Probab. Theory Relat. Fields 107, 283 – 311 (1997)

Probability Theory and Related Fields

Gaussian approximation of local empirical processes indexed by functions

Uwe Einmahl[1,*], David M. Mason[2,]**

[1] Department of Mathematics, Indiana University, Bloomington, IN 47405, USA
[2] Department of Mathematical Sciences, University of Delaware, Newark, DE 19716, USA
(e-mail: davidm@brahms.udel.edu)

Received: 11 January 1995 / In revised form: 12 July 1996

Summary. An extended notion of a local empirical process indexed by functions is introduced, which includes kernel density and regression function estimators and the conditional empirical process as special cases. Under suitable regularity conditions a central limit theorem and a strong approximation by a sequence of Gaussian processes are established for such processes. A compact law of the iterated logarithm (LIL) is then inferred from the corresponding LIL for the approximating sequence of Gaussian processes. A number of statistical applications of our results are indicated.

Mathematics Subject Classification (*1991*): 60F15, 62G05

1 Introduction and statements of main results

Deheuvels and Mason (1994) introduced a notion of a local uniform empirical process indexed by a class of sets and proved a functional law of the iterated logarithm. Such local processes are very useful in the study of statistics which are functions of the observations in a suitable neighborhood of a point. For instance, Deheuvels and Mason (1994) show how the pointwise Bahadur–Kiefer representation for the sample quantile and the law of the iterated logarithm for kernel density estimators follow readily from their results.

In this paper we extend the notion of the local empirical process to allow us to include kernel regression function estimators and the conditional empirical process within our setup. We then establish a weak convergence result and a strong invariance principle for such local processes. From our strong invariance principle we derive a general compact law of the iterated logarithm, which

* Research partially supported by the SFB 343, Universität Bielefeld, and an NSF Grant
** Research partially supported by the Alexander von Humboldt Foundation, the SFB 343, Universität Bielefeld and an NSF Grant

yields, among other results, the Deheuvels and Mason (1994) functional law of iterated logarithm as a special case.

Local empirical processes occur implicitly in the work of Kim and Pollard (1990) on cube root asymptotics and of Nolan and Marron (1989) on automatic bandwidth selection. This is indicated in the continuation of Example 2 below. Local empirical-type processes related to ours arise naturally in certain interval censoring and deconvolution problems. Refer, especially, to Part II of Groeneboom and Wellner (1992). Also for another approach to the study of the local behavior of the empirical process indexed by functions, along with further remarks on applications see Pollard (1995).

Let us begin by fixing some notation. Let $X, X_1, X_2, \ldots,$ be a sequence of i.i.d. $\mathbb{R}^d$ valued random vectors with distribution $\mathbb{P}$ on the Borel subsets $\mathscr{B}$ of $\mathbb{R}^d$. Given any $x \in \mathbb{R}^d$ and any measurable set $J \subseteq \mathbb{R}^d$, we set for any invertible bimeasurable transformation $h\colon \mathbb{R}^d \to \mathbb{R}^d$,

$$A(h) = x + hJ\ . \tag{1.1}$$

Let $\{h_n\}$ denote a sequence of invertible bimeasurable transformations from $\mathbb{R}^d$ to $\mathbb{R}^d$ and assume with $A_n = A(h_n)$ and $a_n = \mathbb{P}(A_n)$, $n \geqq 1$,

$$a_n > 0 \quad \text{for all } n \geqq 1\,, \tag{A.i}$$

$$na_n \to \infty \quad \text{as } n \to \infty\,, \tag{A.ii}$$

and for some $0 \leqq a \leqq 1$,

$$a_n \to a \quad \text{as } n \to \infty\ . \tag{A.iii}$$

For each integer $n \geqq 1$, let $k(n) = [na_n]$, where $[x]$ denotes the integer part of x, and let P_n be the probability measure on $(\mathbb{R}^d, \mathscr{B})$ defined by

$$P_n(B) = \mathbb{P}(x + h_n(J \cap B))/a_n, \quad B \in \mathscr{B}\ . \tag{1.2}$$

Let $\mathscr{F}$ denote a class of square $\mathbb{P}$-integrable functions on $\mathbb{R}^d$ with supports contained in J. To avoid measurability problems we shall assume that there exists a countable subclass $\mathscr{F}_c$ of $\mathscr{F}$ and a measurable set D with $P_n(D) = 0$ for all $n \geqq 0$ such that for any $x_1, \ldots, x_m \in \mathbb{R}^d - D$ and $f \in \mathscr{F}$ there exists a sequence $\{f_j\} \subset \mathscr{F}_c$ satisfying

$$\lim_{j\to\infty} f_j(x_k) = f(x_k), \quad k = 1, \ldots, m\,, \tag{S.i}$$

$$\lim_{j\to\infty} P_n(f_j) = P_n(f) \quad \text{for each } n \geqq 1 \tag{S.ii}$$

and

$$\lim_{j\to\infty} P_n(f_j^2) = P_n(f^2) \quad \text{for each } n \geqq 1\ . \tag{S.iii}$$

Given each integer $n \geqq 1$ and invertible bimeasurable transformation $h\colon \mathbb{R}^d \to \mathbb{R}^d$, we introduce the *local empirical process* at $x \in \mathbb{R}^d$ indexed by $\mathscr{F}$

$$L_n(f,h) = \sum_{i=1}^{n} \frac{f(h^{-1}(X_i - x)) - Ef(h^{-1}(X_i - x))}{\sqrt{n\mathbb{P}(A(h))}} \tag{1.3}$$

and define the *local empirical distribution function* at x indexed by $\mathscr{F}$ by

$$\lambda_n(f,h) = \sum_{i=1}^{n} f(h^{-1}(X_i - x)) / (n\mathbb{P}(A(h))) . \tag{1.4}$$

We could readily extend our setup formally by replacing $(\mathbb{R}^d, \mathscr{B})$ by a general measure space. However, to keep the exposition as simple as possible we will restrict ourselves to $(\mathbb{R}^d, \mathscr{B})$, where all of our examples live. Our setup allows us to consider the following interesting examples, among others, as special cases.

Example 1. Let $U_1, U_2, \ldots,$ be independent uniform $[0,1]^d$, $d \geqq 1$, random variables. Choose an $x \in [0,1]^d$ and a subclass $\mathscr{C}$ of the Borel subsets of $J := [r,s]^d$, where $r < s$ with $s - r = 1$. Setting $\mathscr{F} = \{1_C\colon C \in \mathscr{C}\}$, where each 1_C is the indicator function of C and defining $h\colon \mathbb{R}^d \to \mathbb{R}^d$ by $h(x_1,\ldots,x_d) = (a^{1/d}x_1,\ldots,a^{1/d}x_d)$ with $0 < a \leqq 1$, we get whenever $x + hJ \subset [0,1]^d$

$$L_n(1_C,h) = \sum_{i=1}^{n} \frac{1(U_i \in x + hC) - a|C|}{\sqrt{na}} , \tag{1.5}$$

where $|C|$ denotes the Lebesgue measure of C. This is a version of the local uniform $[0,1]^d$ empirical process first studied by Deheuvels and Mason (1994).

Example 2. Let $X_1, X_2, \ldots$ be i.i.d. real valued random variables with a density g continuous and positive in a neighborhood of a fixed x. Set $J = [-\frac{1}{2}, \frac{1}{2}]$ and define $h\colon \mathbb{R} \to \mathbb{R}$ by $h(u) = \gamma u$ with $0 < \gamma \leqq 1$. Further set $\mathscr{F} = \{K\}$, where K is a kernel function satisfying

$$\int_{-\infty}^{\infty} K(u)\, du = 1 , \tag{K.1}$$

$$K \text{ is of bounded variation} , \tag{K.2}$$

$$K(u) = 0 \quad \text{if } |u| > \tfrac{1}{2} . \tag{K.3}$$

Then

$$\begin{aligned} \hat{g}_n(x) &:= \sum_{i=1}^{n} K((X_i - x)/\gamma) / (n\gamma) \\ &= \lambda_n(K,h)\mathbb{P}(A(h)) / \gamma \end{aligned} \tag{1.6}$$

is the usual kernel density estimator of $g(x)$ with window size γ.

Example 3. Let $(X_1, Y_1), (X_2, Y_2), \ldots,$ be i.i.d. G with joint density g_{XY} and marginal densities g_X and g_Y. Choose $J = [-\frac{1}{2}, \frac{1}{2}] \times \mathbb{R}$, $h(u,v) = (\gamma u, v)$ with $0 < \gamma \leqq 1$, and $A(h) = (x,0) + hJ$. Further, set $R(u,v) = vK(u)$ for $(u,v) \in \mathbb{R}^2$, where K is a kernel function as in Example 2 and $\mathscr{F} = \{R\}$. Now

$$\lambda_n(R,h) = \hat{r}_n(x)\hat{g}_n(x)\gamma / \mathbb{P}(A(h)) , \tag{1.7}$$

where $\hat{g}_n(x)$ is the kernel density estimator of the marginal density $g_X(x)$ and $\hat{r}_n(x)$ is the kernel regression estimator

$$\hat{r}_n(x) = \sum_{i=1}^{n} Y_i K((X_i - x)/\gamma) / (n\gamma\hat{g}_n(x)), \tag{1.8}$$

of $r(x) = E(Y \mid X = x)$.

Example 4. Keeping the notation of Example 3, now choose the class of functions $\mathscr{F} = \{f_y\colon y \in \mathbb{R}\}$, where

$$f_y(u, v) = 1(v \leqq y)K(u), \quad (u, v) \in \mathbb{R}^2. \tag{1.9}$$

Then

$$\lambda_n(f_y, h) = F_n(y \mid x)\hat{g}_n(x)\gamma / \mathbb{P}(A(h)), \quad y \in \mathbb{R}, \tag{1.10}$$

where

$$F_n(y \mid x) = \frac{1}{n\gamma} \sum_{i=1}^{n} 1(Y_i \leqq y)K((X_i - x)/\gamma) / \hat{g}_n(x), \tag{1.11}$$

is the conditional empirical distribution first intensively studied by Stute (1986a, b).

In order to state our main weak convergence result we must introduce some further notation and assumptions. Here we shall borrow heavily from Sheehy and Wellner (1992), especially from their Sect. 3.

For integers $m \geqq 1$ and $n \geqq 1$ define the empirical process indexed by $\mathscr{F}$

$$\alpha_m^{(n)}(f) = \left(\sum_{i=1}^{m} f(Y_i^{(n)}) - mP_n(f) \right) \Big/ \sqrt{m}, \quad f \in \mathscr{F}, \tag{1.12}$$

where $Y_1^{(n)}, \ldots, Y_m^{(n)}$ are assumed to be i.i.d. P_n. Let $\mathscr{F}' = \{f - g\colon f, g \in \mathscr{F}\}$, $\mathscr{F}^2 = \{f^2\colon f \in \mathscr{F}\}$, $(\mathscr{F}')^2 = \{(f-g)^2\colon f, g \in \mathscr{F}\}$ and $\mathscr{G} = \mathscr{F} \cup \mathscr{F}^2 \cup \mathscr{F}' \cup (\mathscr{F}')^2$.

For any functional T defined on a subset $\mathscr{H}$ of the real valued functions on J we denote

$$\|T\|_{\mathscr{H}} = \sup\{|T(f)|\colon f \in \mathscr{H}\}. \tag{1.13}$$

We shall require the following additional assumptions on the class of functions $\mathscr{F}$ and the sequence of probability measures $\{P_n\}$.

(F.i) $\mathscr{F}$ has a uniformly square integrable envelope function F, namely,

$$\lim_{\lambda \to \infty} \limsup_{n \to \infty} P_n F^2 1_{(F \geqslant \lambda)} = 0. \tag{1.14}$$

(F.ii) There exists a probability measure P_0 such that

$$\|P_n - P_0\|_{\mathscr{G}} \to 0 \quad \text{as } n \to \infty. \tag{1.15}$$

For any $f, g \in L^2(\mathbb{R}^d, \mathscr{B})$, and $n \geqq 0$, set

$$\rho_n^2(f, g) = \operatorname{Var}_{P_n}(f - g), \tag{1.16}$$

and

(1.17) $$e_n^2(f,g) = E_{P_n}(f-g)^2 \, .$$

Set for each $n \geqq 0$ and $\delta > 0$

(1.18) $$\mathscr{F}_n'(\delta) = \{(f,g) \in \mathscr{F} \times \mathscr{F} : \; \rho_n(f,g) \leqq \delta\} \, ,$$

and for any real valued functional T on $\mathscr{F}$ set

(1.19) $$\|T\|_{\mathscr{F}_n'(\delta)} = \sup\{|T(f) - T(g)| : \; \rho_n(f,g) \leqq \delta\} \, .$$

Then our next assumption on $\mathscr{F}$ is,
(F.iii) $(\mathscr{F}, \rho_0)$ is totally bounded and for all $\varepsilon > 0$

(1.20) $$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} P(\|\alpha_{k(n)}^{(n)}\|_{\mathscr{F}_n'(\delta)} > \varepsilon) = 0 \, .$$

We shall denote by

(1.21) $$(B_0(f))_{f \in \mathscr{F}}$$

a P_0-Brownian bridge indexed by $\mathscr{F}$, that is, B_0 is a Gaussian process indexed by $\mathscr{F}$ with mean zero and covariance function

(1.22) $$\mathrm{Cov}\,(B_0(f), B_0(g)) = P_0(fg) - P_0 f P_0 g, \quad f,\, g \in \mathscr{F} \, .$$

We will assume that B_0 has uniformly ρ_0 continuous sample paths, whenever such a version of B_0 exists. Note, in particular, that this is the case under the above assumptions since they imply that $\mathscr{F}$ is P_0-pregaussian.

Let Z be a standard normal random variable independent of B_0 and for each $0 \leqq a \leqq 1$ introduce the Gaussian process indexed by $\mathscr{F}$

(1.23) $$W(f;a) := B_0(f) + \sqrt{1-a}\, Z P_0(f), \quad f \in \mathscr{F} \, .$$

For use later on we note that we can assume that B_0 has been extended to be a Gaussian process indexed by the larger class of all measurable real valued functions on $\mathbb{R}^d$ such that $P_0 f^2 < \infty$ with covariance function given by (1.22). This can be justified using the Kolmogorov consistency theorem. Refer to page 5 of Ibragimov and Rozanov (1978) for details. Thus we can assume that the Gaussian process $W(f;a)$ is also well-defined on this class.

We are now prepared to state our main weak convergence result. As in Sheehy and Wellner (1992), we use the notion of weak convergence in the sense of Hoffman–Jørgensen and it will be denoted by the symbol '$\Rightarrow$'.

Theorem 1.1 *Under assumptions* $(A), (S)$ *and* (F) *the sequence of local empirical processes satisfies*

(1.24) $$(L_n(f, h_n))_{f \in \mathscr{F}} \Rightarrow (W(f;a))_{f \in \mathscr{F}} \, .$$

Remark. 1.1 Corollary 3.1 of Sheehy and Wellner (1992) implies that whenever (S), (F.i) and (F.ii) hold and $\mathscr{F}$ is sparse in the sense of Pollard (1982), then (F.iii) is also satisfied. In particular, each of the classes $\mathscr{F}$ in Examples 2–4 are sparse and satisfy (S), (F.i) and (F.ii).

Let $Lx = \log(x \vee e)$ and $LLx = L(Lx)$ for all x. We shall now state our strong approximation result.

Theorem 1.2 *Assume* (S) *and* (A), *where in* (A.iii), $a = 0$. *Also assume that*

(A.iv) $\quad a_n \sim d_n$ *where* $d_n \searrow 0, nd_n \nearrow \infty$ *and* $nd_n/LL_n \to \infty$ *as* $n \to \infty$.

Further assume (F.ii) *and replace* (F.i) *by*

(F.iv) $\quad |f| \leqq K$ *for all* $f \in \mathscr{F}$ *and some* $K \geqq 1$.

In addition, assume

(F.v) *for each* $n \geqq 1$ *and* $m \geqq n$, $f \circ h_n^{-1}$ *and* $f \circ h_m^{-1} \circ h_n \in \mathscr{F}$;

(F.vi) *there exists a sequence of positive constants* $(b_n)_{n \geqq 1}$ *such that for all* $f, g \in \mathscr{F}$ *and* $n \geqq 1$

$$\int_{\mathbb{R}^d} f(h_n x) g(h_n x)\, dP_0(x) = b_n^{-1} \int_{\mathbb{R}^d} f(x) g(x)\, dP_0(x), \tag{1.25}$$

where

$$a_n/b_n \to 1, \quad \text{as } n \to \infty. \tag{1.26}$$

Further assume

(F.vii) $$\|\alpha_{k(n)}^{(n)}\|_{\mathscr{F}}/\sqrt{LLn} \xrightarrow{P} 0, \quad \text{as } n \to \infty,$$

(F.viii) $$\mathscr{F} \text{ is } P_0\text{-pregaussian}.$$

Then one can construct $X_1, X_2, \ldots, i.i.d.\, \mathbb{P}$ *and a sequence* $W_1, W_2, \ldots$ *of independent* P_0*-Brownian motions indexed by* $\mathscr{F}$, *such that with probability one as* $n \to \infty$.

$$\sup_{f \in \mathscr{F}} \left| L_n(f, h_n) - \frac{1}{\sqrt{nb_n}} \sum_{i=1}^{n} W_i(f \circ h_n^{-1}) \right| \Big/ \sqrt{LLn} \to 0. \tag{1.27}$$

Notice that due to assumption (F.vi) for each $n \geqq 1$

$$\tilde{W}_n := \left(\sum_{i=1}^{n} (W_i(f \circ h_n^{-1}))/\sqrt{nb_n} \right)_{f \in \mathscr{F}} \stackrel{\mathscr{D}}{=} (W_1(f))_{f \in \mathscr{F}}. \tag{1.28}$$

Thus $(\tilde{W}_n)_{n \geqq 1}$ is a sequence of P_0-Brownian motions taking values in the separable Banach space $B = C_u(\mathscr{F}, e_0)$, the space of uniformly e_0 continuous functions of $\mathscr{F}$.

Let $\mathscr{K}$ be the unit ball of the reproducing kernel Hilbert space pertaining to W_1. Further, let $B(\mathscr{F})$ denote the class of all bounded functionals on $\mathscr{F}$ equipped with the supremum norm

$$d(\phi, \varphi) = \sup_{f \in \mathscr{F}} |\phi(f) - \varphi(f)|, \;\; \phi, \varphi \in B(\mathscr{F}). \tag{1.29}$$

Our next result gives as a corollary to Theorem 1.2 the compact law of the iterated logarithm for the local empirical process. Its proof is a consequence of the strong approximation (1.27) and the fact that the same result holds for the sequence $(\tilde{W}_n/\sqrt{2LLn})_{n\geqq 1}$, which will be shown in Sect. 3.

Corollary 1.1 *In addition to the assumptions of Theorem* 1.2 *assume*
(F.ix) *for all* f, g *such that* $P_0(f-g)^2 < \infty$, $n \geqq 1$, *and* $2n \geqq m \geqq n$, *we have*

(1.30)

$$\int_{\mathbb{R}^d} (f(h_m^{-1}(h_n(x))) - g(h_m^{-1}(h_n(x))))^2 \, dP_0(x) \leqq M \int_{\mathbb{R}^d} (f(x) - g(x))^2 \, dP_0(x)\,,$$

where $M > 0$,
(F.x) *for every compact set* $A \subset \mathbb{R}^d$ *and* $\delta > 0$ *there exists a* $q_0 > 1$ *such that for all* $1 < q < q_0$ *with* $n_k = [q^k]$

(1.31)
$$\max_{n_k \leqq m \leqq n_{k+1}} \sup_{x \in A} |x - h_m^{-1}(h_{n_k}(x))| \leqq \delta$$

for all large enough k *depending on* A *and* $\varepsilon > 0$. *Then with probability one the sequence of processes*

(1.32)
$$(L_n(f, h_n)/\sqrt{2LLn})_{f \in \mathscr{F}}$$

is relatively compact in $B(\mathscr{F})$ *with set of limit points equal to* $\mathscr{K}$.

Remark. 1.2 Whenever $[na_n]/LL_n$ is bounded in n, the almost sure limiting behavior of the local empirical process is much different. In this case it is more appropriate to replace the independent Wiener processes by independent Poisson processes to obtain a useful strong approximation. See Deheuvels and Mason (1990) and (1995).

Example 1 (*Contd.*) For each integer $n \geqq 1$ set $h_n(x_1, \ldots, x_d) = (a_n^{1/d} x_1, \ldots, a_n^{1/d} x_d)$, where $a_n \searrow 0, na_n \nearrow \infty$ and $na_n/\sqrt{LLn} \to \infty$. Also assume that $x + \varepsilon[r,s]^d \subset [0,1]^d$ for all small enough $\varepsilon > 0$. Then for all n sufficiently large $P_n = P_0 =$ uniform $[r,s]^d$. Now let $\mathscr{F}$ be a uniformly bounded class of real valued measurable functions on $\mathbb{R}^d$ satisfying

(C.1)
$$f(\lambda \cdot) \in \mathscr{F} \quad \text{for all } \lambda \geqq 1\,,$$

(C.2)
$$\mathscr{F} \text{ satisfies } (S)\,,$$

(C.3)
$$\|\alpha_{k(n)}^{(0)}\|_{\mathscr{F}}/\sqrt{LLn} \xrightarrow{P_0} 0, \quad \text{as } n \to \infty\,,$$

(C.4)
$$\mathscr{F} \text{ is } P_0\text{-pregaussian}\,.$$

In particular (C.1)–(C.4) are satisfied whenever $\mathscr{F} = \{1_C \colon C \in \mathscr{C}\}$ where $\mathscr{C}$ is a P_0-Donsker class of sets satisfying (S) and $\gamma C \in \mathscr{C}$ whenever $C \in \mathscr{C}$

and $0 < \gamma \leqq 1$. It is trivial to verify that all of the conditions of Theorem 1.2 and Corollary 1.1 hold. Hence Theorem 1.1 of Deheuvels and Mason (1994) is a special case of our Corollary 1.1. We mention here that Arcones (1994) has formulated a compact law of the iterated logarithm for the local uniform $[0,1]^d$ empirical process indexed by a uniformly bounded class of functions. Also we note that when $d = 1$, $x = 0$ and $\mathscr{F} = \{1_{[0,t]}\colon 0 \leqq t \leqq 1\}$, Example 1 specializes to the uniform $[0,1]$ tail empirical process. Therefore Theorem 1.2 yields the Mason (1988) strong approximation to this process.

We are now going to show that our strong approximation and compact law of the iterated logarithm apply to Examples 2 and 4. We need some facts about VC subgraph classes. For the basic definition refer to Giné and Zinn (1984) or Pollard (1984).

Fact 1.1 (Dudley 1978; Alexander 1987). *Let ϕ be a monotone function on $\mathbb{R}$ and $\mathscr{G}$ be a finite dimensional class of real valued functions defined on a set S then the class $\{\phi(g)\colon g \in \mathscr{G}\}$ is a VC subgraph class.*

Fact 1.2 (Pollard (1984)). *Let $\mathscr{C}$ be a VC class of sets. Then $\{1_C\colon C \in \mathscr{C}\}$ is a VC subgraph class.*

Let $\mathscr{F}$ be a class of measurable real valued functions on $\mathbb{R}^d$ with envelope function F, and let $\mathscr{Q}$ be a probability measure on $\mathbb{R}^d$ such that $\mathscr{Q}(F^2) < \infty$. For any $u > 0$ let $N(u(\mathscr{Q}(F^2))^{1/2}, \mathscr{F}, e_{\mathscr{Q}})$ = minimum $m \geqq 1$ for which there exist functions $f_1, \ldots, f_m$, not necessarily in $\mathscr{F}$, such that for all $f \in \mathscr{F}$, $\min_{1 \leqq i \leqq m} e_{\mathscr{Q}}(f, f_i) < u(\mathscr{Q}(F^2))^{1/2}$, where $e_{\mathscr{Q}}(f, f_i) = (\mathscr{Q}(f - f_i)^2)^{1/2}$.

Fact 1.3 (Alexander 1987). *If $\mathscr{F}$ is VC subgraph class of measurable real valued functions on $\mathbb{R}^d$ with bounded envelope function F, then there exist constants $C > 0$ and $v > 0$ such that*

$$N(u(\mathscr{Q}(F^2))^{1/2}, \mathscr{F}, e_{\mathscr{Q}}) \leqq Cu^{-v} \tag{1.33}$$

for all $0 < u < 1$ and probability measures $\mathscr{Q}$ on $\mathbb{R}^d$.

Fact 1.4 *Suppose $\mathscr{F}_1$ and $\mathscr{F}_2$ are two classes of uniformly bounded measurable real valued functions on $\mathbb{R}^d$ such that for constants $C_1 > 0$, $C_2 > 0$, $v_1 > 0$ and $v_2 > 0$*

$$N(uM_i, \mathscr{F}_i, e_{\mathscr{Q}}) \leqq C_i u^{-v_i}, \quad i = 1, 2\,, \tag{1.34}$$

for all $0 < u < 1$ and probability measures $\mathscr{Q}$ on $\mathbb{R}^d$. Then there exist constants $C_3 > 0$ and $C_4 > 0$ such that

$$N(u(M_1 + M_2), \mathscr{F}_1 + \mathscr{F}_2, e_{\mathscr{Q}}) \leqq C_3 u^{-(v_1 + v_2)} \tag{1.35}$$

and

$$N(uM_1M_2, \mathscr{F}_1\mathscr{F}_2, e_{\mathscr{Q}}) \leqq C_4 u^{-(v_1 + v_2)} \tag{1.36}$$

for all $0 < u < 1$ and probability measures $\mathscr{Q}$ on $\mathbb{R}^d$, where $\mathscr{F}_1 + \mathscr{F}_2 = \{f + g\colon f \in \mathscr{F}_1, g \in \mathscr{F}_2\}$ and $\mathscr{F}_1\mathscr{F}_2 = \{fg\colon f \in \mathscr{F}_1, g \in \mathscr{F}_2\}$.

Proof. Trivial.

Fact 1.5 *Let K be a real valued function of bounded variation on $[a,b], -\infty < a < b < \infty$, and equal to zero on $\mathbb{R} - [a,b]$, and let $\mathscr{C}$ be a VC class of subsets of $\mathbb{R}$. Then for the class of real valued functions*

$$\mathscr{F}_{K,\mathscr{C}} = \{K(xt)1_C(y)\colon -\infty < t < \infty, C \in \mathscr{C}\} \tag{1.37}$$

there exist constants $C > 0$ and $v > 0$ such that (1.33) *holds for all probability measures $\mathscr{Q}$ on $\mathbb{R}^2$.*

Proof. Write $K = K_1 - K_2$, where K_1 and K_2 are two bounded nondecreasing functions on $\mathbb{R}$ and then apply Facts 1.1–1.4. □

Our next fact is a special case of Theorem 3.1 of Alexander (1987).

Fact 1.6 *Let $\{P_n\}_{n \geqq 0}$ be a sequence of probability measures on $\mathbb{R}^d$ and $\mathscr{F}$ be a class of measurable real valued functions on $\mathbb{R}^d$ bounded by M such that (S) and* (F.ii) *hold. Assume there exist constants $C > 0$ and $v > 0$ such that for all probability measures Q and $0 < u < 1$*

$$N(uM^{1/2}, \mathscr{F}, e_{\mathscr{Q}}) \leqq Cu^{-v} , \tag{1.38}$$

then for any sequence of integers $m(n) \to \infty$

$$(\alpha_{m(n)}^{(n)}(f))_{f \in \mathscr{F}} \Rightarrow (B_0(f))_{f \in \mathscr{F}} \quad \textit{as } n \to \infty . \tag{1.39}$$

Example 2 (Cont.) Set $h_n(u) = \tau_n u$ for $n \geqq 1$, where $\tau_n \searrow 0$, $n\tau_n \nearrow$ and $n\tau_n/LLn \to \infty$ as $n \to \infty$. In this case P_0 is uniform $[-\frac{1}{2}, \frac{1}{2}]$. We now choose

$$\mathscr{F} = \{K_t(u) := K(ut)\colon t > 0\} . \tag{1.40}$$

Since X_1 is assumed to have a density continuous and positive in a neighborhood of x, it is readily established using Scheffé's theorem that P_n converges to P_0 in total variation. Therefore (F.ii) holds. Also one sees that (S) is satisfied. Thus by Fact 1.5 with $\mathscr{C} = \{(-\infty, \infty)\}$ and by Fact 1.6, (1.39) holds, which implies (F.vii) and (F.viii). The rest of the assumptions of Theorem 1.2 and Corollary 1.1 are easily checked. After a little manipulation one infers from Corollary 1.1 that with probability one

$$\limsup_{n \to \infty} \pm \sqrt{n\tau_n} \frac{(\hat{g}_n(x) - E\hat{g}_n(x))}{\sqrt{2LLn}} = \left(g(x) \int_{-\infty}^{\infty} K^2(u)\,\mathrm{d}u \right)^{1/2} . \tag{1.41}$$

This result was first proved in Deheuvels and Mason (1994).

We mention that the following process defined for $t > 0$,

$$\left\{ \sum_{i=1}^{n} K(t(X_i - x)/\tau_n) - na_n P_n(K_t) \right\} \Big/ \sqrt{2na_n}$$

has potential use in the study of the rate of consistency of automatic bandwidth estimators. For motivation refer to the paper of Nolan and Marron (1989) and the references therein. In the special case when K is the indicator function of the interval $[-\frac{1}{2}, \frac{1}{2}]$, this process arises naturally in the study of cube asymptotics. See, especially, the derivations of the limiting distribution of the shorth in Shorack and Wellner (1986) and Kim and Pollard (1990). Applications such as these will be addressed elsewhere.

Example 4 (*Cont.*) Set $h_n(u,v) = (\tau_n u, v)$, $n \geqq 1$, where $\tau_n \searrow 0$, $n\tau_n \nearrow \infty$ and $n\tau_n/LLn \to \infty$ as $n \to \infty$. Assume that g_{XY} is continuous on $(x-\varepsilon, x+\varepsilon) \times \mathbb{R}$ for some $\varepsilon > 0$ and $g_X(x) > 0$. In this case $P_0 = P_{0,1} \times P_{0,2}$, where $P_{0,1}$ is uniform $[-\frac{1}{2}, \frac{1}{2}]$ and $P_{0,2}(B) = P(Y_1 \in B \,|\, X_1 = x)$ for $B \in \mathscr{B}$. Choose

$$(1.42) \qquad \mathscr{F} = \{f_y(ut, v) = K(ut)1(v \in (-\infty, y]) \colon t \geqq 1, y \in \mathbb{R}\}\,.$$

By Facts 1.5 and 1.6, (F.vii) and (F.viii) hold. Since P_n converges to P_0 in total variation, (F.ii) is also satisfied. It is straightforward to show that the other assumptions of Theorem 1.2 and Corollary 1.1 hold.

Here

$$\frac{L_n(f_y, h_n)}{\sqrt{2LLn}} = \sum_{i=1}^{n} \frac{1(Y_i \leqq y)K((X_i - x)/\tau_n) - na_nP_n(f_y)}{\sqrt{2na_nLL_n}}\,,$$

and we get that

$$\limsup_{n\to\infty} \sup_{y\in\mathbb{R}} \pm \frac{L_n(f_y, h_n)}{\sqrt{2LLn}} = \left(\int_{-\infty}^{\infty} K^2(u)\,\mathrm{d}u\right)^{1/2} \quad \text{a.s.}$$

This is the LIL version of the Stute (1986b) Glivenko–Cantelli theorem for the conditional empirical process.

The remainder of our paper is organized as follows. We begin in Sect. 2 by proving a coupling inequality for the empirical process. Our method of proof is somewhat similar to the one employed by Dudley and Philipp (1983). We first establish a coupling inequality for multidimensional random vectors, where we use a result of Zaitsev (1987) on the rate of convergence in the multidimensional central limit theorem in combination with the Strassen–Dudley theorem. We then employ a recent inequality of Talagrand (1994) to approximate the empirical process by a suitable finite dimensional process. In Sect. 3, we prove the main results. The basic idea is that the local empirical process behaves to some extent like a randomly stopped empirical process (see Proposition 3.1 below), which allows us to reduce the approximation problem to one involving the usual empirical process, which in turn can be solved using our coupling inequality.

2 A useful coupling inequality

In this section we establish a useful coupling inequality for the empirical process, which will be essential for the proof of our strong invariance principle.

For probability measures P and Q on the Borel subsets of $\mathbb{R}^d$ and $\delta > 0$, let

$$\lambda(P, Q, \delta) := \sup\{\max P(A) - Q(A^\delta), Q(A) - P(A^\delta)): A \subset \mathbb{R}^d, \text{ closed}\}, \tag{2.1}$$

where A^δ denotes the closed δ-neighborhood of A,

$$A^\delta := \left\{x \in \mathbb{R}^d : \inf_{y \in A} |x - y| \leqq \delta\right\} \tag{2.2}$$

with $| \cdot |$ being the Euclidean norm on $\mathbb{R}^d$.

Further, let $X_1, \ldots, X_m$ $m \geqq 1$, be independent mean zero random vectors satisfying for some $M > 0$

$$|X_i| \leqq M, \quad 1 \leqq i \leqq m. \tag{2.3}$$

Denote the distribution of $X_1 + \cdots + X_m$ by $\mathbb{P}_m$ and let $\mathbb{Q}_m$ be the d-dimensional normal distribution with mean zero and covariance matrix

$$\operatorname{cov}(X_1) + \cdots + \operatorname{cov}(X_m). \tag{2.4}$$

The following inequality follows from the work of Zaitsev (1987).

Fact 2.1 *For all integers $m \geqq 1$,*

$$\lambda(\mathbb{P}_m, \mathbb{Q}_m, \delta) \leqq c_1 \exp(-c_2 \delta / M), \tag{2.5}$$

where c_1 and c_2 are positive constants depending only on d.

Using the Strassen–Dudley theorem (see Dudley 1968), Fact 2.1 and standard arguments from measure theory such as Lemmas 1.2.2 and 1.2.3 in Dudley (1984), we readily infer the next fact.

Fact 2.2 *Let $X_1, \ldots, X_m$ be independent mean zero d-dimensional random vectors satisfying* (2.3)*. If the underlying probability space $(\Omega, \mathscr{F}, P)$ is rich enough, one can define independent normally distributed mean zero random vectors $V_1, \ldots, V_m$ with $\operatorname{cov}(V_i) = \operatorname{cov}(X_i)$, $1 \leqq i \leqq m$, such that*

$$P\left(\left|\sum_{i=1}^m (X_i - V_i)\right| \geqq \delta\right) \leqq c_1 \exp(-c_2 \delta / M). \tag{2.6}$$

The following maximal version of the Bernstein inequality follows from Doob's maximal inequality and the proof of the usual Bernstein inequality on page 14 of Dudley (1984).

Fact 2.3 *Let $\xi_1, \ldots, \xi_m$ be independent mean zero random variables satisfying*

$$|\xi_i| \leqq M, \quad 1 \leqq i \leqq m. \tag{2.7}$$

Then for all $t \geqq 0$

$$P\left(\max_{1 \leqq j \leqq m} \left|\sum_{i=1}^j \xi_i\right| \geqq t\right) \leqq 2 \exp\left(-t^2 \Big/ \left(2B_m + \frac{2M}{3} t\right)\right), \tag{2.8}$$

where $B_m := \sum_{i=1}^m E\xi_i^2$.

Let $|\cdot|_+$ be the maximal norm on $\mathbb{R}^d$. Using Facts 2.2 and 2.3 we shall establish the following simple coupling inequality for sums of independent d-dimensional random vectors.

Proposition 2.1 *Let $X_j = (X_j^{(1)}, \ldots, X_j^{(d)})$, $1 \leqq j \leqq n$, be independent mean zero random vectors satisfying* (2.3) *and let $\sigma^2 > 0$ be such that*

$$E(X_j^{(i)})^2 \leqq \sigma^2, \quad 1 \leqq i \leqq d, \ 1 \leqq j \leqq n \,. \tag{2.9}$$

Let $1 \leqq L \leqq n$ be an integer and $x > 0$ be fixed. If the underlying probability space is rich enough, one can construct independent normally distributed mean zero random vectors $V_1, \ldots, V_n$ with $\operatorname{cov}(V_i) = \operatorname{cov}(X_i)$, $1 \leqq i \leqq n$, *such that*

$$P\left(\max_{1 \leqq j \leqq n} \left|\sum_{i=1}^{j} (X_i - V_i)\right|_+ \geqq x\right) \leqq c_3(L+1)\left(\exp(-c_4 x/ML) + \exp(-Lx^2/64n\sigma^2)\right), \tag{2.10}$$

where c_3 and c_4 are positive constants depending only on d.

Proof. Set $m = [n/L]$,

$$U_j := \sum_{i=(j-1)m+1}^{jm} X_i, \quad 1 \leqq j \leqq L \tag{2.11}$$

and

$$W_j := \sum_{i=(j-1)m+1}^{jm} V_i, \quad 1 \leqq j \leqq L \,, \tag{2.12}$$

where $V_1, \ldots, V_n$ are constructed using Fact 2.2 to be mean zero independent normally distributed random vectors with $\operatorname{cov}(V_i) = \operatorname{cov}(X_i)$, $1 \leqq i \leqq n$, such that

$$P(|U_j - W_j| \geqq x/2L) \leqq c_1 \exp(-c_2 x/2ML) \tag{2.13}$$

for $1 \leqq j \leqq L$. Since $|x|_+ \leqq |x|$, $x \in \mathbb{R}^d$, (2.13) clearly implies that for $1 \leqq j \leqq L$,

$$P(|U_j - W_j|_+ \geqq x/2L) \leqq c_1 \exp(-c_2 x/2ML) \,. \tag{2.14}$$

It follows that

$$P\left(\max_{1 \leqq k \leqq n} \left|\sum_{i=1}^{k} (X_i - V_i)\right|_+ \geqq x\right) \leqq c_1 L \exp(-c_2 x/2ML) + \Delta_{L,1} + \Delta_{L,2} \,, \tag{2.15}$$

where

$$\Delta_{L,1} := P\left(\max_{0 \leqq j \leqq L} \max_{1 \leqq k \leqq m} \left|\sum_{i=jm+1}^{jm+k} X_i\right|_+ \geqq x/4\right),$$

and

$$\Delta_{L,2} := P\left(\max_{0\leqq j\leqq L}\max_{1\leqq k\leqq m}\left|\sum_{i=jm+1}^{jm+k} V_i\right|_+ \geqq x/4\right).$$

To bound $\Delta_{L,1}$ we note that

$$\Delta_{L,1} \leqq d\max_{1\leqq v\leqq d}\sum_{j=0}^{L} P\left(\max_{1\leqq k\leqq m}\left|\sum_{i=jm+1}^{jm+k} X_i^{(v)}\right| \geqq x/4\right),$$

which, in turn, by Fact 2.3 is

$$\begin{aligned}&\leqq 2d(L+1)\exp(-x^2/(32m\sigma^2+8Mx/3))\\ &\leqq 2d(L+1)(\exp(-x^2/64m\sigma^2)+\exp(-3x/16ML)).\end{aligned}$$

Using the fact that the V_i's have a symmetric distribution, along with a standard exponential inequality for the tail probabilities of the normal distribution we get similarly

$$\Delta_{L,2} \leqq 2d(L+1)\exp(-x^2/32m\sigma^2).$$

Setting $c_3 = c_1 \vee (4d)$ and $c_4 = (c_2/2)\wedge(3/16)$ completes the proof of (2.10). □

Recall definition (1.3), (1.12), (1.16) and (1.18). The following inequality is readily inferred from Theorem 3.5, Talagrand (1994).

Fact 2.4 *Assume that* $\mathscr{F}$ *satisfies* (S) *and for some* $K > 0$

$$|f| \leqq K,\quad f\in\mathscr{F}. \tag{2.16}$$

Then there exists an absolute constant $A > 0$ *such that for all* $t > 0$, $\delta > 0$, $m \geqq 1$ *and* $n \geqq 1$,

$$\begin{aligned}&P(\|\sqrt{m}\alpha_m^{(n)}(f)\|_{\mathscr{F}_n'(\delta)} \geqq t + AE\|\sqrt{m}\alpha_m^{(n)}(f)\|_{\mathscr{F}_n'(\delta)})\\ &\quad\leqq \exp(-t^2/mA^2\delta^2)+\exp(-t/KA).\end{aligned} \tag{2.17}$$

Proof. To see (2.17) for $K = 1$, we note that an inspection of Talagrand's proof shows that his Theorem 3.5 is also valid if we replace his H by

$$\bar{H} := E\left\|\sum_{i=1}^{m}\varepsilon_i(f(Y_i^{(n)})-P_n(f))\right\|_{\mathscr{G}}, \tag{2.18}$$

where $\mathscr{G} = \{(f+2)/4\colon f\in\mathscr{F}_n'(\delta)\}$. Noticing that

$$\bar{H} \leqq 2E\|\sqrt{m}\alpha_m^{(n)}(f)\|_{\mathscr{F}_n'(\delta)}, \tag{2.19}$$

we obtain (2.17) after some straightforward manipulation. □

Using a well known inequality for Gaussian random variables (see e.g. Ledoux and Talagrand, 1991, Lemma 3.1), we readily obtain the following inequality.

Fact 2.5 *Assume that for a probability measure P_0 the class $\mathscr{F}$ is P_0-pregaussian and let for $f, g \in \mathscr{F}$*

$$\rho_0^2(f,g) = \mathrm{Var}_{P_0}(f-g)\ . \tag{2.20}$$

Let $\bar{B}, \bar{B}_1, \bar{B}_2, \ldots,$ be independent P_0-Brownian bridges indexed by $\mathscr{F}$. For all $t > 0$, $\delta > 0$ and $m \geqq 1$:

$$P\left(\left\|\sum_{j=1}^{m}\bar{B}_j\right\|_{\mathscr{F}'(\delta)} \geqq 2\sqrt{m}E\|\bar{B}\|_{\mathscr{F}'(\delta)} + t\right) \leqq \exp(-t^2/2m\delta^2)\ , \tag{2.21}$$

where

$$\mathscr{F}'(\delta) = \{f-g \colon \rho_0(f,g) \leqq \delta\}\ . \tag{2.22}$$

We can now state and prove the main result of this section.

Proposition 2.2 *Let $\mathscr{F}$ be a class of real valued functions satisfying (S) and (2.16) and let $\{P_n\}_{n\geqq 1}$ be a sequence of probability measures on $(\mathbb{R}^d, \mathscr{B})$ for which* (F.ii) *holds for a probability measure P_0. Assume further that $\mathscr{F}$ is P_0-pregaussian. Then given any $0 < \delta < 1$ and sequence of positive integers $\{k_n\}_{n\geqq 1}$ there exists an $n(\delta)$ such that for every $n \geqq n(\delta)$ and $u > 0$ one can construct empirical processes $(\alpha_m^{(n)}(f))_{f\in\mathscr{F}}$, $1 \leqq m \leqq k_n$ and independent P_0-Brownian Bridges $(\bar{B}_m(f))_{f\in\mathscr{F}}$, $1 \leqq m \leqq k_n$ such that*

$$\begin{aligned} P\left(\max_{1\leqq m\leqq k_n}\left\|\sqrt{m}\alpha_m^{(n)} - \sum_{i=1}^{m}\bar{B}_i\right\|_{\mathscr{F}} \geqq u + \beta_{n,k_n}(\delta)\right) \\ \leqq K_1\{\exp(-K_2u/K) + \exp(-u^2/K_nA_0^2\delta^2K^2)\}\ , \end{aligned} \tag{2.23}$$

where

$$\beta_{n,k_n}(\delta) := A_1\sqrt{k_n}(E\|\bar{B}\|_{\mathscr{F}} + E\|\alpha_{k_n}^{(n)}\|_{\mathscr{F}_n'(\delta)})\ , \tag{2.24}$$

$K_1 = K_1(\delta)$, $K_2 = K_2(\delta)$ are constants depending on δ only, and A_0 and A_1 are absolute constants.

Proof. First note that since $\mathscr{F}$ is P_0-pregaussian, it is totally bounded with respect to ρ_0. Recalling (1.15), we can find for any $0 < \delta < 1$ a subclass $\{f_1, \ldots, f_r\}$ of $\mathscr{F}$ where r depends on δ such that for all $n \geqq n_1(\delta)$, for some $n_1(\delta)$,

$$\min_{1\leqq i\leqq r}\rho_n(f, f_i) \leqq \delta\ . \tag{2.25}$$

So if the sequence $(\alpha_m^{(n)}(f))_{f\in\mathscr{F}}$, $1 \leqq m \leqq k_n$, $n \geqq n_1(\delta)$, is given, we set for $1 \leqq j \leqq k_n$,

$$X_j^{(i)} = f_i(Y_j^{(n)}) - P_n(f_i), \quad i = 1, \ldots, r\ ,$$

and we clearly have for $1 \leqq m \leqq k_n$

$$(\sqrt{m}\alpha_m^{(n)}(f_1), \ldots, \sqrt{m}\alpha_m^{(n)}(f_r)) = \left(\sum_{i=1}^{m}X_i^{(1)}, \ldots, \sum_{i=1}^{m}X_i^{(r)}\right)\ . \tag{2.26}$$

Using Proposition 2.1 first we define normally distributed random vectors V_i, $1 \leqq i \leqq k_n$, such that

$$\Delta_1 := P\left(\max_{1\leqq m\leqq k_n}\left|\sum_{i=1}^m (X_i - V_i)\right|_+ > \frac{u}{4}\right) \leqq c_3(r)(1+\delta^{-2})(\exp(-c_4(r)\delta^2 u/4K\sqrt{r}) + \exp(-u^2/1025k_n\delta^2K^2)) , \tag{2.27}$$

where we apply (2.3) with $\sigma = K$, $L = [\delta^{-2}]$ and $M = \sqrt{r}K$.

Next, let

$$\Sigma = \operatorname{cov}(B(f_1), \ldots, B(f_r))$$

and

$$\Sigma_n = \operatorname{cov}(X_1^{(1)}, \ldots, X_1^{(r)}) .$$

We can assume that for $1 \leqq i \leqq k_n$,

$$(V_i^{(1)}, \ldots, V_i^{(r)}) = \Sigma_n^{1/2}\bar{Z}_i ,$$

where $\bar{Z}_1, \ldots, \bar{Z}_{k_n}$ are independent standard normal r-vectors.

Set

$$W_i = \Sigma^{1/2}\bar{Z}_i, \quad i = 1, \ldots, k_n . \tag{2.28}$$

Clearly $(W_i^{(1)}, \ldots, W_i^{(r)}) \stackrel{d}{=} (\bar{B}_i(f_1), \ldots, \bar{B}_i(f_r))$, $i = 1, \ldots, k_n$, where $\bar{B}_1, \ldots, \bar{B}_{k_n}$ are independent P_0-Brownian bridges. Therefore, without loss of generality, we can assume that

$$(W_i^{(1)}, \ldots, W_i^{(r)}) = (\bar{B}_i(f_1), \ldots, \bar{B}_i(f_r)), \quad 1 \leqq i \leqq k_n , \tag{2.29}$$

Also since $\Sigma_n \to \Sigma$ as $n \to \infty$, we have by using symmetry of the $\bar{Z}_i$'s and a standard bound on the tail of a normal distribution that for all large n

$$\Delta_2 := P\left(\max_{1\leqq m\leqq k_n}\left|\sum_{i=1}^m (V_i - W_i)\right|_+ \geqq u/4\right) \leqq \exp(-u^2/\delta^2 k_n) \tag{2.30}$$

for all $u > 0$ and $n \geqq n_2(\delta)$.

It is easy to see now that for all $n \geqq n(\delta) = n_1(\delta) \vee n_2(\delta)$ large enough so that $\mathscr{F}_n'(\delta) \subset \mathscr{F}'(2\delta)$ (by (1.15)),

$$\begin{aligned} &P\left(\max_{1\leqq m\leqq k_n}\left\|\sqrt{m}\alpha_m^{(n)} - \sum_{i=1}^m \bar{B}_i\right\|_{\mathscr{F}} \geqq u + \beta_{n,k_n}(\delta)\right) \\ &\leqq \Delta_1 + \Delta_2 \\ &\quad + P\left(\max_{1\leqq m\leqq k_n}\|\sqrt{m}\alpha_m^{(n)}\|_{\mathscr{F}_n'(\delta)} \geqq \frac{u}{4} + A_1\sqrt{k_n}E\|\alpha_{k_n}^{(n)}\|_{\mathscr{F}_n'(\delta)}\right) \\ &\quad + P\left(\max_{1\leqq m\leqq k_n}\left\|\sum_{i=1}^m \bar{B}_i\right\|_{\mathscr{F}'(2\delta)} \geqq \frac{u}{4} + A_1\sqrt{k}_nE\|\bar{B}\|_{\mathscr{F}}\right) \\ &=: \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4 . \end{aligned} \tag{2.31}$$

To bound Δ_3, we note by the Ottaviani inequality, (see, for instance Dudley 1984, Lemma 3.2.7),

$$\Delta_3 \leqq 2P\left(\|\sqrt{k_n}\alpha_{k_n}^{(n)}\|_{\mathscr{F}_n'(\delta)} \geqq \frac{u}{4} + (A_1 - 2)\sqrt{k_n}E\|\alpha_{k_n}^{(n)}\|_{\mathscr{F}_n'(\delta)}\right) , \tag{2.32}$$

where we use the fact that by Jensen's inequality

$$\sqrt{m}E\|\alpha_m^{(n)}\|_{\mathscr{F}_n'(\delta)} \leqq \sqrt{k_n}E\|\alpha_{k_n}^{(n)}\|_{\mathscr{F}_n'(\delta)}, \quad 1 \leqq m \leqq k_n .$$

Now applying (2.17), we get

$$\Delta_3 \leqq 2\exp(-u^2/16k_nA^2\delta^2) + 2\exp(-u/4KA) , \tag{2.33}$$

provided we have chosen $A_1 \geqq A + 2$.

Finally using Fact 2.5 along with Lévy's inequality, Lemma 3.2.11 of Dudley (1984), we find that

$$\Delta_4 \leqq 2\exp(-u^2/128k_n\delta^2) , \tag{2.34}$$

provided of course we have chosen $A \geqq 2$. Combining (2.31), (2.27), (2.30), (2.33) and (2.34) we obtain (2.23). □

Remark. 2.1 Though we will use Proposition 2.2 only for empirical measures on the Euclidean space, it might be worthwhile to point out that the above proof works for empirical measures on general probability spaces. Thus the conclusion of Proposition 2.2 holds as well in this more abstract setting.

3 Proofs of main results

In our proofs we shall make repeated use of the following proposition. Given any integer $n \geqq 1$ and invertible transformation $h_n : \mathbb{R}^d \to \mathbb{R}^d$, let $Y_1^{(n)}, Y_2^{(n)}, \ldots,$ be i.i.d. P_n. Set for $f \in \mathscr{F}, n \geqq 1$ and $j \geqq 1$

$$S_j^{(n)}(f) = \sum_{i=1}^{j} f(h_n^{-1}(X_i - x)) - ja_nP_n(f) , \tag{3.1}$$

and

$$T_j^{(n)}(f) = \sqrt{j}\alpha_j^{(n)}(f) . \tag{3.2}$$

Further, independently of $Y_1^{(n)}, Y_2^{(n)}, \ldots,$ let $\varepsilon_1, \varepsilon_2, \ldots,$ be i.i.d. Bernoulli (a_n) random variables and set $\nu(j) = \varepsilon_1 + \cdots + \varepsilon_j, j \geqq 1$.

Proposition 3.1 *With the above notation and assumptions for all $n \geqq 1$*

$$(S_j^{(n)}(f))_{j\geqq 1} \stackrel{d}{=} (T_{\nu(j)}^{(n)}(f) + (\nu(j) - ja_nP_n(f)))_{j\geqq 1} \tag{3.3}$$

as vectors indexed by $\mathscr{F}$.

Proof. Let $\eta_1, \eta_2, \ldots,$ be i.i.d. Q_n, independent of the $Y_i^{(n)}$'s and the ε_i's, where

$$Q_n(B) = \mathbb{P}(B \cap A_n^C) / \mathbb{P}(A_n^C) \tag{3.4}$$

for all $B \in \mathscr{B}$. Then it is readily checked that the random variables X_j defined for any $j \geqq 1$ to be,

$$X_j = (h_n Y^{(n)}_{v(j)} + x)\varepsilon_j + \eta_{j-v(j)}(1-\varepsilon_j) \tag{3.5}$$

are i.i.d. $\mathbb{P}$. Recalling that all the functions $f \in \mathscr{F}$ are zero outside J, we see that

$$(S^{(n)}_j(f))_{j \geqq 1} \stackrel{d}{=} \left(\sum_{i=1}^{v(j)} f(Y^{(n)}_i) - ja_n P_n(f) \right)_{j \geqq 1} . \qquad \square$$

Before we proceed with the proof of Theorem 1.1, we remark that it is based on methods and ideas of proving weak convergence theorems for random sample size empirical processes that originate with Pyke (1968), and which have been recently generalized in Klaassen and Wellner (1992). See, in particular, the proof of their Theorem 4.

Proof of Theorem 1.1. First, by Proposition 3.1, for each integer $n \geqq 1$,

$$\{L_n(f,h_n) : \ f \in \mathscr{F}\} \stackrel{d}{=} \left\{ \sqrt{\frac{v(n)}{na_n}} \alpha^{(n)}_{v(n)}(f) + \frac{(v(n)-na_n)P_n(f)}{\sqrt{na_n}} : \ f \in \mathscr{F} \right\} . \tag{3.6}$$

Using assumption (A) it is easily shown that, with a as in (A.iii),

$$(v(n) - na_n)/\sqrt{na_n} \stackrel{d}{\to} \sqrt{1-a}Z \quad \text{as } n \to \infty . \tag{3.7}$$

Furthermore, (S) and (F) imply

$$\alpha^{(n)}_{k(n)} \Rightarrow B_0 , \tag{3.8}$$

where B_0 is a P_0-Brownian bridge. (see Sheehy and Wellner (1992, Theorem 3.1).

Also, obviously by (3.7)

$$v(n)/na_n \stackrel{P}{\to} 1 \quad \text{as } n \to \infty . \tag{3.9}$$

Therefore, since the two sequences in (3.7) and (3.8) are independent of each other, we see by (3.6), (3.7) and (3.9) to finish the proof of Theorem 1.1, it suffices to prove that for every $c > 0$, as $n \to \infty$,

$$\max_{|k(n)-m| \leqq c\sqrt{k(n)}} \| \sqrt{k(n)}\alpha^{(n)}_{k(n)} - \sqrt{m}\alpha^{(n)}_m \|_{\mathscr{F}} / \sqrt{na_n} \stackrel{P}{\to} 0 . \tag{3.10}$$

To establish (3.10), in turn, it is enough to verify that for all $c > 0$, as $n \to \infty$,

$$\max_{1 \leqq m \leqq c\sqrt{k(n)}} \| \sqrt{m}\alpha^{(n)}_m \| / \sqrt{k(n)} \stackrel{P}{\to} 0 . \tag{3.11}$$

This will be accomplished by a number of lemmas.

Lemma 3.1 *Under assumptions* (A), (S) *and* (F) *there exists a constant* M *such that for all* $n \geqq 1$

$$E\|\alpha^{(n)}_{k(n)}\|^2_{\mathscr{F}} \leqq M . \tag{3.12}$$

Proof. Let $\tilde{Y}_1^{(n)}, \ldots, \tilde{Y}_{k(n)}^{(n)}$ be independent copies of $Y_1^{(n)}, \ldots, Y_{k(n)}^{(n)}$ with corresponding empirical process $\tilde{\alpha}_{k(n)}^{(n)}$. Since $E\alpha_{k(n)}^{(n)}(f) = 0$ for all $f \in \mathscr{F}$, by Jensen's inequality it is enough to show that there exists a constant $\bar{M} > 0$ such that for all $n \geqq 1$

$$E\|\alpha_{k(n)}^{(n)} - \tilde{\alpha}_{k(n)}^{(n)}\|_{\mathscr{F}}^2 =: E_n(\mathscr{F}) < \bar{M} . \tag{3.13}$$

Set

$$t_n := \inf\{t > 0 : P(\|\alpha_{k(n)}^{(n)} - \tilde{\alpha}_{k(n)}^{(n)}\|_{\mathscr{F}} > t) \leqq 1/72\} .$$

Applying the Hoffmann–Jørgensen inequality, cf. Ledoux and Talagrand (1991), page 156, we get

$$E_n(\mathscr{F}) \leqq 18 \left(E \max_{1 \leqq i \leqq k(n)} \|f(Y_i^{(n)}) - f(\tilde{Y}_i^{(n)})\|_{\mathscr{F}}^2 / k(n) + t_n^2 \right),$$

which, in turn, is

$$\leqq 18 \left(E \left(\max_{1 \leqq i \leqq k(n)} Z_{i,n}^2 / k(n) \right) + t_n^2 \right),$$

where $Z_{i,n} := F(Y_i^{(n)}) + F(\tilde{Y}_i^{(n)}),\ i = 1, \ldots, k(n)$.

By assumption (F.i), we have $EZ_{i,n}^2 < \tilde{M},\ 1 \leqq i \leqq k(n)$, for some $\tilde{M} < \infty$ and it follows that

$$E \max_{1 \leqq i \leqq k(n)} Z_{i,n}^2 / k(n) \leqq \sum_{i=1}^{k(n)} EZ_{i,n}^2 / k(n) < \tilde{M} . \tag{3.14}$$

Finally, noting that by (1.15) and (1.20), t_n is bounded, we obtain (3.12). □

Lemma 3.2 *Let $\Delta_n(\delta), 0 < \delta < 1, n \geqq 1$, be a set of non-negative random variables such that for all $\varepsilon > 0$*

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} P(\Delta_n(\delta) > \varepsilon) = 0 \tag{3.15}$$

and for some constant R

$$E\Delta_n^2(\delta) < R \quad \text{for all } n \geqq 1 \quad \text{and } 0 < \delta < 1 . \tag{3.16}$$

Then

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} E\Delta_n(\delta) = 0 . \tag{3.17}$$

Proof. The proof is trivial. Note that for any $\varepsilon > 0$

$$E\Delta_n(\delta) \leqq \varepsilon + R^{1/2}(P(\Delta_n(\delta) > \varepsilon))^{1/2} . \qquad \square$$

Lemma 3.3 *Let $m(n)$ be any sequence of positive integers such that*

$$m(n) / k(n) \to 0 \quad \text{as } n \to \infty . \tag{3.18}$$

Then

$$\limsup_{n \to \infty} \max_{1 \leqq m \leqq m(n)} E\|\sqrt{m}\alpha_m^{(n)}\|_{\mathscr{F}} / \sqrt{k(n)} = 0 . \tag{3.19}$$

Proof. By (F.ii) and (F.iii) for each $0 < \delta < 1$ there exist $f_1, \ldots, f_{r(\delta)} \in \mathscr{F}$ such that for all $1 \leqq m \leqq m(n)$

$$E\|\sqrt{m}\alpha_m^{(n)}\|_{\mathscr{F}} / \sqrt{k(n)} \leqq \sum_{i=1}^{r(\delta)} E\frac{|\sqrt{m}\alpha_m^{(n)}(f_i)|}{\sqrt{k(n)}} + E\Delta_{n,m}(\delta)\,, \tag{3.20}$$

where

$$\Delta_{n,m}(\delta) = \|\sqrt{m}\alpha_m^{(n)}\|_{\mathscr{F}_n'(\delta)} / \sqrt{k(n)}\,.$$

Now for all n large enough so that $m(n) \leqq k(n)$, we have by Jensen's inequality for all $1 \leqq m \leqq m(n)$

$$E\Delta_{m,n}(\delta) \leqq E\Delta_n(\delta)\,, \tag{3.21}$$

where

$$\Delta_n(\delta) = \|\alpha_{k(n)}^{(n)}\|_{\mathscr{F}_n'(\delta)}\,.$$

Further, by Cauchy–Schwarz and (F.i) for all $1 \leqq m \leqq m(n)$

$$E\frac{|\sqrt{m}\alpha_m^{(n)}(f_i)|}{\sqrt{k(n)}} \leqq \frac{\sqrt{m(n)}}{\sqrt{k(n)}}\sqrt{P_n(F^2)} \quad \text{for } i = 1, \ldots, r(\delta)\,. \tag{3.22}$$

Noting that

$$E\Delta_n^2(\delta) \leqq 4E\|\alpha_{k(n)}^{(n)}\|_{\mathscr{F}}^2\,, \tag{3.23}$$

we see that (3.19) follows from (3.21)–(3.23), (1.20) and Lemmas 3.1 and 3.2. □

The proof of (3.11) is now an easy consequence of Lemma 3.3 and Ottaviani's inequality. □

Proof of Theorem 1.2. For notational convenience we assume $K = 1$ in (F.iv). First we record an inequality for the empirical process which is obtained by combining Lemmas 3.2.1 and 3.2.7 of Dudley (1984) with his Remark 3.2.5.

Fact 3.1 *We have for all* $x \geqq 0$ *and* $n \geqq 1$

$$P\left(\max_{1 \leqq k \leqq n} \|T_k^{(n)}\|_{\mathscr{F}} \geqq x + 3E\|T_n^{(n)}\|_{\mathscr{F}}\right) \leqq 2(\exp(-x^2/12n) + \exp(-x/4))\,. \tag{3.24}$$

Lemma 3.4 *For any* $x \geqq 0$ *and* $na_n \geqq 1$

$$P\left(\max_{1 \leqq k \leqq n} \|T_{v(k)}^{(n)} - T_{[ka_n]}^{(n)}\|_{\mathscr{F}} \geqq x + \gamma_n\right) \leqq 4na_n\left\{\exp\left(\frac{-x^2}{192\sqrt{na_nLL_n}}\right) + \exp(-x/8)\right\} + 2(Ln)^{-1.5} + 2\exp(-2\sqrt{na_nLLn})\,, \tag{3.25}$$

where $\gamma_n := 12E\|T^{(n)}_{l_n}\|_{\mathscr{F}}, l_n := [4\sqrt{na_nLLn}]$ *and* $v(k), 1 \leqq k \leqq n$, *is as in Proposition* 3.1.

Proof. Obviously,

$$(3.26) \qquad P\left(\max_{1\leqq k\leqq n}\|T^{(n)}_{v(k)} - T^{(n)}_{[ka_n]}\|_{\mathscr{F}} \geqq x+\gamma_n\right)$$
$$\leqq P\left(\max_{1\leqq k\leqq n}|v(k)-ka_n| \geqq 3\sqrt{na_nLLn}\right)$$
$$+P\left(\max_{1\leqq k\leqq [na_n]}\max_{|l-k|\leqq l_n}\|T^{(n)}_l - T^{(n)}_k\|_{\mathscr{F}} \geqq x+\gamma_n\right)$$
$$=: \Delta_{n,1}+\Delta_{n,2}\ .$$

Using Fact 2.3 we see that

$$(3.27) \qquad \Delta_{n,1} \leqq 2(Ln)^{-1.5} + 2\exp(-2\sqrt{na_nLLn})\ .$$

Next observe that

$$(3.28) \qquad \Delta_{n,2} \leqq 2na_nP\left(\max_{1\leqq k\leqq 2l_n}\|T^{(n)}_k\|_{\mathscr{F}} \geqq \frac{x}{2}+\frac{\gamma_n}{2}\right)\ .$$

Since we have

$$(3.29) \qquad E\|T^{(n)}_{2l_n}\|_{\mathscr{F}} \leqq 2E\|T^{(n)}_{l_n}\|_{\mathscr{F}}\ ,$$

we get via Fact 3.1

$$(3.30) \qquad \Delta_{n,2} \leqq 4na_n\left\{\exp\left(-\frac{x^2}{48l_n}\right)+\exp\left(-\frac{x}{8}\right)\right\}\ .$$

Combining this with our bound for $\Delta_{n,1}$, we get (3.25). □

From Lemma 3.4 and Proposition 2.2 we shall derive the following crucial lemma.

Lemma 3.5 *Given* $0<\delta<1$, *there exists an* $m(\delta)$ *such that for each* $n \geqq m(\delta)$ *one can construct independent* P_0*-Brownian bridges* $\bar{B}_j, 1 \leqq j \leqq [na_n]$ *satisfying*

(3.31)

$$P\left(\max_{1\leqq k\leqq n}\|T^{(n)}_{v(k)} - \sum_{j=1}^{[ka_n]}\bar{B}_j\|_{\mathscr{F}} \geqq 2A_0\delta\sqrt{na_nLLn}+\gamma_n+\beta_{n,[na_n]}(\delta)\right)$$
$$\leqq K_3\{na_n\exp(-K_4\sqrt{na_nLLn})+(Ln)^{-1.5}\},$$

where K_3 *and* K_4 *are constants depending only on* δ *and* $\beta_{n,[na_n]}(\delta)$ *is defined as in* (2.24).

Proof. Just use the inequality

$$P\left(\max_{1\leqq k\leqq n}\|T^{(n)}_{v(k)}-\sum_{i=1}^{[ka_n]}\bar{B}_i\|_{\mathscr{F}}\geqq 2A_0\delta\sqrt{na_nLLn}+\gamma_n+\beta_{n,[na_n]}(\delta)\right)$$
$$\leqq P\left(\max_{1\leqq k\leqq n}\|T^{(n)}_{v(k)}-T^{(n)}_{[ka_n]}\|_{\mathscr{F}}\geqq \frac{A_0}{2}\delta\sqrt{na_nLLn}+\gamma_n\right)$$
$$+P\left(\max_{1\leqq k\leqq n}\|T^{(n)}_{[ka_n]}-\sum_{i=1}^{[ka_n]}\bar{B}_i\|_{\mathscr{F}}\geqq \frac{3}{2}A_0\delta\sqrt{na_nLLn}+\beta_{n,[na_n]}(\delta)\right),$$

and recall that we assume $K=1$. □

Remark. 3.1 From the proof it is clear that we can choose the Brownian bridges $\bar{B}_j, 1\leqq j\leqq [na_n]$, independent of $v(k), 1\leqq k\leqq [na_n]$, which we will also assume from now on.

Next, let $\{\bar{K}(t,f), t\geqq 0, f\in\mathscr{F}\}$ be a Kiefer process indexed by $\mathscr{F}$ such that

$$\bar{K}(l,f)=\sum_{i=1}^{l}\bar{B}_i(f),\quad f\in\mathscr{F},\quad 1\leqq l\leqq [na_n]\,. \tag{3.32}$$

In view of Remark 3.1, we can assume that $\bar{K}$ is independent of $v(k), 1\leqq k\leqq n$.

Lemma 3.6 *For each $n\geqq 1$ and $x>0$*

$$P\left(\max_{1\leqq k\leqq n}\|\bar{K}(ka_n,f)-\bar{K}([ka_n],f)\|_{\mathscr{F}}\geqq x+2E\|B\|_{\mathscr{F}}\right)\leqq (2na_n+2)\exp(-x^2/2)\,. \tag{3.33}$$

Proof. It is easy to see that

$$\max_{1\leqq k\leqq n}\|\bar{K}(ka_n,f)-\bar{K}([ka_n],f)\|_{\mathscr{F}}\leqq \max_{0\leqq k\leqq [na_n]}\max_{0\leqq y\leqq 1}\|\bar{K}(k+y,f)-\bar{K}(k,f)\|_{\mathscr{F}}\,.$$

Since the Kiefer process $\bar{K}$ has stationary independent increments, we clearly have that the above probability is

$$\leqq ([na_n]+1)P\left(\max_{0\leqq y\leqq 1}\|\bar{K}(y,f)\|_{\mathscr{F}}\geqq x+2E\|B\|_{\mathscr{F}}\right),$$

which by Lévy's inequality is

$$\leqq 2(na_n+1)P(\|\bar{B}_1\|_{\mathscr{F}}\geqq x+2E\|\bar{B}_1\|_{\mathscr{F}})$$

from which the assertion follows, using Lemma 3.1 of Ledoux and Talagrand (1991). □

Setting for any $n \geqq 1$

$$B_j(f) = \frac{1}{\sqrt{a_n}}(\bar{K}(ja_n, f) - \bar{K}((j-1)a_n, f)), \quad 1 \leqq j \leqq n ,$$

we obtain independent P_0-Brownian bridges which are also independent of $v(k), 1 \leqq k \leqq n$.

Combining Lemmas 3.5 and 3.6 we get the following essential result.

Proposition 3.2 *Given* $0 < \delta < 1$, *there exists an* $m(\delta)$ *such that for each* $n \geqq m(\delta)$, *one can construct independent* P_0*-Brownian bridges* $B_j, 1 \leqq j \leqq n$ *which are independent of* $v(k)$, $1 \leqq k \leqq n$, *such that*

$$P\left(\max_{1 \leqq k \leqq n} \|T^{(n)}_{v(k)} - \sqrt{a_n}\sum_{i=1}^{k} B_j\|_{\mathscr{F}} \right.$$
$$\left. \geqq 3A_0\delta\sqrt{na_nLLn} + \beta_{n,[na_n]}(\delta) + \gamma_n + 2E\|B\|_{\mathscr{F}}\right)$$
$$\leqq K_5\{na_n \exp(-K_6\sqrt{na_nLLn}) + (Ln)^{-1.5}\} ,$$

where K_5 *and* K_6 *are positive constants depending on* δ *only.*

We now approximate $v(k), 1 \leqq k \leqq n$, where we use once more Proposition 2.1. Employing more sophisticated strong approximation techniques it would be possible to get much better inequalities. However, the subsequent Lemma 3.7 will be sufficient for our purposes.

Lemma 3.7 *Given* $0 < \delta < 1$ *there exists an* $m(\delta)$ *such that for each* $n \geqq m(\delta)$, *one can construct independent standard normal random variables* $Z_i, 1 \leqq i \leqq n$ *such that*

$$P\left(\max_{1 \leqq k \leqq n} |v(k) - \sum_{i=1}^{k} Z_i\sqrt{a_n(1-a_n)}| \geqq 10\delta\sqrt{na_nLLn}\right)$$
$$\leqq K_7\{\exp(-K_8\sqrt{na_nLLn}) + (Ln)^{-1.5}\} ,$$

where K_7, K_8 *are positive constants.*

Proof. Apply Proposition 2.1 with $\sigma^2 = a_n, M = 1$ and $L = [\delta^{-2}]$. □

Remark. 3.2 It is clear that we can choose the Z_i's independent of the Brownian bridges $B_i, 1 \leqq i \leqq n$, defined in Proposition 3.2.

Lemma 3.8 *We have for all* $n \geqq 1$ *and* $x > 0$

$$P\left(\max_{1 \leqq k \leqq n}\left|\left(\sqrt{a_n(1-a_n)} - \sqrt{a_n}\right)\sum_{i=1}^{k} Z_i\right| \geqq x\right)$$
$$\leqq 4\exp(-x^2/2na_n^3) .$$

Proof. By Lévy's inequality the above probability is bounded above by

$$2P\left(\left|\left(\sqrt{a_n(1-a_n)} - \sqrt{a_n}\right)\sum_{i=1}^{n} Z_j\right| \geqq x\right) ,$$

which by a standard exponential inequality is

$$\leqq 4\exp(-x^2/2na_n(1-\sqrt{1-a_n})^2)\,.$$

Using the trivial fact that $|1-\sqrt{1-a_n}| \leqq a_n$, we can finish the proof. □

Now set

$$\bar{W}_i(f) := B_i(f) + Z_i P_0(f), \quad f \in \mathscr{F}, \quad 1 \leqq i \leqq n\,. \tag{3.34}$$

Due to the independence of Z_i and B_i, we obtain in this way independent P_0-Brownian motions indexed by $\mathscr{F}$. Recall (1.29) and the definition of b_n in (F.vi).

Lemma 3.9 *Given* $0 < \delta < 1$ *there exists an* $m(\delta)$ *such that for each* $n \geqq m(\delta)$ *and all* $t > 0$

$$P\left(\max_{1\leqq k\leqq n}\left\|\left(\sqrt{a_n}-\sqrt{b_n}\right)\sum_{i=1}^{k}\bar{W}_i\right\|_{\mathscr{F}} \geqq 2\sqrt{na_n}E\|\bar{W}_1\|_{\mathscr{F}}+t\right)$$
$$\leqq 2\exp(-t^2/2\delta^2 na_n)\,.$$

Proof. The proof follows from (1.26), Lévy's inequality and the following fact: for all $x > 0$

$$P\left(\left\|\sum_{i=1}^{n}\bar{W}_i\right\|_{\mathscr{F}} > 2\sqrt{n}E\|\bar{W}_1\|_{\mathscr{F}}+x\right) \leqq \exp\left(-x^2/2n\sup_{f\in\mathscr{F}}P_0(f^2)\right), \tag{3.35}$$

which can be readily derived from Lemma 3.1 of Ledoux and Talagrand (1981). Noting that $\sup_{f\in\mathscr{F}}P_0(f^2) \leqq 1$, we are done. □

We can now infer from Propositions 3.1 and 3.2 and Lemmas 3.7–3.9:

Proposition 3.3 *Given* $0 < \delta < 1$, *there exists an* $m(\delta)$ *such that for each* $n \geqq m(\delta)$, *one can construct independent* P_0*-Brownian motions* $\bar{W}_i$, $1 \leqq i \leqq n$, *indexed by* $\mathscr{F}$ *so that*

(3.36)
$$P\left(\max_{1\leqq k\leqq n}\left\|S_k^{(n)}-\sqrt{b_n}\sum_{i=1}^{k}\bar{W}_i\right\|_{\mathscr{F}} \geqq \tilde{A}\left(\delta\sqrt{na_nLLn}+\gamma_n+\tilde{\beta}_{n,[na_n]}\right)\right)$$
$$\leqq K_9\left(na_n\exp\left(-K_{10}\sqrt{na_nLLn}\right)+(Ln)^{-1.5}+(Ln)^{-(\delta/a_n)^2}\right),$$

where $\tilde{A} \geqq 1$ *is an absolute constant*, K_9 *and* K_{10} *are constants depending on* δ *only and*

$$\tilde{\beta}_{n,[na_n]} = E\|T_{[na_n]}^{(n)}\|_{\mathscr{F}} + \sqrt{na_n}E\|B\|_{\mathscr{F}} + \sqrt{na_n}E\|\bar{W}_1\|_{\mathscr{F}}\,. \tag{3.37}$$

Next set for $1 \leqq i \leqq n$

$$W_i(f) = \sqrt{b_n}\,\bar{W}_i(f\circ h_n), \quad f \in \mathscr{F}\,, \tag{3.38}$$

and observe that by (1.25), W_i, $1 \leqq i \leqq n$, are again independent P_0-Brownian motions indexed by $\mathscr{F}$. It is easy now to see that inequality (3.36) is still valid if we replace $\sqrt{b_n}\,\bar{W}_i(f)$ by $W_i(f,h_n) := W_i(f \circ h_n^{-1})$, $1 \leqq i \leqq n$. (Recall (F.v).)

We now have all of the necessary tools to finish the proof of Theorem 1.2. It is enough to show that for any given $0 < \delta < \frac{1}{2}$ there is a construction possible such that with probability one

$$\limsup_{n\to\infty} \left\| S_n^{(n)}(f) - \sum_{i=1}^{n} W_i(f,h_n) \right\|_{\mathscr{F}} \Big/ \sqrt{na_n LLn} \leqq D\sqrt{\delta}\,, \tag{3.39}$$

where D is a positive constant not depending on δ. Statement (1.27) then follows from (3.39) by a known argument of Major (1976).

Set $m_k := 2^{k-1}$, $n_k := m_{k+1} - 1$, $k = 1, 2, \ldots$ Using Proposition 3.3 in combination with (3.38), we can construct independent P_0-Brownian motions W_i, $m_k \leqq i \leqq n_k$ such that for large k

$$P(\Delta_k \geqq 1.5\tilde{A}\delta c_{m_k}) \leqq 2K_9(Lm_k)^{-1.5}\,, \tag{3.40}$$

where $c_n = \sqrt{na_n LLn}$, and

$$\Delta_k := \max_{m_k \leqq n \leqq n_k} \left\| S_n^{(m_k)}(f) - S_{n_{k-1}}^{(m_k)}(f) - \sum_{i=m_k}^{n} W_i(f,h_{m_k}) \right\|_{\mathscr{F}} .$$

Notice that we are using the following fact from assumption (F.vii) that $(E\|T_{[na_n]}^{(n)}\|_{\mathscr{F}} + \gamma_n)/c_n \to 0$ as $n \to \infty$, by arguing as in Lemma 3.1.

Due to condition (F.v) we have for each $m_k \leqq n \leqq n_k$ and $1 \leqq \ell \leqq k$,

$$\left\| S_n^{(n)}(f) - \sum_{i=1}^{n} W_i(f,h_n) \right\|_{\mathscr{F}} \leqq \sum_{i=1}^{\ell} \Delta_{k+1-i} + Z_k(1) + Z_k(2)\,,$$

where

$$Z_k(1) := \max_{1 \leqq m \leqq n_{k-\ell}} \left\| S_m^{(m_k)}(f) \right\|_{\mathscr{F}}\,,$$

$$Z_k(2) := \max_{1 \leqq m \leqq n_{k-\ell}} \left\| \sum_{i=1}^{m} W_i(f,h_{m_k}) \right\|_{\mathscr{F}} .$$

It is easy now to see that

$$\begin{aligned} &\limsup_{n\to\infty} \left\| S_n^{(n)}(f) - \sum_{i=1}^{n} W_i(f,h_n) \right\|_{\mathscr{F}} \Big/ c_n \\ &\quad \leqq \limsup_{k\to\infty} \sum_{i=1}^{\ell} \Delta_{k+1-i}/c_{m_k} \\ &\qquad + \limsup_{k\to\infty} Z_k(1)/c_{m_k} + \limsup_{k\to\infty} Z_k(2)/c_{m_k}\,. \end{aligned} \tag{3.41}$$

Using (3.40) in conjunction with the Borel–Cantelli lemma, we readily obtain for each $1 \leqq i \leqq \ell$ with probability one

$$\limsup_{k\to\infty} \Delta_{k+1-i}/c_{m_k} \leqq \tfrac{3}{2}\tilde{A}\,\delta\,, \tag{3.42}$$

In view of Proposition 3.3, we can assume that for any large k, there are independent P_0-Brownian motions $\bar{W}_i$, $1 \leqq i \leqq m_k$ such that

$$P\left(\left|Z_k(1) - \sqrt{b_{m_k}} \max_{1 \leqq m \leqq n_{k-\ell}} \left\|\sum_{i=1}^m \bar{W}_i\right\|_{\mathscr{F}}\right| \geqq \tfrac{3}{2}\tilde{A}\,\delta c_{m_k}\right) \leqq 2K_9(Lm_k)^{-1.5}\,. \tag{3.43}$$

Moreover, by Lévy's inequality and (3.35), we have for all large k

$$P\left(\max_{1 \leqq m \leqq n_{k-\ell}} \left\|\sum_{i=1}^m \bar{W}_i\right\|_{\mathscr{F}} \geqq \frac{\tilde{A}}{2}\delta c_{m_k}/b_{m_k}^{1/2}\right) \leqq 2(Lm_k)^{-\tilde{A}^2\delta^2 2^{\ell/17}}\,. \tag{3.44}$$

Setting $l := [7\delta^{-1/2}]$ and recalling that $\tilde{A} \geqq 1$, we can infer from (3.43) and (3.44) by a Borel–Cantelli argument that with probability one

$$\limsup_{k\to\infty} Z_k(1)/c_{m_k} \leqq 2\tilde{A}\,\delta\,. \tag{3.45}$$

Finally note that (3.44) also implies that with probability one

$$\limsup_{k\to\infty} Z_k(2)/c_{m_k} \leqq \frac{\tilde{A}\,\delta}{2}\,. \tag{3.46}$$

Combining (3.41), (3.42), (3.45) and (3.46), we obtain statement (3.39) with $D = 13\tilde{A}$, which finishes the proof of Theorem 1.2. □

Proof of Corollary 1.1. Recall the notation from the introduction, namely (1.18), e_0, ρ_0, $\tilde{W}_n$ and $\mathscr{K}$.

Step 1. Let $n_k := [q^k]$, where $q > 1$. We claim that with probability one

$$d(\tilde{W}_{n_k}/\sqrt{2Lk}, \mathscr{K}) \to 0, \quad \text{as } k \to \infty, \text{ and} \tag{3.47}$$

$$\text{the set of limit points of } \{\tilde{W}_{n_k}/\sqrt{2Lk}\}_{k \geqq 1} \text{ is equal to } \mathscr{K}. \tag{3.48}$$

In view of Theorem 4.1 of Carmona and Kôno (1976) it suffices to show that for any linear functional $H \in B^*$ and $k \geqq 1$

$$\lim_{m\to\infty} E(H(\tilde{W}_{n_k})H(\tilde{W}_{n_{k+m}})) = 0\,. \tag{3.49}$$

To see (3.49), note that by independence

$$\begin{aligned} &E(H(\tilde{W}_{n_k}) \cdot H(\tilde{W}_{n_{k+m}})) \\ &\quad = \sum_{i=1}^{n_k} E(H(W_{i,k})H(W_{i,k+m})) \Big/ \sqrt{n_k n_{k+m} b_{n_k} b_{n_{k+m}}}\,, \end{aligned}$$

where $W_{i,k+m} = (W_i(f, h_{n_{k+m}}))_{f\in\mathscr{F}}$, $m \geqq 0$.

Moreover, we have

$$\begin{aligned} E(H(W_{i,k})H(W_{i,k+m})) &\leqq (E(H^2(W_{i,k})))^{1/2}(E(H^2(W_{i,k+m})))^{1/2} \\ &\leqq \|H\|^2(E\|W_{i,k}\|^2_{\mathscr{F}} E\|W_{i,k+m}\|^2_{\mathscr{F}})^{1/2}\,. \end{aligned}$$

Recalling (3.38), we readily obtain that

$$E(H(\tilde{W}_{i,k})H(\tilde{W}_{i,k+m})) \leqq \|H\|^2 E\|W_1\|^2_{\mathscr{F}}(b_{n_k}b_{n_{k+m}}n_k/n_{k+m})^{1/2} ,$$

and consequently (3.49).

Step 2. Let $n_k = n_k(q) = [q^k]$ and set

$$\Delta_k(q) := \max_{n_k \leqq n \leqq n_{k+1}} \|\tilde{W}_n - \tilde{W}_{n_k}\|_{\mathscr{F}} .$$

To complete the proof it is obviously enough to show that

$$\limsup_{n\to\infty} \Delta_k(q) \Big/ \sqrt{2Lk} \leqq c(q) \text{ a.s.} , \tag{3.50}$$

where $\lim_{q\downarrow 1} c(q) = 0$. Towards this end we note that

$$\Delta_k(q) \leqq \Delta_{k,1}(q) + \Delta_{k,2}(q) + \Delta_{k,3}(q) ,$$

where

$$\Delta_{k,1}(q) := \max_{n_k \leqq n \leqq n_{k+1}} \left| \frac{1}{\sqrt{n_k}} - \frac{\sqrt{b_{n_k}}}{\sqrt{nb_n}} \right| \frac{1}{\sqrt{b_{n_k}}} \left\| \sum_{i=1}^{n_k} W_i(f, h_{n_k}) \right\|_{\mathscr{F}} ,$$

$$\Delta_{k,2}(q) := \max_{n_k \leqq n \leqq n_{k+1}} \left\| \sum_{i=1}^{n} (W_i(f, h_n) - W_i(f, h_{n_k})) \right\|_{\mathscr{F}} \Big/ \sqrt{n_k b_{n_k}} ,$$

$$\Delta_{k,3}(q) := \max_{n_k \leqq n \leqq n_{k+1}} \left\| \sum_{i=n_k+1}^{n} W_i(f, h_{n_k}) \right\|_{\mathscr{F}} \Big/ \sqrt{n_k b_{n_k}} .$$

Since (3.47) holds we have using (A.iv) and (1.26)

$$\limsup_{k\to\infty} \Delta_{k,1}(q) / \sqrt{2Lk} \leqq 1 - q^{-1/2} \text{ a.s.} \tag{3.51}$$

Notice that

$$\Delta_{k,2}(q) = \max_{n_k \leqq n \leqq n_{k+1}} \left\| \sum_{i=1}^{n} (W_i(f_{n,k}, h_{n_k}) - W_i(f, h_{n_k})) \right\|_{\mathscr{F}} \Big/ \sqrt{n_k b_{n_k}} ,$$

where $f_{n,k} = f \circ h_n^{-1} \circ h_{n_k}$.

Next we record the fact which follows from Lemma 3.10 below that

$$\limsup_{n\to\infty} \sup_{f\in\mathscr{F}} \max_{n_k \leqq n \leqq n_{k+1}} e_0(f_{n,k}, f) \leqq \delta(q) , \tag{3.52}$$

where $\delta(q) \to 0$ as $q \downarrow 1$.

Now for any $\delta > 0$ set

$$\tilde{\mathscr{F}}(\delta) = \{(f,g) \in \mathscr{F} \times \mathscr{F} \colon\ e_0(f,g) \leqq \delta\} \tag{3.53}$$

and for any real valued function T on $\mathscr{F}$ set

$$\|T\|_{\tilde{\mathscr{F}}(\delta)} = \sup\{|T(f) - T(g)| \colon\ e_0(f,g) \leqq \delta\} . \tag{3.54}$$

Recalling (F.v), we now see that

$$\Delta_{k,2}(q) \leqq \max_{n_k \leqq n \leqq n_{k+1}} \left\| \sum_{i=1}^{n} W_i(f, h_{n_k}) \right\|_{\tilde{\mathscr{F}}(\delta(q))} \Big/ \sqrt{n_k b_{n_k}} \tag{3.55}$$

$$\stackrel{d}{=} \max_{n_k \leqq n \leqq n_{k+1}} \left\| \sum_{i=1}^{n} W_i \right\|_{\tilde{\mathscr{F}}(\delta(q))} \Big/ \sqrt{n_k} \,.$$

Using Lévy's inequality and the exponential inequality for P_0-Brownian motions given in (3.35), we find that with probability one

$$\limsup_{k \to \infty} \Delta_{k,2}(q)/\sqrt{2Lk} \leqq \sqrt{q}\delta(q) \,. \tag{3.56}$$

Finally noting that

$$\Delta_{k,3}(q) \stackrel{d}{=} \max_{1 \leqq n \leqq n_{k+1}-n_k} \left\| \sum_{i=1}^{n} W_i \right\|_{\mathscr{F}} \Big/ \sqrt{n_k} \,,$$

we obtain by a similar argument that with probability one

$$\limsup_{k \to \infty} \Delta_{k,3}(q)/\sqrt{2Lk} \leqq \sqrt{q-1} \,. \tag{3.57}$$

Combining (3.51), (3.56) and (3.57) we obtain (3.50) as soon as we have proved the following lemma.

Lemma 3.10 *Let $\mathscr{F}$ be a class of functions satisfying*
(i) *support* $(f) \subset J$, $f \in \mathscr{F}$
(ii) $|f| \leqq K$, $f \in \mathscr{F}$ *for some* $K > 0$
(iii) $\mathscr{F}$ *is totally bounded for* ρ_0.
If, in addition, (F.ix) *and* (F.x) *hold for a given sequence of bimeasurable invertible transformations* $h_n : \mathbb{R}^d \to \mathbb{R}^d$, *then*

$$\limsup_{k \to \infty} \sup_{f \in \mathscr{F}} \max_{n_k \leqq n \leqq n_{k+1}} e_0(f_{n,k}, f) \leqq \delta(q) \,, \tag{3.58}$$

with $f_{n,k} = f \circ h_n^{-1} \circ h_{n_k}$, *where* $\delta(q) \to 0$ *as* $q \downarrow 1$.

Proof. First note that (ii) and (iii) imply that $\mathscr{F}$ is also totally bounded for e_0. Thus for any given $\varepsilon > 0$ we can find $\mathscr{F}(\varepsilon) = \{f_1, \ldots, f_m\} \in \mathscr{F}$ such that for all $f \in \mathscr{F}$

$$\min_{1 \leqq i \leqq m} e_0(f, f_i) < \varepsilon \,. \tag{3.59}$$

Moreover by Lusin's theorem each function $f_i \in \mathscr{F}(\varepsilon)$ can be approximated by a continuous g_i bounded by K such that

$$e_0(f_i, g_i) < \varepsilon, \quad 1 \leqq i \leqq m \,. \tag{3.60}$$

Denote this class of functions by $G(\varepsilon) = \{g_1, \ldots, g_m\}$.

Next we choose a compact set $A \subset \mathbb{R}^d$ such that $P_0(A^C) \leqq \varepsilon/K^2$. By uniform continuity of each of the functions g_i on A we can select a $\delta > 0$ such that $|g_i(x) - g_i(y)| < \varepsilon$ whenever $x \in A$ and $|x - y| \leqq \delta$.

Now choose q_0 as in (F.x) such that for all $1 < q < q_0$, (1.31) holds. Hence for all large k and $n_k \leqq n \leqq n_{k+1}$ for $1 \leqq i \leqq m$

$$\begin{aligned} & e_0^2(g_i \circ h_n^{-1} \circ h_{n_k}, g_i) \\ & \quad \leqq \varepsilon^2 + \int_{\mathbb{R}^d} g_i^2(x) 1_{A^C}(x)\, dP_0(x) + \int_{\mathbb{R}^d} g_i^2(h_n^{-1}(h_{n_k}(x)) 1_{A^C}(x)\, dP_0(x) \\ & \quad \leqq \varepsilon^2 + 2\varepsilon\,. \end{aligned}$$

Therefore for all large k and $n_k \leqq n \leqq n_{k+1}$, whenever $1 < q < q_0$, we have uniformly in $f \in \mathscr{F}$, using (F.ix) in combination with (3.59), that

$$\begin{aligned} e_0(f, f_{k,n}) & \leqq e_0(f, g_i) + e_0(f_{k,n}, g_i \circ h_n^{-1} \circ h_{n_k}) \\ & \quad + e_0(g_i, g_i \circ h_n^{-1} \circ h_{n_k}) \leqq 2\varepsilon + M\varepsilon + \sqrt{\varepsilon^2 + 2\varepsilon}\,, \end{aligned}$$

where g_i is selected so that $e_0(f, g_i) < 2\varepsilon$. Since $\varepsilon > 0$ can be chosen arbitrarily close to zero, this completes the proof. □

Acknowledgements. The authors would like to thank the University of Bielefeld, University of Düsseldorf and the University of Paris VI for their hospitality, and, in particular, Professors Götze, Janssen and Deheuvels. This paper was written while the authors were visiting these universities in the spring of 1994. They are also indebted to a referee for a number of good suggestions for additional statistical applications of their results.

References

Alexander, K.S.: Central limit theorems for stochastic processes under random entropy conditions. Probab. Theory Relat. Fields **75**, 351–378 (1987)

Arcones, M.: Necessary and sufficient conditions for the law of the iterated logarithm of local processes. preprint (1994)

Carmona, R., Kôno, N.: Convergence en loi et lois du logarithme itéré pour les vecteurs gaussians. Z. Wahrscheinlichkeitstheorie verw. Gebiete **36**, 241–267 (1976)

Deheuvels, P., Mason, D.M.: Nonstandard functional laws of the iterated logarithm for tail empirical and quantile processes. Ann. Probab. **18**, 1693–1722 (1990)

Deheuvels, P., Mason, D.M.: Functional Laws of the iterated logarithm for local empirical processes indexed by sets. Ann. Probab. **22**, 1619–1661 (1994)

Deheuvels, P., Mason, D.M.: Nonstandard local empirical processes indexed by sets. J. Statist. Plann. Inference **45**, 91–112 (1995)

Dudley, R.M.: Distances of probability measures and random variables. Ann. Math. Statist. **39**, 1563–1572 (1968)

Dudley, R.M.: Central limit theorems for empirical measures. Ann. Probab. **6**, 899–829 (1978)

Dudley, R.M.: A course on empirical processes. École d' Éte de Probabilités de Saint-Flour XII-1982 (Lect. Notes in Math. Vol. **1097**, pp. 2–142) Berlin: Springer Verlag 1984

Dudley. R.M., Philipp, W.: Almost sure invariance principles for sums of Banach space valued random elements and empirical processes. Z. Wahrscheinlichkeitstheorie verw. Gebiete **62**, 509–552 (1983)

Giné, E., Zinn, J.: Some limit theorems for empirical processes. Ann. Probab. **12**, 929–989 (1984)

Groeneboom, P., Wellner, J.A.: Information bounds and nonparametric maximum likelihood estimation. Basel: Birkäuser Verlag, 1992

Ibragimov, I.A., Rozanov, Y.A.: Gaussian random processes. New York, Springer Verlag 1978
Kim, J., Pollard, D.: Cube root asymptotics. Ann. Statist. **18**, 191–219 (1990)
Klaassen, C.A.J., Wellner, J.A.: Kac empirical processes and the bootstrap. In: Probability on Banach spaces, 8, eds. R.M. Dudley, M.G. Hahn, J. Kuelbs, 411–429, Boston: Birkäuser Verlag 1992
Ledoux, M., Talagrand, M.: Probability in Banach Spaces. New York: Springer Verlag 1991
Major. P.: Approximation of partial sums of *i.i.d.r.v.s* when the summands have only two moments. Z. Wahrscheinlichkeitstheorie verw. Gebiete **35**, 221–229 (1976)
Mason, D.M.: A strong invariance theorem for the tail empirical process. Ann. Inst. Henri Poincaré Probab. Statist. **24**, 491–506 (1988)
Nolan, D., Marron, J.S.: Uniform consistency of automatic and location-adaptive delta-sequence estimators. Probab. Theory Relat. Fields **80**, 619–632 (1989)
Pollard, D.: A central limit theorem for empirical processes. J. Austral. Math. Soc. Ser. A **33**, 235–248 (1982)
Pollard, D.: Convergence of stochastic processes. New York: Springer Verlag 1984
Pollard, D.: Uniform ratio limit theorems for empirical process. Scand. J. Statist. **22**, 271–278 (1995)
Pyke, R.: The weak convergence of the empirical with random sample size. Proc. Cambridge Phil. Soc. **64**, 155–160 (1968)
Sheehy, A., Wellner, J.A.: Uniform Donsker classes of functions. Ann. Probab. **20**, 1983–2030 (1992)
Shorack, G., Wellner, J.A.: Empirical processes with applications to statistics. New York: Wiley, 1986
Stute, W.: Conditional empirical processes. Ann. Statist. **14**, 638–647 (1986a)
Stute, W.: On the almost sure convergence of conditional empirical distribution functions. Ann. Probab. **14**, 891–901 (1986b)
Talagrand, M.: Sharper bounds for Gaussian and empirical processes. Ann Probab. **22**, 28–76 (1994)
Zaitsev, A.Yu.: On the Gaussian approximation of convolutions under multidimensional analogues of S.N. Bernstein's inequality conditions. Probab. Theory Relat. Fields. **74**, 534–566 (1987)