Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 119 (2009) 1541–1560
stochastic processes and their applications
www.elsevier.com/locate/spa

Gaussian approximation of the empirical process under random entropy conditions

Adel Settati*

University de Rennes 1, IRMAR, 32, rue de mirabeau L 323, 35700 Rennes, France

Received 18 March 2008; received in revised form 30 July 2008; accepted 1 August 2008
Available online 5 August 2008

Abstract

We obtain rates of strong approximation of the empirical process indexed by functions by a Brownian bridge under only random entropy conditions. The results of Berthet and Mason [P. Berthet, D.M. Mason, Revisiting two strong approximation results of Dudley and Philipp, in: High Dimensional Probability, in: IMS Lecture Notes-Monograph Series, vol. 51, 2006, pp. 155–172] under bracketing entropy are extended by combining their method to properties of the empirical entropy. Our results show that one can improve the universal rate $v_n = o(\sqrt{\log \log n})$ from Dudley and Philipp [R.M. Dudley, W. Philipp, Invariance principles for sums of Banach space valued random elements and empirical processes, Z. Wahrsch. Verw. Gebiete 62 (1983) 509–552] into $v_n \to 0$ at a logarithmic rate, under a weak random entropy assumption which is close to necessary. As an application the results of Koltchinskii [V.I. Kolchinskii, Komlós–Major–Tusnády approximation for the general empirical process and Haar expansions of classes of functions, J. Theoret. Probab. 7 (1994) 73–118] are revisited when the conditions coming in addition to random entropy are relaxed.

MSC: primary 60F15; secondary 60F05; 62G30

Keywords: Empirical processes; Strong invariance principle; Random entropy

1. Introduction

Suppose that $(\mathcal{X}, \mathcal{A})$ is a measurable space, $(X_n)_{n \geqslant 1}$ is a sequence of independent random elements defined on a probability space $(\Omega, \Sigma, \mathbb{P})$ taking values in $(\mathcal{X}, \mathcal{A})$ and having the

* Tel.: +33 02 23 23 51 64; fax: +33 02 23 23 67 90.
E-mail address: settati_adel@yahoo.fr.

0304-4149/$ - see front matter
doi:10.1016/j.spa.2008.08.001

distribution $P = \mathbb{P}^X$. Let $\mathcal{M}$ be the set of all measurable functions from $(\mathcal{X}, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\mathbb{R})$ stands for the Borel σ-field. First, define the P-empirical process indexed by a class $\mathcal{F} \subset \mathcal{M}$ to be

$$\alpha_n(f) = \sqrt{n}(P_n(f) - P(f)), \quad f \in \mathcal{F},$$

where P_n is the empirical measure,

$$P_n f = \int f \mathrm{d}P_n = \frac{1}{n}\sum_{k=1}^{n} f(X_i), \quad \text{and} \quad Pf = \int_{\mathcal{X}} f \mathrm{d}P.$$

If it is assumed that $\mathcal{F} \subset L_2(\mathcal{X}, dP)$ then the finite-dimensional distributions of the sequence of random functions $(\alpha_n)_{n\geqslant 1}$ converge weakly as $n \to \infty$ to the finite-dimensional distributions of a mean zero Gaussian random function B with the same matrix of covariance as $(\alpha_n(f))_{f\in\mathcal{F}}$, that is

$$\langle f, g\rangle = \mathrm{cov}(B(f), B(g)) = E(f(X)g(X)) - E(f(X))E(g(X)), \quad f, g \in \mathcal{F}.$$

In the context of the present article the process B will always admit a version which is almost surely bounded and continuous with respect to the intrinsic semi-metric

$$d_p(f, g) = \sqrt{E(f(X) - g(X))^2}, \quad f, g \in \mathcal{F}.$$

We call B a P-Brownian bridge indexed by $\mathcal{F}$. A strong Gaussian approximation holds when it is possible to construct a version of the i.i.d. random variables X_n and versions B_n of B on the same underlying probability space $(\Omega, \Sigma, \mathbb{P})$ in such a way that

$$\|\alpha_n - B_n\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |\alpha_n(f) - B_n(f)| = O(v_n) \quad \mathbb{P}\text{-a.s.} \tag{1}$$

where $(v_n)_{n>0}$ is a deterministic sequence. Dudley and Philipp [12], see also p. 306 in Dudley [11], showed that (1) holds with $v_n = o(\sqrt{\log\log n})$ whenever $\mathcal{F}$ is P-Donsker. Heuristically we can expect that v_n is not much better than diverging at a rate $\sqrt{\log\log n}$ for some classes of functions satisfying only a minimal sufficient condition to be P-Donsker. Now, in light of the results by Giné and Zinn [14] for classes of sets – see also Talagrand [28] for classes of functions – weak assumptions on the random entropy of $\mathcal{F}$ are close to such a minimal sufficient assumption. This why the aim of the present article is to investigate the problem of Gaussian approximation under only random entropy conditions. In Section 2 our assumptions are discussed and properties of random entropy are recalled.

Most of the known strong approximation results deal with fast rates $v_n \to 0$ in (1) and then require that the P-Donsker class $\mathcal{F}$ satisfy stronger assumptions. Let us first recall a few of them. Let $\mathcal{B}(I)$ denote the Borel σ-algebra of $I = [0, 1]$ and let μ denote the Lebesgue measure on I. In the classical case $\mathcal{X} = I$, $\mathcal{A} = \mathcal{B}(I)$, $P = \mu$ and $\mathcal{F}$ is the class of indicators of half intervals, i.e. $\mathcal{F} = \{I_{[0x)}, x \in I\}$ the so called KMT theorem of Komlós–Major–Tusnády [22] states that one can construct on the same probability space $(\Omega, \Sigma, \mathbb{P})$ the i.i.d. random variables X_n and versions B_n of the standard Brownian bridge B such that, for all $t > 0$ and $n \geq 1$, we have

$$\mathbb{P}\left(\sup_{x\in I} |\alpha_n(1_{[0,x[}) - B_n(x)| \geqslant \frac{t + C\log n}{\sqrt{n}}\right) \leqslant A\exp\{-\theta t\}, \tag{2}$$

where $A > 0$, $C > 0$, $\theta > 0$ are explicit constants. This implies that (1) holds with $v_n = \log n/\sqrt{n}$. See Csörgő and Révész [9], Csörgő and Horváth [8], Mason and van Zwet [25] and Bretagnolle and Massart [7] for further improvements. Borisov [6], Koltchinskii [19,20] and [21], Massart [26], Rio [27] and Einmahl and Mason [13] among others essentially extended the KMT result to the general case of empirical processes defined on classes of measurable sets and functions. In particular, Koltchinskii [18,20] and [21] obtained some sharp results under random entropy conditions, especially for classes of indicators of measurable sets $\mathcal{F} = \{1_C : C \in \mathcal{C}\}, \mathcal{C} \subset \mathcal{A}$. Already Dudley and Philipp [12] proved that one can improve $v_n = o(\sqrt{\log\log n})$ in (1) into $v_n = (\log n)^{-\theta}$ for some $\theta > 0$ if one assumes mild entropy bounds for $\mathcal{F}$. For classes of functions, this was refined by Rio [27] and Koltchinskii [21] who showed that for many Donsker classes satisfying a suitable entropy bound – in particular, VC classes – (1) holds with v_n decreasing to zero at a polynomial rate $n^{-\theta}$ for some $\theta > 0$. Next to entropy conditions, these authors had to impose further regularity conditions on the class $\mathcal{F}$. For example Rio [27] in Theorem 1.1 obtains $v_n = n^{-1/2d}$ in the case of VC-classes of functions of uniformly bounded variation and probability measures P with a bounded Lebesgue density on I^d. Koltchinskii [21] considered uniformly bounded classes of functions $\mathcal{F}$ having a good representation under P on a Haar expansion of the space $\mathcal{X}$ and satisfying a random entropy condition. This allowed him to take advantage of KMT and establish the following KMT-type approximation of $\alpha_n(f)$, $f \in \mathcal{F}$. One can construct on the same probability space $(\Omega, \Sigma, \mathbb{P})$ the i.i.d. random variables X_n and versions B_n of the standard Brownian bridge B such that, for all $t > 0$ and $n \geq 1$, we have

$$\mathbb{P}(\|\alpha_n - B_n\|_{\mathcal{F}} \geqslant \delta_n(t + C_1 \log n)) \leqslant A_1 \exp\{-B_1 t\} \tag{3}$$

where the constants $A_1 > 0$, $B_1 > 0$ and $C_1 > 0$ depend only on $\mathcal{F}$ and δ_n depends on the random entropy of $\mathcal{F}$ and its Haar representation under P – see Theorems 3.1–3.4 in Koltchinskii [21]. As a corollary if the functions of $\mathcal{F}$ are q-times differentiable for $q > 1$, then Theorem 11.1 of [21] yields (3) with $\delta_n = n^{-(q-1)/(2q+2)}$. Then (1) holds true with $v_n = \delta_n \log n$.

It is also of interest to study (1) when most of additional assumptions such as Haar approximability, bounded variation and bounded density are removed. In this direction an alternative to the KMT approach was recently developed in Berthet and Mason [2,3] by using a coupling inequality of Zaitsev [33,34] under either general uniform and bracketing entropy conditions on $\mathcal{F}$ or some more restrictive analytical conditions. They basically obtained the following kind of Gaussian approximation which is slightly weaker than (1) but strong enough to be widely applied. For each $\lambda > 0$ there exists a $\sigma > 0$ such that for each integer $n \geqslant 1$ one can construct on the same probability space $(\Omega, \Sigma, \mathbb{P})$ the i.i.d. random variables X_n and a versions B_n of B such that

$$\mathbb{P}\left(\|\alpha_n - B_n\|_{\mathcal{F}} \geqslant \sigma v_n\right) \leqslant n^{-\lambda} \tag{4}$$

where v_n depends only on the entropy of $\mathcal{F}$ – see Propositions 1 and 2 in Berthet and Mason [2] – or on some representation of $\mathcal{F}$ — see [3]. Note that (4) provides a rate in the weak invariance principle.

In this paper we combine the coupling technique developed in Berthet and Mason [2] with properties of the random entropy of Giné and Zinn [14] to derive strong invariance principles in the form (4). Our results are stated in Section 3 and proved in Sections 4 and 5. They are somewhat intermediate between the results of Dudley and Philipp [12] and the results of Koltchinskii [21], and close to those of Berthet and Mason [2] under bracketing entropy. Our random entropy condition (F.3) on P and $\mathcal{F}$ is relatively mild compared to Koltchinskii [21] and Rio [27] but stronger than Dudley and Philipp [12]. In particular, in Theorem 3.1 we show that one can improve in (4) the pessimistic rate $v_n = o(\sqrt{\log\log n})$ from Dudley and Philipp [12] into $v_n \to 0$ at a logarithmic rate, under solely our weakest version of (F.3). On the opposite in Theorem 3.4 we show that under the strongest version of (F.3) the rate v_n in (4) is polynomial. We further show how mixing several kind of entropy assumptions can improve the rates of Gaussian approximation. As a further motivation we refer to Berthet and Saumard [4] where Gaussian approximation in the indexed by functions setting and for fixed n appears to be a key step towards sharp results in M-estimation.

2. Preliminaries

2.1. Empirical entropy

Assume that (S, d) is a totally bounded metric space, S_ε is a finite ε-net of S with respect to d and $\varepsilon > 0$. Define a map $\lambda_\varepsilon : S \mapsto S_\varepsilon$ such that $d(s, \lambda_\varepsilon s) \le \varepsilon$, $s \in S$. Let $N(\varepsilon), \varepsilon > 0$ be a decreasing function such that $\mathrm{card}(S_\varepsilon) \le N(\varepsilon), \varepsilon > 0$. Then

$$N_d(S, \varepsilon) \le N(\varepsilon),$$

where $N_d(S, \varepsilon)$ is the smallest number of points in such a ε-net for the space S. Let $H_d(S, \varepsilon)$ denote the ε-entropy of S with respect to d,

$$H_d(S, \varepsilon) = \log N_d(S, \varepsilon) \le H(\varepsilon) = \log N(\varepsilon). \tag{5}$$

For an arbitrary probability Q on $(\mathcal{X}, \mathcal{A})$ and $p \in [1, +\infty)$ let $d_{Q,p}$ denote the metric of the space $L_p(\mathcal{X}, dQ)$. Throughout we shall use extensively the projection mapping λ_ε under $d = d_{P,2}$ together with $N_{d_{P,2}}$ and $H_{d_{P,2}}$. Likewise let $H_{n,1}(\mathcal{F}, \varepsilon)$ denote the ε-entropy of $\mathcal{F}$ with respect to $d = d_{P_n,1}$. This is the random entropy of $\mathcal{F}$ under P.

Vapnik and Chervonenkis [31,32], Koltchinskii [18–21], Le Cam [23], Giné and Zinn [14–16], Talagrand [29] investigated the character and the rates of convergence of the sequence of empirical measures P_n to the theoretical distribution P in relation with the asymptotic behavior of the empirical entropy $H_{n,1}(\mathcal{F}, \varepsilon)$, when $n \to \infty$ and $\varepsilon \to 0$. In particular, assuming that $\mathcal{F}$ is uniformly bounded, if $\alpha \in (\frac{1}{2}, 1]$ then the condition

$$H_{n,1}(\mathcal{F}, \varepsilon n^{\alpha-1}) = o(n^\alpha), \quad n \to \infty, \varepsilon > 0 \quad \mathbb{P}\text{-a.s.} \tag{6}$$

is sufficient – and also necessary when $\alpha = 1$, this is the Glivenko Cantelli property – to have, see Koltchinskii [21],

$$\|P_n - P\|_{\mathcal{F}} = o(n^{\alpha-1}), \quad n \to \infty \quad \mathbb{P}\text{-a.s.} \tag{7}$$

One could conclude heuristically that the rate of convergence of the sequence of empirical distributions to the theoretical one also depends on the value of α in (6). We will strengthen (6) by requiring that $\alpha < 1/2$ in which case (7) becomes $O(n^{-1/2})$ in law – this is the Donsker property, see Dudley [11] and Van der Vaart And Wellner [30]. This will give us a rate in the Donsker theorem. Now, the random entropy condition (6) with $\alpha = 1/2$ in probability has been shown to be almost necessary to the Donsker property of uniformly bounded pregaussian classes in Giné and Zinn [14] for classes of sets and Talagrand [28] for classes of functions. We refer to Giné and Zinn [14] for the precise meaning of "$\mathcal{F}$ is P-pregaussian" and "$\mathcal{F}$ is P-Donsker".

Theorem 2.1 (*Giné and Zinn [14], Theorems 5.1 and 5.4*). *Let $\mathcal{F}$ be a class of uniformly bounded measurable functions, then $\mathcal{F}$ is a functional P-Donsker class if both:*

(i) *$\mathcal{F}$ is P-pregaussian*
(ii) *There exists $\delta, \sigma, \varepsilon_0 > 0$ such that for all $0 < \varepsilon < \varepsilon_0$,*

$$\lim_{n\to\infty} \mathbb{P}\left(H_{n,1}(\delta\varepsilon n^{-\frac{1}{2}}, \mathcal{F}) > \sigma\varepsilon n^{\frac{1}{2}}\right) = 0.$$

Moreover, if $\mathcal{F}$ is a class of indicators of measurable sets then the condition (ii) *is also necessary.*

By Dudley and Philipp [12] conditions (i) and (ii) in turn imply that (1) holds with $v_n = o(\sqrt{\log\log n})$. Now, in the assumption (F.3) everywhere used in this article we replace ε of (ii) by $\varphi_n/\sqrt{n} \to 0$ for some φ_n and obtain that $v_n \to 0$ in (4). In light of Theorem 2.1 and Assertion 4, which says that (F.3) implies (i), we deduce that the random entropy assumption (F.3) is not only sufficient for the Donsker property of uniformly bounded pregaussian classes, but also close to necessary. In that sense we show that one can have $v_n \to 0$ in (1) under solely a weak condition of the kind (ii).

In Section 2.3 we recall the relationship between $H_{n,1}$ and $H_{[],1}$, which we define now. Given a class $\mathcal{F} \subset \mathcal{M}$, $\varepsilon > 0$ and $i = 1, 2$, let $N_{[],i}(\mathcal{F}, \varepsilon)$ denote the smallest number $N \geqslant 1$ such that we can find functions f_j^-, f_j^+, $j = 1, \dots, N$ satisfying the conditions

$$f_j^- \leqslant f_j^+, \quad d_{P,i}(f_j^+, f_j^-) < \varepsilon, \quad j = 1, \dots, N,$$

and the fact that for any $f \in \mathcal{F}$ there exists j, $1 \leqslant j \leqslant N$ such that $f_j^- \leqslant f \leqslant f_j^+$. Then

$$H_{[],i}(\mathcal{F}, \varepsilon) = \log N_{[],i}(\mathcal{F}, \varepsilon), \quad \varepsilon > 0$$

is the ε-entropy with bracketing of the class $\mathcal{F}$.

2.2. *Basic assumptions*

We shall assume that $\mathcal{F}$ satisfies conditions among the following.

(F.1) For some $M > 0$, for all $f \in \mathcal{F}$, $\|f\|_{\mathcal{X}} = \sup_{x\in\mathcal{X}} |f(x)| \leqslant M$.

(F.2) There exists a countable subclass $\mathcal{F}_\infty$ of $\mathcal{F}$ such that we can find for any function $f \in \mathcal{F}$ a sequence of functions $(f_m)_{m\geqslant 1}$ in $\mathcal{F}_\infty$ for which $\lim_{m\to\infty} f_m(x) = f(x)$ for all $x \in \mathcal{X}$.

(F.3) There exists $(\varphi_n)_{n\geqslant 1}$ and $\alpha \in (0, 1/2)$, $\varphi_n = O(n^\alpha)$, and $(\beta_n)_{n\geqslant 1}$, $\liminf_{n\to\infty} \beta_n > 1$, such that for all $\lambda > 1$, we can find $\sigma_0 > 0$ such that for all $n > 0$

$$\mathbb{P}(H_{n,1}(\mathcal{F}, n^{-1}\sigma_0\varphi_n) \geqslant \sigma_0\varphi_n) \leqslant \beta_n^{-\lambda}. \tag{8}$$

(F.4) For some $\alpha \in (0, 1/2)$ we have

$$H_{d_{P,2}}(\mathcal{F}, \delta) = O(\delta^{-\frac{\alpha}{1-\alpha}}), \quad \delta \to 0.$$

(F.5) For some $\alpha \in (0, 1/2)$ we have

$$H_{[],1}(\mathcal{F}, \delta) = O(\delta^{-\frac{\alpha}{1-\alpha}}), \quad \delta \to 0.$$

(F.6) For some $\rho > 0$, there exist $v > 0$ and $\theta > 0$, such that we have

$$H_{d_{P,2}}(\mathcal{F}, \delta) \leq v \log^\rho \left(\frac{\theta}{\delta}\right), \quad \delta \to 0.$$

(F.7) For some $\rho > 0$, there exist $\vartheta > 0$, such that we have

$$H_{[],1}(\mathcal{F}, \delta) \leq \vartheta \log^{\rho}\left(\frac{1}{\delta}\right), \quad \delta \to 0.$$

Assumption (F.1), which allows to apply some key inequalities, and (F.2) which avoids using outer probability measures, are classical. Our main conditions are of the type (F.3)–(F.7) and are discussed in the next paragraph. The constant θ in (F.6) is used only in Assertion 3 and does not appear in our rates.

2.3. The relationship between the conditions

The aim of this section is to remind that the conditions on empirical entropy of type (F.3) imply a certain type of asymptotic behavior of the entropy of classes of functions $\mathcal{F}$ with respect to the metric $d_{P,i}$. Conversely, if one uses the metric entropy with bracketing $H_{[],i}$ instead of the usual metric entropy then the opposite statements also would hold, see Koltchinskii [20]. The relationship between the empirical entropy and the metric entropy of a class $\mathcal{F}$ with respect to $d_{P,i}$ is recalled in Lemmas 2.2 and 2.3 and the subsequent assertions.

Let $(\varphi_n)_{n\geqslant 1}$ be an increasing sequence of positive real numbers and $\alpha \in (0, 1/2)$ such that

$$\varphi_n = O(n^{\alpha}), \quad n \to \infty.$$

Assume (F.1) and let $M(\mathcal{F}, \delta)$, called the capacity of $\mathcal{F}$, denote the maximal size of a finite subset $\mathcal{F}' \subset \mathcal{F}$ such that

$$\min\{d_{P,1}(f, g) : f \neq g, f, g \in \mathcal{F}'\} \geqslant 2\delta.$$

For all $\tau > 0$, let $[\tau]$ be the greatest integer smaller than $\tau > 0$. We put

$$m(\delta) = [17\delta^{-1} \log M(\mathcal{F}, \delta)].$$

The following lemma gives a necessary condition to (F.3) in terms of the capacity of $\mathcal{F}$. Keep in mind the strong relation between the metric entropy and the capacity,

$$H_{d_{P,1}}(\mathcal{F}, 2\delta) \leq \log M(\mathcal{F}, \delta) \leq H_{d_{P,1}}(\mathcal{F}, \delta) \tag{9}$$

see e.g. Lifshits [24] where conditions are formulated in terms of capacity.

Lemma 2.2. *Under* (F.3) *we have*

$$\log M(\mathcal{F}, \delta) \leq \sigma_0 \varphi_{m(\delta)} + \frac{\delta}{17}, \quad \delta \to 0.$$

Proof. This is easily deduced from the proof of Lemma 8.1 in Koltchinskii [21]. □

The following two immediate consequences of Lemma 2.2 will be useful.

Assertion 1. (i) *If* (F.3) *is true with* $\varphi_n = O(n^{\alpha})$, $\alpha \in (0, 1/2)$ *then*

$$H_{d_{P,1}}(\mathcal{F}, \delta) = O(\delta^{-\frac{\alpha}{1-\alpha}}), \quad \delta \to 0. \tag{10}$$

(ii) *If* (F.3) *is true with* $\varphi_n = \log^{\rho} n$, $\rho > 0$ *then*

$$H_{d_{P,1}}(\mathcal{F}, \delta) \leq 2\sigma_0 \log^{\rho}\left(\frac{34}{\delta}\right), \quad \delta \to 0. \tag{11}$$

Next we state a sufficient condition for (F.3) to hold, in terms of the bracketing entropy in $L_1(\mathcal{X}, dP)$.

Lemma 2.3. *Let* $(\varphi_n)_{n\geq 1}$ *and* $(\beta_n)_{n\geq 1}$ *be two sequences such that* $0 < k \log \beta_n \leq \varphi_n$, $k > 0$ *and* $\varphi_n \leq n^\alpha$, $\alpha \in (0, 1/2)$. *If there exists* $\kappa < 1/4$ *and* $\sigma_1 > 0$ *such that*

$$H_{[],1}(\mathcal{F}, \sigma_1 n^{-1}\varphi_n) \leq \kappa\sigma_1\varphi_n, \tag{12}$$

then (F.3) *is satisfied with* $\sigma_0 = \max(\sigma_1, (\kappa - 1/4)k\lambda)$.

Proof. It is similar to the proof of Theorem 4.1 in Giné and Zinn [14] and Section 8 in Koltchinskii [21]. □

The following consequences of Lemma 2.3 provide sufficient conditions to (F.3).

Assertion 2. (i) *If* (F.5) *holds then* (F.3) *is satisfied with* $\varphi_n = n^\alpha$ *and* $\beta_n = n$.
(ii) *If* (F.7) *holds then* (F.3) *is satisfied with* $\varphi_n = \log^\rho n$ *and* $\beta_n = \exp(\log^\rho n)$.

The following statement easily follows from Assertion 1, (F.1) and the fact that

$$H_{d_{P,2}}(\mathcal{F}, (2M\delta)^{\frac{1}{2}}) \leq H_{d_{P,1}}(\mathcal{F}, \delta) \leq H_{[],1}(\mathcal{F}, 2\delta).$$

Assertion 3. (i) *If* (F.3) *is true then*

$$H_{d_{P,2}}(\mathcal{F}, \delta) = O(\delta^{-\frac{2\alpha}{1-\alpha}}), \quad \delta \to 0. \tag{13}$$

(ii) (F.3) *with* $\varphi_n = \log^\rho n$ *implies* (F.6) *with* $v = 2^{\rho+1}\sigma_0$ *and* $\theta = 8M^{1/2}$.
(iii) (F.5) *implies also* (13).
(iv) (F.7) *implies* (F.6) *with* $v = 2^\rho \vartheta$ *and* $\theta = M^{1/2}$.

In all this paper we suppose that the condition (F.3) holds with $\varphi_n = O(n^\alpha)$, which implies that (13) is always true, so if $\alpha \in (0, 1/2)$ then there exists $C > 0$ such that

$$\begin{aligned}\int_0^1 \sqrt{H_{d_{P,2}}(\mathcal{F}, \delta)}\mathrm{d}\delta &< C \int_0^1 \delta^{-\frac{\alpha}{1-\alpha}}\mathrm{d}\delta \\ &= C\frac{1-\alpha}{1-2\alpha} \\ &< \infty,\end{aligned}$$

and this implies the existence of a separable, bounded, $d_{P,2}$-uniformly continuous modification of the Gaussian random function B, see Dudley [10]. More precisely we have the following Assertion.

Assertion 4. *If* (F.1) *and* (F.3) *hold, then* $\mathcal{F}$ *is* P*-pregaussian.*

3. Main results

We now state our Gaussian approximation results under various entropy conditions. First, the polynomial entropy case. Thereafter, as well as in the remainder of the paper, we assume that $\alpha \in (0, 1/2)$ while $\sigma = \sigma(\lambda)$ denote constants depending on a fixed λ but not necessarily the same at each occurrence.

Theorem 3.1. *If* (F.1), (F.2) *and* (F.3) *hold with* $\varphi_n = n^\alpha$, $\beta_n = n$, *then for each* $\lambda > 0$ *there exists a* $\sigma > 0$ *such that for each integer* $n \geqslant 1$ *one can construct on the same probability space* $(\Omega, \Sigma, \mathbb{P})$ *the i.i.d. random variables* $X_1, \dots, X_n$ *and a version* B_n *of* B *such that*

$$\mathbb{P}\left(\|\alpha_n - B_n\|_{\mathcal{F}} \geqslant \sigma(\log n)^{-\gamma}\right) \leqslant n^{-\lambda}, \tag{14}$$

where

$$\gamma = \frac{1-2\alpha}{2\alpha}. \tag{15}$$

Moreover, if (F.4) *is further assumed then*

$$\gamma = \frac{2-3\alpha}{2\alpha}. \tag{16}$$

As a consequence of Theorem 3.1, (4) holds with $v_n = (\log n)^{-\gamma}$ which is better than the rate $o(\sqrt{\log\log n})$ obtained by Dudley and Philipp [12] under no condition on the P-Donsker class $\mathcal{F}$. By applying y Assertions 1–3, it is easy to check that Theorem 3.1 leads to the following corollary which is in accordance with Proposition 2 in Berthet and Mason [2].

Corollary 3.2. *Under* (F.1), (F.2) *then for each* $\lambda > 0$ *there exists a* $\sigma > 0$ *such that for each integer* $n \geqslant 1$ *one can construct on the same probability space* $(\Omega, \Sigma, \mathbb{P})$ *the i.i.d. random variables* $X_1, \dots, X_n$ *and a version* B_n *of* B *such that*

$$\mathbb{P}\left(\|\alpha_n - B_n\|_{\mathcal{F}} \geqslant \sigma(\log n)^{-\gamma}\right) \leqslant n^{-\lambda},$$

where

$$\gamma = \begin{cases} \dfrac{1-2\alpha}{2\alpha} & \textit{under } (\mathrm{F}.5) \\ \dfrac{2-3\alpha}{2\alpha} & \textit{under } (\mathrm{F}.4) \textit{ and } (\mathrm{F}.5). \end{cases}$$

The following corollary interpolates between (15) and (16) when a kind of margin condition is assumed in addition to (F.3), in the sense that the variance is controlled by the expectation as in Berthet and Saumard [4].

Corollary 3.3. *If* (F.1), (F.2) *and* (F.3) *hold with* $\varphi_n = n^\alpha$, $\alpha \in (0, 1/2)$, $\beta_n = n$, *and for* $\gamma_0 \in (1, 2)$, $R > 0$,

$$d_{P,2}(f, g)^{\gamma_0} \leq R d_{P,1}(f, g), \quad f, g \in \mathcal{F} \tag{17}$$

then for each $\lambda > 0$ *there exists a* $\sigma > 0$ *such that for each integer* $n \geqslant 1$ *one can construct on the same probability space* $(\Omega, \Sigma, \mathbb{P})$ *the i.i.d. random variables* $X_1, \dots, X_n$ *and a version* B_n *of* B *such that*

$$\mathbb{P}\left(\|\alpha_n - B_n\|_{\mathcal{F}} \geqslant \sigma(\log n)^{-\gamma}\right) \leqslant n^{-\lambda}, \tag{18}$$

where

$$\gamma = \frac{1-\alpha}{\gamma_0\alpha} - \frac{1}{2}.$$

Now let us consider the case of a logarithmic growth of the entropy $H_{d_{P,i}}(\mathcal{F}, .)$. In the first statement of the following theorem we assume that (F.3) is satisfied with $\varphi_n = \log n$ and $\beta_n = n$. By Assertion 3 this implies that (F.6) holds with $\rho = 1$, $v = 4\sigma_0$ and $\theta = 8M^{1/2}$.

Theorem 3.4. *If* (F.1), (F.2) *and* (F.3) *are satisfied with* $\varphi_n = \log n$ *and* $\beta_n = n$ *then for each* $\lambda > 0$ *there exists a* $\sigma > 0$ *such that for each integer* $n \geqslant 1$ *one can construct on the same probability space* $(\Omega, \Sigma, \mathbb{P})$ *the i.i.d. random variables* $X_1, \ldots, X_n$ *and a version* B_n *of* B *such that*

$$\mathbb{P}\left(\|\alpha_n - B_n\|_{\mathcal{F}} \geqslant \sigma n^{-\frac{1}{2+5v}} (\log n)^{\frac{5v}{4+10v}}\right) \leqslant n^{-\lambda},$$

where the constant $v > 0$ *comes from* (F.6).

Moreover, if (F.6) *is further assumed with* $\rho < 1$, $v > 0$ *and* $\theta > 0$ *we have*

$$\mathbb{P}\left(\|\alpha_n - B_n\|_{\mathcal{F}} \geqslant \sigma n^{-\frac{1}{2}} \exp\left(\frac{5v}{2^{\rho+1}} \log^{\rho} n\right) \log n\right) \leqslant n^{-\lambda}.$$

If (F.1), (F.2) *and* (F.3) *are assumed with* $\varphi_n = (\log n)^{\rho}$, $\rho > 1$ *and* $\beta_n = n$ *then for any* $c < \frac{1}{2^{\rho}5}$

$$\mathbb{P}\left(\|\alpha_n - B_n\|_{\mathcal{F}} \geqslant \sigma \exp(-(c \log n)^{1/\rho})\right) \leqslant n^{-\lambda}.$$

The rate obtained in the previously mentioned results of Koltchinskii and Rio are strictly better than Theorem 3.4 at the expense of assuming either a good Haar approximation or the VC-property and uniformly bounded variation of the class $\mathcal{F}$. In particular, Theorem 3.4 of Koltchinskii [21] provides under a stronger version of (F.3) a rate that can be as fast as $n^{-1/2}(\log n)^{\rho}$ in case of a very sharp Haar representation. On the opposite, when relaxing any additional conditions our Theorem 3.4 still yields a rate which can be polynomial and thus far better than (14).

Likewise, by applying Assertions 1–3, it is easily checked that Theorem 3.4 leads to the following corollary which extends Proposition 1 in Berthet and Mason [2]. Recall that (F.7) implies (F.6) with $v = 2^{\rho}\vartheta$ and $\theta = M^{1/2}$.

Corollary 3.5. *Under* (F.1), (F.2) *and* (F.7) *then for each* $\lambda > 0$ *there exists a* $\sigma > 0$ *such that for each integer* $n \geqslant 1$ *one can construct on the same probability space* $(\Omega, \Sigma, \mathbb{P})$ *the i.i.d. random variables* $X_1, \ldots, X_n$ *and a version* B_n *of* B *such that*

$$\mathbb{P}(\|\alpha_n - B_n\|_{\mathcal{F}} \geqslant \sigma v_n) \leqslant n^{-\lambda},$$

where

$$v_n = \begin{cases} n^{-\frac{1}{2}} \exp\left(\dfrac{5v}{2^{\rho+1}} \log^{\rho} n\right) \log n, & \text{if } \rho < 1, \\ \exp(-(c \log n)^{1/\rho}), \quad \text{any } c < \dfrac{1}{52^{\rho}} & \text{if } \rho > 1, \\ n^{-\frac{1}{2+5v}} (\log n)^{\frac{5v}{4+10v}}, & \text{if } \rho = 1. \end{cases}$$

4. Proof of a more general result

4.1. General case

In what follows there exists a probability space $(\Omega, \Sigma, \mathbb{P})$ which is rich enough for all our needs. We make use of $H_{d_{P,2}}(\mathcal{F}, .)$ and λ_δ as defined in Section 2.1 under $d = d_{P,2}$. Our main statement is fairly general and can be stated as follows. Recall that σ may change at each occurrence.

Proposition 4.1. *Suppose that* (F.1), (F.2) *and* (F.3) *hold with* $(\varphi_n)_{n\geqslant 1}$ *and* $(\beta_n)_{n\geqslant 1}$ *satisfying*

$$\log\log n = O(\beta_n), \qquad \log\beta_n = O(\varphi_n), \quad n \to \infty. \tag{19}$$

Let $(\delta_n)_{n\geqslant 1}$ *and* $(\phi_n)_{n\geqslant 1}$ *be such that for each* $\lambda > 0$ *there exists a* $\sigma > 0$ *such that for each* $n \geqslant 1$

$$\mathbb{P}\left(\|\alpha_n \circ \lambda_{\sqrt{\frac{\varphi_n}{n}}} - \alpha_n \circ \lambda_{\delta_n}\|_{\mathcal{F}} \geq \sigma\phi_n\right) \leq \beta_n^{-\lambda}. \tag{20}$$

Let $(\psi_n)_{n\geqslant 1}$ *is a sequence satisfying that for each* $\lambda > 0$ *there exists* $\sigma > 0$ *such that for each* $n \geqslant 1$,

$$\exp\left(\frac{5}{2}a_n\right)(2a_n + \lambda\log\beta_n) \leq \sigma\psi_n, \quad \textit{where } a_n = H_{d_{P,2}}(\mathcal{F}, \delta_n). \tag{21}$$

Then for each $\lambda > 0$ *there exists a* $\sigma > 0$ *such that for each integer* $n \geqslant 1$ *one can construct on the same probability space* $(\Omega, \Sigma, \mathbb{P})$ *the i.i.d. random variables* $X_1, \ldots, X_n$ *and a version* B_n *of* B *such that*

$$\mathbb{P}(\|\alpha_n - B_n\|_{\mathcal{F}} \geq \sigma\upsilon_n) \leq \beta_n^{-\lambda}, \tag{22}$$

where σ *depends only on* λ, $\mathcal{F}$ *and* P, *and*

$$\upsilon_n = \max\left(\frac{\varphi_n}{\sqrt{n}}, \frac{\psi_n}{\sqrt{n}}, \phi_n, \delta_n\max(\sqrt{a_n}, \sqrt{\log\beta_n})\right).$$

The technical assumption (20) is necessary to derive the results of Section 3 since it is not always sufficient to apply Proposition 4.1 with $\delta_n = \sqrt{\varphi_n/n}$ and $\phi_n = 0$. The reason is that for certain choices of β_n and φ_n we cannot find a sequence ψ_n satisfying (21) such that $\psi_n = O(\sqrt{n})$ and thus $\upsilon_n \to 0$. Let us discuss this point.

On one hand, assuming that $\upsilon_n \to \infty$ as $n \to \infty$, condition (21) implies that

$$\exp\left(\frac{5}{2}a_n\right) \leq \sigma\psi_n \leq \sqrt{n},$$

and hence

$$a_n = H_{d_{P,2}}(\mathcal{F}, \delta_n) = O(\log n), \tag{23}$$

which means that Proposition 4.1 is only applicable when a_n is logarithmic in n. An example is given by Theorem 3.4 where the entropy has a logarithmic growth.

On the other hand, if we take $\delta_n = \sqrt{\varphi_n/n} = n^{(\alpha-1)/2}$ where $\alpha < 1/2$ and $\phi_n = 0$ then according to (13) and (F.1) there exists $K > 0$ such that $H_{d_{P,2}}(\mathcal{F}, \delta) \leq K\delta^{-2\alpha/(1-\alpha)}$. Hence $a_n = H_{d_{P,2}}(n^{(\alpha-1)/2}) = O(n^{-\alpha})$ is not logarithmic and Proposition 4.1 is not applicable without changing the projection $\sqrt{\varphi_n/n}$. Indeed, whenever $H_{d_{P,2}}(\mathcal{F}, \sqrt{\varphi_n/n})$ is not logarithmic, we shall turn the ray $\sqrt{\varphi_n/n}$ into $\delta_n = r(\log n)^{-\beta}$ where $\beta > 0$ and $r > 0$ by making use of a chaining technique. An example is given by Theorem 3.1 where the entropy has a polynomial growth.

4.2. *Proof of* Proposition 4.1

First case. Let us start with the easiest case $\delta_n = \sqrt{\varphi_n/n}$, so that (20) trivially holds. Consider a finite ε-net $\mathcal{F}_\varepsilon$ of $\mathcal{F}$ with respect to $d_{P,2}$

$$\mathcal{F}_\varepsilon = \{f_k : 1 \le k \le N_{d_{P,2}}(F, \varepsilon)\}$$

and the projection λ_ε under $d_{P,2}$ such that $d_{P,2}(\lambda_\varepsilon f, f) < \varepsilon$, $f \in \mathcal{F}$. Given $\varepsilon > 0$ and $n \ge 1$, our aim is to construct a probability space $(\Omega, \Sigma, \mathbb{P})$ on which sit $X_1, \ldots, X_n$ and a version B_n of the P-Brownian bridge B indexed by $\mathcal{F}$ such that for $\mathcal{F}_\varepsilon$ defined as above and for some $\delta_1 > 0$, $\delta_2 > 0$ and $\delta_3 > 0$ to be chosen such that $\max(\delta_1, \delta_2, \delta_3) = \sigma\upsilon_n$ from (22) we have

$$\begin{aligned}\mathbb{P}\left(\|\alpha_n - B_n\|_{\mathcal{F}} \ge \max\left(\delta_1, \delta_2, \delta_3\right)\right) &\le \mathbb{P}\left(\|\alpha_n - \alpha_n \circ \lambda_\varepsilon\|_{\mathcal{F}} \ge \delta_1\right) \\ &\quad + \mathbb{P}\left(\|\alpha_n \circ \lambda_\varepsilon - B_n \circ \lambda_\varepsilon\|_{\mathcal{F}} \ge \delta_2\right) + \mathbb{P}\left(\|B_n \circ \lambda_\epsilon - B_n\|_{\mathcal{F}} \ge \delta_3\right) \qquad (24)\end{aligned}$$

with all these probabilities simultaneously as small as $\beta_n^{-\lambda}$ for a suitably chosen $\varepsilon > 0$.

Our construction is based on some estimates of the approximation of the empirical process by the Brownian bridge on finite classes of functions. These estimates rely on the following coupling of Zaitzev [33] as in Berthet and Mason [2,3]. Let $|.|_N$, $N \ge 1$, denote the usual Euclidean norm on $\mathbb{R}^N$.

Coupling inequality. *Let $Y_1, \ldots, Y_n$ be independent mean zero random variables in $\mathbb{R}^N$, $N > 0$, such that for some $B_1 > 0$,*

$$|Y_i|_N \le B_1, \quad i = 1, \ldots, n.$$

If $(\Omega, \Sigma, \mathbb{P})$ is rich enough then for each $\delta > 0$, one can define independent normally distributed mean zero random variables $Z_1, \ldots, Z_n$ with Z_i and Y_i having the same covariance matrix for $i = 1, \ldots, n$, such that for universal constants C_1 and C_2

$$\mathbb{P}\left\{\left|\sum_{i=1}^{n}(Y_i - Z_i)\right|_N \ge \delta\right\} \le C_1 N^2 \exp\left(\frac{-C_2\delta}{N^2 B_1}\right). \qquad (25)$$

Proof. See Zaitzev [33] and Einmahl and Mason [13]. □

Consider the n i.i.d. mean zero random variables in $\mathbb{R}^N$, where $N = N_{d_{P,2}}(\mathcal{F}, \varepsilon)$,

$$Y_i = \frac{1}{\sqrt{n}}\left(f_1(X_i) - E(f_1(X)), \ldots, f_N(X_i) - E(f_N(X))\right), \quad 1 \le i \le n.$$

Since $f_k \in \mathcal{F}$ and (F.1) holds, we have

$$|Y_i|_N \le 2M\sqrt{\frac{N}{n}}, \quad 1 \le i \le n.$$

Therefore, by the coupling inequality (25) we can construct $Y_1, \ldots, Y_n$ i.i.d. $Y := (Y^1, \ldots, Y^N)$ having the previously defined law and $Z_1, \ldots, Z_n$ i.i.d. $Z := (Z^1, \ldots, Z^N)$ mean zero Gaussian variables on the same probability space such that for positive constants C_1 and C_3 we have

$$\mathbb{P}\left(\|\alpha_n \circ \lambda_\varepsilon - B_n \circ \lambda_\varepsilon\|_{\mathcal{F}} \geq \delta_2\right) \leq \mathbb{P}\left\{\left|\sum_{i=1}^{n}(Y_i - Z_i)\right|_N \geq \delta_2\right\} \leq C_1 N^2 \exp\left(\frac{-C_3\sqrt{n}\delta_2}{N^{5/2}}\right), \tag{26}$$

where $\mathrm{cov}(Z^l, Z^k) = \mathrm{cov}(Y^l, Y^k) = \langle f_l, f_k\rangle$ and

$$B_n(f_k) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i^k.$$

Moreover by Lemma A1 of Berkes and Philipp [1] this space can be extended to include a P-Brownian bridge B_n indexed by $\mathcal{F}$ taking the above values on $\mathcal{F}_\varepsilon$ and we can again extend this space to support $X_1, \ldots, X_n$ such that the Y_i have the above representation in terms of X_i. The terms in (24) are now using this B_n and this α_n. Notice that the probability space on which $Y_1, \ldots, Y_n, Z_1, \ldots, Z_n, B_n$ and $X_1, \ldots, X_n$ eventually sit together depends on $n \geq 1$ and on the choice of $\varepsilon > 0$ and $\delta_2 > 0$ to be made later.

Next we control the modulus of continuity of α_n in the right hand term of (24) by making use of the following inequality. We can not apply Talagrand inequality as in Berthet and Mason [2] since we do not assume uniform or bracketing entropy conditions in order to work only under the weaker (F.3). Let $\log_2(x)$ denote the inverse of 2^x.

Lemma 4.2. *Let $\mathcal{C}$ be a class satisfying the conditions* (F.1) *with $M = 1/2$ and* (F.2). *If*

$$n \sup_{g\in\mathcal{C}} P(g^2) \leq u, \tag{27}$$

then

$$\mathbb{P}\left(\sqrt{n}\|\alpha_n\|_{\mathcal{C}} \geq u\right) \leq 8(\log_2\log_2(2nu^{-1}) + 1)\exp(-2^{11}u) + \mathbb{P}\left(H_{n,1}(\mathcal{C}, 2^{-4}un^{-1}) \geq 2^{-11}u\right).$$

Proof. Similar statements were proved in Lemma 3 in Koltchinskii [17] and in Theorem 3.1 in Giné and Zinn [14]. □

We shall apply this inequality to the class $\mathcal{C} = \{(4M)^{-1}(f - \lambda_\varepsilon f), f \in \mathcal{F}\}$. Observe that the class $\mathcal{C}$ satisfies (F.1), (F.2) and for any $\delta > 0$,

$$H_{n,1}(\mathcal{C}, \delta) \leq 2H_{n,1}\left(\mathcal{F}, \frac{M\delta}{2}\right), \quad \mathbb{P}\text{-a.s.}$$

which ensures that the condition (F.3) is also fulfilled by $\mathcal{C}$. Moreover we have (27) with $u = n\varepsilon^2/16M^2 = \sigma n\varepsilon^2$. Here, as well as in the remainder of the proofs, σ denotes a constant, not necessarily the same at each occurrence. Let $(\varphi_n)_{n\geqslant 1}$ and $(\beta_n)_{n\geqslant 1}$ be two sequences satisfying (19). By taking $u = \sigma\varphi_n$ and $\varphi_n = n\varepsilon^2$ in Lemma 4.2 we have, for σ large enough

$$\mathbb{P}\left(\sqrt{n}\|\alpha_n - \alpha_n \circ \lambda_{\sqrt{\frac{\varphi_n}{n}}}\|_{\mathcal{F}} \geq \sigma\varphi_n\right) \leq 8\left(\log_2\log_2(2n(\sigma\varphi_n)^{-1}) + 1\right)\exp(-2^{11}\sigma\varphi_n) + \mathbb{P}\left(H_{n,1}\left(\mathcal{C}, 2^{-4}\sigma\varphi_n n^{-1}\right) \geq 2^{-11}\sigma\varphi_n\right). \tag{28}$$

From (19), (28) and (F.3), we deduce that for each $\lambda > 0$ there exists $\sigma > 0$ such that for $n > 0$,

$$\mathbb{P}\left(\|\alpha_n - \alpha_n \circ \lambda_{\sqrt{\frac{\varphi_n}{n}}}\|_{\mathcal{F}} \geq \sigma \frac{\varphi_n}{\sqrt{n}}\right) \leq \beta_n^{-\lambda}. \tag{29}$$

Having now determined $\delta_1 = \sigma\varphi_n/\sqrt{n}$ and $\varepsilon = \sqrt{\varphi_n/n}$ in (24) let us take $\delta_2 = \sigma\psi_n/\sqrt{n}$ in (26) where ψ_n is as in (21), so that

$$\begin{aligned}
&\mathbb{P}\left(\|\alpha_n \circ \lambda_{\sqrt{\frac{\varphi_n}{n}}} - B_n \circ \lambda_{\sqrt{\frac{\varphi_n}{n}}}\|_{\mathcal{F}} \geq \sigma \frac{\psi_n}{\sqrt{n}}\right) \\
&\leq C_1 \exp\left\{2H_{d_{P,2}}\left(\mathcal{F}, \sqrt{\frac{\varphi_n}{n}}\right) - C_3\sigma\psi_n \exp\left(-\frac{5}{2}H_{d_{P,2}}\left(\mathcal{F}, \sqrt{\frac{\varphi_n}{n}}\right)\right)\right\} \\
&\leq C_1 \exp\left\{2a_n - C_3\sigma\psi_n \exp\left(-\frac{5}{2}a_n\right)\right\}.
\end{aligned}$$

According to (21), for each $\lambda > 0$ there exists $\sigma > 0$ such that, for any $n > 0$,

$$\mathbb{P}\left(\|\alpha_n \circ \lambda_{\sqrt{\frac{\varphi_n}{n}}} - B_n \circ \lambda_{\sqrt{\frac{\varphi_n}{n}}}\|_{\mathcal{F}} \geq \sigma \frac{\psi_n}{\sqrt{n}}\right) \leq \beta_n^{-\lambda}. \tag{30}$$

It remains to control the third term in (24) with some well known inequality for the modulus of continuity of B. Given a class $\mathcal{C}$ write $\sigma_{\mathcal{C}}^2 = \sup_{h\in\mathcal{C}} E(B^2(h))$. Recall the following concentration probability estimate for $\|B\|_{\mathcal{C}}$ due to Borell [5], then the moment bound of Dudley [10].

Borell's inequality. *For all $t > 0$,*

$$\mathbb{P}(\|B\|_{\mathcal{C}} - E(\|B\|_{\mathcal{C}}) > t) \leq 2\exp\left(-\frac{t^2}{2\sigma_{\mathcal{C}}^2}\right). \tag{31}$$

Gaussian moment inequality. *For some universal constant $A_0 > 0$ and all $\delta > 0$,*

$$E\left(\sup_{d_{P,2}(f,g)<\delta}\{|B(f) - B(g)| : f, g \in \mathcal{F}\}\right) \leq A_0 \int_{[0,\delta]} \sqrt{H_{d_{P,2}}(\mathcal{F}, \varepsilon)}\mathrm{d}\varepsilon < \infty. \tag{32}$$

Let us now consider $\mathcal{C} = \{f - g, d_{P,2}(f, g) \leq \sqrt{\varphi_n/n}, f, g \in \mathcal{F}\}$ for which we have

$$\sigma_{\mathcal{C}}^2 = \sup_{d_2(f,g)\leq\sqrt{\frac{\varphi_n}{n}}} E(B(f-g)^2) \leq \sup_{d_2(f,g)\leq\sqrt{\frac{\varphi_n}{n}}} d_{P,2}^2(f, g) \leq \frac{\varphi_n}{n},$$

and, by (32), for $A_1 > A_0$,

$$E(\|B\|_{\mathcal{C}}) \leq A_0 \int_{[0,\sqrt{\frac{\varphi_n}{n}}]} \sqrt{H_{d_{P,2}}(\mathcal{F}, \varepsilon)}\mathrm{d}\varepsilon \leq A_1\sqrt{\frac{\varphi_n}{n}}\sqrt{a_n}.$$

Hence (31) gives, for all $t > 0$,

$$\mathbb{P}\left(\sqrt{n}\|B - B \circ \lambda_{\sqrt{\frac{\varphi_n}{n}}}\|_{\mathcal{F}} \geq A_1\sqrt{\frac{\varphi_n}{n}}\sqrt{a_n} + t\right) \leq 2\exp\left(-\frac{nt^2}{2\varphi_n}\right).$$

By taking

$$t = \sigma\sqrt{\frac{\varphi_n}{n}}\max\left(\sqrt{a_n}, \sqrt{\log\beta_n}\right)$$

we finally get that for each $\lambda > 0$ there exists $\sigma > 0$ such that for $n > 0$, the third term in (24) satisfies

$$\mathbb{P}\left(\|B_n - B_n \circ \lambda_{\sqrt{\frac{\varphi_n}{n}}}\|_{\mathcal{F}} \geq \sigma\sqrt{\frac{\varphi_n}{n}}\max\left(\sqrt{a_n}, \sqrt{\log\beta_n}\right)\right) \leq \beta_n^{-\lambda}, \tag{33}$$

where $\sigma > 0$ depends only on $\mathcal{F}$, φ_n and β_n from (19). Combining (24), (29), (30) and (33), it becomes clear that there exists a $\sigma > 0$ such that (22) holds with $\phi_n = 0$.

Second case. If $\delta_n \neq \sqrt{\varphi_n/n}$ in (20) we use the following decomposition

$$\begin{aligned}\|\alpha_n - B_n\|_{\mathcal{F}} \leq{} & \|\alpha_n - \alpha_n \circ \lambda_{\sqrt{\frac{\varphi_n}{n}}}\|_{\mathcal{F}} + \|\alpha_n \circ \lambda_{\sqrt{\frac{\varphi_n}{n}}} - \alpha_n \circ \lambda_{\delta_n}\|_{\mathcal{F}} \\ & + \|\alpha_n \circ \lambda_{\delta_n} - B_n \circ \lambda_{\delta_n}\|_{\mathcal{F}} + \|B_n - B_n \circ \lambda_{\delta_n}\|_{\mathcal{F}}\end{aligned} \tag{34}$$

where we can immediately bound the first term by using (29) and the second term by using (20). Condition (21) then allows us to replace $\sqrt{\varphi_n/n}$ by δ_n in the computation leading to (30), so we get for each $\lambda > 0$ that there exists $\sigma > 0$ such that for $n > 0$,

$$\mathbb{P}\left(\|\alpha_n \circ \lambda_{\delta_n} - B_n \circ \lambda_{\delta_n}\|_{\mathcal{F}} \geq \sigma\frac{\psi_n}{\sqrt{n}}\right) \leq \beta_n^{-\lambda}. \tag{35}$$

As for (33) we have

$$\mathbb{P}\left(\|B_n - B_n \circ \lambda_{\delta_n}\|_{\mathcal{F}} \geq \sigma\delta_n\max(\sqrt{a_n}, \sqrt{\log\beta_n})\right) \leq \beta_n^{-\lambda}. \tag{36}$$

From (20), (29) and (34)–(36) we conclude that for each $\lambda > 0$ there exists a $\sigma > 0$ such that (22) holds. $\square$

5. Other proofs

Now we make Proposition 4.1 more explicit in the special cases of Section 3.

5.1. *Proof of* Theorem 3.1

Proof of the first assertion. According to the explanation given in Section 4.1, the proof is an application of Proposition 4.1. For this we shall fit δ_n to the ray of projection $\sqrt{\varphi_n/n}$ by using a chaining argument. From (21) and (i) of Assertion 1 we conclude that $a_n = H_{d_{P,2}}(\mathcal{F}, \delta_n)$ is logarithmic in n, and polynomial in δ_n. Therefore, δ_n is logarithmic in n. Let us put

$$\delta_n = r(\log n)^{-\beta}, \quad \beta > 0, r > 0. \tag{37}$$

But if $\beta_n = n$ while δ_n and a_n are logarithmic in n, the term $\delta_n \max(\sqrt{a_n}, \sqrt{\log\beta_n})$ in (22) is logarithmic in n, so there is no restriction to look for a $\gamma > 0$ for which we can take

$$\frac{\varphi_n}{\sqrt{n}} = \frac{\psi_n}{\sqrt{n}} = \phi_n = (\log n)^{-\gamma}, \quad \gamma > 0. \tag{38}$$

Also note that (F.3) being satisfied with $\varphi_n = n^\alpha$ and $\beta_n = n$ it is also satisfied with $\varphi_n = \sqrt{n}(\log n)^{-\gamma}$ and $\beta_n = n$ since $H_{n,1}(\mathcal{F}, \varepsilon)$ is a decreasing function of ε. Now, from

(13) and (37) we have for each $\lambda > 0$

$$\begin{aligned} &\exp\left(\frac{5}{2}a_n\right)(2a_n + \lambda \log \beta_n) \\ &\quad \le \exp\left(\frac{5}{2}Kr^{-\frac{2\alpha}{1-\alpha}}(\log n)^{\frac{2\alpha\beta}{1-\alpha}}\right)\left(2Kr^{-\frac{2\alpha}{1-\alpha}}(\log n)^{\frac{2\alpha\beta}{1-\alpha}} + \lambda \log n\right), \end{aligned} \tag{39}$$

so by choosing

$$\beta = \frac{1-\alpha}{2\alpha}$$

then r large enough, it is clear that (21) holds with $\psi_n = \sqrt{n}(\log n)^{-\gamma}$ for any desired $\gamma > 0$ and some fixed $\sigma = \sigma(r, \gamma)$. The remainder of the proof consists in checking condition (20). Thus, we intend to show that for each $\lambda > 0$ there exists $\sigma = \sigma(\lambda, \alpha, r) > 0$ such that

$$P_1 = \mathbb{P}\left(\|\alpha_n \circ \lambda_{n^{-1/4}(\log n)^{-\frac{\gamma}{2}}} - \alpha_n \circ \lambda_{r(\log n)^{-\beta}}\|_{\mathcal{F}} \ge \sigma(\log n)^{-\gamma}\right) \le n^{-\lambda}. \tag{40}$$

Let H and N be as in (5) with $d = d_{P,2}$. From (13) we can take

$$H(u) = \log N(u) = Ru^{-\frac{2\alpha}{1-\alpha}}. \tag{41}$$

Here $R > 0$ again denotes a constant changing at each occurrence. Let

$$\begin{aligned} &k = [\log_2(rn^{\frac{1}{4}}(\log n)^{-\beta+\frac{\gamma}{2}})], \quad \delta_k = \delta_n = r(\log n)^{-\beta}, \\ &\delta_j = 2^j n^{-\frac{1}{4}}(\log n)^{-\frac{\gamma}{2}}, \quad j = 0, 1, 2, \dots, k-1, \end{aligned}$$

then set

$$\begin{aligned} &\beta_j^2 = \max\left(H(\delta_j), \log(k - j + N(r(\log n)^{-\beta}))\right), \\ &S = \sum_{j=0}^{k-1}\delta_j\beta_j, \qquad a_j = \frac{\delta_j\beta_j}{S}, \quad j = 0, 1, 2, \dots, k-1. \end{aligned}$$

We have

$$\begin{aligned} P_1 &\le \sum_{j=0}^{k-1}\mathbb{P}\left(\|\alpha_n \circ \lambda_{\delta_{j+1}} - \alpha_n \circ \lambda_{\delta_j}\|_{\mathcal{F}} \ge \sigma a_j(\log n)^{-\gamma}\right) \\ &\le \sum_{j=0}^{k-1} N(\delta_{j+1})N(\delta_j) \sup_{d(f,g)<3\delta_j} \mathbb{P}\left(|\alpha_n(f) - \alpha_n(g)| \ge \sigma a_j(\log n)^{-\gamma}\right). \end{aligned} \tag{42}$$

By using Bernstein's inequality – see page 102 of van der Vaart and Wellner [30] – there exists $G_1 > 0$ and $G_2 > 0$ such that the j-th term in (42) is less than

$$G_1 N^2(\delta_j)\exp\left(-\frac{G_2\sigma^2 a_j^2}{\delta_j^2}(\log n)^{-2\gamma}\right) \le G_1\exp\left(\beta_j^2(2 - \frac{G_2\sigma^2}{S^2}(\log n)^{-2\gamma})\right). \tag{43}$$

Let us prove that there exists $R > 0$ such that

$$S \le R(\log n)^{-\beta\frac{1-2\alpha}{1-\alpha}}. \tag{44}$$

On one hand, since $\delta_j = 2^{-(k-j-1)} r(\log n)^{-\beta}$, $j = 0, 1, 2, \ldots, k-1$ and $\log k$ is log-logarithmic in n, we have

$$\begin{aligned}
&\sum_{j=0}^{k-1} \delta_j \left(\log\left(k - j + N(r(\log n)^{-\beta})\right)\right)^{\frac{1}{2}} \\
&\quad \le r(\log n)^{-\beta} H^{\frac{1}{2}}(r(\log n)^{-\beta}) \sum_{j=0}^{k-1} 2^{-(k-j-1)} \left(1 + \frac{\log(k-j)^{\frac{1}{2}}}{H^{\frac{1}{2}}(r(\log n)^{-\beta})}\right) \\
&\quad \le R(\log n)^{-\beta} H^{\frac{1}{2}}(r(\log n)^{-\beta}).
\end{aligned} \tag{45}$$

On the other hand, since $\delta_j = 2(\delta_j - \delta_{j-1})$, $j = 0, 1, 2, \ldots, k-1$, it holds

$$\sum_{j=0}^{k-1} \delta_j H^{\frac{1}{2}}(\delta_j) = 2\sum_{j=0}^{k-1} (\delta_j - \delta_{j-1}) H^{\frac{1}{2}}(\delta_j) \le 2\int_0^{r(\log n)^{-\beta}} H^{\frac{1}{2}}(u)\mathrm{d}u. \tag{46}$$

It follows from (45) and (46) that there exists R such that

$$\begin{aligned}
S &\le R(\log n)^{-\beta} H^{\frac{1}{2}}(r(\log n)^{-\beta}) + 2\int_0^{r(\log n)^{-\beta}} H^{\frac{1}{2}}(u)\mathrm{d}u \\
&\le R(\log n)^{-\beta} H^{\frac{1}{2}}(r(\log n)^{-\beta})
\end{aligned}$$

which, by (41), implies (44). We get by taking

$$\beta = \frac{1-\alpha}{2\alpha}, \qquad \gamma = \frac{1-2\alpha}{2\alpha} \tag{47}$$

from (42)–(44) that

$$\begin{aligned}
P_1 &\le \sum_{j=0}^{k-1} 2\exp\left(\beta_j^2 (2 - R\sigma^2)\right) \\
&\le 2\sum_{j=0}^{k-1} \left(k - j + N(r(\log n)^{-\beta})\right)^{2-R\sigma^2} \\
&\le 2\int_{N(r(\log n)^{-\beta})}^{\infty} u^{2-R\sigma^2}\mathrm{d}u \\
&\le \frac{2}{-3+R\sigma^2} \exp\left(R(3 - R\sigma^2)\log n\right).
\end{aligned} \tag{48}$$

We conclude by (48) that for each $\lambda > 0$ there exists $\sigma > 0$ such that (40) holds with γ and β are as in (47). Since (39) and (40) correspond to (20) and (21) of Proposition 4.1 it follows that the rate of approximation in (14) is given by

$$\begin{aligned}
v_n &= \max\left((\log n)^{-\gamma}, (\log n)^{-\beta} \max((\log n)^{\frac{\beta\alpha}{1-\alpha}}, \sqrt{\log n})\right) \\
&= (\log n)^{-\frac{1-2\alpha}{2\alpha}}.
\end{aligned}$$

The first assertion of Theorem 3.1 is then checked. □

Proof of the second assertion. If (F.4) is further satisfied then in (39), (41) and (48) we can replace $2\alpha/(1-\alpha)$ by $\alpha/(1-\alpha)$, so we can improve our choice of β and γ by taking in (38), (39) and (48)

$$\beta = \frac{1-\alpha}{\alpha}, \qquad \gamma = \frac{2-3\alpha}{2\alpha}.$$

By the above arguments and Proposition 4.1 we conclude that

$$\begin{aligned} v_n &= \max\left((\log n)^{-\gamma}, (\log n)^{-\beta} \max((\log n)^{\frac{\beta\alpha}{2(1-\alpha)}}, \sqrt{\log n})\right) \\ &= (\log n)^{-\frac{2-3\alpha}{2\alpha}}. \quad \square \end{aligned}$$

Proof of Corollary 3.3. . If (17) is satisfied then $H_{d_{P,2}}(\mathcal{F}, \varepsilon) \le H_{d_{P,1}}(\mathcal{F}, (\varepsilon/R)^{\gamma_0})$, so from (13), if (F.3) is satisfied with $\varphi_n = n^\alpha$ we have

$$H_{d_{P,2}}(\mathcal{F}, \varepsilon) = O(\varepsilon^{-\frac{\alpha\gamma_0}{1-\alpha}})$$

which implies that in (39), (41) and (48) we can replace $2\alpha/(1-\alpha)$ by $\alpha\gamma_0/(1-\alpha)$, and thus improve (38), (39) and (48) by taking

$$\beta = \frac{1-\alpha}{2\alpha\gamma_0}, \qquad \gamma = \frac{1-\alpha}{\alpha\gamma_0} - \frac{1}{2}.$$

We then obtain, by (22),

$$\begin{aligned} v_n &= \max\left((\log n)^{-\gamma}, (\log n)^{-\beta} \max((\log n)^{\frac{\beta\alpha\gamma_0}{2(1-\alpha)}}, \sqrt{\log n})\right) \\ &= (\log n)^{-\frac{1-\alpha}{\gamma_0\alpha}+\frac{1}{2}}. \quad \square \end{aligned}$$

5.2. *Proof of* Theorem 3.4

Proof of the first assertion. In this case we have $\varphi_n = \log n$ and $\beta_n = n$. Let us put

$$\delta_n = n^{-\gamma} (\log n)^{-\beta}, \qquad \gamma \ge 0, \beta \in \mathbb{R}. \tag{49}$$

We shall approximate the ray $\sqrt{\varphi_n/n}$ by showing that for each $\gamma < 1/2$, $\beta \in \mathbb{R}$ and $\lambda > 0$, there exists $\sigma(\lambda, \gamma, \beta) > 0$ such that

$$P_2 = \mathbb{P}\left(\|\alpha_n \circ \lambda_{n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}} - \alpha_n \circ \lambda_{n^{-\gamma}(\log n)^{-\beta}}\|_{\mathcal{F}} \ge \sigma n^{-\gamma} (\log n)^{-\beta+\frac{1}{2}}\right) \le n^{-\lambda}. \tag{50}$$

Let again

$$\begin{aligned} &H(u) = v \log u, \qquad k = [n^{-\gamma+\frac{1}{2}} (\log n)^{-\beta-\frac{1}{2}}], \\ &\delta_k = n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}}, \qquad \delta_j = 2^j \delta_k, \quad j = 0, 1, 2, \ldots, k-1, \\ &\beta_j^2 = \max\left(H(\delta_j), \log(k - j + N(n^{-\gamma} \log^{-\beta} n))\right), \\ &S = \sum_{j=0}^{k-1} \delta_j \beta_j, \qquad a_j = \frac{\delta_j \beta_j}{S}, \quad j = 0, 1, 2, \ldots, k-1. \end{aligned}$$

As for (44) one can easily show that

$$\begin{aligned} S &\leq R \int_0^{n^{-\gamma}(\log n)^{-\beta}} \left(\log \frac{1}{x}\right)^{\frac{1}{2}} \mathrm{d}x \\ &\leq R n^{-\gamma} (\log n)^{-\beta+\frac{1}{2}} \end{aligned}$$

which implies, as in (48), that for a constant $G > 0$ we have

$$\begin{aligned} P_2 &\leq 2 \int_{N(n^{-\gamma} \log^{-\beta} n)}^{\infty} u^{2-G\sigma^2} \mathrm{d}u \\ &\leq \frac{2}{G\sigma^2 - 3} (n^{\gamma} (\log n)^{\beta})^{v(3-G\sigma^2)}. \end{aligned}$$

As a consequence for each $\gamma < 1/2$, $\beta \in \mathbb{R}$ and $\lambda > 0$, there exists $\sigma > 0$ such that (50) and hence (20) hold. Moreover if we take $\psi_n = n^{-\gamma+1/2}(\log n)^{-\beta+1/2}$ and δ_n as in (49) then (21) is equivalent to

$$\begin{aligned} &\exp\left(2a_n - \sigma\psi_n \exp\left(-\frac{5}{2}a_n\right)\right) \\ &\quad = \exp\left(2v\log(n^{\gamma}(\log n)^{\beta}) - \sigma n^{-\gamma+\frac{1}{2}-\frac{5}{2}\gamma v}(\log n)^{-\beta+\frac{1}{2}-\frac{5}{2}\beta v}\right) \end{aligned}$$

so if we put

$$\gamma = \frac{1}{2+5v}, \qquad \beta = \frac{1}{2+5v},$$

we have

$$\exp\left(2a_n - \sigma\psi_n \exp\left(-\frac{5}{2}a_n\right)\right) \leq \exp\left((K-\sigma)\log n\right), \tag{51}$$

where $K > 0$. It follows from (50), (51) and Proposition 4.1 that the rate of convergence in (22) is explicitly

$$\begin{aligned} v_n &= \max\left(\frac{\log n}{\sqrt{n}}, n^{-\gamma}(\log n)^{-\beta+\frac{1}{2}}, n^{-\gamma}(\log n)^{-\beta} \max\left((\log(n^{-\gamma}(\log n)^{-\beta}))^{\frac{1}{2}}, (\log n)^{\frac{1}{2}}\right)\right) \\ &= n^{-\frac{1}{(2+5v)}}(\log n)^{\frac{5v}{(4+10v)}}. \quad \square \end{aligned}$$

Proof of the second assertion. Letting now

$$\delta_n = \sqrt{\frac{\varphi_n}{n}} = \sqrt{\frac{\log n}{n}}$$

and assuming that (F.6) is satisfied with $\rho < 1$ it follows that (20) is satisfied with $\phi_n = 0$ and (21) with $\beta_n = n$ is implied by

$$\begin{aligned} &\exp\{2a_n - \sigma\psi_n \exp(-(5/2)a_n)\} \\ &\quad \leq \exp\left\{2v(\log(\sqrt{n}))^{\rho} - \sigma\psi_n \exp\left(-\frac{5v}{2}(\log(\sqrt{n}))^{\rho}\right)\right\} \\ &\quad \leq n^{-\lambda}. \end{aligned}$$

Hence (21) holds true with

$$\psi_n = \exp\left(\frac{5v}{2^{1+\rho}}(\log n)^\rho\right)\log n$$

and the rate (22) given by Proposition 4.1 is

$$\begin{aligned} v_n &\le K \max\left(\frac{\log n}{\sqrt{n}},\ \frac{1}{\sqrt{n}}\exp\left(\frac{5v}{2^{1+\rho}}(\log n)^\rho\right)\log n, \sqrt{\frac{\log n}{n}}\right. \\ &\quad \left.\times \max\left(\left(\log\left(\frac{\sqrt{n}}{(\log n)^{\frac{1}{2}}}\right)\right)^{\frac{\rho}{2}}, \sqrt{\log n}\right)\right) \\ &\le K\frac{1}{\sqrt{n}}\exp\left(\frac{5v}{2^{1+\rho}}(\log n)^\rho\right)\log n. \quad \square \end{aligned}$$

Proof of the third assertion. It is very similar to the proof of the first assertion. $\square$

Acknowledgments

The author thanks his Ph.D. supervisor Philippe Berthet for constant care and encouragement during the preparation of this manuscript. He is also indebted to an anonymous referee who helped motivating and commenting our results in light of [12,14,27].

References

[1] I. Berkes, W. Philipp, Approximation theorems for independent and weakly dependent random vectors, Ann. Probab. 7 (1979) 29–54.
[2] P. Berthet, D.M. Mason, Revisiting two strong approximation results of Dudley and Philipp, in: High Dimensional Probability, in: IMS Lecture Notes-Monograph Series, vol. 51, 2006, pp. 155–172.
[3] P. Berthet, D.M. Mason, Strong invariance principles for the empirical process indexed by functions. 2007. Preprint.
[4] P. Berthet, A. Saumard, Stability of empirical risk minimizers through Brownian approximation. 2007. Preprint.
[5] C. Borell, The Brunn–Minkowski inequality in Gauss space, Invent. Math. 30 (1975) 207–216.
[6] I.S. Borisov, Rate of convergence in the invariance principle for empirical measures, in: Proc. 1st World Congress of Bernoulli Soc., VNU Sci. Press, Amsterdam, 1987, pp. 833–836.
[7] J. Bretagnolle, P. Massart, Hungarian constructions from the nonasymptotic viewpoint, Ann. Probab. 17 (1989) 39–256.
[8] M. Csörgő, L. Horváth, Rényi-type empirical processes, J. Multivariate Anal. 41 (1992) 338–358.
[9] M. Csörgő, P. Révész, Strong approximations in probability and statistics, in: Probability and Mathematical Statistics, Academic Press, Inc., 1981.
[10] R.M. Dudley, The sizes of compact subsets of Hilbert space and continuity of Gaussian processes, J. Funct. Anal. 1 (1967) 290–330.
[11] R.M. Dudley, Uniform central limit theorems, in: Cambridge Studies in Advanced Mathematics, vol. 63, Cambridge University Press, 1999.
[12] R.M. Dudley, W. Philipp, Invariance principles for sums of Banach space valued random elements and empirical processes, Z. Wahrscheinlichkeitstheor. Verwandte. Geb. 62 (1983) 509–552.
[13] U. Einmahl, D.M. Mason, Gaussian approximation of local empirical processes indexed by functions, Probab. Theory Related Fields 107 (1997) 283–311.
[14] E. Giné, J. Zinn, Some limit theorems for empirical processes. With discussion, Ann. Probab. 12 (1984) 929–998.
[15] E. Giné, J. Zinn, Empirical processes indexed by Lipschitz functions, Ann. Probab. 14 (1986) 1329–1338.
[16] E. Giné, J. Zinn, Lectures on the central limit theorem for empirical processes, in: Probability and Banach Spaces, (Zaragoza, 1985), in: Lecture Notes in Math., Springer, 1986, pp. 50–113.

[17] V.I. Kolchinskii, On the central limit theorem for empirical measures, Theory Probab. Math. Statist. 24 (1981) 71–82.
[18] V.I. Kolchinskii, Functional limit theorems and empirical entropy II, Teor. Veroyatn. Mat. Stat. 34 (1986) 73–85 (in Russian).
[19] V.I. Kolchinskii, Rates of convergence in the invariance principle for empirical processes, in: New Trends in Probability and Statistics, Vol. 1, (Bakuriani, 1990), VSP, 1991, pp. 167–182.
[20] V.I. Kolchinskii, On the accuracy of the approximation of an empirical process by a Brownian bridge, Sibirsk. Mat. Zh. (translation in Siberian Math. J.) 32 (1991) 48–60 (in Russian).
[21] V.I. Kolchinskii, Komlós–Major–Tusnády approximation for the general empirical process and Haar expansions of classes of functions, J. Theoret. Probab. 7 (1994) 73–118.
[22] J. Komlós, O. Major, G. Tusnády, An approximation of partial sums of independent rv's and the sample df I, Z. Wahrscheinlichkeitstheor. Verwandte. Geb. 32 (1979) 111–131.
[23] L. Le Cam, A remark on empirical measures, A Festschrift for Erich L. Lehmann, Wadsworth Statist./Probab. Ser. (1983) 305–327.
[24] M.A. Lifshits, Gaussian random functions, in: Mathematics and its Applications, Kluwer Academic Publishers, 1995, p. 322.
[25] D.M. Mason, W.R. van Zwet, A refinement of the KMT inequality for the uniform empirical process, Ann. Probab. 15 (3) (1993) 871–884.
[26] P. Massart, Strong approximation for multivariate empirical and related processes, via KMT constructions, Ann. Probab. 17 (1989) 266–291.
[27] E. Rio, Local invariance principles and their application to density estimation, Probab. Theory Related Fields 98 (1) (1994) 21–45.
[28] M. Talagrand, Donsker classes and random geometry, Ann. Probab. 15 (4) (1987) 1327–1338.
[29] M. Talagrand, Donsker classes of sets, Probab. Theory Related Fields 78 (1988) 169–191.
[30] A.W. Van der Vaart, J.A. Wellner, Weak Convergence and Empirical Processes, Springer, 1996.
[31] V.N. Vapnik, A.Ja. Cervonenkis, The uniform convergence of frequencies of the appearance of events to their probabilities, Teor. Vero. Primen. 16 (1971) 264–279 (in Russian).
[32] V.N. Vapnik, A.Ya. Chervonenkis, Necessary and sufficient conditions for the uniform convergence of empirical means to their true values, Teor. Veroyatnost. i Primenen. 26 (1981) 543–563 (in Russian).
[33] Y. Zaitzev, Estimates for the Lévy–Prokhorov distance in the multidimensional central limit theorem for random vectors with finite exponential moments, Teor. Veroyatn. Primen. 31 (1986) 246–265 (in Russian).
[34] Y. Zaitzev, On the Gaussian approximation of convolutions under multidimensional analogues of S.N. Bernstein's inequality conditions, Probab. Theory Related Fields 74 (1987) 535–566.