be set to zero.

Proof. We apply Theorem 10.13 in van de Geer [16] with $\mathcal{P}_n^* = \mathcal{P}_{H(n)}(t, \zeta, D)$ and p_n^* as in Lemma 3. [If $H(n) = \infty$, choose $p_n^* = p_0$.] Expression 10.69 in van de Geer [16] is readily verified for p_n^* since $p_n^*(x) > \zeta > 0$ holds for every $x \in \mathbb{T}$ and since p_0 is bounded.

For two densities p, q, denote by $h^2(p,q) = \int_{\mathbb{T}} (p^{1/2} - q^{1/2})^2 d\lambda$ the usual Hellinger distance. By using the equality

$$\int_{\mathbb{T}} (\sqrt{p} - \sqrt{q})^2 d\lambda = \int_{\mathbb{T}} \left[(p-q)(\sqrt{p}+\sqrt{q})^{-1}\right]^2 d\lambda$$

we conclude that there exist universal constants $0 < c \le C < \infty$ so that

$$ch(p,q) \le \|p-q\|_{2,\lambda} \le Ch(p,q) \tag{24}$$

holds for $p, q \in \mathcal{P}(t, \zeta, D)$ with $t > 1/2$ and $\zeta > 0$ (cf. Part 2 of Proposition 3).

Now, by convexity of $\mathcal{P}_{H(n)}(t, \zeta, D)$, and since $p_n^* \in \mathcal{P}_{H(n)}(t, \zeta, D)$ holds for every n, the set

$$\tilde{\mathcal{P}}_{H(n)} = \{(p + p_n^*)/2 : p \in \mathcal{P}_{H(n)}(t, \zeta, D)\}$$

is contained in $\mathcal{P}_{H(n)}(t, \zeta, D)$ and hence norm-bounded in $\mathsf{W}_2^t(\mathbb{T})$. Consequently, using (24), the Hellinger-bracketing metric entropy $H_{[]}(\varepsilon, \bar{\mathcal{P}}_{H(n)}, \|\cdot\|_{2,\lambda})$ (with $\bar{\mathcal{P}}_{H(n)} = \{\sqrt{p} : p \in \tilde{\mathcal{P}}_{H(n)}\}$) can be bounded by the $\mathcal{L}^2(\mathbb{T}, \lambda)$-bracketing metric entropy of a bounded subset of $\mathsf{W}_2^t(\mathbb{T})$. The latter is of order $\varepsilon^{-1/t}$ (where $\varepsilon \to 0$ denotes bracket size) by using Part 2 of Corollary 2 in Nickl and Pötscher [13]. [This corollary can be applied with $\Omega = [0,1)$, $\beta = 0$, $d = 1$, $\mu = \lambda | [0,1)$, $r = p = q = 2$, upon noting that $W_2^s(\mathbb{R})$ coincides with the Besov space $B_{22}^s(\mathbb{R})$, and upon noting that $\mathsf{W}_2^t(\mathbb{T})$ can be identified with a bounded subset of the Banach space of restrictions of functions in $W_2^s(\mathbb{R})$ to $[0,1)$, cf. also Remark 3 in Nickl and Pötscher [13].] Theorem 10.13 in van de Geer [16] with $\Psi(\delta) = A\delta^{1-1/2t}$ for some suitable constant $0 < A < \infty$ now gives

$$h^2(p_n^{MLE}, p_0) = O_{\mathbb{P}}(n^{-2t/(2t+1)}) + h^2(p_n^*, p_0).$$

Using (24) and Part 3 of Lemma 3, we obtain

$$\begin{aligned} \left\|p_n^{MLE} - p_0\right\|_{2,\lambda} &= O_{\mathbb{P}}(n^{-t/(2t+1)}) + O(\|p_n^* - p_0\|_{2,\lambda}) \\ &= O_{\mathbb{P}}(n^{-t/(2t+1)}) + o(H(n)^{-t}). \end{aligned} \tag{25}$$

This proves (23) for the case $s = 0$.

The remaining cases $0 < s \le t$ follow from interpolation properties of Sobolev spaces: Expression 3.6.1/3 (and 3.5.1/13, 3.5.4/18,19) in Schmeisser and Triebel [14] implies the interpolation inequality

$$\|f\|_{s,2,\lambda} \le C \|f\|_{t,2,\lambda}^{s/t} \|f\|_{2,\lambda}^{(t-s)/t}$$

for $f \in \mathsf{W}_2^t(\mathbb{T})$ and $0 < C < \infty$. The result (23) now follows from (25) and the bound $\|p_n^{MLE} - p_0\|_{t,2,\lambda} \le 2D$ (as well as from $(a+b)^c \le 2^c(a^c + b^c)$ for $a, b, c > 0$). □

4.3. Proof of Theorem 2

Throughout Section 4.3, we shall assume that the conditions of Theorem 2 hold. In case $H(n) = \infty$ for every n, we shall use the convention that $Z_n = O(H(n)^{-\alpha})$ for some random element Z_n and $\alpha > 0$ implies $Z_n = 0$ for every n. Define the set

$$\mathcal{U}_{H(n),t,B} = \{f \in \langle e_k \rangle_{H(n)} : \quad \|f\|_{t,2,\lambda} \le B\},$$

with the convention that $\mathcal{U}_{\infty,t,B}$ equals the closed ball in $\mathsf{W}_2^t(\mathbb{T})$ of radius B. The following lemma is the key to prove Theorem 2.

Lemma 4. *We have that*

$$\left\|\sqrt{n}(\mathbb{P}_n^{MLE} - \mathbb{P}_n)\right\|_{\infty,\mathcal{U}_{H(n),t,B}} = O_{\mathbb{P}^*}(n^{-(t-k)/(2t+1)}) + O_{\mathbb{P}}(H(n)^{-(t-k)}) + o(n^{1/2}H(n)^{-2t})$$

for every $k > 1/2$ and $0 < B < \infty$.

Proof. **Step 1:**

By a point *internal* to $\mathcal{P}_{H(n)}(t, \zeta, D)$ we mean a probability density function $p \in \mathcal{P}_{H(n)}(t, \zeta, D)$ that satisfies $\|p\|_{t,2,\lambda} < D$ as well as $p(x) > \zeta$ for every $x \in \mathbb{T}$. Let p_n^* be the sequence from Lemma 3, which is an internal point for every n. If $H(n) = \infty$, choose $p_n^* = p_0$ for every n, which is also an internal point by Condition 1. Set

$$\mathcal{U}_{H(n),t,\eta,0} = \left\{ f \in \mathcal{U}_{H(n),t,\eta} : \int_{\mathbb{T}} f d\lambda = 0 \right\}$$

for $0 < \eta \leq D - \|p_n^*\|_{t,2,\lambda}$ which is possible since p_n^* is an internal point. Define for $0 < \varepsilon \leq 1$ the function

$$h_n^*(w) := (1-\varepsilon)\, p_n^{MLE} + \varepsilon p_n^* + \varepsilon w \qquad (w \in \mathcal{U}_{H(n),t,\eta,0}).$$

We now show that, for η small enough but positive,

$$\{h_n^*(w) : w \in \mathcal{U}_{t,\eta,0}\} \subseteq \mathcal{P}_{H(n)}(t, \zeta, D) \tag{26}$$

holds. To see this, observe the following three facts: First, by the triangle inequality

$$\begin{aligned}\|h_n^*(w)\|_{t,2,\lambda} &\leq (1-\varepsilon)\, D + \varepsilon\, \|p_n^*\|_{t,2,\lambda} + \varepsilon\eta \\ &\leq D\end{aligned}$$

holds for every $0 < \varepsilon \leq 1$, every n and every $w \in \mathcal{U}_{H(n),t,\eta,0}$ by definition of η. This verifies the Sobolev-norm condition for containment of $h_n^*(w)$ in $\mathcal{P}_{H(n)}(t, \zeta, D)$. Second, since $\|w\|_\infty \leq C_t\|w\|_{t,2,\lambda} \leq C_t\eta$ holds by Part 1 of Proposition 3 and since $\inf_{x\in\mathbb{T}} p_0(x) > \zeta$ holds by continuity of p_0 on the compact metric space $\mathbb{T}$, it follows that $p_n^*(x) + w(x) \geq \zeta + \beta - C_t\eta$ holds for some $\beta > 0$. This implies for $0 < \eta \leq \beta/C_t$ small enough that

$$\begin{aligned}(h_n^*(w))(x) &= (1-\varepsilon)\, p_n^{MLE}(x) + \varepsilon(p_n^* + w)(x) \\ &\geq \zeta\end{aligned}$$

holds for every $x \in \mathbb{T}$ and every $w \in \mathcal{U}_{H(n),t,\eta,0}$. Third, since w integrates to zero, $h_n^*(w)$ is a density for every $w \in \mathcal{U}_{H(n),t,\eta,0}$.

Consequently, in view of (26), since p_n^{MLE} is a maximizer of $L_n(\cdot)$ over $\mathcal{P}_{H(n)}(t, \zeta, D)$, and since $L_n(\cdot)$ is Fréchet differentiable at p_n^{MLE} by Proposition 4, the derivative of $L_n(\cdot)$ at p_n^{MLE} in the direction of $h_n^*(w) = p_n^{MLE} + \varepsilon(w - p_n^{MLE} + p_n^*)$, every $w \in \mathcal{U}_{H(n),t,\eta,0}$, has to be nonpositive, that is, we have that,

$$DL_n\left(p_n^{MLE}\right)\left(w - p_n^{MLE} + p_n^*\right) \leq 0 \qquad \text{for every } w \in \mathcal{U}_{t,\eta,0}.$$

Since $\mathcal{U}_{t,\eta,0}$ also contains $-w$, we conclude that

$$|DL_n\left(p_n^{MLE}\right)(w)| \leq DL_n\left(p_n^{MLE}\right)\left(p_n^{MLE} - p_n^*\right) \qquad \text{for every } w \in \mathcal{U}_{t,\eta,0} \tag{27}$$

holds.

Step 2:

Define the operator

$$\Pi(f) = (f - \mathbb{P}f)p_0 \tag{28}$$

from $\mathsf{L}^\infty(\mathbb{T})$ into $\mathsf{L}^\infty(\mathbb{T}) \cap \{g : \int_{\mathbb{T}} g\,d\lambda = 0\}$. Note that $\mathsf{W}_2^t(\mathbb{T})$ is a multiplication algebra for $t > 1/2$, that is $\|fg\|_{t,2,\lambda} \leq M\|f\|_{t,2,\lambda}\|g\|_{t,2,\lambda}$ holds for some $0 < M < \infty$. [This is easily seen to follow from the fact that $\mathsf{W}_2^t(\mathbb{R})$ is a multiplication algebra and upon noting that $\mathsf{W}_2^t(\mathbb{T})$ can be identified with a bounded subset of the (factor) Banach space of restrictions of functions in $\mathsf{W}_2^s(\mathbb{R})$ to $[0, 1)$.] Hence, for every $f \in \mathcal{U}_{H(n),t,B}$, we have

$$\begin{aligned} \|\Pi(f)\|_{t,2,\lambda} &\leq M \|f - \mathbb{P}f\|_{t,2,\lambda} \|p_0\|_{t,2,\lambda} \\ &\leq MD(\|f\|_{t,2,\lambda} + \|\mathbb{P}f\|_{t,2,\lambda}) \\ &\leq MD(B + \|C_t B\|_{t,2,\lambda}) \\ &\leq MDB(1 + C_t) < \infty \end{aligned} \tag{29}$$

by using Proposition 3. Now with η as in Step 1 define

$$s(\Pi(f)) = \eta \|\Pi(f)\|_{t,2,\lambda}^{-1} \Pi(f)$$

if $\Pi(f) \neq 0$, and set $s(\Pi(f)) = 0$ otherwise. Then it follows that $s(\Pi(f)) \in \mathcal{U}_{H(n),t,\eta,0}$ for every $f \in \mathcal{U}_{H(n),t,B}$.

Step 3:

Inserting $s(\Pi(f))$ for w in (27) we obtain that

$$\left|DL_n(p_n^{MLE})(s(\Pi(f)))\right| \leq DL_n(p_n^{MLE})(p_n^{MLE} - p_n^*) \quad \text{for every } f \in \mathcal{U}_{t,B} \tag{30}$$

holds. Using Proposition 4, we see that the expected value of the likelihood derivative at p_0 equals zero along the directions $\{g : \int_{\mathbb{T}} g\,d\lambda = 0\}$ and thus in particular along the direction $p_n^{MLE} - p_n^*$:

$$\mathbb{P}DL\,(p_0)\,(p_n^{MLE} - p_n^*) = \int_{\mathbb{T}} \left(p_n^{MLE} - p_n^*\right) p_0^{-1} d\mathbb{P} = \left\|p_n^{MLE}\right\|_{1,\lambda} - \|p_n^*\|_{1,\lambda} = 0 \tag{31}$$

since both p_n^{MLE} and p_n^* are probability densities. Thus, we have from (30) that

$$DL_n\left(p_n^{MLE}\right)(s(\Pi(f))) = (DL_n\left(p_n^{MLE}\right) - \mathbb{P}DL\,(p_0))(p_n^{MLE} - p_n^*)$$

holds for every $f \in \mathcal{U}_{H(n),t,B}$. Let now k be as in the lemma. W.l.o.g. we may restrict ourselves to the case $k \leq t$. Choose a real j, $1/2 < j < k$. Let $\mathcal{U}_{j,1}$ denote the unit ball of $\mathsf{W}_2^j(\mathbb{T})$ which is a universal Donsker class. By Proposition 4 we obtain

$$\begin{aligned}
&\left|\left(DL_n\left(p_n^{MLE}\right)-\mathbb{P}DL\left(p_0\right)\right)\left(p_n^{MLE}-p_n^*\right)\right| \\
&\leq \left|\left(DL_n\left(p_n^{MLE}\right)-\mathbb{P}DL\left(p_n^{MLE}\right)\right)\left(p_n^{MLE}-p_0\right)\right|+ \\
&\left|\left(\mathbb{P}DL\left(p_n^{MLE}\right)-\mathbb{P}DL\left(p_0\right)\right)\left(p_n^{MLE}-p_0\right)\right|+ \\
&\left|\left(DL_n\left(p_n^{MLE}\right)-\mathbb{P}DL\left(p_n^{MLE}\right)\right)\left(p_n^*-p_0\right)\right|+ \\
&\left|\left(\mathbb{P}DL\left(p_n^{MLE}\right)-\mathbb{P}DL\left(p_0\right)\right)\left(p_n^*-p_0\right)\right| \\
&\leq \sup_{p\in\mathcal{P}(t,\zeta,D)}\left\|DL_n(p)-\mathbb{P}DL(p)\right\|_{\infty,\mathcal{U}_j}\left\|p_n^{MLE}-p_0\right\|_{j,2,\lambda} \\
&+\zeta^{-1}\left\|p_n^{MLE}-p_0\right\|_{2,\lambda}^2 \\
&+\sup_{p\in\mathcal{P}(t,\zeta,D)}\left\|DL_n(p)-\mathbb{P}DL(p)\right\|_{\infty,\mathcal{U}_j}\left\|p_n^*-p_0\right\|_{j,2,\lambda} \\
&+\zeta^{-1}\left\|p_n^*-p_0\right\|_{2,\lambda}^2 \\
&=: Z_n'.
\end{aligned}$$

It follows from Lemma 1, Part 3 of Lemma 3 and Proposition 5 that

$$\begin{aligned}
(32)\quad Z_n' &= O_{\mathbb{P}^*}(n^{-1/2-(t-j)/(2t+1)})+o_{\mathbb{P}}(n^{-1/2}H(n)^{-(t-j)}) \\
&\quad +O_{\mathbb{P}}(n^{-2t/(2t+1)})+o(H(n)^{-2t}) \\
&\quad +O_{\mathbb{P}}(n^{-1/2})o(H(n)^{-(t-j)})+o(H(n)^{-2t}) \\
&= O_{\mathbb{P}^*}(n^{-1/2-(t-k)/(2t+1)})+O_{\mathbb{P}}(n^{-1/2}H(n)^{-(t-k)})+o(H(n)^{-2t}).
\end{aligned}$$

Multiplying by $\eta^{-1}\|\Pi(f)\|_{t,2,\lambda}$ we conclude that

$$\left|DL_n\left(p_n^{MLE}\right)(\Pi(f))\right| \leq \eta^{-1}\left\|\Pi(f)\right\|_{t,2,\lambda} Z_n' \qquad \text{for every } f\in\mathcal{U}_{H(n),t,B}$$

holds. Since Z_n' does not depend on $f\in\mathcal{U}_{H(n),t,B}$, we conclude from (29) that

$$(33)\qquad \sup_{f\in\mathcal{U}_{H(n),t,B}}\left|DL_n\left(p_n^{MLE}\right)(\Pi(f))\right| \leq Z_n$$

holds, where Z_n is of the same order as Z_n' in (32).

Step 4:

Expression (33) now allows to quantify the difference between the empirical process $\sqrt{n}(\mathbb{P}_n-\mathbb{P})$ and the MLE-process $\sqrt{n}(\mathbb{P}_n^{MLE}-\mathbb{P})$ on the class $\mathcal{U}_{H(n),t,B}$. We have

$$\begin{aligned}
&\left\|\sqrt{n}(\mathbb{P}_n^{MLE}-\mathbb{P}_n)\right\|_{\infty,\mathcal{U}_{H(n),t,B}} \\
&= \sqrt{n}\sup_{f\in\mathcal{U}_{H(n),t,B}}\left|(\mathbb{P}_n^{MLE}-\mathbb{P})f-(\mathbb{P}_n-\mathbb{P})f\right| \\
&= \sqrt{n}\sup_{f\in\mathcal{U}_{H(n),t,B}}\left|\int_{\mathbb{T}}(f-\int_{\mathbb{T}}fd\mathbb{P})\left(p_n^{MLE}-p_0\right)d\lambda-(\mathbb{P}_n-\mathbb{P})f\right|.
\end{aligned}$$

Observe furthermore that

$$-\mathbb{P}D^2L\left(p_0\right)\left(p_n^{MLE}-p_0,g\right)=\int_{\mathbb{T}}gp_0^{-1}\left(p_n^{MLE}-p_0\right)d\lambda$$

holds for $g \in \mathsf{L}^\infty(\mathbb{T})$ by Proposition 4, and hence – recalling definition (28) – we obtain

$$\text{(34)} \qquad \left\|\sqrt{n}(\mathbb{P}_n^{MLE} - \mathbb{P}_n)\right\|_{\infty,\mathcal{U}_{H(n),t,B}} = \sqrt{n} \sup_{f\in\mathcal{U}_{H(n),t,B}} \left|-\mathbb{P}D^2L\left(p_0\right)\left(p_n^{MLE} - p_0, \Pi(f)\right) - DL_n\left(p_0\right)\left(\Pi(f)\right)\right|.$$

Next, by the mean value theorem, we have

$$DL_n\left(p_n^{MLE}\right)\left(\Pi(f)\right) = DL_n\left(p_0\right)\left(\Pi(f)\right) + D^2L_n\left(\bar{p}_n\right)\left(p_n^{MLE} - p_0, \Pi(f)\right)$$

where the mean value $\bar{p}_n \equiv \bar{p}_n(f)$ lies, for every $f \in \mathcal{U}_{H(n),t,B}$, on the line segment between p_n^{MLE} and p_0, which is contained in $\mathcal{P}(t,\zeta,D)$ for every $f \in \mathcal{U}_{H(n),t,B}$. This gives

$$\text{(35)} \qquad DL_n\left(p_n^{MLE}\right)\left(\Pi(f)\right) - \mathbb{P}D^2L\left(p_0\right)\left(p_n^{MLE} - p_0, \Pi(f)\right) = DL_n\left(p_0\right)\left(\Pi(f)\right) + \left(D^2L_n\left(\bar{p}_n\right) - \mathbb{P}D^2L\left(p_0\right)\right)\left(p_n^{MLE} - p_0, \Pi(f)\right).$$

Note now that the set $\Pi(\mathcal{U}_{H(n),t,B}) = \{\Pi(f) = (f - \mathbb{P}f)p_0 : f \in \mathcal{U}_{H(n),t,B}\}$ is a $\mathbb{P}$-Donsker class by (29). Let k be as in Step 3. Again, w.l.o.g. we may restrict ourselves to the case $k \le t$. Choose a real j, $1/2 < j < k$. Let $\mathcal{U}_{j,1}$ denote the unit ball of $\mathsf{W}_2^j(\mathbb{T})$ which is a $\mathbb{P}$-Donsker class. Using Lemma 1, Propositions 3, 4, 5 as well as (29), we obtain

$$\begin{aligned}
&\sup_{f\in\mathcal{U}_{H(n),t,B}} \left|\left(D^2L_n\left(\bar{p}_n\right) - \mathbb{P}D^2L\left(p_0\right)\right)\left(p_n^{MLE} - p_0, \Pi(f)\right)\right| \\
\le &\sup_{f\in\mathcal{U}_{H(n),t,B}} \left|\left(D^2L_n\left(\bar{p}_n\right) - \mathbb{P}D^2L\left(\bar{p}_n\right)\right)\left(p_n^{MLE} - p_0, \Pi(f)\right)\right| + \\
&\sup_{f\in\mathcal{U}_{H(n),t,B}} \left|\left(\mathbb{P}D^2L\left(\bar{p}_n\right) - \mathbb{P}D^2L\left(p_0\right)\right)\left(p_n^{MLE} - p_0, \Pi(f)\right)\right| \\
\le &\sup_{p\in\mathcal{P}(t,\zeta,D)} \left\|D^2L_n\left(p\right) - \mathbb{P}D^2L\left(p\right)\right\|_{\infty,\mathcal{U}_{j,1},\Pi(\mathcal{U}_{H(n),t,B})} \left\|p_n^{MLE} - p_0\right\|_{j,2,\lambda} + \\
&2\zeta^{-3}C_tD \sup_{f\in\mathcal{U}_{H(n),t,B}} \left\|\Pi(f)\right\|_\infty \left\|p_n^{MLE} - p_0\right\|_{2,\lambda} \left\|\bar{p}_n - p_0\right\|_{2,\lambda} \\
=: &\, W_n
\end{aligned}$$

where again

$$\text{(36)} \qquad W_n = O_{\mathbb{P}^*}(n^{-1/2-(t-k)/(2t+1)}) + O_{\mathbb{P}}(n^{-1/2}H(n)^{-(t-k)}) + o(H(n)^{-2t}).$$

Here we have used the simple fact that

$$\left\|\bar{p}_n - p_0\right\|_{2,\lambda} = \left\|\xi(f)p_n^{MLE} + (1 - \xi(f))p_0 - p_0\right\|_{2,\lambda} = \xi(f)\left\|p_n^{MLE} - p_0\right\|_{2,\lambda}$$

for some $0 \le \xi(f) \le 1$. This together with (33) and (35) gives

$$\sup_{f\in\mathcal{U}_{H(n),t,B}} \left|-\mathbb{P}D^2L\left(p_0\right)\left(p_n^{MLE} - p_0, \Pi(f)\right) - DL_n\left(p_0\right)\left(\Pi(f)\right)\right| \le W_n + Z_n.$$

Inserting this into (34) shows that

$$\left\|\sqrt{n}(\mathbb{P}_n^{MLE} - \mathbb{P}_n)\right\|_{\infty,\mathcal{U}_{H(n),t,B}} \le \sqrt{n}(W_n + Z_n)$$

holds. The proof of the lemma is now complete after collecting the terms in (32) and (36). □

We note that Lemma 4 parallels Theorem 1 in Nickl [12] in the sense that the set $\mathcal{U}_{H(n),t,B}$ over which $\mathbb{P}_n^{MLE}$ and $\mathbb{P}_n$ are asymptotically close depends on the set $\mathcal{P}_{H(n)}(t,\zeta,D)$ over which the likelihood is maximized (via t and $H(n)$). [In Nickl [12], there was no dependence on $H(n)$ since no sieve was used.]

Proof of Theorem 2. Define $\hat{\nu}_n = \sqrt{n}(\mathbb{P}_n^{MLE} - \mathbb{P})$. We set w.l.o.g. $\mathcal{F}$ equal to $\mathcal{U}_{s,1}$ the unit ball of $\mathsf{W}_2^s(\mathbb{T})$. For $f \in \mathcal{U}_{s,1}$, its Fourier series truncated at $K(n)$ will be denoted by $u_n(f) = \sum_{k\in\mathbb{Z}\cap[-K(n),K(n)]} Ff(k)e_k$, where the sequence $K(n)$ of positive integers is chosen such that $K(n) \le H(n)$ holds for every n. We start from

$$\begin{aligned}\sqrt{n}\left\|\mathbb{P}_n^{MLE} - \mathbb{P}_n\right\|_{\infty,\mathcal{U}_{s,1}} &= \|\hat{\nu}_n - \nu_n\|_{\infty,\mathcal{U}_{s,1}} \\ &\le \sup_{f\in\mathcal{U}_{s,1}} |\hat{\nu}_n(f - u_n(f))| \\ &\quad + \sup_{f\in\mathcal{U}_{s,1}} |\hat{\nu}_n(u_n(f)) - \nu_n(u_n(f))| \\ &\quad + \sup_{f\in\mathcal{U}_{s,1}} |\nu_n(u_n(f) - f)| \\ &= I + II + III.\end{aligned}$$

We now obtain bounds for these three expressions to prove the theorem.

Bound for I: Observe first that

$$\begin{aligned}\sup_{f\in\mathcal{U}_{s,1}} \|u_n(f) - f\|_{2,\lambda}^2 &\le \sup_{k\in\mathbb{Z}\setminus[-K(n),K(n)]} \langle k\rangle^{-2s} \sup_{f\in\mathcal{U}_{s,1}} \sum_{k\in\mathbb{Z}\setminus[-K(n),K(n)]} \langle k\rangle^{2s} |Ff(k)|^2 \\ &= O(K(n)^{-2s})o(1)\end{aligned}$$

holds by the same reasoning as in (20). Hence, by Proposition 5 and Cauchy-Schwarz's inequality, the first term can be bounded by

$$\begin{aligned}&n^{1/2} \sup_{f\in\mathcal{U}_{s,1}} \left|\int_{\mathbb{T}} (p_n^{MLE} - p_0)(u_n(f) - f)d\lambda\right| \\ &\le n^{1/2}\left\|p_n^{MLE} - p_0\right\|_{2,\lambda} \sup_{f\in\mathcal{U}_{s,1}} \|u_n(f) - f\|_{2,\lambda} \\ &= n^{1/2}(o_{\mathbb{P}}(n^{-t/(2t+1)}K(n)^{-s}) + o(H(n)^{-t}K(n)^{-s})).\end{aligned}$$

Bound for II: We first treat the case $s < t$. Observe that

$$\begin{aligned}\sup_{f\in\mathcal{U}_{s,1}} \|u_n(f)\|_{t,2,\lambda}^2 &= \sup_{f\in\mathcal{U}_{s,1}} \sum_{k\in\mathbb{Z}\cap[-K(n),K(n)]} |Ff(k)|^2 \langle k\rangle^{2s} \langle k\rangle^{2t-2s} \\ &\le \sup_{k\in\mathbb{Z}\cap[-K(n),K(n)]} \langle k\rangle^{2t-2s} \sup_{f\in\mathcal{U}_{s,1}} \|f\|_{s,2,\lambda}^2 = O(K(n)^{2t-2s})\end{aligned}$$

holds by Hölder's inequality. Consequently, the term II equals zero in the trivial case where $u_n(f) = 0$ for every $f \in \mathcal{F}$ and otherwise is equal to

$$\begin{aligned}&\sup_{\substack{f\in\mathcal{U}_{s,1},\\ u_n(f)\neq 0}} \|u_n(f)\|_{t,2,\lambda} \left|(\hat{\nu}_n - \nu_n)(u_n(f)/\|u_n(f)\|_{t,2,\lambda})\right| \\ &= O_{\mathbb{P}^*}(n^{-(t-k)/(2t+1)}K(n)^{t-s}) + O_{\mathbb{P}}(H(n)^{-(t-k)}K(n)^{t-s}) \\ &\quad + o(n^{1/2}H(n)^{-2t}K(n)^{t-s})\end{aligned} \tag{37}$$

for every $k > 1/2$ by Lemma 4, since $u_n(f)/\|u_n(f)\|_{t,2,\lambda}$ is contained, for every n, in the set $\mathcal{U}_{H(n),t,1}$ (recalling that $K(n) \le H(n)$ holds for every n). If $s \ge t$, then

$\sup_f \|u_n(f)\|_{t,2,\lambda} \le \sup_f \|f\|_{t,2,\lambda} = O(1)$ holds and we obtain that the expression in (37) is of order $O_{\mathbb{P}^*}(n^{-(t-k)/(2t+1)}) + O_{\mathbb{P}}(H(n)^{-(t-k)}) + o(n^{1/2}H(n)^{-2t})$ by Lemma 4.

Bound for III: Finally, the third term is equal to

$$\sqrt{n} \sup_{f \in \mathcal{U}_{s,1}} |(\mathbb{P}_n - \mathbb{P})(u_n(f) - f)| \le \sqrt{n} \|\mathbb{P}_n - \mathbb{P}\|_{\infty, \mathcal{U}_{j,1}} \sup_{f \in \mathcal{U}_{s,1}} \|u_n(f) - f\|_{j,2,\lambda}$$
$$= o_{\mathbb{P}}(K(n)^{-(s-j)})$$

for every $j > s > 1/2$ by the same reasoning as in (20) above and since $\mathcal{U}_{j,1}$ is a universal Donsker class.

Collecting terms, we obtain for $s < t$ that

$$\begin{aligned}\sqrt{n} \left\|\mathbb{P}_n^{MLE} - \mathbb{P}_n\right\|_{\infty, \mathcal{U}_{s,1}} &= o_{\mathbb{P}}(n^{1/2 - t/(2t+1)} K(n)^{-s}) + o(n^{1/2} H(n)^{-t} K(n)^{-s}) \\ &\quad + O_{\mathbb{P}^*}(n^{-(t-k)/(2t+1)} K(n)^{t-s}) \\ &\quad + O(H(n)^{-(t-k)} K(n)^{t-s}) \\ &\quad + o(n^{1/2} H(n)^{-2t} K(n)^{t-s}) + o_{\mathbb{P}}(K(n)^{-(s-k)})\end{aligned}$$

holds for every $k > s > 1/2$.

If $s \ge t$ we arrive at

$$\begin{aligned}\sqrt{n} \left\|\mathbb{P}_n^{MLE} - \mathbb{P}_n\right\|_{\infty, \mathcal{U}_{s,1}} &= o_{\mathbb{P}}(n^{1/2 - t/(2t+1)} K(n)^{-s}) + o(n^{1/2} H(n)^{-t} K(n)^{-s}) \\ &\quad + O_{\mathbb{P}^*}(n^{-(t-k)/(2t+1)}) + O_{\mathbb{P}}(H(n)^{-(t-k)}) \\ &\quad + o(n^{1/2} H(n)^{-2t}) + o_{\mathbb{P}}(K(n)^{-(s-k)})\end{aligned}$$

again for every $k > 1/2$. This completes the proof of the first claim of the theorem. The second claim follows immediately from the fact that $\mathcal{F}$ is a universal Donsker class above. □

Acknowledgement

The author would like to thank B. M. Pötscher for helpful comments on subjects related to this article.

References

[1] Bickel, J. P. and Ritov, Y. (2003). Nonparametric estimators which can be 'plugged-in'. *Ann. Statist.* **31** 1033–1053.

[2] Birgé, L. and Massart, P. (1993). Rates of convergence of minimum contrast estimators. *Prob. Theory Related Fields* **97** 113–150.

[3] Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves: Exponential bounds and rates of convergence. *Bernoulli* **4** 329–375.

[4] Devroye, L. and Lugosi, G. (2001). *Combinatorial Methods in Density Estimation.* Springer, New York.

[5] Dudley, R. M. (1999). *Uniform Central Limit Theorems.* Cambridge University Press, Cambridge, England.

[6] Folland, G. B. (1999). *Real Analysis.* Wiley, New York.

[7] Giné, E. (1975). Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev-norms. *Ann. Statist.* **3** 1243–1266.

[8] Giné, E. and Nickl, R. (2008). Uniform central limit theorems for kernel density estimators. *Probab. Theory Related Fields* **141** 333–387.

[9] Giné, E. and Nickl, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probab. Theory Related Fields* **143** 569–596.

[10] Giné, E. and Nickl, R. (2009). Uniform limit theorems for wavelet density estimators. *Ann. Probab.* To appear.

[11] Marcus, D. J. (1985). Relationships between Donsker classes and Sobolev spaces. *Z. Wahrsch. Verw. Gebiete* **69** 323–330.

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

[13] Nickl, R. and Pötscher, B. M. (2007). Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov- and Sobolev-type. *J. Theoret. Probab.* **20** 177–199.

[14] Schmeisser, H.-J. and Triebel, H. (1987). *Topics in Fourier Analysis and Function Spaces.* Wiley, New York.

[15] van de Geer, S. (1995). The method of sieves and minimum contrast estimators. *Math. Methods Statist.* **4** 20–38.

[16] van de Geer, S. (2000). *Empirical Processes in M-Estimation.* Cambridge University Press, Cambridge, England.

[17] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

[18] Wong, W. H. and Shen, X. (1995). Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *Ann. Statist.* **23** 339–362.