PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. IV, Fasc. 1 (1984), p. 1-12

# LIMIT THEOREMS FOR EMPIRICAL PROCESSES INDEXED BY CLASSES OF SETS ALLOWING A FINITE-DIMENSIONAL PARAMETRIZATION

BY

PETER GAENSSLER (MUNICH)

*Abstract.* Let $\xi_1, \xi_2, \ldots$ be independent identically distributed random variables defined on some probability space $(\Omega, \mathscr{A}, P)$ and taking their values in a measurable space $(X, \mathscr{B})$ according to the probability distribution $\mu$ on $\mathscr{B}$ defined by $\mu(B) := P(\xi_1 \in B)$, $B \in \mathscr{B}$. Let

$$\mu_n := n^{-1}(\varepsilon_{\xi_1} + \ldots + \varepsilon_{\xi_n})$$

be the empirical measure on $\mathscr{B}$ based on $\xi_1, \ldots, \xi_n$ and, given a class $\mathscr{C} \subset \mathscr{B}$, let

$$\beta_n(C) := n^{1/2}(\mu_n(C) - \mu(C)), \quad C \in \mathscr{B},$$

be the empirical $\mathscr{C}$-process, considered as a stochastic process indexed by $\mathscr{C}$. Various properties of $\beta_n$ as $n \to \infty$ are studied.

Dudley [6] gave general conditions for the convergence in law of $\beta_n = (\beta_n(C))_{C \in \mathscr{C}}$ to a certain Gaussian process indexed by $\mathscr{C}$ (in which case $\mathscr{C}$ is called a *$\mu$-Donsker class*). It is the purpose of the present paper to give some further illustrations of some of the main results of Dudley.

Along this it is shown that Dudley's sufficient condition (for $\mathscr{C}$ being a $\mu$-Donsker class) based on metric entropy with inclusion applies to classes $\mathscr{C}$ allowing, in a certain sense, a finite-dimensional parametrization (Section 2).

If $(X, \mathscr{B})$ is the Euclidean space $\boldsymbol{R}^k$, $k \geqslant 1$, with its Borel $\sigma$-algebra $\mathscr{B}(\boldsymbol{R}^k)$, two specific examples considered in Section 3 are the class $\mathscr{C} = \mathscr{K}(\boldsymbol{R}^k)$ consisting of all closed Euclidean balls and the class $\mathscr{C} = \{(-\infty, \boldsymbol{t}]: \boldsymbol{t} \in \boldsymbol{R}^k\}$. Both are $\mu$-Donsker classes for any probability measure $\mu$ on $\mathscr{B}(\boldsymbol{R}^k)$; even more, it is shown that $\mathscr{K}(\boldsymbol{R}^k)$ is a Vapnik-Chervonenkis class and also $\mu\in$-Suslin in the sense of Dudley [6] and, therefore, $\mathscr{K}(\boldsymbol{R}^k)$ is a Strassen log log class in the sense of Kuelbs and Dudley [12].

We hope that our presentation may contribute to make further propaganda for Dudley's central limit theory for empirical measures.

**1. Introduction and preliminaries.** Let $\xi_1, \xi_2, \dots$ be independent identically distributed random variables defined on some probability space $(\Omega, \mathscr{A}, P)$ and taking their values in a measurable space $(X, \mathscr{B})$ according to the probability distribution or law $\mu$ on $\mathscr{B}$ defined by $\mu(B) := P(\xi_1 \in B)$, $B \in \mathscr{B}$. Throughout it will be tacitly assumed that $(\Omega, \mathscr{A}, P)$ is the countable product $(X^N, \mathscr{B}^N, \mu^N)$ of copies of $(X, \mathscr{B}, \mu)$ and that the $\xi_i$'s are the coordinate projections (canonical model).

Let $\mu_n := n^{-1}(\varepsilon_{\xi_1} + \dots + \varepsilon_{\xi_n})$ be the empirical measure on $\mathscr{B}$ based on the first $n$ observations, where $\varepsilon_\xi(B) := 1_B \circ \xi$, $B \in \mathscr{B}$.

Given a class $\mathscr{C} \subset \mathscr{B}$, let $\beta_n(C) := n^{1/2}(\mu_n(C) - \mu(C))$, $C \in \mathscr{C}$, be the *empirical $\mathscr{C}$-process*, considered as a stochastic process indexed by $\mathscr{C}$.

From the central limit theorem in finite-dimensional vector spaces we know that, at least when restricted to a finite subclass of $\mathscr{C}$, $\beta_n$ converges in law to a Gaussian process $G_\mu = (G_\mu(C))_{C \in \mathscr{C}}$, where

$$E(G_\mu(C)) = 0 \quad \text{for all } C \in \mathscr{C}$$

and

$$\operatorname{cov}(G_\mu(C_1), G_\mu(C_2)) = \mu(C_1 \cap C_2) - \mu(C_1)\mu(C_2), \quad C_1, C_2 \in \mathscr{C}.$$

In the classical case where $(X, \mathscr{B})$ is the unit interval $[0, 1]$ with its Borel $\sigma$-algebra $\mathscr{B}([0, 1])$ and $\mu$ is the Lebesgue measure on $[0, 1]$, the empirical $\mathscr{C}$-process for $\mathscr{C} := \{[0, t] : t \in [0, 1]\}$ coincides with the so-called *uniform empirical process* $\alpha_n(t) := n^{1/2}(F_n(t) - t)$, $t \in [0, 1]$, $F_n$ being the empirical distribution function based on $\xi_1, \dots, \xi_n$. In this case the corresponding limiting process $G_\mu$ is the Brownian bridge $B^\circ = (B^\circ(t))_{t \in [0,1]}$ with paths in the space $C[0, 1]$ of all continuous functions on $[0, 1]$. Furthermore, *Donsker's invariance principle* states that not only convergence in law of the finite-dimensional distributions of $\alpha_n$ to those of $B^\circ$ takes place but — even more — that $\alpha_n$, considered as a sequence of random elements in the space $D[0, 1]$ of all right continuous functions on $[0, 1]$ with left-hand limits, converges in law to $B^\circ$.

Here we follow Dudley's definition (see [3] and [4]) of weak convergence (convergence in law for random elements) in non-separable metric spaces $S = (S, \varrho)$:

(1.1) Given a sequence of random elements $\eta_n$ in $S$, $n = 0, 1, 2, \dots$, i.e. $\mathscr{A}_n$, $\mathscr{B}_b$-measurable mappings from some basic probability spaces $(\Omega_n, \mathscr{A}_n, P_n)$ into $S$ with laws $\mathscr{L}\{\eta_n\}$ defined on $\mathscr{B}_b$, where $\mathscr{B}_b$ is the $\sigma$-algebra generated by the open $\varrho$-balls in $S$, $\eta_n$ is said to *converge in law* to $\eta_0$ ($\eta_n \xrightarrow{\mathscr{L}} \eta_0$) if

(i) $\mathscr{L}\{\eta_0\}$ concentrates on a separable subspace $S_0$ of $S$;

(ii) $\mathrm{E}(f\circ\eta_n)\to\mathrm{E}(f\circ\eta_0)$ for every continuous, bounded, and $\mathscr{B}_b$-measurable function $f\colon S\to \boldsymbol{R}$.

In the before-mentioned invariance principle of Donsker, $\eta_n=\alpha_n$ and $\eta_0 = B^\circ$ may be viewed as random elements in $S=D[0, 1]$ endowed with the supremum metric $\varrho$, and convergence in law of $\alpha_n$ to $B^\circ$ is to be understood as $\alpha_n \xrightarrow{\mathscr{L}} B^\circ$ in the sense of (1.1), $S_0$ being the separable subspace $C[0, 1]$ of $(D[0, 1], \varrho)$ (cf. [11], Section 2.1).

Going back to the general empirical $\mathscr{C}$-process $\beta_n$, in order to obtain a *functional limit theorem* (*invariance principle*) like $\beta_n \xrightarrow{\mathscr{L}} G_\mu$ we have to look first for appropriate spaces $S$ and $S_0$ serving as sample spaces for $\beta_n$ and $G_\mu$. For this Dudley [6] introduced the spaces $C^b(\mathscr{C}, \mu)$, respectively $S_0 = \mathscr{U}C^b(\mathscr{C}, \mu)$ and $S=D_0(\mathscr{C}, \mu)$, defined as follows:

$$C^b(\mathscr{C}, \mu) := \{f\colon \mathscr{C}\to \boldsymbol{R},\ f \text{ bounded and continuous with respect to the } L_2(\mu)\text{-norm on } \mathscr{C}\},$$

where f is called *continuous* with respect to the $L_2(\mu)$-norm on $\mathscr{C}$ if, for any sequence $(C_n)_{n=0,1,\dots}$ in $\mathscr{C}$, $f(C_n)\to f(C_0)$ provided that

$$\int_X |1_{C_n}-1_{C_0}|^2\, d\mu = \mu(C_n \triangle C_0)\to 0 \quad \text{as } n\to\infty;$$

$$\mathscr{U}C^b(\mathscr{C}, \mu) := \{f\colon \mathscr{C}\to \boldsymbol{R},\ f \text{ bounded and uniformly continuous with respect to the } L_2(\mu)\text{-norm on } \mathscr{C}\}.$$

Note that in the classical case of the uniform empirical process we have $\mathscr{U}C^b(\mathscr{C}, \mu) = C[0, 1]$ whereas $\mathscr{U}C^b(\mathscr{C}, \mu)\neq C[0, 1]$ in general, i.e. if $\mu$ is not the Lebesgue measure. Indeed, consider, for example, $\mu := \varepsilon_{t_0}$ for some $t_0\in[0, 1]$ and define

$$f(C) := \begin{cases} 0 & \text{for } C=[0, t],\ t<t_0,\\ 1 & \text{for } C=[0, t],\ t\geqslant t_0;\end{cases}$$

then $f\in \mathscr{U}C^b(\mathscr{C}, \mu)$ but $f\notin C[0, 1]$.

The other space, i.e. $D_0(\mathscr{C}, \mu)$, is now defined as the linear space generated by $C^b(\mathscr{C}, \mu)$ and all functions $C\to\varepsilon_x(C)$, $x\in X$.

Note that the empirical $\mathscr{C}$-process $\beta_n$, $n\in N$, takes its values in $D_0(\mathscr{C}, \mu)$.

Let $\mathscr{B}_b = \mathscr{B}_b(D_0(\mathscr{C}, \mu), \varrho)$ be the $\sigma$-algebra in $D_0(\mathscr{C}, \mu)$ generated by the open $\varrho$-balls, where $\varrho$ is the supremum metric in $D_0(\mathscr{C}, \mu)$, i.e.

$$\varrho(f, g) := \sup_{C\in\mathscr{C}} |f(C)-g(C)|, \qquad f, g\in D_0(\mathscr{C}, \mu).$$

Then, in order to look at $\beta_n$ as random elements in $S=D_0(\mathscr{C}, \mu)$ it is necessary to find conditions on $\mathscr{C}$ and $\mu$ such that

$(\mathrm{M}_0)$ $\beta_n: \Omega \to D_0(\mathscr{C}, \mu)$ is measurable from the measure-theoretic completion of $(\Omega, \mathscr{A}, P)$ to $(D_0(\mathscr{C}, \mu), \mathscr{B}_b)$ for each $n \in N$.

If $(\mathrm{M}_0)$ holds, then $\mathscr{C}$ is called *empirically measurable for* $\mu$ ($\mathscr{C}$ is $\mu$-EM).

As remarked by Dudley [6], a countable class $\mathscr{C} \subset \mathscr{B}$ is always $\mu$-EM for any $\mu$; more generally, if $\mathscr{C}$ has a countable subclass $\mathscr{D}$ such that for all $C \in \mathscr{C}$ there are $D_n \in \mathscr{D}$ with $1_{D_n}(x) \to 1_C(x)$ for all $x \in X$, then $\mathscr{C}$ is $\mu$-EM for any $\mu$. As an example, in $X = R^k$, $k \geqslant 1$, we have

(1.2) the class $\mathscr{C} = \mathscr{K}(R^k)$ of all closed Euclidean balls or the class $\mathscr{C} = \mathscr{I}(R^k)$ of all half-intervals $(-\infty, t]$, $t \in R^k$, is $\mu$-EM for any probability measure $\mu$ on the Borel $\sigma$-algebra $\mathscr{B}(R^k)$ in $R^k$.

Besides $(\mathrm{M}_0)$ Dudley [6] introduced a second much stronger ($\mu\in$-Suslin) measurability condition playing an essential role for $\mu$-Donsker classes $\mathscr{C}$ to be Strassen log log classes for $\mu$ in the sense of Kuelbs and Dudley [12] (see (3.3) below). Let us recall here the corresponding basic definitions.

A measurable space $(Y, \mathscr{F})$ is called a *Suslin-measurable space* if there is a metric on $Y$ for which $\mathscr{F}$ is the $\sigma$-algebra of Borel sets and such that there is a continuous function from a Polish space onto $Y$.

Given a probability space $(X, \mathscr{B}, \mu)$, a class $\mathscr{C} \subset \mathscr{B}$, and a $\sigma$-algebra $\mathscr{S}$ of subset of $\mathscr{C}$, we say that $(X, \mathscr{B}; \mathscr{C}, \mathscr{S})$ is $\mu\in$-*Suslin* if the following conditions hold:

(i) $(X, \mathscr{B})$ and $(\mathscr{C}, \mathscr{S})$ are both Suslin-measurable spaces;

(ii) the $\in$-relation, $\{\langle x, C\rangle : x \in C\}$, is a measurable subset of $X \times \mathscr{C}$ for the product $\sigma$-algebra of $\mathscr{B}$ and $\mathscr{S}$ generated by the rectangles $B \times S$, $B \in \mathscr{B}$, $S \in \mathscr{S}$;

(iii) for the pseudometric $d_\mu$ on $\mathscr{C}$ defined by

$$d_\mu(C_1, C_2) := \mu(C_1 \triangle C_2), \quad C_1, C_2 \in \mathscr{C},$$

all $d_\mu$-open sets belong to $\mathscr{S}$.

If the condition

$(\mathrm{M}_1)$ $(X, \mathscr{B}; \mathscr{C}, \mathscr{S})$ is $\mu\in$-Suslin for some $\mathscr{S}$

holds, we say that $(\mathrm{M}_1)$ *is satisfied for* $\mathscr{C}$ *and* $\mu$.

It is shown in [6] that $(\mathrm{M}_1)$ implies $(\mathrm{M}_0)$. Furthermore, with a similar proof as that of Propositions (4.5) and (4.3) in [6] we obtain

(1.3) *Let* $X = (X, e)$ *be a locally compact, separable metric space and* $\mathscr{G}$ *a collection of continuous real functions on* $X$. *Suppose that* $\mathscr{E}$ *is a* $\sigma$-*algebra of subsets of* $\mathscr{G}$ *such that, for each* $x \in X$, $g \to g(x)$ *is* $\mathscr{E}$-*measurable. For* $g \in \mathscr{G}$ *let*

$$\mathrm{pos}_0(g) := \{x \in X : g(x) \geqslant 0\},$$

*whence*

$$\mathrm{pos}_0(\mathscr{G}) := \{\mathrm{pos}_0(g) : g \in \mathscr{G}\} \subset \mathscr{F},$$

*where $\mathscr{F}$ is the collection of all closed subsets of X. Let $\mathscr{B}(\mathscr{F})$ be the Effros-Borel structure on $\mathscr{F}$, i.e. $\mathscr{B}(\mathscr{F})$ is the $\sigma$-algebra of subsets of $\mathscr{F}$ generated by* all sets of the form $\{F \in \mathscr{F} : F \subset H\}$, $H \in \mathscr{F}$. *Then the map* $g \to \mathrm{pos}_0(g)$ *is* $\mathscr{E}$, *$\mathscr{B}(\mathscr{F})$-measurable. Furthermore, if $(\mathscr{G}, \mathscr{E})$ is a Suslin-measurable space, then* $(\mathrm{M}_1)$ *is satisfied for* $\mathrm{pos}_0(\mathscr{G})$ *and any probability measure $\mu$ on the Borel $\sigma$-algebra $\mathscr{B}(X)$ in X.*

If $(\mathrm{M}_0)$ holds for $\mathscr{C}$ and $\mu$, if, in addition, $G_\mu = (G_\mu(C))_{C \in \mathscr{C}}$ has a version taking its values in the space $S_0 = \mathscr{U}C^b(C, \mu)$, and if $\mathscr{U}C^b(\mathscr{C}, \mu)$ is separable with respect to the supremum metric $\varrho$, then $G_\mu$ as well as $\beta_n$ may be regarded as a random element in $D_0(\mathscr{C}, \mu)$ and (1.1) makes sense for $\eta_n = \beta_n$ and $\eta_0 = G_\mu$, in which case, i.e. if $\beta_n \xrightarrow{\mathscr{L}} G_\mu$, we say that $\mathscr{C}$ is a *$\mu$-Donsker class.*

**2. Empirical processes indexed by classes of sets allowing a finite-dimensional parametrization.** Let again $(X, \mathscr{B}, \mu)$ be a probability space and $\mathscr{C} \subset \mathscr{B}$. For each $\varepsilon > 0$ let $N_I(\varepsilon, \mathscr{C}, \mu)$ be the smallest $n$ such that for some $A_1, \ldots, A_n \in \mathscr{B}$ (not necessarily in $\mathscr{C}$) and for every $C \in \mathscr{C}$ there exist $i, j$ with $A_i \subset C \subset A_j$ and $\mu(A_j \setminus A_i) < \varepsilon$. On the other hand, let $N(\varepsilon, \mathscr{C}, \mu)$ be the smallest $n$ such that

$$\mathscr{C} = \bigcup_{j=1}^{n} \mathscr{C}_j$$

for some classes $\mathscr{C}_j$ with $\sup\{d_\mu(A, B) : A, B \in \mathscr{C}_j\} \leqslant 2\varepsilon$ for each $j$. Then $\log N(\varepsilon, \mathscr{C}, \mu)$ is called a *metric entropy* and $\log N_I(\varepsilon, \mathscr{C}, \mu)$ is called a *metric entropy with inclusion.* It is easy to see that

(2.1) $N(\varepsilon, \mathscr{C}, \mu) < \infty$ *for each* $\varepsilon > 0$ *iff* $(\mathscr{C}, d_\mu)$ *is totally bounded, in which case* $(\mathscr{U}C^b(\mathscr{C}, \mu), \varrho)$ *is separable. Furthermore,* $N(\varepsilon, \mathscr{C}, \mu) \leqslant N_I(\varepsilon, \mathscr{C}, \mu)$ *for each* $\varepsilon > 0$, *any* $\mathscr{C} \subset \mathscr{B}$, *and any* $\mu$.

As shown in [4] and [5], p. 71, $G_\mu = (G_\mu(C))_{C \in \mathscr{C}}$ has a version taking its values in $\mathscr{U}C^b(\mathscr{C}, \mu)$ if

$$(\mathrm{E}_0) \qquad \int_0^1 (\log N(x^2, \mathscr{C}, \mu))^{1/2} dx < \infty,$$

in which case $(\mathscr{U}C^b(\mathscr{C}, \mu), \varrho)$ is also a separable space (cf. (2.1)).

Replacing $(\mathrm{E}_0)$ by the stronger entropy condition

$$(\mathrm{E}_1) \qquad \int_0^1 (\log N_I(x^2, \mathscr{C}, \mu))^{1/2} dx < \infty,$$

Dudley ([6], Theorem 5.1) obtained one of his main results:

(2.2) *If* $(\mathrm{M}_0)$ *and* $(\mathrm{E}_1)$ *are fulfilled for a class* $\mathscr{C} \subset \mathscr{B}$ *and a probability measure* $\mu$ *on* $\mathscr{B}$, *then* $\mathscr{C}$ *is a* $\mu$*-Donsker class.*

To illustrate the broad scope of applicability of (2.2) we will first prove an invariance principle for empirical $\mathscr{C}$-processes indexed by classes $\mathscr{C}$ allowing a finite-dimensional parametrization in the sense of the following theorem:

(2.3) THEOREM. *Let $X$ be a locally compact, separable metric space, $\mathscr{B} = \mathscr{B}(X)$ the $\sigma$-algebra of Borel sets in $X$, and let $K$ be a compact subset of $\mathbf{R}^l$, $l \geq 1$. Suppose that $f: X \times K \to \mathbf{R}$ is a function satisfying the following three conditions (with respect to a given probability measure $\mu$ on $\mathscr{B}$):*

*(i) $f_z := f(\cdot, z): X \to \mathbf{R}$ is continuous for each $z \in K$;*

*(ii) $f_\cdot(x): K \to \mathbf{R}$ is "uniformly Lipschitz", i.e.*

$$M := \sup_{x \in X} \sup \{|f_z(x) - f_{z'}(x)| \, |z-z'|^{-1}: z \neq z', \; z, z' \in K\} < \infty,$$

*where $|z-z'|$ denotes the Euclidean distance between $z$ and $z'$;*

*(iii) $\mu(\{f_z \in [-\varepsilon, \varepsilon)\}) = \mathcal{O}(\varepsilon)$ uniformly in $z \in K$.*

*Let $\mathscr{C} \subset \mathscr{B}$ be defined by $\mathscr{C} := \{\{f_z \geq 0\}: z \in K\}$.*

*Then $\mathscr{C}$ is a $\mu$-Donsker class; furthermore, $(\mathrm{M}_1)$ is satisfied for $\mathscr{C}$ and $\mu$.*

Proof. Without loss of generality we may and do assume that $K = [0, 1]^l$. For each $m \in \mathbf{N}$ let

$$I_m := \{t \in [0, 1]: t = i/2^m, \; i = 0, 1, \ldots, 2^m\}.$$

We will first show that $(\mathrm{E}_1)$ holds true. For this let $0 < \varepsilon < 1$ be given and choose $m = m(\varepsilon)$ so that

$$(+) \qquad 2^{-m} \leq \varepsilon/M < 2^{-m+1},$$

where $M$ $(0 < M < \infty)$ is the constant given by (ii). Now, for each $z \in K = [0, 1]^l$ there exists $z_{i_0} \in I_m^l$ such that $|z - z_{i_0}| \leq 2^{-m} \leq \varepsilon/M$, whence according to (ii) we have

$$-\varepsilon \leq f_z(x) - f_{z_{i_0}}(x) \leq \varepsilon \quad \text{for all } x \in X,$$

which implies

$$\underline{A}_{z_{i_0}} := \{f_{z_{i_0}} \geq \varepsilon\} \subset \{f_z \geq 0\} \subset \{f_{z_{i_0}} \geq -\varepsilon\} =: \bar{A}_{z_{i_0}}.$$

Furthermore, $\bar{A}_{z_{i_0}} \setminus \underline{A}_{z_{i_0}} = \{f_{z_{i_0}} \in [-\varepsilon, \varepsilon)\}$, and so by (iii) we get

$$\mu(\bar{A}_{z_{i_0}} \setminus \underline{A}_{z_{i_0}}) < C_1 \varepsilon$$

for some constant $C_1$ independent of $z$ and $z_{i_0}$. It follows that

$$\mathscr{A}_m := \{\underline{A}_{z_i}, \bar{A}_{z_i}, z_i \in I_m^l\} \subset \mathscr{B}(X)$$

(according to (i)) and that for each $C \in \mathscr{C}$ there exist sets $A_i, A_j \in \mathscr{A}_m$ with $A_i \subset C \subset A_j$ and $\mu(A_j \setminus A_i) < C_1 \varepsilon$. Since $\mathrm{card}(\mathscr{A}_m) \leq 2 \cdot 2^{ml}$ and since $2^m < 2M\varepsilon^{-1}$ by $(+)$, we obtain $N_I(\varepsilon, \mathscr{C}, \mu) \leq C_2 \varepsilon^{-l}$ for some constant $C_2$, which implies $(\mathrm{E}_1)$.

It remains to show that $(M_1)$ is satisfied for $\mathscr{C}$ and $\mu$ (which implies $(M_0)$ yielding together with $(E_1)$ the assertion that $\mathscr{C}$ is a $\mu$-Donsker class according to (2.2)). But for $\mathscr{G} := \{f_z\colon z \in K\}$ and $\mathscr{E} := \mathscr{G} \cap \mathscr{B}(C(X))$, where $\mathscr{B}(C(X))$ is the Borel $\sigma$-algebra in the space $C(X)$ of all continuous functions on $X$ endowed with the metric $d$ defined by

$$d(f, g) := \sum_{n \in N} 2^{-n} \min\{1, \sup_{x \in K_n} |f(x) - g(x)|\},$$

$(K_n)_{n \in N}$ being a monotone increasing sequence of compact sets in $X$ with $\bigcup_{n \in N} \operatorname{int}(K_n) = X$ (note that every locally compact, separable metric space is a $\sigma$-compact space), the assumptions of (1.3) are fulfilled. Since in the present situation $(\mathscr{G}, \mathscr{E})$ is also a Suslin-measurable space (as the continuous image of $K$), the assertion follows from (1.3).

(2.4) Examples. (a) Let $(X, \mathscr{B}, \mu) := ([0, 1]^k, \mathscr{B}([0, 1]^k), \lambda_k)$, $k \geqslant 1$, $\lambda_k$ being the $k$-dimensional Lebesgue measure, and let $\mathscr{C} \subset \mathscr{B}$ be the class of all closed Euclidean balls in $X = [0, 1]^k$. Then $\mathscr{C}$ is a $\lambda_k$-Donsker class and $(M_1)$ is satisfied for $\mathscr{C}$ and $\lambda_k$.

In fact, take

$$K := \{z = (y, r)\colon y \in [0, 1]^k, r \in [0, r_y], r_y := \sup\{u\colon B(y, u) \subset [0, 1]^k\}\},$$

where $B(y, u) := \{x \in \mathbb{R}^k\colon e(x, y) \leqslant u\}$ ($e$ denoting the Euclidean distance in $[0, 1]^k$), and define $f\colon [0, 1]^k \times K \to \mathbb{R}$ by

(2.5) $f(x, z) := e(x, \complement(\operatorname{int} B(y, r))) - e(x, B(y, r)) \quad (z = (y, r) \in K)$, i.e. $f(x, z) = r - e(x, y)$, where $\complement(\operatorname{int}(B(y, r)))$ denotes the complement of the interior of $B(y, r)$.

Then $\{\{f_z \geqslant 0\}\colon z \in K\}$ is the class of all closed Euclidean balls in $X = [0, 1]^k$ and it is easy to verify conditions (i)-(iii) of Theorem (2.3), giving the result.

(b) As a further example one may consider the same probability space $(X, \mathscr{B}, \mu)$ as in (a) and $\mathscr{C} = \{[0, \boldsymbol{t}]\colon \boldsymbol{t} \in [0, 1]^k\}$ to get a $\lambda_k$-Donsker class satisfying $(M_1)$ (cf. [13], [16], and [1]).

(c) One can show that we can apply Theorem (2.3) also in the case $\mu = \lambda_k$ and

$$\mathscr{C} = \{(C + z) \cap [0, 1]^k\colon z \in [0, 1]^k\},$$

$C$ being a fixed closed and convex subset of $X = [0, 1]^k$, $k \geqslant 1$ (cf. [14]), taking as in (2.5)

$$f(x, z) := e(x, \complement(\operatorname{int} C_z)) - e(x, C_z), \quad x, z \in [0, 1]^k,$$

with $C_z := C + z$.

In fact, it is easy to show that both conditions (i) and (ii) of Theorem (2.3) are fulfilled assuming only that $C$ is closed. As to (iii), in the present situation we have to note first that $\{f_z \in [-\varepsilon, \varepsilon)\} \subset C_z^\varepsilon \setminus_\varepsilon C_z$ for all $z \in [0, 1]^k$ (where $A^\varepsilon := \{x: e(x, A) \leqslant \varepsilon\}$ and ${}_\varepsilon A := \{x: e(x, \complement A) > \varepsilon\}$) and we can then use the fact that, for the class $\mathscr{C}_k$ of *all* convex Borel sets in $[0, 1]^k$,

$$\sup_{C \in \mathscr{C}_k} \lambda_k(C^\varepsilon \setminus_\varepsilon C) \leqslant c_k \varepsilon \quad \text{for } \varepsilon \downarrow 0$$

with some constant $c_k$ depending only on $k$ (cf. [10]). Consequently, we infer that

$$\sup_{z \in [0,1]^k} \lambda_k(C_z^\varepsilon \setminus_\varepsilon C_z) = \mathcal{O}(\varepsilon),$$

which is equivalent to $\lambda_k(C^\varepsilon \setminus_\varepsilon C) = \mathcal{O}(\varepsilon)$ (due to the translation invariance of $\lambda_k$), whence (iii) is also fulfilled.

This consideration also shows that the set of all translates of a fixed closed (not necessarily convex) set $C$ is a $\lambda_k$-Donsker class provided that $C$ has a smooth boundary in the sense that $\lambda_k(C^\varepsilon \setminus_\varepsilon C) = \mathcal{O}(\varepsilon)$. On the other hand, we know from [8] that the set of *all* closed and convex subsets of $[0, 1]^k$ is *not* a $\lambda_k$-Donsker class for $k \geqslant 3$.

(2.6) Remarks. Theorem (2.3) still applies if in the above-mentioned examples $\lambda_k$ is replaced by any probability measure $\mu$ having a bounded density with respect to $\lambda_k$. On the other hand, the assumption of $K$ being compact is rather restrictive. In the next section we will show that in case of the examples (a) and (b) we can get rid of these restriction, and the whole Euclidean space $\mathbb{R}^k$ is then allowed to be the sample space; cf. also the final remarks (3.9). Moreover, it will follow (cf. (3.3) below) that under the conditions of Theorem (2.3) $\mathscr{C}$ is also a Strassen log log class for $\mu$.

**3. The empirical $\mathscr{C}$-process indexed by some classes $\mathscr{C}$.** Let $(X, \mathscr{B})$ be the Euclidean space $\mathbb{R}^k$, $k \geqslant 1$, with its Borel $\sigma$-algebra $\mathscr{B} = \mathscr{B}(\mathbb{R}^k)$ and let $\mu$ be any probability measure on $\mathscr{B}(\mathbb{R}^k)$. Let $\mathscr{K}(\mathbb{R}^k)$ be the class of all closed Euclidean balls and put

$$\mathscr{I}(\mathbb{R}^k) := \{(-\infty, \boldsymbol{t}]: \boldsymbol{t} \in \mathbb{R}^k\}.$$

Note that both classes are already $\mu$-EM (cf. (1.2)); in the same way as at the end of the proof of Theorem (2.3) it can be shown that even

(3.1) $(M_1)$ *is satisfied for* $\mu$ *and* $\mathscr{C} = \mathscr{K}(\mathbb{R}^k)$, or $\mathscr{C} = \mathscr{I}(\mathbb{R}^k)$, *respectively*.

It is easy to show that $(E_1)$ holds for $\mathscr{C} = \mathscr{I}(\mathbb{R}^k)$ and any $\mu$, whence $\mathscr{I}(\mathbb{R}^k)$ is a $\mu$-Donsker class (proved first by Donsker [2] for $k = 1$ and by Dudley [3] for $k \geqslant 1$). In fact:

(a) For $k = 1$, consider for any $\varepsilon$ $(0 < \varepsilon \leqslant 1)$ the partition

$$-\infty =: t_0 < t_1 < \ldots < t_{m-1} < t_m := \infty$$

of $\boldsymbol{R}$, where $t_{i+1} := \sup\{t \in \boldsymbol{R}: \mu((t_i, t]) \leqslant \varepsilon/2\}$. Since $\mu((t_i, t_{i+1}]) \geqslant \varepsilon/2$ and $\mu(\boldsymbol{R}) = 1$, we have $m-1 \leqslant 2/\varepsilon$.

Then, taking as $A_i$'s in the definition of $N_I(\varepsilon, \mathscr{I}(\boldsymbol{R}), \mu)$ all sets of the form $\emptyset$, $(-\infty, t_1)$, $(-\infty, t_1]$, $(-\infty, t_2)$, $(-\infty, t_2]$, $\ldots$, $(-\infty, t_{m-1})$, $(-\infty, t_{m-1}]$, $\boldsymbol{R}$, we obtain

$$\min\{n: (\exists A_1, \ldots, A_n \in \mathscr{B}(\boldsymbol{R}))(\forall C \in \mathscr{I}(\boldsymbol{R}))(\exists i, j)$$
$$A_i \subset C \subset A_j \text{ and } \mu(A_j \backslash A_i) < \varepsilon\} \leqslant 2(m-1)+2 = 2m \leqslant 4/\varepsilon + 2 \leqslant 6/\varepsilon.$$

Consequently, $\log N_I(\varepsilon^2, \mathscr{I}(\boldsymbol{R}), \mu) \leqslant \log 6/\varepsilon^2$, showing that $(E_1)$ holds for $k = 1$.

(b) For $k > 1$ the result is an immediate consequence of (a) and inequality $(*)$ of the following lemma:

(3.2) Lemma. *Let $(X, \mathscr{B})$ be a measurable space and let $\mu$ be a probability measure on the product $\sigma$-algebra $\bigotimes_1^k \mathscr{B}$ in $X^k$, $k \geqslant 1$, with marginal laws $\pi_i \mu$ on $\mathscr{B}$, $i = 1, \ldots, k$. Let $\mathscr{C}_i \subset \mathscr{B}$, $i = 1, \ldots, k$, be given classes of sets and*

$$\mathscr{C} := \{\mathop{\mathrm{X}}_{i=1}^{k} C_i: C_i \in \mathscr{C}_i, i = 1, \ldots, k\}.$$

*Then*

$$(*) \qquad N_I(\varepsilon, \mathscr{C}, \mu) \leqslant \prod_{i=1}^{k} N_I(\varepsilon/k, \mathscr{C}_i, \pi_i \mu).$$

Proof. We may and do assume that $n_i := N_I(\varepsilon/k, \mathscr{C}_i, \pi_i \mu) < \infty$ for each $i = 1, \ldots, k$. Then there exist $A_{i1}, \ldots, A_{in_i} \in \mathscr{B}$ such that for any $C_i \in \mathscr{C}_i$ there exist $r_i, s_i \in \{1, \ldots, n_i\}$ with $A_{ir_i} \subset C_i \subset A_{is_i}$ and $\pi_i \mu(A_{is_i} \backslash A_{ir_i}) < \varepsilon/k$, $i = 1, \ldots, k$. Consequently, we obtain

$$\mathop{\mathrm{X}}_{i=1}^{k} A_{ir_i} \subset \mathop{\mathrm{X}}_{i=1}^{k} C_i \subset \mathop{\mathrm{X}}_{i=1}^{k} A_{is_i} \quad \text{and} \quad \mu\Big(\mathop{\mathrm{X}}_{i=1}^{k} A_{is_i} \backslash \mathop{\mathrm{X}}_{i=1}^{k} A_{ir_i}\Big) \leqslant \sum_{i=1}^{k} \mu(B_i),$$

where $B_i := X \times \ldots \times X \times (A_{is_i} \backslash A_{ir_i}) \times X \times \ldots \times X$, and thus

$$\sum_{i=1}^{k} \mu(B_i) = \sum_{i=1}^{k} \pi_i \mu(A_{is_i} \backslash A_{ir_i}) < \varepsilon.$$

Since there are at most $n_1 n_2 \ldots n_k$ approximating sets of the from

$$\mathop{\mathrm{X}}_{i=1}^{k} A_{it_i} \in \bigotimes_1^k \mathscr{B},$$

inequality $(*)$ holds true.

Next, one of the main theorems in [12] states that for any probability space $(X, \mathscr{B}, \mu)$ we have the following result:

(3.3) *If* $(M_1)$ *is satisfied for a class* $\mathscr{C} \subset \mathscr{B}$ *and* $\mu$ *and if* $\mathscr{C}$ *is a* $\mu$*-Donsker class, then* $\mathscr{C}$ *is a Strassen* log log *class for* $\mu$, *i.e. with probability one the set*

$$\left\{\left(\frac{\beta_n(C)}{(2\log\log n)^{1/2}}\right)_{C\in\mathscr{C}} : n \geqslant n_0\right\}$$

*is relatively compact (with respect to the supremum metric* $\varrho$ *in* $D_0(\mathscr{C}, \mu)$*) with the limit set*

$$B_{\mathscr{C}} := \{C \to \int_C f d\mu,\ C\in\mathscr{C} : f \in B\},$$

*where*

$$B := \{f \in L^2(X, \mathscr{B}, \mu) : \int f d\mu = 0 \text{ and } \int |f|^2 d\mu \leqslant 1\}.$$

Since we have shown before that for $(X, \mathscr{B}, \mu) = (R^k, \mathscr{B}(R^k), \mu)$ the class $\mathscr{C} = \mathscr{I}(R^k)$ satisfies $(M_1)$ and is $\mu$-Donsker for any $\mu$, we obtain from (3.3) the results of Finkelstein [9] and Richter [15], namely:

(3.4) $\mathscr{I}(R^k)$ *is a Strassen* log log *class for any probability measure* $\mu$ *on* $\mathscr{B}(R^k)$.

That the same holds true for $\mathscr{C} = \mathscr{K}(R^k)$ can be derived from the following result of Dudley ([6], Theorem 7.1) and Kuelbs and Dudley ([12], Corollary 2.4), respectively:

(3.5) (i) *If* $(M_1)$ *is satisfied for a class* $\mathscr{C}$ *and* $\mu$ *and if* $\mathscr{C}$ *is a Vapnik-Chervonenkis class (VCC), then* $\mathscr{C}$ *is a* $\mu$*-Donsker class.*

(ii) *If* $(M_1)$ *is satisfied for* $\mu$ *and a VCC* $\mathscr{C}$, *then* $\mathscr{C}$ *is a Strassen* log log *class for* $\mu$.

In this context, given a probability space $(X, \mathscr{B}, \mu)$, a class $\mathscr{C} \subset \mathscr{B}$ is called a *Vapnik-Chervonenkis class* if there exists a natural number $s$ such that for all $F \subset X$ with card $(F) = s$ there exists a subset $F_0$ of $F$ which *cannot* be represented in the form $F_0 = F \cap C$ for some $C \in \mathscr{C}$, i.e. "$\mathscr{C}$ does not cut *all* subset of $F$".

In view of (3.5) and (3.1) it remains to show that $\mathscr{C} = \mathscr{K}(R^k)$ is a VCC in order to get

(3.6) $\mathscr{K}(R^k)$ *is a* $\mu$*-Donsker and a Strassen* log log *class for any probability measure* $\mu$ *on* $\mathscr{B}(R^k)$.

The following nice proof for $\mathscr{K}(R^k)$ being a VCC was communicated to us by Flemming Topsøe; it is based on the auxiliary results (3.7) and (3.8):

(3.7) Radon's Theorem (cf. Valentine [17], Theorem 1.26). *For any* $F \subset R^k$, $k \geqslant 1$, *with* card $(F) \geqslant k+2$ *there exists a partition of* $F$ *into two disjoint subsets* $F_1$ *and* $F_2$ *such that* $\text{Co}(F_1) \cap \text{Co}(F_2) \neq \emptyset$, *where* $\text{Co}(F_i)$ *denotes the convex hull of* $F_i$.

(3.8) *For any two closed Euclidean balls $C_1$, $C_2$ in $\mathbf{R}^k$, $k \geqslant 1$, we have*

$$\mathrm{Co}\,(C_1 \backslash C_2) \cap \mathrm{Co}\,(C_2 \backslash C_1) = \emptyset.$$

Now, to prove that $\mathscr{K}(\mathbf{R}^k)$ is a VCC it suffices to show that for each $F \subset \mathbf{R}^k$ with $\mathrm{card}(F) = k+2$ there exists a subset $F_0$ of $F$ which cannot be represented in the form $F_0 = F \cap C$ for some $C \in \mathscr{K}(\mathbf{R}^k)$. For an indirect proof, suppose that for each $F_0 \subset F$ there exists $C \in \mathscr{K}(\mathbf{R}^k)$ such that $F_0 = F \cap C$. This implies that for the $F_i$'s of (3.7) which decompose the given $F$ there exist $C_i \in \mathscr{K}(\mathbf{R}^k)$ such that $F_i = F \cap C_i$, $i = 1, 2$. Since $F_1 \cap F_2 = \emptyset$, we have $F_1 \subset C_1 \backslash C_2$ and $F_2 \subset C_2 \backslash C_1$, and therefore

$$\mathrm{Co}(C_1 \backslash C_2) \cap \mathrm{Co}(C_2 \backslash C_1) \supset \mathrm{Co}(F_1) \cap \mathrm{Co}(F_2) \neq \emptyset,$$

which contradicts (3.8).

(3.9) Remarks. The statement that $\mathscr{K}(\mathbf{R}^k)$ is a VCC follows also from Theorem 7.2 of Dudley [6]. Furthermore, Dudley [8] even showed that the so-called Vapnik-Chervonenkis number $V(\mathscr{K}(\mathbf{R}^k))$ equals $k+2$; cf. also [19] for an easier proof of the latter. In the same paper [19] Wenocur and Dudley gave another proof for $\mathscr{I}(\mathbf{R}^k)$ being a VCC (see also Proposition (7.12) of Dudley [6]) by computing its Vapnik-Chervonenkis number $V(\mathscr{I}(\mathbf{R}^k))$ to be $k+1$; therefore the argument used above for $\mathscr{K}(\mathbf{R}^k)$ is applicable to $\mathscr{I}(\mathbf{R}^k)$ as well.

**Acknowledgement.** Thanks to David Barclay who initiated parts of the present paper when working together on this subject. The author is also grateful to the referee for some helpful comments on an earlier version of the manuscript.

## References

[1] P. J. Bickel and M. J. Wichura, *Convergence criteria for multiparameter stochastic processes and some applications*, Ann. Math. Statist. 42 (1971), p. 1656-1670.

[2] M. D. Donsker, *Justification and extension of Doob's heuristic approach to Kolmogorov-Smirnov theorems*, ibidem 23 (1952), p. 277-281.

[3] R. M. Dudley, *Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces*, Illinois J. Math. 10 (1966), p. 109-126.

[4] – *Measures on non-separable metric spaces*, ibidem 11 (1967), p. 449-453.

[5] – *Sample functions of the Gausian process*, Ann. Probability 1 (1973), p. 66-103.

[6] – *Central limit theorems for empirical measures*, ibidem 6 (1978), p. 899-929.

[7] – *Lower layers in $\mathbf{R}^2$ and convex sets in $\mathbf{R}^3$ are not GB-classes*, Lecture Notes in Math. 703 (1979), p. 97-102.

[8] – *Balls in $\mathbf{R}^k$ do not cut all subsets of $k+2$ points*, Advances in Math. 31 (1979), p. 306-308.

[9] H. Finkelstein, *The law of the iterated logarithm for empirical distributions*, Ann. Math. Statist. 42 (1971), p. 607-615.

[10] P. Gaenssler, *On certain properties of convex sets in Euclidean spaces with probabilistic implications*, unpublished manuscript, 1981.
[11] – and W. Stute, *Empirical processes: A survey of results for independent and identically distributed random variables*, Ann. Probability 7 (1979), p. 193-243.
[12] J. Kuelbs and R. M. Dudley, Log log *laws for empirical measures*, ibidem 8 (1980), p. 405-418.
[13] G. Neuhaus, *On weak convergence of stochastic processes with multi-dimensional time parameter*, Ann. Math. Statist. 42 (1971), p. 1285-1295.
[14] R. Pyke, *Recent developments in empirical processes*, Advances in Appl. Probability 11 (1979), p. 267-268.
[15] H. Richter, *Das Gesetz vom iterierten Logarithmus für empirische Verteilungsfunktionen im* $R^k$, Manuscripta Math. 11 (1974), p. 291-303.
[16] M. L. Straf, *Weak convergence of stochastic processes with several parameters*, Proc. Sixth Berkeley Symp. Math. Statist. Prob. 2, p. 187-221, University of California Press, 1971.
[17] F. A. Valentine, *Convex sets*, McGraw-Hill, New York 1964.
[18] V. N. Vapnik and A. Ya. Chervonenkis, *On uniform convergence of the frequencies of events to their probabilities*, Theor. Probability Appl. 16 (1971), p. 264-280.
[19] R. S. Wenocur and R. M. Dudley, *Some special Vapnik-Chervonenkis classes*, Discrete Math. 33 (1981), p. 313-318.

Mathematical Institute
University of Munich
Theresienstrasse 39
D-8000 Munich 2, W-Germany

*Received on 5. 9. 1981*