IMS Collections
High Dimensional Probability V: The Luminy Volume
Vol. 5 (2009) 84–102

DOI: 10.1214/09-IMSCOLL507

Uniform Central Limit Theorems for pregaussian classes of functions

Dragan Radulović[1] **and Marten Wegkamp**[2,*]

Florida Atlantic University and Florida State University

Abstract: We study weak convergence of general (smoothed) empirical processes indexed by classes of functions $\mathcal{F}$ under minimal conditions. We present a general result that, applied to specific situations, enables us to prove uniform central limit theorems under P-pregaussian assumption on $\mathcal{F}$ only.

1. Introduction

Let $X_1, X_2 \ldots$ be a sequence of i.i.d. random variables with common probability measure P on $\mathbb{R}$, and let $\mathcal{F}$ be a class of functions. $\mathcal{F}$ is called a P-Donsker class if the empirical process

$$\mathbb{G}_n(f) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{f(X_i) - \mathbb{E}f(X_i)\}, \ f \in \mathcal{F} \tag{1}$$

converges weakly to a tight Gaussian limit in $\ell^\infty(\mathcal{F})$, the space of all bounded functions on $\mathcal{F}$. A necessary condition is that $\mathcal{F}$ is P-pregaussian: there exists a tight Gaussian process in $\ell^\infty(\mathcal{F})$ that is indexed by $\mathcal{F}$, see [3, 5, 12]. Of course a P-Donsker class is P-pregaussian, but the converse is not true. There exist Gaussian processes that cannot be obtained as a limit of $\{\mathbb{G}_n(f),\ f \in \mathcal{F}\}$ as pointed out by ([4, Section 3] (see also [1, page 178 and Section 3.8], [5, Example 7.5], [7] and [10]). The obvious mechanism behind this defficiency is the discrepancy between the discrete measure P_n and underlying probability measure P, that is not necessarily discrete. However, if P is a discrete measure satisfying some regularity conditions, the Borisov-Durst result [3, page 244] states that $\{\mathbb{G}_n(f),\ f \in \mathcal{F}\}$ converges weakly to a tight Gaussian process for all P-pregaussian classes $\mathcal{F}$. On the other hand, [9, 10] demonstrated that the same phenomenon holds under certain regularity conditions if P has a density $p(x)$ for the smoothed empirical process

$$\widehat{\mathbb{G}}_n(f) = \sqrt{n} \int_{\mathbb{R}} f(x)\{\widehat{p}_n(x) - p(x)\}\, dx,\ f \in \mathcal{F}$$

where $\widehat{p}_n(x)$ is a histogram or kernel density estimator of $p(x)$. Although the results presented in [9, 10] constitute the main principle, they are limited in scope. Theorem 1 in Section 2 expands on this idea by considering the more general setting

*Research is supported in part by NSF grant DMS 0706829.
[1]Department of Mathematics, Florida Atlantic University, Boca Raton, FL 33431-0991.
[2]Department of Statistics, Florida State University, Tallahassee, FL 32306-4330.
Keywords and phrases: kernel density estimator, Fourier series density estimator, empirical processes, uniform central limit theorems, weak convergence.

where

$$\widehat{\mathbb{G}}_n(f) = \sqrt{n} \int_{\mathbb{R}} f(x)\, d\{\widehat{P}_n(x) - P(x)\}, \ f \in \mathcal{F} \tag{2}$$

based on some estimator $\widehat{P}_n$ of P. Our approach differs from the recent paper [6] in that we explicitly use closeness of the estimator $\widehat{P}_n$ to P. This enables us to avoid imposing conditions on $\mathcal{F}$.

We apply Theorem 1 to Fourier and kernel type density estimators. For example, using a Fourier series density estimator $\widehat{p}_n(x)$, Theorem 3 below establishes weak convergence of (2) under very general conditions. There is a trade-off between the smoothness of density $p(x)$ and the class $\mathcal{F}$, see [6, Section 4]. Loosely speaking, if $p(x)$ is a bit more than twice differentiable and is bounded away from zero and infinity, then the only requirement on the bounded class $\mathcal{F}$ is that it is P-pregaussian, which nicely matches Borisov-Durst analog for discrete measures. Theorem 5 generalizes the case for densities that are not necessarily bounded away from zero.

As for kernel density estimators, [11, 13] require further conditions on $\mathcal{F}$, and [9, 10] show that $\{\widehat{\mathbb{G}}_n(f),\ f \in \mathcal{F}\}$ converges weakly under smoothness conditions on $p(x)$ only. Theorem 10 in Section 4 generalizes this result for densities with possibly unbounded support and not necessarily bounded away from zero.

The paper is organized as follows: Section 2 describes a general strategy for establishing weak convergence of $\{\widehat{\mathbb{G}}_n(f)\, f \in \mathcal{F}\}$ and presents a general result in Theorem 1. Sections 2 and 3 specialize Theorem 1 to Fourier series density estimators and kernel density estimators, respectively. Proofs of the results can be found at the end of each section.

2. General result

Given a P-pregaussian class $\mathcal{F}$ of functions, we are interested in weak convergence in $\ell^\infty(\mathcal{F})$ of the smoothed empirical process $\{\widehat{\mathbb{G}}_n(f),\ f \in \mathcal{F}\}$ defined in (2) to a limiting Gaussian process. For this we need that

- the (finite dimensional) vectors $(\widehat{\mathbb{G}}_n(f_1), \ldots, \widehat{\mathbb{G}}_n(f_k))$ converge weakly to a multivariate normal distribution for all $f_1, \ldots, f_k \in \mathcal{F}$, $k = 1, 2, \ldots$ and
- $\widehat{\mathbb{G}}_n(f)$ is stochastically equicontinuous with respect to the $L^2(P)$ semi-metric.

Let $\widehat{P}_n$ be an estimator of P such that $\widehat{P}_n f = \int f(x)\, d\widehat{P}_n(x)$ can be written as

$$\int_{\mathbb{R}} f(x)\, d\widehat{P}_n(x) = \frac{1}{n} \sum_{i=1}^{n} Y_{n,i}(f), \tag{3}$$

where $Y_{n,1}(f), Y_{n,2}(f), \ldots$ are i.i.d. random variables, linear in their argument $f \in \mathcal{F}$. In case that the bias of $\{\widehat{\mathbb{G}}_n(f),\ f \in \mathcal{F}\}$ is asymptotically negligible, we only need to consider the centered process

$$\widehat{\mathbb{G}}_n^0(f) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{Y_{n,i}(f) - \mathbb{E}Y_{n,i}(f)\}, \ f \in \mathcal{F}. \tag{4}$$

The following result is our basic tool to establish weak convergence. It modifies results obtained in [5] and parallels Theorem 3 in [6].

Theorem 1. *Consider the centered empirical process $\{\widehat{\mathbb{G}}_n^0(f) : f \in \mathcal{F}\}$ based on the general estimator $\widehat{P}_n$ defined in (3). Assume that*

(A1) $\mathbb{G}_n^0(f) \to N(0, \sigma_f^2)$ *in distribution, as* $n \to \infty$, *with* $\sigma_f^2 = Pf^2 - (Pf)^2$.

(A2) there exists $c \geq 1$ *such that, for all* $\delta > 0$,

$$\limsup_{n\to\infty} \sup_{f\in\mathcal{F}:\ K_n n^{-1/2} \leq Pf^2 \leq \delta} \frac{\mathbb{E}[Y_{n,1}^2(f)]}{Pf^2} \leq c$$

where $K_n \geq |Y_{n,i}(f)|$ *for all* i *and* $f \in \mathcal{F}$ *and* $K_n n^{-1/2} \to 0$.

(A3)

$$\mathbb{E}\left[\sup_{Pf^2 \leq K_n^2 n^{-1/2}} |\widehat{\mathbb{G}}_n^0(f)|\right] \to 0$$

as $n \to \infty$.

Then $\{\widehat{\mathbb{G}}_n^0(f), f \in \mathcal{F}\}$ *converges weakly in* $\ell^\infty(\mathcal{F})$ *to a tight Gaussian process* $\mathbb{G}_P$ *with covariance structure*

$$\mathrm{Cov}(\mathbb{G}_P(f), \mathbb{G}_P(g)) = \mathbb{E}[f(X)g(X)] - \mathbb{E}[f(X)]\mathbb{E}[g(X)].$$

Theorem 1 significantly simplifies the problem of stochastic equicontinuity of $\{\widehat{\mathbb{G}}_n(f),\ f \in \mathcal{F}\}$ since it only requires that the supremum in (A3) is taken over balls $P(f-g)^2 \leq K_n n^{-1/2}$ instead of $P(f-g)^2 \leq \delta$. Controling the supremum still remains a formidable task and in general one needs to impose some additional assumptions on the class of the functions $\mathcal{F}$. This is the very source of various requirements one finds in the literature: Bracketing classes, VC classes, VC -hull classes, etc.

The main advantage of this theorem becomes apparent once it is applied to the situations where both the estimator $\widehat{P}_n$ as well as the measure P are of the same type. For example, a simple "two-line" proof that invokes Theorem 1 recovers the Borisov-Durst type result for discrete P. To appreciate why, consider a discrete measure P defined on the integers by the probabilities $p_k = P\{X = k\}$ and $p_k \geq p_{k+1}$ (possibly after rearranging the original atoms) and $\sum_k \sqrt{p_k} < \infty$, and let $p_{n,k} = n^{-1}\sum_{i=1}^n I_{\{X_i = k\}}$ for $k \in \mathbb{N}$. In order to establish weak convergence of the empirical process $\{\mathbb{G}_n(f),\ f \in \mathcal{F}\}$ for a P-pregaussian, uniformly bounded class of functions $\mathcal{F}$ ($\|f\|_\infty \leq 1$ for all f), we check assumptions (A1), (A2) and (A3). While (A1) and (A2) are trivially met, it remains then to verify (A3). First we observe that

$$\mathbb{E}\left[\sup_{Pf^2 \leq n^{-1/2}} \sqrt{n}\left|\sum_{k=1}^{\infty} f(k)(p_{n,k} - p_k)\right|\right] \leq$$
$$\mathbb{E}\left[\sup_{Pf^2 \leq n^{-1/2}} \sqrt{n}\left|\sum_{k=1}^{M_n} f(k)(p_{n,k} - p_k)\right|\right] + 2\sqrt{n}\sum_{k=M_n}^{\infty} p_k$$

The first part on the right, after multplying and dividing by $\sqrt{p_k}$ and applying the Cauchy Schwarz inequality, is bounded by

$$\left\{\sup_{Pf^2 \leq n^{-1/2}} \sqrt{n}\sum_{k=1}^{M_n} f^2(k)p_k\right\}^{1/2} \left\{\sqrt{n}\sum_{k=1}^{M_n} \frac{\mathbb{E}\left[(p_{n,k} - p_k)^2\right]}{p_k}\right\}^{1/2} \leq \sqrt{\frac{M_n}{n^{1/2}}}.$$

The result now follows easily since the condition $\sum \sqrt{p_k} < \infty$ implies that we can find a sequence M_n such that both $\sqrt{n}\sum_{k=M_n}^{\infty} p_k$ as well as $n^{-1/2}M_n$ converge to zero. This little computation is simple and fully exploits the fact that we used a discrete measure P_n to estimate the underlying discrete measure P.

The next step is to apply this idea to absolutely continuous measures P. We will consider the density estimators $\widehat{p}_n(x)$ that possess the following property:

$$\int_{\mathbb{R}} f(x)\widehat{p}_n(x)\,dx = \frac{1}{n}\sum_{i=1}^{n} Y_{n,i}(f), \tag{5}$$

where $Y_{n,i}(f)$ satisfies the above assumptions. The following corollary formalizes the above computation and applies it to the smoothed empirical process. Every nonparametric density estimator $\widehat{p}_n(x)$ has some inevitable bias $p(x) - \bar{p}_n(x)$, where

$$\bar{p}_n(x) = \mathbb{E}[\widehat{p}_n(x)]$$

and the centered process $\{\widehat{\mathbb{G}}_n^0(f),\ f \in \mathcal{F}\}$ is defined by

$$\widehat{\mathbb{G}}_n^0(f) = \sqrt{n}\int_{\mathbb{R}} f(x)\{\widehat{p}_n(x) - \bar{p}_n(x)\}\,dx = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{Y_{n,i}(f) - \mathbb{E}Y_{n,i}(f)\}. \tag{6}$$

Corollary 2. *Consider the centered process $\{\widehat{\mathbb{G}}_n^0(f),\ f \in \mathcal{F}\}$ based on general density estimator $\widehat{p}_n(x)$ that satisfies (5). Assume that the class $\mathcal{F}$ is $P-$pregaussian, that the assumptions A1 and A2 are satisfied, and*

$$\lim_{n\to\infty} K_n\sqrt{n}\int \frac{\mathrm{Var}(\widehat{p}_n(x))}{p(x)}\,dx = 0. \tag{7}$$

Then $\{\widehat{\mathbb{G}}_n^0(f),\ f \in \mathcal{F}\}$, converges weakly to a tight Gaussian process $\{\mathbb{G}_P(f),\ f \in \mathcal{F}\}$ that has covariance structure $\mathrm{Cov}(\mathbb{G}_P(f), \mathbb{G}_P(g)) = Pfg - PfPg$.

Examples. Many popular density estimators can be written in the form (5) as we show in the following examples. For kernel density estimators with kernel K and bandwidth h_n, simply take

$$Y_{n,i}(f) = \int \frac{f(x)}{h_n} K\left(\frac{x - X_i}{h_n}\right) dx,$$

while for Fourier series based density estimators, put

$$Y_{n,i}(f) = \int f(t) D_m(X_i - t)\,dt$$

where $D_m(t)$ is the Dirichlet kernel

$$D_m(t) = \frac{\sin((m + 1/2)t)}{\sin(t/2)}.$$

The histogram density estimator $p_{n,H}(x)$ with binwidth h_n fits in this framework for

$$Y_{n,j}(f) = \sum_{i\in\mathbb{Z}} I_{\{ih_n \le X_j < (i+1)h_n\}} \int_{ih_n}^{(i+1)h_n} \frac{f(x)}{h_n}\,dx$$

and finally, the wavelet based density estimator $p_{n,W}(x)$ is of the form (5) with

$$Y_{n,j}(f) = \sum_{L\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} \phi_{k,L}(X_j)\int f(x)\phi_{k,L}(x)\,dx + \sum_{k\in\mathbb{Z}}\sum_{l>L} \psi_{k,l}(X_j)\int f(x)\psi_{k,l}(x)\,dx$$

for an appropriate scaling function $\phi(x)$ and mother wavelet $\psi(x)$.

Proofs

Proof of Theorem 1. Assumption (A1) and the Cramer-Wold device yield the finite dimensional convergence of the process $\{\widehat{\mathbb{G}}_n^0(f),\ f \in \mathcal{F}\}$. To prove that $\{\widehat{\mathbb{G}}_n^0(f),\ f \in \mathcal{F}\}$ is stochastically equicontinuous is more delicate. Let $\mathbb{G}(f) + (Pf)Z$, where $\{\mathbb{G}(f),\ f \in \mathcal{F}\}$ is a centered Gaussian process with all its sample paths bounded and uniformly continuous with respect to the semi metric $\rho(f,g) = P(f - Pf)(g - Pg)$ and Z is $N(0,1)$, independent of $\mathbb{G}$. The existence of $\mathbb{G}$ is guaranteed by the P-pregaussian condition on $\mathcal{F}$. Let $\delta_n = K_n^{1/2} n^{-1/4}$ and let $m_n = N(\delta_n, L^2(P), \mathcal{F})$ be the δ_n-covering number of $\mathcal{F}$ in $L^2(P)$. Sudakov's lower bound, see [5, Theorem 2.16], implies that there exists a finite set $\{g_1, \ldots, g_{m_n}\} \subseteq \mathcal{F}$ such that

$$P(g_i - g_j)^2 > \delta_n^2 \ \text{ for } 1 \le i \ne j \le m_n$$

and

$$\log(m_n) \le \alpha_n \delta_n^{-2} \ \text{ for some sequence } \alpha_n \to 0.$$

By the triangle inequality,

$$|\widehat{\mathbb{G}}_n^0(f) - \widehat{\mathbb{G}}_n^0(g)| \le |\widehat{\mathbb{G}}_n^0(f) - \widehat{\mathbb{G}}_n^0(g_i)| + |\widehat{\mathbb{G}}_n^0(g) - \widehat{\mathbb{G}}_n^0(g_j)| + |\widehat{\mathbb{G}}_n^0(g_i) - \widehat{\mathbb{G}}_n^0(g_j)|$$

for all $f, g \in \mathcal{F}_\delta$ and where we may take g_i, g_j such that $P(f - g_i)^2 \le \delta_n^2$ and $P(g - g_j)^2 \le \delta_n^2$ as $\{g_1, \ldots, g_{m_n}\}$ form a δ_n covering net of $\mathcal{F}$ in $L_2(P)$. For n large enough, as $\delta_n \to 0$, we have

$$P(g_i - g_j)^2 \le 4\delta^2$$

and consequently,

$$\sup_{f \in \mathcal{F}_\delta} |\widehat{\mathbb{G}}_n^0(f))| \le 2 \sup_{f \in \mathcal{F}_{\delta_n}} |\widehat{\mathbb{G}}_n^0(f)| + \max_{f \in \mathcal{G}_{2\delta}} |\widehat{\mathbb{G}}_n^0(f)|, \tag{8}$$

where

$$\mathcal{G}_{2\delta} = \left\{ g_i - g_j :\ \delta_n^2 < P(g_i - g_j)^2 \le 4\delta^2,\ 1 \le i < j \le m_n \right\}.$$

Since in condition (3) we assumed that

$$2\mathbb{E}^* \left[\sup_{f \in \mathcal{F}_{\delta_n}} |\widehat{\mathbb{G}}_n^0(f)| \right] \to 0,$$

it remains to bound $\mathbb{E}[\max_{f \in \mathcal{G}_{2\delta}} |\widehat{\mathbb{G}}_n^0(f)|]$. We argue as in [5, pages 952, 953]. For n large enough, condition (2) implies that $\mathbb{E}[Y_{n,1}^2(f)] \le 2cPf^2$ for $Pf^2 > \delta_n^2$. Observe that, by Chebyshev's inequality, uniformly in $f \in \mathcal{G}_{2\delta}$, for n large enough

$$\mathbb{P}\left\{ |\widehat{\mathbb{G}}_n^0(f)| \le 2\varepsilon \right\} \ge 1 - \frac{\frac{1}{n}\sum_{i=1}^n \mathbb{E}[Y_{n,i}^2(f)]}{4\varepsilon^2} \ge 1 - \frac{2cPf^2}{4\varepsilon^2} \ge \frac{1}{2} \tag{9}$$

for $c\delta^2 \le \varepsilon^2$. Using a standard symmetrization argument [8, page 14], we obtain

$$\mathbb{P}\left\{ \max_{f \in \mathcal{G}_{2\delta}} |\widehat{\mathbb{G}}_n^0(f)| \ge 4\varepsilon \right\} \le 4\mathbb{P}\left\{ \max_{f \in \mathcal{G}_{2\delta}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \sigma_i Y_{n,i}(f) \right| \ge \varepsilon \right\}$$

where $\sigma_1, \ldots, \sigma_n$ is a Rademacher sequence, independent of $X_1, \ldots, X_n$. Define the set

$$A_n = \left\{ \frac{1}{n} \sum_{i=1}^n Y_{n,i}^2(g_k - g_\ell) \le (1 + 2c)P(g_k - g_\ell)^2,\ 1 \le k, \ell \le m_n \right\}.$$

Since $Y_{n,1}^2(f), \ldots, Y_{n,n}^2(f)$ are independent, bounded (by K_n^2) random variables, we may apply Bernstein's inequality and obtain, for all $f \in \mathcal{G}_{2\delta}$, and n large enough,

$$\begin{aligned}\mathbb{P}\left\{\sum_{i=1}^n Y_{n,i}^2(f) > (1+2c)Pf^2\right\} &\leq \mathbb{P}\left\{\sum_{i=1}^n \left[Y_{n,i}^2(f) - \mathbb{E}[Y_{n,i}^2(f)]\right] > nPf^2\right\} \\ &\leq \exp\left(-\frac{1}{2}\frac{n(Pf^2)^2}{\mathbb{E}Y_{n,1}^4(f) + \frac{2}{3}K_n^2 Pf^2}\right) \\ &\leq \exp\left(-\frac{n}{2(c+\frac{2}{3})K_n^2}Pf^2\right).\end{aligned}$$

Consequently, we find by the union bound,

$$\mathbb{P}(A_n^c) \leq m_n^2 \exp\left\{-\frac{n^{1/2}}{2K_n(c+1)}\right\} \to 0.$$

Next, observe that

$$\begin{aligned}&\mathbb{E}\left[\max_{f\in\mathcal{G}_{2\delta}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \sigma_i Y_{n,i}(f)\right| I_{A_n}\middle| X_1,\ldots,X_n\right] \\ &\leq \mathbb{E}\left[\max_{f\in\mathcal{G}_{2\delta}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i Y_{n,i}(f)\right| I_{A_n}\middle| X_1,\ldots,X_n\right]\end{aligned}$$

where $Z_1, \ldots, Z_n$ are i.i.d. $N(0,1)$, and

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i Y_{n,i}(f)\right)^2 I_{A_n}\middle| X_1,\ldots,X_n\right] \leq \frac{1}{n}\sum_{i=1}^n Y_{n,i}^2(f) I_{A_n} \leq (1+2c)Pf^2.$$

This means that the Gaussian process $\{n^{-1/2}\sum_{i=1}^n Z_i Y_{n,i}(f) I_{A_n},\ f \in \mathcal{F}\}$, conditionally given $X_1, \ldots, X_n$, is tighter than the tight Gaussian process $\{\mathbb{G}(f) + (Pf)Z,\ f \in \mathcal{F}\}$. By Fernique's comparison result between Gaussian processes [5, Theorem 2.17], we find that

$$\mathbb{P}\left\{\max_{f\in\mathcal{G}_{2\delta}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \sigma_i Y_{n,i}(f)\right| \geq \varepsilon, A_n\right\}$$

converges to 0 as $n \to \infty$. This concludes the proof of the asymptotic equicontinuity of $\widehat{\mathbb{G}}_n^0$ and the theorem follows. □

Proof for Corollary 2. We prove the corollary by showing that condition (7) implies (A3) in Theorem 1. This follows easily as

$$\begin{aligned}&\mathbb{E}\left[\sup_{Pf^2\leq K_n n^{-1/2}}\left|\widehat{\mathbb{G}}_n^0(f)\right|\right] = \\ &\mathbb{E}\left[\sup_{Pf^2\leq K_n n^{-1/2}}\sqrt{n}\left|\int_{\{x:\,p(x)>0\}} f(x)\{\widehat{p}_n(x)-p(x)\}\,dx\right|\right] \leq \\ &\sup_{Pf^2\leq K_n n^{-1/2}}\sqrt{n\int f^2(x)p(x)\,dx\cdot\mathbb{E}\left[\int\frac{\{\widehat{p}_n(x)-p(x)\}^2}{p(x)}\,dx\right]^{1/2}} \\ &\leq \left\{K_n\sqrt{n}\int\frac{\mathrm{Var}(\widehat{p}_n(x))}{p(x)}dx\right\}^{1/2} \to 0\end{aligned}$$

proving the result. $\square$

3. Fourier density estimators

Throughout this section we assume that density $p(x)$ is defined on the interval $[0, 2\pi]$ and we consider classes $\mathcal{F}$ of periodic functions $f : [0, 2\pi] \to \mathbb{R}$ with envelope function

$$F(x) = \sup_{f\in\mathcal{F}} |f(x)|.$$

If $x, t \in [0, 2\pi]$ and $x + t > 2\pi$, we set $f(x + t) = f(x + t - 2\pi)$. The Fourier coefficients of p and f are

$$a_k = \frac{1}{2\pi}\int_0^{2\pi} e^{-ikx} p(x)\, dx$$

and

$$b_{k,f} = \frac{1}{2\pi}\int_0^{2\pi} e^{-ikx} f(x)\, dx.$$

Given a sample $X_1, \ldots, X_n$ and an integer sequence m depending on the sample size n, the Fourier density estimator is defined by

$$p_{n,m}(x) = \sum_{|k|\le m} a_k^n e^{ikx}$$

where

$$a_k^n = \frac{1}{2\pi}\cdot\frac{1}{n}\sum_{j=1}^{n} e^{-ikX_j}.$$

The smoothed empirical process is denoted by

$$\widehat{\mathbb{G}}_n(f) = \sqrt{n}\int f(x)\{p_{n,m}(x) - p(x)\}\, dx,\ \ f \in \mathcal{F}$$

while $\{\mathbb{G}_P(f),\ f \in \mathcal{F}\}$ stands for a Gaussian process with covariance structure

$$\mathrm{Cov}(\mathbb{G}_P(f), \mathbb{G}_P(g)) = Pfg - PfPg.$$

Throughout this section we work under the following set of assumptions:
(B1) There exists $c < \infty$ such that for $\alpha \ge 0$ and $\beta \ge 0$

$$\sup_{f\in\mathcal{F}} |b_{k,f}| \le \frac{c}{|k|^\beta} \quad \text{and} \quad |a_k| \le \frac{c}{|k|^\alpha}$$

and
(B2) $\alpha + \beta > 2$.

Theorem 3. *Let $\mathcal{F}$ be a P-pregaussian class of functions. Assume (B1) and (B2) and that $d \le p(x) \le 1/d$ for some $0 < d < \infty$. Then, $\{\widehat{\mathbb{G}}_n(f)\, f \in \mathcal{F}\}$ converges weakly to $\{\mathbb{G}_P(f),\ f \in \mathcal{F}\}$ in $\ell^\infty(\mathcal{F})$, provided either*

(a) $||F||_\infty \le 1$ or $\beta > 1$, and $m = n^\gamma$ with $1/\{2(\alpha + \beta - 1)\} < \gamma < 1/2$,
(b) $\beta \le 1$ and $\alpha + 2\beta > 3$ and $m = n^\gamma$ with $1/\{2(\alpha + \beta - 1)\} < \gamma < 1/\{2(2 - \beta)\}$.

Corollary 4. *If $\alpha > 2$ and $d \le p(x) \le 1/d$ for some $0 < d < \infty$, then, for uniformly bounded classes of functions $\mathcal{F}$, the following two statements are equivalent:*

$$\{\widehat{\mathbb{G}}_n(f),\, f \in \mathcal{F}\} \to \{\mathbb{G}_P(f),\, f \in \mathcal{F}\} \quad \text{weakly in } \ell^\infty(\mathcal{F}) \tag{10}$$

$$\mathcal{F} \text{ is } P\text{-pregaussian.} \tag{11}$$

Proof. That (10) implies (11) follows from the very definition of P-pregaussian. The implication (11) $\to$ (10) follows since $\alpha > 2$ implies that we only need that $b_{k,f}$ are uniformly bounded. This in turn is easily seen as

$$|b_{k,f}| \le \frac{1}{2\pi} \int_0^{2\pi} |e^{-ikx}||f(x)|dx \le 1.$$

□

Remark. The condition $\alpha + \beta > 2$ captures the trade-off between the smoothness of underlying density p and the smoothness of the functions f. It is well known that for $\beta < 1$ the class of functions $\mathcal{F}$ is not $P-$Donsker, which means that Theorem 3 establishes uniform central limit theorems for classes that are no longer P-Donsker as soon as p has more than one derivative ($\alpha > 1$). Pushing the conditions to the extreme by considering $\beta > 2$, $\widehat{\mathbb{G}}_n$ converges weakly for any density p. This follows since all L^1 functions have uniformly bounded Fourier coefficients so that we can formally set $\alpha = 0$.

The obvious restriction of Theorem 3 is its assumption on the density p. We now argue that by their very nature, Fourier series density estimators are intrinsically ill equiped when dealing with densities close to zero. Namely, the classical representation using Dirichlet kernel reveals that

$$p_{n,m}(t) = \frac{1}{n} \sum_{i=1}^n \frac{\sin(m+1/2)(X_i - t)}{\sin((X_i - t)/2)},$$

which means that the probability mass, although centered at X_i's is spread out over the whole interval. This follows from $||D_m||_{L^1} \to \infty$ as $m \to \infty$. Thus, for example, if $p(x) \approx e^{-1/x}$ for $x \approx 0$, $\min_{i \le n} X_i$ is of the order $1/\ln n$, which means that for $x \approx 1/n$, $p_{n,m}(x) \approx 1/n$ while $p(x) \approx \exp(-n)$.

We impose the following assumptions to ensure that $p(x)$ behaves nicely around zero.

(B3) $\int p^{-1}(x)\, dx < \infty$ and $\|p\|_\infty < \infty$

(B4) The following four integrals $\int_0^{2\pi} p^2(x)\, dx$, $\int_0^{2\pi} F^2(x)\, dx$, $\int_0^{2\pi} F^2(x)p^2(x)\, dx$ and $\int_0^{2\pi} F^2(x)p(x)\, dx$ are all finite.

Theorem 5. *Assume (B1), (B3) and (B4) with $\alpha/2 + \beta > 2$. Then, for $m = n^\gamma$ with*

$$\frac{1}{\alpha/2 + \beta - 1 + \max(1-\beta, 0)} < \gamma < \frac{1}{2\max(2-\beta, 1)}$$

the smoothed empirical process $\{\widehat{\mathbb{G}}_n(f),\, f \in \mathcal{F}\}$ converges weakly to a tight Gaussian process $\{\mathbb{G}_P(f),\, f \in \mathcal{F}\}$ in $\ell^\infty(\mathcal{F})$.

Proofs

For the remainder of this section we simplify the notation: Integration $\frac{1}{2\pi}\int_0^{2\pi}$ will be denoted by $\int$ and the density estimator $p_{n,m}$ by p_n. The usual notation for absolute value $|\cdot|$ is used for both real and complex functions, with the obvious understanding that $|a+bi| = \sqrt{a^2+b^2}$. We write $a \lesssim b$ if $a \leq Cb$ for some fixed constant C not depending on a or b. Finally the normalizing factor $\frac{1}{2\pi}$ in definition of a_k^n is dropped.

A simple computation shows that

$$
\begin{aligned}
\widehat{\mathbb{G}}_n(f) &= \sqrt{n}\int f(x)\{p_n(x) - \mathbb{E}[p_n(x)]\}\,dx \\
&= \sqrt{n}\int f(t)\left\{\sum_{|k|\leq m} e^{ikt}(a_k^n - \mathbb{E}[a_k^n])\right\} dt \\
&= \sqrt{n}\int f(t)\left\{\sum_{|k|\leq m} e^{ikt}\frac{1}{n}\sum_{j=1}^n (e^{-ikX_j} - \mathbb{E}[e^{-ikX_j}])\right\} dt \\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^n \sum_{|k|\leq m}\int f(t)e^{-ikt}\left\{e^{ikX_j} - \mathbb{E}[e^{ikX_j}]\right\} dt \\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^n \{Y_{n,j}(f) - \mathbb{E}[Y_{n,j}(f)]\}
\end{aligned}
$$

for

$$Y_{n,j}(f) = \sum_{|k|\leq m} e^{ikX_j} b_{k,f} = \sum_{|k|\leq m} e^{ikX_j} b_{k,f} = (D_m * f)(X_j) = \int f(t) D_m(X_j - t)\,dt$$

where

$$D_m(t) = \frac{\sin(m+1/2)t)}{\sin(t/2)}.$$

The proof of Theorem 3 requires a few auxiliary lemmas.

Lemma 6. *Under B1, we have*

$$||Y_{n,i}||_\infty \lesssim m^{\max(1-\beta,0)}. \tag{12}$$

If, in addition, $0 < \delta \leq p(x) < \infty$, *we have*

$$\mathbb{E}\left[Y_{n,j}^2(f)\right] \lesssim Pf^2 \text{ and } ||Y_{n,i}||_\infty \lesssim \ln n ||f||_\infty. \tag{13}$$

Proof. The first assertion follows from

$$||Y_{n,i}||_\infty = \left\|\sum_{|k|\leq m} e^{ikX_j} b_{k,f}\right\|_\infty \lesssim \sum_{|k|\leq m, k\neq 0} k^{-\beta} \lesssim m^{\max(1-\beta,0)}.$$

For the second claim we reason as follows. Since $0 < \delta \leq p(x) < \infty$, we have

$$
\begin{aligned}
\mathbb{E}\left[Y_{n,1}^2(f)\right] &= \int \left(\sum_{|k|\leq m} e^{ikX_1} b_{k,f}\right)^2 p(x)\,dx \\
&\lesssim \int \left(\sum_{|k|\leq m} e^{ikX_1} b_{k,f}\right)^2 dx \\
&= \int \sum_{|k|\leq m} e^{ikX_1} b_{k,f} \overline{\sum_{|k|\leq m} e^{ikX_1} b_{k,f}}\,dx
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{|k|\le m}\sum_{|l|\le m} b_{k,f} b_{-l,f} \int e^{i(k-l)x}\,dx \\
&= \sum_{|k|\le m} b_{k,f} b_{-k,f} \le \sum_{|k|<\infty} b_{k,f}^2 \\
&= \int f^2(x)dx \\
&\lesssim Pf^2
\end{aligned}
$$

We write $\|f\|_q = \left\{\int |f(x)|^q\,dx\right\}^{1/q}$ and we obtain, for $q > 1$,

$$
\begin{aligned}
\left(\mathbb{E}\left[Y_{n,i}^q(f)\right]\right)^{1/q} &\le \delta^{-1}||Y_{n,i}||_q \\
&\le \delta^{-1}||D_m||_1||f||_q \\
&\le \delta^{-1}\ln(n)||f||_q
\end{aligned}
$$

by Young's inequality. Take $q = 2$ and $q = \infty$ to conclude the proof. □

Lemma 7. *Assume B1, B2 and B4. Then, for $m = n^\gamma$ with $1/\{2(\alpha+\beta-1)\} < \gamma < 1/2$, we have*

$$
\lim_{n\to\infty}\sup_f \left|\sqrt{n}\int f(t)\{\mathbb{E}[p_n(t)] - p(t)\}\,dt\right| = 0. \tag{14}
$$

Proof. First observe that

$$
\begin{aligned}
\mathbb{E}[p_n(t)] &= \mathbb{E}\left[\sum_{|k|\le m} a_k^n e^{ikt}\right] \\
&= \sum_{|k|\le m} e^{ikt}\frac{1}{n}\sum_{j=1,n}\mathbb{E}[e^{-ikX_j}] \\
&= \sum_{|k|\le m} e^{ikt}\mathbb{E}[e^{-ikX_1}] \\
&= \sum_{|k|\le m} e^{ikt}\int e^{-ikx}p(x)\,dx \\
&= \sum_{|k|\le m} e^{ikt}a_k.
\end{aligned}
$$

Consequently,

$$
\int f(x)\mathbb{E}[p_n(x)]\,dx = \sum_{|k|\le m} a_k \int f(x)e^{ikx} = \sum_{|k|\le m} a_k b_{-k,f}.
$$

B4 allows us to invoke Parseval's identity:

$$
\int p(x)f(x)\,dx = \sum_{k\in\mathbf{Z}} a_k b_{-k,f} \tag{15}
$$

whence

$$\begin{aligned}
&\sup_f \sqrt{n} \left| \int f(x)\{\mathbb{E}[p_n(x)] - p(x)\}\, dx \right| \\
&= \sqrt{n} \sup_f \left| \sum_{|k|>m} a_k b_{-k,f} \right| \\
&\lesssim \sqrt{n} \sum_{k=m}^{\infty} \frac{1}{k^{\alpha+\beta}} \\
&\lesssim \frac{n^{1/2}}{m^{\alpha+\beta-1}} \\
&= \frac{n^{1/2}}{n^{\gamma(\alpha+\beta-1)}} \to 0,
\end{aligned}$$

by B1 and B2. □

Lemma 8. *Assume B1, B2 and B4. Then, for any integer m,*

$$\sup_f \left| \mathbb{E}[Y_{n,1}^2(f)] - Pf^2 \right| \lesssim \frac{1}{m^{\alpha/2+\beta-1}} \tag{16}$$

Proof. Again B4 allows us to invoke Parseval's identity:

$$Pf^2 = \sum_{k \in Z} c_{k,f} b_{-k,f},$$

where $c_{k,f} = \int f(x)p(x)e^{-ikx}dx$. Another application of Parseval's identity yields

$$c_{k,f} = \int f(x)\overline{p(x)e^{ikx}}dx = \int p(x)e^{-ikx} f(x)dx = \sum_{l \in Z} g_l b_{-l,f} = \sum_{l \in Z} a_{k+l} b_{-l,f}.$$

Since g_l is the l-th Fourier coefficient of $p(x)e^{-ikx}$, we have $g_l = a_{k+l}$. and hence

$$Pf^2 = \sum_{k \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} a_{k+l} b_{-l,f} b_{-k,f} = \sum_{k \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} a_{-(k+l)} b_{l,f} b_{k,f}.$$

On the other hand,

$$\begin{aligned}
\mathbb{E}[Y_{n,1}^2(f)] &= \mathbb{E}\left[\sum_{|k| \le m} e^{ikX_1} b_{k,f} \right]^2 \\
&= \mathbb{E}\left[\overline{\sum_{|k| \le m} e^{ikX_1} b_{k,f}} \sum_{|k| \le m} e^{ikX_1} b_{k,f} \right] \\
&= \mathbb{E}\left[\sum_{|l| \le m} e^{-ilX_1} b_{-l,f} \sum_{|k| \le m} e^{ikX_1} b_{k,f} \right] \\
&= \sum_{|k| \le m} \sum_{|l| \le m} b_{k,f} b_{l,f} E e^{ilX_1} e^{ikX_1} \\
&= \sum_{|k| \le m} \sum_{|l| \le m} b_{k,f} b_{l,f} \int e^{ilx} e^{ikx} p(x)\, dx
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{|k|\le m} \sum_{|l|\le m} b_{k,f} b_{l,f} \int e^{i(l+k)x} p(x)\,dx \\
&= \sum_{|k|\le m} \sum_{|l|\le m} b_{k,f} b_{l,f} a_{-(l+k)}.
\end{aligned}
$$

This yields further that

$$
\begin{aligned}
Pf^2 - \mathbb{E}[Y_{1,f}^2] &= \sum_{k\in Z} \sum_{l\in Z} a_{-(k+l)} b_{k,f} b_{l,f} - \sum_{|k|\le m} \sum_{|l|\le m} b_{k,f} b_{l,f} a_{-(l+k)} \\
&= \sum_{|(l,k)\in([-m,m]*[-m,m])^c} b_{k,f} b_{l,f} a_{-(l+k)}
\end{aligned}
$$

so that, using B2,

$$
\begin{aligned}
\sup_f \left|\mathbb{E}[Y_{1,f}^2] - Pf^2\right| &\lesssim \sum_{|(l,k)\in([-m,m]*[-m,m])^c} \frac{1}{|k|^\beta |j|^\beta |l+k|^\alpha} \\
&\lesssim m^{1-\beta-\alpha/2},
\end{aligned}
$$

with the obvious convention that the summation is not taken for $k=0$, $l=0$ and $l+k=0$. The final estimate on the right is obtained by breaking the summation over k,l into six regions: $I=\{(k,l):\ l\in(m,\infty),\ k\in[1,\infty)\}$, $II=\{(k,l):\ l\in(m,\infty),\ k\in[-\infty,0]\}$, $III=\{(k,l):\ l\in(-\infty,-m),\ k\in[1,\infty)\}$, $IV=\{(k,l):\ l\in(-\infty,-m),\ k\in[-\infty,0)\}$, $V=\{(k,l):\ l\in(-m,m),\ k\in[m,\infty)\}$, and $VI=\{(k,l): l\in(-m,m),\ k\in[-\infty,-m)\}$ and obtain for each region a similar estimate. The computation is straightforward but lengthy and for this reason we present only one typical case, namely region I. Since

$$
\frac{1}{(l+k)^\alpha} = \frac{1}{(l+k)^{\alpha/2}} \frac{1}{(l+k)^{\alpha/2}} \le \frac{1}{l^{\alpha/2}} \frac{1}{k^{\alpha/2}},
$$

we have

$$
\begin{aligned}
\sum_{l=m}^{\infty} \sum_{k=1}^{\infty} \frac{1}{k^\beta l^\beta (l+k)^\alpha} &\le \sum_{l=m}^{\infty} \sum_{k=1}^{\infty} \frac{1}{k^\beta l^\beta l^{\alpha/2} k^{\alpha/2}} \\
&= \sum_{l=m}^{\infty} \frac{1}{l^{\beta+\alpha/2}} \sum_{k=1}^{\infty} \frac{1}{k^{\beta+\alpha/2}} \\
&\lesssim \sum_{l=m}^{\infty} \frac{1}{l^{\beta+\alpha/2}} \\
&\lesssim m^{1-\beta-\alpha/2},
\end{aligned}
$$

since $\beta+\alpha/2>1$. The estimates for the regions $II,\dots,VI$ are very similar. □

Lemma 9. *Assume B3 or* $0<\delta\le p(x)$. *Then we have*

$$
\mathbb{E}\left[\sup_{Pf^2\le K_n n^{-1/2}} \left|\sqrt{n}\int f(x)\{p_n(x)-\mathbb{E}[p_n(x)]\}\,dx\right|\right] \lesssim \sqrt{\frac{mK_n}{\sqrt{n}}}.
$$

Proof. We start with the simple observation that $0 < \delta \le p(x)$ implies $\int 1/p(x)\,dx < \infty$. Next, we bound

$$\begin{aligned}
&\mathbb{E}\left[\sup_{Pf^2 \le K_n n^{-1/2}} \left|\sqrt{n}\int f(x)\{p_n(x) - \mathbb{E}[p_n(x)]\}\frac{\sqrt{p(x)}}{\sqrt{p(x)}}\,dx\right|\right] \le \\
&\mathbb{E}\left[\sup_{Pf^2 \le K_n n^{-1/2}} \sqrt{nPf^2}\sqrt{\int \frac{(p_n(x) - \mathbb{E}[p_n(x)])^2}{p(x)}\,dx}\right] \le \\
&\sqrt{K_n n^{1/2}\int \frac{\mathrm{Var}(p_n(x))}{p(x)}\,dx}.
\end{aligned}$$

Using the Dirichlet kernel representation, we get

$$\mathrm{Var}(p_n(x)) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1,n} D_m(x - X_i)\right) \le \frac{1}{n}\mathbb{E}\left[D_m^2(x - X_1)\right],$$

and consequently, for $q(x) = 1/p(x)$,

$$\begin{aligned}
K_n n^{1/2}\int \frac{\mathrm{Var}(p_n(x))}{p(x)}\,dx &\le K_n n^{-1/2}\int q(x)\int D_m^2(x-t)p(t)\,dt\,dx \\
&= K_n n^{-1/2}\int\int D_m^2(x-t)p(t)q(x)\,dx\,dt \\
&= K_n n^{-1/2}\int p(t)\int q(x)D_m^2(x-t)\,dx\,dt \\
&\lesssim K_n n^{-1/2}\int\int D_m^2(t-x)q(x)\,dx\,dt \\
&= K_n n^{-1/2}||D_m^2 * q||_1 \\
&\le K_n n^{-1/2}||D_m^2||_1||q||_1 \\
&\lesssim K_n n^{-1/2} m
\end{aligned}$$

since $||D_m^2||_1 \lesssim m$ and $||q||_1 = \int 1/p(x)\,dx < \infty$. □

Proof of Theorem 3. We verify conditions (A1), (A2) and (A3) from Theorem 1. Clearly, $Y_{n,i}(f)$ are i.i.d. and linear in f. By Lemma 6, $\|Y_{n,i}(f)\|_\infty \le m = o(\sqrt{n})$, so that Lemmas 7 and 8 easily yield $\mathbb{E}[Y_{n,i}(f)]^k \to Pf^k$ for $k = 1, 2$. The Central limit theorem implies (A1).

Second, since by assumption $p(x)$ is bounded away from zero, (B3) and (B4) are automatically satisfed and we can apply Lemmas 6–9. Lemma 7 states that we only need to consider centered process $\widehat{\mathbb{G}}_n^0$.

For case (a) of Theorem 3, Lemma 6 implies that $||Y_{n,i}||_\infty \lesssim \ln n$ as well as $\mathbb{E}[Y_{n,i}^2(f)] \lesssim Pf^2$. A1 and A2 are satisfied by Lemma 6 and Lemma 7 with $K_n = \ln^2(n)$. Lemma 9 implies that

$$\mathbb{E}\left[\sup_{Pf^2 \le (\ln^2 n)n^{-1/2}} \left|\sqrt{n}\int f(x)\{p_n(x) - \mathbb{E}[p_n(x)]\}\,dx\right|\right]$$

is bounded by

$$\frac{m\ln^2(n)}{n^{1/2}} = n^{\gamma - 1/2}\ln^2 n \to 0$$

since $\gamma < 1/2$ by assumption.

Case (b) of Theorem 3 is handled in similar way. Since now $\beta < 1$, $\alpha + 2\beta > 3$ implies (B2). Again, as $p(x)$ is bounded away from zero, (B4) is satisfied so that we can invoke Lemmas 6–9. Lemma 6 implies that $||Y_{n,i}||_\infty \lesssim K_n = m^{1-\beta}$ as well as $\mathbb{E}[Y_{n,i}^2(f)] \lesssim Pf^2$, whence (A1) and (A2) follow from Lemmas 6 and 7. Since

$$\frac{mK_n}{n^{1/2}} = n^{\gamma(1-\beta)-1/2} \to 0$$

as $\gamma(1-\beta) < 1/2$ by the assumption, Lemma 9 concludes our proof. $\square$

Proof of Theorem 5. Since $\alpha/2 + \beta > 2$ implies (B2), we can apply Lemmas 6–9. Lemma 7 implies that we only consider the centered process $\widehat{\mathbb{G}}_n^0$, while Lemma 6 implies that $||Y_{n,i}||_\infty \lesssim K_n$ for $K_n \lesssim m^{\max(1-\beta,0)}$. Lemma 8 yields that

$$\left|\mathbb{E}[Y_{n,i}^2(f)] - Pf^2\right| \lesssim \frac{1}{m^{\alpha/2+\beta-1}} \lesssim \frac{K_n}{n^{1/2}}$$

since $\gamma > 1/\{\alpha/2 + \beta - 1 + \max(1-\beta, 0)\}$. Thus (A1) and (A2) are met. Finally, condition (A3) follows from Lemma 9 and the following estimate

$$\frac{K_n m}{n^{1/2}} = \frac{m^{\max(2-\beta,1)}}{n^{1/2}} = n^{\gamma\max(2-\beta,1)-1/2} \to 0$$

since $\gamma < 1/\{2\max(2-\beta, 1)\}$. $\square$

4. Kernel density estimators

Let $X_1, X_2, \ldots$ be i.i.d. with density $p(x)$ and let $\widehat{p}_n(x)$ be a kernel density estimator of $p(x)$,

$$\widehat{p}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right), \tag{17}$$

where the kernel K satisfies, for some $d > 1$,

$$\begin{aligned} K \geq 0, \quad & \int K(z)\,dz = 1, \quad \sup_z K(z) < \infty, \\ \int zK(z)\,dz = 0, \quad & \int z^2 K(z)\,dz < \infty \end{aligned} \tag{18}$$

and the sequence of bandwidths h_n converges to zero as $n \to \infty$. Let $\mathcal{F}$ be a P-pregausian class of uniformly bounded functions,

$$||f||_\infty \leq 1 \text{ for all } f \in \mathcal{F}.$$

We impose conditions on the kernel K, the bandwidth h_n, the tails of $p(x)$ and we require smoothness of the convolutions $(f * \widetilde{p})(x)$ and $(f^2 * \widetilde{p})(x)$, where $\widetilde{p}(x) = p(-x)$, to establish weak convergence of $\{\widehat{\mathbb{G}}_n(f),\, f \in \mathcal{F}\}$. Let $C_M^s(\mathbb{R})$ be the class of functions $g : \mathbb{R} \to \mathbb{R}$ with $|g^{(\lfloor s \rfloor)}(x) - g^{(\lfloor s \rfloor)}(y)| \leq M|x - y|^{s-\lfloor s \rfloor}$.

Theorem 10. *Let $\beta = 2 \wedge s > 1$. Assume that both $(f * \widetilde{p})$ and $(f^2 * \widetilde{p})$ are in $C_M^s(\mathbb{R})$ for some $M < \infty$ for all $f \in \mathcal{F}$, and that*

$$\lim_{t\to\infty} t^\alpha \mathbb{P}\{|X| \geq t\} < \infty \tag{19}$$

for some $\alpha > \beta/(\beta-1)$ *and*

$$nh_n^{2\beta} \to 0 \text{ and } n^{\frac{\alpha-1}{\alpha}} h_n^2 \to \infty \text{ as } n \to \infty. \tag{20}$$

Then $\{\widehat{\mathbb{G}}_n(f),\, f \in \mathcal{F}\}$ *converges weakly to a tight Gaussian process in* $\ell^\infty(\mathcal{F})$.

Remark. Notice that if $p(x)$ has exponential tails, corresponding roughly to the case $\alpha = +\infty$ in (19), the condition on the bandwidth becomes $n^{-1/2} \ll h_n \ll n^{-1/(2\beta)}$. Hence we need some minimal smoothness $s > 1$.

Remark. We require that $\mathcal{F}$ is a P-pregaussian class of uniformly bounded functions. The condition that $(f * \widetilde{p}) \in C^s(\mathbb{R})$ follows if $p \in C^s(\mathbb{R})$, independent of f. On the other hand, it follows from the proof of [6, Theorem 5] that, if $f \in BV(\mathbb{R})$ and $p \in C^{s_p}(\mathbb{R})$, then $(f * \widetilde{p}) \in C^{1+s_p}(\mathbb{R})$. However, note that any bounded subset of $BV(\mathbb{R})$ is P-Donsker.

Remark. How does our result compare to results recently obtained by [6]? Minor differences are the conditions on $p(x)$ and h_n. Whereas we assume condition (19) on the density $p(x)$, [6] requires that $\sup_x \sqrt{1+x^2}\, p(x) < \infty$. For non-negative kernels K, we share the same upper bound for h_n that makes the bias of $\widehat{\mathbb{G}}_n$ disappear. The paper [6] allows for a smaller lower bound in some special cases described in [6, Theorems 9, 10]. Our method relies on closeness of $\widehat{p}_n(x)$ to $p(x)$, which we achieve by requiring that at least $nh_n^2 \to \infty$. The restriction on h_n is the (small) price to pay for our more general approach that allows for *arbitrary* P-pregaussian classes $\mathcal{F}$. The proof of Theorem 10 does not require any structure of $\mathcal{F}$ in contrast to the approach taken in [6]. This is the main difference between the two papers. Notice that too small bandwidths h_n force the smoothed measure $d\widehat{P}_n(x) = \widehat{p}_n(x)\, dx$ to be close to the empirical measure $P_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$ and this should be prevented for non P-Donsker classes $\mathcal{F}$. Higher order kernels are allowed in [6]. Our proof relies on non-negative $\widehat{p}_n(x)$ (and thus on $K \ge 0$).

Nickl [7] shows that certain subsets of certain Besov spaces are P-pregaussian, yet not P-Donsker for all P that have bounded densities with respect to the Lebesgue measure. Our result immediately implies that the smoothed empirical process indexed by any class of truncated functions of these Besov spaces converges, since truncation does not increase the metric entropy numbers. (It is possible that truncation makes the classes in fact smaller, perhaps even P-Donsker, but we do not need to verify this to apply our result.)

The remainder of the section is devoted to the proof of Theorem 10.

Proof. Throughout we will use the notation $K_h(x) = (1/h)K(x/h)$ and $\beta = 2 \wedge s$. We first show that the bias of $\widehat{\mathbb{G}}_n$ is asymptotically negligible. Recall that $\bar{p}_n(x) = \mathbb{E}[\widehat{p}_n(x)]$. Using a standard Taylor argument with the smoothness condition $(f*\widetilde{p}) \in C^s(\mathbb{R})$, we find

$$\begin{aligned}
&\sqrt{n}\left|\int f(x)\{\bar{p}_n(x) - p(x)\}\, dx\right| \\
&= \sqrt{n}\left|\int f(x) \int K(z)\{p(x - h_n z) - p(x)\}\, dz\, dx\right| \\
&= \sqrt{n}\left|\int K(z)\{(f*\widetilde{p})(zh_n) - (f*\widetilde{p})(0)\} dz\right|
\end{aligned}$$

$$= \sqrt{n}\left|\int K(z)zh_n(f*\widetilde{p})'(0)\,dz + \int K(z)zh_n\{(f*\widetilde{p})'(\xi) - (f*\widetilde{p})'(0)\}dz\right|$$
$$\le \sqrt{n}\left|\int zK(z)h_nM|\xi h_n|^{\beta-1}\,dz\right|$$

where ξ is between 0 and z. The right hand side is of order $\sqrt{n}h_n^\beta$ by assumption (18) which is asymptotically negligible by assumption (20).

Consequently, $\widehat{\mathbb{G}}_n - \widehat{\mathbb{G}}_n^0 \to 0$ in probability, as $n\to\infty$. We now verify conditions (1)–(3) of Theorem 1.

Note that

$$\sqrt{n}\int_{\mathbb{R}} f(x)\{\widehat{p}_n(x) - \bar{p}_n(x)\}\,dx = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{Y_{n,i}(f) - \mathbb{E}Y_{n,i}(f)\}$$

with

$$Y_{n,i}(f) = \int f(x)K_{h_n}(x-X_i)\,dx, \quad i=1,\dots,n.$$

Each term is bounded as

$$|Y_{n,i}(f)| = \left|\int f(x)K_{h_n}(x-X_i)\,dx\right| = \left|\int f(X_i+zh_n)K(z)\,dz\right| \le \int K(z)\,dz = 1.$$

The Lebesgue density theorem implies the pointwise convergence $(f*K_{h_n})(x)\to f(x)$ as $n\to\infty$ and an application of the dominated convergence theorem yields

$$\lim_{n\to\infty}\mathbb{E}\left[\{(f*K_{h_n})(X_1)\}^j\right] = \mathbb{E}[f^j(X_1)] = Pf^j,\ j=1,2.$$

This, coupled with the fact that f and K are bounded, verifies condition (A1) of Theorem 1.

Next we establish condition (A2) with $c=1$ and $K_n=1$. For all $n\ge 1$,

$$\begin{aligned}
\mathbb{E}[Y_{n,1}^2(f)] &= \mathbb{E}\left[\int f(x)K_{h_n}(x-X_i)\,dx\right]^2\\
&= \int\left\{\int f(x+zh_n)K(z)\,dz\right\}^2 p(x)\,dx\\
&\le \int\left\{\int f^2(x+zh_n)K(z)\,dz\right\}p(x)\,dx\\
&= \int\{(f^2*\widetilde{p})(-zh_n)\}K(z)\,dz\\
&= Pf^2 + O(h_n^\beta)\\
&= Pf^2 + o(n^{-1/2}),
\end{aligned}$$

uniformly in $f\in\mathcal{F}$.

Finally we verify condition (A3). We abbreviate $\mathcal{F}_{n^{-1/4}}$ by $\mathcal{F}_n$. Let $B_n = \{x : \bar{p}_n(x)\ne 0\}$ so that

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}_n}\left|\int_{B_n^c} f(x)\{\widehat{p}_n(x)-\bar{p}_n(x)\}\,dx\right|\right] \le 2\int_{B_n^c}\mathbb{E}[|\widehat{p}_n(x)|]\,dx = 2\int_{B_n^c}\bar{p}_n(x)\,dx = 0$$

and

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}_n}\left|\sqrt{n}\int_{\mathbb{R}} f(x)\{\widehat{p}_n(x)-\bar{p}_n(x)\}\,dx\right|\right]$$
$$=\mathbb{E}\left[\sup_{f\in\mathcal{F}_n}\left|\sqrt{n}\int_{B_n} f(x)\{\widehat{p}_n(x)-\bar{p}_n(x)\}\,dx\right|\right].$$

By condition (20) on the bandwidth, there exists a sequence $L_n\to\infty$ such that

$$\frac{n}{L_n^{2\alpha}}\to 0 \text{ and } \frac{L_n^2}{nh_n^2}\to 0. \tag{21}$$

(We can take L_n slightly larger than $n^{1/(2\alpha)}$.) Then

$$\begin{aligned}
&\mathbb{E}\left[\sup_{f\in\mathcal{F}_{n^{-1/4}}}|\widehat{\mathbb{G}}_n^0(f)|\right]\\
&\le 2\int_{B_n\cap[-L_n,L_n]^c}\sqrt{n}\mathbb{E}[|\widehat{p}_n(x)|]\,dx\\
&\quad+\mathbb{E}\left[\sup_{f\in\mathcal{F}_{n^{-1/4}}}\left|\int_{B_n\cap[-L_n,L_n]} f(x)\sqrt{n}\{\widehat{p}_n(x)-\bar{p}_n(x)\}\,dx\right|\right]\\
&:= I+II.
\end{aligned}$$

The first term I equals

$$\begin{aligned}
I&=2\sqrt{n}\int_{B_n\cap[-L_n,L_n]^c}\int K(z)p(x-h_nz)\,dz\,dx\\
&\le 2\sqrt{n}\int K(z)(\mathbb{P}\{X\ge L_n-zh_n\}+\mathbb{P}\{X\le -L_n-zh_n\})\,dz.
\end{aligned}$$

Split the integration into two parts: $|z|\le L_n/(2h_n)$ and its complement, and obtain

$$\begin{aligned}
I&\le 2\sqrt{n}\int_{|z|\le L_n/(2h_n)} K(z)(\mathbb{P}\{X\ge L_n-zh_n\}+\mathbb{P}\{X\le -L_n-zh_n\})\,dz\\
&\quad+2\sqrt{n}\int_{|z|\ge L_n/(2h_n)} K(z)(\mathbb{P}\{X\ge L_n-zh_n\}+\mathbb{P}\{X\le -L_n-zh_n\})\,dz\\
&\le 2\sqrt{n}\mathbb{P}\left\{|X|\ge\frac{1}{2}L_n\right\}+4\sqrt{n}\int_{|z|\ge L_n/(2h_n)}K(z)\,dz\\
&\le 2\sqrt{n}\mathbb{P}\left\{|X|\ge\frac{1}{2}L_n\right\}+4\sqrt{n}\frac{(2h_n)^\beta}{L_n^\beta}\int|z|^\beta K(z)\,dz.
\end{aligned}$$

By assumptions (19), (20) and by construction of L_n, the term on the right converges to zero.

Next we show that $II\to 0$ in probability. Using the Cauchy-Schwarz inequality, we find

$$\begin{aligned}
&\left[\int_{B_n\cap[-L_n,L_n]} f(x)\{\widehat{p}_n(x)-\bar{p}_n(x)\}\,dx\right]^2\\
&\le\int f^2(x)\bar{p}_n(x)\,dx\cdot\int_{B_n\cap[-L_n,L_n]}\frac{\{\widehat{p}_n(x)-\bar{p}_n(x)\}^2}{\bar{p}_n(x)}\,dx.
\end{aligned}$$

By condition (18) on the kernel and the smoothness condition on $f^2 * \widetilde{p}$ imply that

$$\begin{aligned}
&\sup_{f\in\mathcal{F}_n}\left|\int f^2(x)\{\bar{p}_n(x)-p(x)\}\,dx\right| \\
&\quad=\sup_{f\in\mathcal{F}_n}\left|\int f^2(x)\left\{\int K(z)p(x-zh_n)\,dz-p(x)\right\}dx\right| \\
&\quad=\sup_{f\in\mathcal{F}_n}\left|\int K(z)\left\{\int f^2(x)p(x-zh_n)-f^2(x)p(x)\,dx\right\}dz\right| \\
&\quad=O(h_n^\beta)=o(n^{-1/2}).
\end{aligned}$$

Next, it follows easily that

$$\begin{aligned}
\mathrm{Var}(\widehat{p}_n(x)) &\le \frac{1}{nh_n^2}\mathbb{E}\left[K^2\left(\frac{x-X_1}{h_n}\right)\right] \\
&\le \frac{1}{nh_n}\|K\|_\infty\mathbb{E}[K_{h_n}(x-X_1)] \\
&= \frac{\|K\|_\infty}{nh_n}\bar{p}_n(x).
\end{aligned}$$

and hence

$$\int_{B_n\cap[-L_n,L_n]}\frac{\mathrm{Var}(\widehat{p}_n(x))}{\bar{p}_n(x)}\,dx \le \frac{2\|K\|_\infty L_n}{nh_n}.$$

Combining the preceding four displays, we obtain

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}_n}\left|\sqrt{n}\int_{B_n\cap[-L_n,L_n]}f(x)\{\widehat{p}_n(x)-\bar{p}_n(x)\}\,dx\right|\right]=O\left(\frac{L_n^2}{nh_n^2}\right)^{1/4}=o(1)$$

as $n\to\infty$, by construction of the sequence L_n. This concludes the verification of condition (A3), and Theorem 10 follows now from Theorem 1. □

References

[1] Araujo, A. and Giné, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables.* Wiley, New York.

[2] Devroye, L. and Gyorfi, L. (1985). *Nonparametric Density Estimation: the L_1 View.* John Wiley & Sons.

[3] Dudley, R. (1999). *Uniform Central Limit Theorems*, Cambridge.

[4] Dudley, R. and Strassen, V. (1969). *The Central Limit Theorem and epsilon-entropy. Lecture Notes in Mathematics* **87** 224–231. Springer.

[5] Giné, E. and Zinn, J. (1984). Some limit theorems for empirical processes. *Annals of Probability* **12** 929–998.

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

[7] Nickl, R. (2006). Empirical and Gaussian processes on Besov classes. In *High Dimensional Probability IV* (E. Giné, V. Koltchinskii, W. Li and J. Zinn, eds.). *IMS Lecture Notes* **51** 185–195.

[8] Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer.

[9] Radulović, D. and Wegkamp, M. (2000). Weak convergence of smoothed empirical processes: Beyond Donsker classes. In *High Dimensional Probability II* (E. Gine, D. Mason and J. Wellner, eds.) 89–106. Birkhauser.

[10] Radulović, D. and Wegkamp, M. (2003). Necessary and sufficient conditions for weak convergence of smoothed empirical processes. *Statistics and Probability Letters* **61** (3) 321–336.

[11] van der Vaart, A. (1994). Weak convergence of smoothed empirical processes. *Scandinavian Journal of Statistics* **21** 501–504.

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

[13] Yukich, J. (1992). Weak convergence of smoothed empirical processes. *Scandinavian Journal of Statistics* **19** 271–279.