4. Functional Central Limit Theorems.

In the last section we have already mentioned Donsker's functional central limit theorem for the uniform empirical process $\alpha_n \equiv (\alpha_n(t))_{t\in[0,1]}$, where $\alpha_n(t) = n^{1/2}(U_n(t) - t)$, $U_n(t)$ being the empirical df based on independent random variables η_i having uniform distribution on the sample space $X = [0,1]$ with its Borel σ-algebra $\mathcal{B} = [0,1] \cap \mathbb{B}$.

In the setting of an empirical $\mathcal{C}$-process $\beta_n \equiv (\beta_n(C))_{C\in\mathcal{C}}$ the uniform empirical process α_n is a very special case taking $\mathcal{C} = \{[0,t]: t \in [0,1]\}$ and identifying $\alpha_n(t)$ with $\beta_n(C) = n^{1/2}(\mu_n(C) - \mu(C))$ for $C = [0,t]$, μ_n being the empirical measure based on $\eta_1,\ldots,\eta_n$ and μ being the uniform distribution on $[0,1]$; note that $\mu_n(C) = U_n(t)$ and $\mu(C) = t$ for $C = [0,t]$.

The present section is concerned with some extensions of Donsker's functional central limit theorem in its form (44)(ii) to more general situations.

FUNCTIONAL CENTRAL LIMIT THEOREMS FOR EMPIRICAL $\mathcal{C}$-PROCESSES:

Let $X = (X,\mathcal{B})$ be an arbitrary measurable space considered as a sample space for a given sequence $\xi_1,\xi_2,\ldots$ of i.i.d. random elements in $(X,\mathcal{B})$, the ξ_i's being defined on some common p-space $(\Omega,\mathcal{F},\mathbb{P})$ with law μ on $\mathcal{B}$. If not stated otherwise we will consider the canonical model

$$(\Omega,\mathcal{F},\mathbb{P}) = (X^{\mathbb{N}}, \mathcal{B}_{\mathbb{N}}, \times_{\mathbb{N}} \mu)$$

with the ξ_i's being the coordinate projections of $X^{\mathbb{N}}$ onto X.

Let $\mu_n(B) = \frac{1}{n} \sum_{i=1}^{n} 1_B(\xi_i)$, $B \in \mathcal{B}$, be the empirical measure based on $\xi_1,\ldots,\xi_n$.

Now, given some subclass $\mathcal{C}$ of $\mathcal{B}$, consider the empirical $\mathcal{C}$-process $\beta_n \equiv (\beta_n(C))_{C\in\mathcal{C}}$, defined by

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

as a stochastic process (on $(\Omega,\mathcal{F},\mathbb{P})$) indexed by $\mathcal{C}$.

As mentioned in Section 1, its covariance structure is given by

$$\mathrm{cov}(\beta_n(C_1),\beta_n(C_2)) = \mu(C_1 \cap C_2) - \mu(C_1)\mu(C_2), \quad C_1,C_2 \in \mathcal{C}.$$

So, the analogue of (44)(ii) would be the statement that (in the sense of (34))

(46) $\beta_n \xrightarrow{\mathcal{L}_b} \mathbb{G}_\mu$, $\mathbb{G}_\mu \equiv (G_\mu(C))_{C\in\mathcal{C}}$ being a mean-zero Gaussian process

with $\mathrm{cov}(G_\mu(C_1),G_\mu(C_2)) = \mu(C_1 \cap C_2) - \mu(C_1)\mu(C_2)$, $C_1,C_2 \in \mathcal{C}$.

But this amounts at first to make a proper choice for a metric space $S = (S,d)$ together with a suitable separable subspace S_o serving as sample spaces for β_n and its limiting process $\mathbb{G}_\mu$, respectively.

Following Dudley (1978) we propose to choose $S_o' \equiv \mathcal{U}^b(\mathcal{C},d_\mu) :=$ $\{\varphi\colon \mathcal{C} \to \mathbb{R}\colon \varphi$ bounded and uniformly d_μ-continuous$\}$, where d_μ is the pseudo-metric defined on $\mathcal{C}$ by

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

($C_1 \,\Delta\, C_2$ denoting the symmetric difference between C_1 and C_2).

Note that, concerning the $\mu(C)$-part of $\beta_n(C)$, $C \to \mu(C)$ is a function belonging to S_o (since $|\mu(C_1) - \mu(C_2)| \leq d_\mu(C_1,C_2)$).

In order to cope also with the $\mu_n(C)$-part of $\beta_n(C)$ (and the factor $n^{1/2}$), let

$$S \equiv D_o(\mathcal{C},\mu) := \{\varphi = \varphi_1 + \varphi_2\colon \varphi_1 \in S_o \text{ and } \varphi_2 = \sum_{i=1}^{k} a_i \varepsilon_{x_i} \text{ for some}$$
$$a_i \in \mathbb{R},\ x_i \in X,\ 1 \leq i \leq k,\ k \in \mathbb{N}\}.$$

Note that S is a linear space containing S_o as a linear subspace;

also $\beta_n(\cdot,\omega) \in S$ for all $\omega \in \Omega$.

Finally, let S (and its subspace S_o) be metrized by the metric $d := \rho$, where ρ is the supremum-metric, i.e.,

$$\rho(\varphi',\varphi'') := \sup_{C\in\mathcal{C}} |\varphi'(C) - \varphi''(C)| \quad \text{for } \varphi',\varphi'' \in S.$$

Note that the closure $D(\mathcal{C},\mu)$ of $D_o(\mathcal{C},\mu)$ in the Banach space $\ell^\infty(\mathcal{C}) = (\ell^\infty(\mathcal{C}),\rho)$ of all bounded real-valued functions on $\mathcal{C}$ can be considered as an extension of $D = D[0,1]$ in the classical case, where $X = [0,1]$, $\mathcal{C} = \{[0,t]: t \in [0,1]\}$, and μ is the uniform distribution on $[0,1]$ or any other distribution on $[0,1]$ with a strictly increasing distribution function; also, in the latter case, $\mathcal{U}^b(\mathcal{C},d_\mu)$ equals $C[0,1]$ after identifying $\varphi([0,t])$ with $x(t)$.

Having made this choice for S_o, S and d, in view of (46) the following problems still remain:

<u>PROBLEM ⓐ (MEASURABILITY):</u> Find conditions under which the β_n's can be viewed as random elements in $(S,\mathcal{A})$ for some σ-algebra $\mathcal{A}$ in S such that one meets the situation of Section 3, i.e.

(47) $$\mathcal{B}_b(S,\rho) \subset \mathcal{A} \subset \mathcal{B}(S,\rho)$$

(with $\mathcal{B}_b(S,\rho)$ being the σ-algebra generated by the open ρ-balls in S, and $\mathcal{B}(S,\rho)$ being the Borel σ-algebra in (S,ρ)).

Taking $\mathcal{A} := \sigma(\{\pi_C: C \in \mathcal{C}\})$, with $\pi_C: S \to \mathbb{R}$ being defined by $\pi_C(\varphi) := \varphi(C)$, $C \in \mathcal{C}$,

<u>β_n is $\mathcal{F},\mathcal{A}$-measurable</u>

(since $\mathcal{F},\sigma(\{\pi_C: C \in \mathcal{C}\})$-measurability of β_n is equivalent with $\mathcal{F},\mathcal{B}$-measurability of $\pi_C(\beta_n) = \beta_n(C)$ for each fixed $C \in \mathcal{C}$, the latter being satisfied since $\beta_n(C)$ is a random variable (on $(\Omega,\mathcal{F},\mathbb{P})$) for each fixed C),

<u>but the first inclusion in (47) fails to hold, in general:</u>

in fact, looking back to (10) in Section 1, it follows that in the example considered there β_n is <u>not</u> even $\mathcal{F},\mathcal{B}_b(S,\rho)$-measurable.

So, we will restrict our consideration to cases where the following measurability condition

$$(M): \mathcal{B}_b(S,\rho) \subset \mathcal{A} := \sigma(\{\pi_C: C \in \mathcal{C}\})$$

is fulfilled, which turns out to be satisfied in important cases of interest; note that (M) implies (47), since the other inclusion there holds trivially due to the ρ-continuity of the π_C's for each fixed $C \in \mathcal{C}$.

LEMMA 20. Suppose that $\mathcal{C}$ fulfills the following condition

(SE): There exists a countable subclass $\mathcal{D}$ of $\mathcal{C}$ such that for any $C \in \mathcal{C}$ there exists a sequence $(D_n)_{n\in\mathbb{N}}$ in $\mathcal{D}$ with $1_{D_n}(x) \longrightarrow 1_C(x)$ for all $x \in X$;

then (M) holds true.

Proof. (SE) implies that for any $C \in \mathcal{C}$ there exists a sequence $(D_n)_{n\in\mathbb{N}}$ in $\mathcal{D}$ such that $\lim_{n\to\infty} d_\mu(D_n,C) = 0$ from which it follows that $\varphi_1(C) = \lim_{n\to\infty} \varphi_1(D_n)$ for every $\varphi_1 \in S_o$; on the other hand, since $1_{D_n}(x) \to 1_C(x)$ for all x is equivalent with $\lim_{n\to\infty} \varepsilon_x(D_n) = \varepsilon_x(C)$ for all x, we obtain $\varphi(C) = \lim_{n\to\infty} \varphi(D_n)$ for every $\varphi \in S$. But from this it follows that for any $\varphi_o \in S$ and any $r > 0$

$$\{\varphi \in S: \rho(\varphi,\varphi_o) \leq r\} = \bigcap_{D\in\mathcal{D}} \{\varphi \in S: |\varphi(D) - \varphi_o(D)| \leq r\} \in \mathcal{A},$$

since $\mathcal{D}$ is countable, implying (M). □

(48) EXAMPLES. (a) Let $(X,\mathcal{B}) = (\mathbb{R}^k,\mathcal{B}_k)$, $k \geq 1$, and let $\mathcal{C}$ be the class $\mathcal{J}_k$ of all lower left orthants or the class $\mathbb{B}_k$ of all closed Euclidean balls in $\mathbb{R}^k$, respectively; then (SE) and therefore (M) holds true for $\mathcal{C} = \mathcal{J}_k$ and $\mathcal{C} = \mathbb{B}_k$, respectively.

(b) If we consider instead e.g., the class $\mathcal{C} := \{C_o + z: z \in \mathbb{R}^k\}$, C_o being a fixed closed Euclidean ball in $\mathbb{R}^k$, then (SE) fails to hold: in fact, no $\mathcal{D} = \{C_o + q: q \in R\}$ with countable $R \subset \mathbb{R}^k$ can serve as a countable subclass of $\mathcal{C}$ with the desired property stated in (SE), since for any fixed $z \in \mathbb{R}^k \setminus R$ and any $D_q \in \mathcal{D}$ there exists a $y_q \in \mathbb{R}^k$ such that

$$1 = 1_{C_o+z}(y_q) \neq 1_{D_q}(y_q) = 0$$

(cf. FIGURE 4).

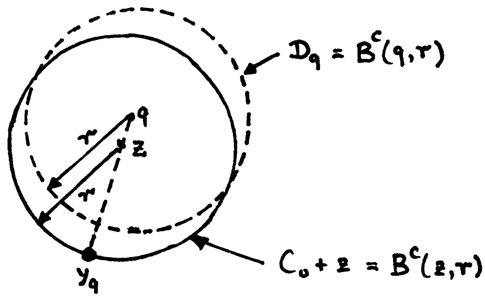

FIGURE 4

We shall see below how to cope also with examples where (SE) fails to hold. For this another measurability assumption (M_o) weaker than (M) will be needed.

It should be noticed (cf. the proof of Lemma 20) that in case of (SE) we have SEPARABILITY of the process $\beta_n \equiv (\beta_n(C))_{C\in\mathcal{C}}$ in the sense that each sample path of β_n is uniquely determined by its values on $\mathcal{D}$.

Let us make some further remarks at this place: first, note that (M) implies

(49) $\mathcal{B}_b(T,\rho) = \sigma(\{\pi_C(T)\colon C \in \mathcal{C}\}) = \mathcal{B}(T,\rho)$ for <u>any</u> separable subspace T of S, with $\pi_C(T)\colon T \to \mathbb{R}$ being defined by $\pi_C(T)(\varphi) := \varphi(C)$.

In fact, the same reasoning which gave us (39) in Section 3 yields

(50) $\mathcal{B}_b(T,\rho) = T \cap \mathcal{B}_b(S,\rho)$ for any separable subspace T of S,

whence (cf. Lemma 11 (iv))

$$\mathcal{B}_b(T,\rho) = T \cap \mathcal{B}_b(S,\rho) \underset{(M)}{\subset} T \cap \mathcal{A} = \sigma(\{\pi_C(T)\colon C \in \mathcal{C}\}) \subset \mathcal{B}(T,\rho) = \mathcal{B}_b(T,\rho),$$

which proves (49).

Next, <u>concerning $S_o \equiv \mathcal{U}^b(\mathcal{C},d_\mu)$</u>, it follows even without imposing (M) that

(49*) $\mathcal{B}_b(S_o,\rho) = \sigma(\{\pi_C(S_o)\colon C \in \mathcal{C}\}) = \mathcal{B}(S_o,\rho)$, <u>provided that $\mathcal{C}$ is totally bounded for d_μ.</u>

In fact, if $\mathcal{C}$ is totally bounded for d_μ, there exists a countable d_μ-dense subset $\mathcal{D}$ of $\mathcal{C}$ implying, due to the d_μ-continuity of functions belonging to S_o, that for any $\varphi_o \in S_o$ and any $r > 0$

$$\{\varphi \in S_o: \rho(\varphi,\varphi_o) \leq r\} = \bigcap_{D\in\mathcal{D}} \{\varphi \in S_o: |\varphi(D) - \varphi_o(D)| \leq r\}$$

$\in \sigma(\{\pi_C(S_o): C \in \mathcal{C}\})$, whence $\mathcal{B}_b(S_o,\rho) \subset \sigma(\{\pi_C(S_o): C \in \mathcal{C}\}) \subset \mathcal{B}(S_o,\rho)$;

on the other hand, using the Stone-Weierstraß theorem, it can be shown that

(51) $S_o \equiv \mathcal{U}^b(\mathcal{C},d_\mu)$ is separable and ρ-closed (i.e. $S_o^c = S_o$),

<u>provided that $\mathcal{C}$ is totally bounded for d_μ.</u>

This proves (49^*).

For later use it is important to note that (49^*) together with (50) and (51) imply

<u>LEMMA 21.</u> Let $\mathcal{C}$ be totally bounded for d_μ and suppose that $\mathbb{G}_\mu \equiv (G_\mu(C))_{C\in\mathcal{C}}$ has all its sample paths in $S_o \equiv \mathcal{U}^b(\mathcal{C},d_\mu)$; then $\mathbb{G}_\mu$ can be viewed as a random element in $(S,\mathcal{B}_b(S,\rho))$ with $L\{\mathbb{G}_\mu\}(S_o) = 1$. Furthermore, $L\{\mathbb{G}_\mu\}$ as well as any other law $\nu \in M_b^1(S)$ with $\nu(S_o) = 1$ is uniquely determined by its fidis (which are the image measures that $\pi_{C_1,\ldots,C_k}(S_o): S_o \to \mathbb{R}^k$ induce on $\mathcal{B}_k$ from ν when ν is viewed as defined on $S_o \cap \mathcal{B}_b(S,\rho) = \mathcal{B}_b(S_o,\rho) = \sigma(\{\pi_C(S_o): C \in \mathcal{C}\})$, where $\pi_{C_1,\ldots,C_k}(S_o)(\varphi) := (\varphi(C_1),\ldots,\varphi(C_k))$.)

This leads us to the next

<u>PROBLEM (b): (EXISTENCE OF A VERSION OF $\overline{\mathbb{G}}_\mu \equiv (\overline{G}_\mu(C))_{C\in\mathcal{C}}$ in $S_o \equiv \mathcal{U}^b(\mathcal{C},d_\mu)$):</u>

Let $\overline{\mathbb{G}}_\mu \equiv (\overline{G}_\mu(C))_{C\in\mathcal{C}}$ be a mean-zero Gaussian process with covariance structure (cf. Section 1, (4))

$$\operatorname{cov}(\overline{G}_\mu(C_1), \overline{G}_\mu(C_2)) = \mu(C_1 \cap C_2) - \mu(C_1)\mu(C_2),\ C_1,C_2 \in \mathcal{C}.$$

Noticing that the fidis of β_n (viewed as a random element in $(S,\mathcal{A})$ with $\mathcal{A} := \sigma(\{\pi_C: C \in \mathcal{C}\})$) are well defined, we have according to (4) of Section 1

(52) $\beta_n \xrightarrow[\text{f.d.}]{L} \overline{\mathbb{G}}_\mu$, $\overline{\mathbb{G}}_\mu$ being viewed as the coordinate process on

$$(\overline{\Omega},\overline{F},\overline{\mathbb{P}}) = (\mathbb{R}^{\mathcal{C}},\mathcal{B}_{\mathcal{C}},L\{\overline{\mathbb{G}}_\mu\})$$

(where $L\{\overline{\mathbb{G}}_\mu\}$ is uniquely determined by the fidis of $\overline{\mathbb{G}}_\mu$ (Kolmogorov's theorem)).

Now, the problem is to find suitable conditions under which there exists a version $\mathbb{G}_\mu$ of $\overline{\mathbb{G}}_\mu$ having all its sample paths in S_o, where here VERSION is to be understood in the sense that $\mathbb{G}_\mu$ and $\overline{\mathbb{G}}_\mu$ have the same fidis (denoted by $\mathbb{G}_\mu \overset{L}{\underset{f.d.}{=}} \overline{\mathbb{G}}_\mu$); in this connection $\mathbb{G}_\mu$ is allowed to be defined on a p-space $(\Omega',F',\mathbb{P}')$ different from $(\overline{\Omega},\overline{F},\overline{\mathbb{P}})$.

It turns out that in order to get a positive result, C is not allowed to be too "large"; cf. R.M. Dudley (1979 a) and also R.M. Dudley (1982).
A proper condition on a class C being not too large to allow for a solution of problem ⓑ is in terms of the so-called metric entropy:

for this, let for any $\varepsilon>0$ $N(\varepsilon,C,\mu)$ be the smallest $n \in \mathbb{N}$ such that
$C = \bigcup_{j=1}^{n} C_j$ for some classes C_j with d_μ-diam$(C_j) := \sup\{d_\mu(C',C''): C',C''\in C_j\} \leq 2\varepsilon$ for each j;
$\log N(\varepsilon,C,\mu)$ is called a METRIC ENTROPY (of C w.r.t. μ).

Obviously, $N(\varepsilon,C,\mu) < \infty$ for each $\varepsilon>0$ iff C is totally bounded for d_μ (in which case $S_o \equiv \mathcal{U}^b(C,d_\mu)$ is separable and ρ-closed, by (51)).

Now, as shown by R.M. Dudley (1967) and (1973), cf. p. 71,

(53) $\overline{\mathbb{G}}_\mu$ has a version $\mathbb{G}_\mu \equiv (G_\mu(C))_{C\in C}$ having all its sample paths in $S_o \equiv \mathcal{U}^b(C,d_\mu)$ provided that

$$(E_o): \int_0^1 (\log N(x^2,C,\mu))^{1/2}\, dx < \infty.$$

But it turns out that (E_o) is not sufficient to ensure (46); in fact, disregarding for the moment measurability questions, the following example shows that (46), i.e. $\beta_n \xrightarrow{L_b} \mathbb{G}_\mu$, fails to hold although (E_o) is satisfied:
Let C be the collection of all finite subsets of $X = [0,1]$ and let μ be the uniform distribution (Lebesgue measure) on $\mathcal{B} = [0,1] \cap \mathbb{B}$; then $d_\mu(C_1,C_2) = 0$

for all $C_1, C_2 \in \mathcal{C}$, whence $N(x^2,\mathcal{C},\mu) \equiv 1$ and therefore (E_o) obviously holds true; also $\mu(C) \equiv 0$ on $\mathcal{C}$ implies that $\mathbb{G}_\mu \equiv 0$, but still (46) fails to hold: (46) would imply $\sup_{C\in\mathcal{C}} |\beta_n(C)| \xrightarrow{L} \sup_{C\in\mathcal{C}} |G_\mu(C)| = 0$ which cannot be true since for the present $\mathcal{C}$ $\sup_{C\in\mathcal{C}} |\beta_n(C)| \equiv n^{1/2} \to \infty$ as $n \to \infty$.

A proper strengthening of (E_o) which will yield (46) and hence also a solution to problem ⓑ is in terms of the so-called metric entropy with inclusion.

For this, let for any $\varepsilon>0$ $N_I(\varepsilon,\mathcal{C},\mu)$ be the smallest $n \in \mathbb{N}$ such that for some $A_1,\ldots,A_n \in \mathcal{B}$ (not necessarily in $\mathcal{C}$), for every $C \in \mathcal{C}$ there exist i,j with $A_i \subset C \subset A_j$ and $\mu(A_j \setminus A_i) < \varepsilon$; $\log N_I(\varepsilon,\mathcal{C},\mu)$ is called a METRIC ENTROPY WITH INCLUSION (of $\mathcal{C}$ w.r.t. μ).

Compared with $N(\varepsilon,\mathcal{C},\mu)$ we have for any $\mathcal{C} \subset \mathcal{B}$ and any μ

(54) $$N(\varepsilon,\mathcal{C},\mu) \leqq N_I(\varepsilon,\mathcal{C},\mu) \text{ for each } \varepsilon>0.$$

For, suppose w.l.o.g. that $n = N_I(\varepsilon,\mathcal{C},\mu) < \infty$; then there exist $A_1,\ldots,A_n \in \mathcal{B}$ such that for any $C \in \mathcal{C}$ there exist i,j with $A_i \subset C \subset A_j$ and $\mu(A_j \setminus A_i) < \varepsilon$. But then, for $i=1,\ldots,n$, $\mathcal{C}_i := \{C \in \mathcal{C}: A_i \subset C \text{ and } d_\mu(A_i,C) < \varepsilon\} \neq \emptyset$, d_μ-diam $(\mathcal{C}_i) \leqq 2\varepsilon$, and $\mathcal{C} = \bigcup_{i=1}^{n} \mathcal{C}_i$ (since for each $C \in \mathcal{C}$ there exist i,j such that $A_i \subset C \subset A_j$ and $\mu(A_j \setminus A_i) < \varepsilon$ which implies $d_\mu(A_i,C) \leqq \mu(A_j \setminus A_i) < \varepsilon$, i.e., $C \in \mathcal{C}_i$). This proves (54).

Now, as shown by R.M. Dudley (1978), Theorem 5.1, the following result holds true:

<u>THEOREM A.</u> (M) together with

$$(E_1): \int_0^1 (\log N_I(x^2,\mathcal{C},\mu))^{1/2}\, dx < \infty$$

imply that $\beta_n \xrightarrow{L_b} \mathbb{G}_\mu$.

The proof of Theorem A is based on the following fundamental characterization theorem (cf. R.M. Dudley (1978)):

<u>THEOREM B.</u> Let $(X,\mathcal{B})$ be an arbitrary measurable space considered as sample space for a given sequence $\xi_1,\xi_2,\ldots$ of i.i.d. random elements ξ_i in $(X,\mathcal{B})$, where the ξ_i's are viewed as coordinate projections of $(\Omega,\mathcal{F},\mathbb{P}) = (X^{\mathbb{N}}, \mathcal{B}_{\mathbb{N}}, \times_{\mathbb{N}} \mu)$ onto X with law $\mathcal{L}\{\xi_i\} \equiv \mu$ on $\mathcal{B}$. Suppose, given some subclass $\mathcal{C}$ of $\mathcal{B}$ together with the empirical $\mathcal{C}$-process $\beta_n \equiv (\beta_n(C))_{C\in\mathcal{C}}$ based on $\xi_1,\ldots,\xi_n$, that (M) is fulfilled.

Then $\beta_n \xrightarrow{\mathcal{L}_b} \mathbb{G}_\mu$ (in which case $\mathcal{C}$ will be called a μ-DONSKER CLASS) if and only if both

(a) $\mathcal{C}$ is totally bounded for d_μ, and

(b) for any $\varepsilon,\eta > 0$ there exists a $\delta = \delta(\varepsilon,\eta)$, $0 < \delta < 1$, and there exists an $n_o = n_o(\varepsilon,\eta,\delta) \in \mathbb{N}$ such that for $n \geq n_o$

$$\mathbb{P}^*(w_{\beta_n}(\delta) > \varepsilon) < \eta,$$

where $w_\varphi(\delta) := \sup\{|\varphi(C_1) - \varphi(C_2)|: d_\mu(C_1,C_2) < \delta,\ C_1,C_2 \in \mathcal{C}\}$ for $\varphi \in S \equiv D_o(\mathcal{C},\mu)$.

(55) <u>REMARK.</u> A comparison with (45) shows the complete analogy with the classical situation $X = [0,1]$, $\mathcal{B} = [0,1] \cap \mathbb{B}$, μ = uniform distribution on $\mathcal{B}$, $\mathcal{C} = \{[0,t]: t \in [0,1]\}$, where β_n can be identified with the uniform empirical process α_n; note that, due to the compactness of the unit interval, $\mathcal{C} = \{[0,t]: t \in [0,1]\}$ is totally bounded for d_μ: given any $\varepsilon>0$ let

$$n_o := \inf\{n: \tfrac{1}{n} \leq 2\varepsilon\} \text{ and } \mathcal{C}_j := \{[0,t]: \tfrac{j-1}{n_o} \leq t < \tfrac{j}{n_o}\};$$

then d_μ-diam $(\mathcal{C}_j) \leq 2\varepsilon$ and $\mathcal{C} = \bigcup_{j=1}^{n_o} \mathcal{C}_j$.

Before proving Theorem B we will show two auxiliary results:

<u>PROPOSITION B_1 (cf. Problem ⓑ above).</u> Suppose that (a) and (b) of Theorem B

are fulfilled; then $\overline{\mathbb{G}}_\mu = (\overline{G}_\mu(C))_{C\in\mathcal{C}}$ has a version in $S_o \equiv \mathcal{U}^b(\mathcal{C},d_\mu)$, i.e., there exists a Gaussian process $\mathbb{G}_\mu \equiv (G_\mu(C))_{C\in\mathcal{C}}$ having all its sample paths in S_o and such that $\mathbb{G}_\mu \overset{L}{\underset{\text{f.d.}}{=}} \overline{\mathbb{G}}_\mu$.

Thus, by Lemma 21, $\mathbb{G}_\mu$ can be viewed as a random element in $(S,\mathcal{B}_b(S,\rho))$ with $L\{\mathbb{G}_\mu\}(S_o) = 1$, where, by (51), S_o is separable and ρ-closed.

<u>Proof of Proposition B_1.</u> As already remarked in connection with problem (b) above, the process $\overline{\mathbb{G}}_\mu$ is viewed as the coordinate process on $(\mathbb{R}^{\mathcal{C}},\mathcal{B}_{\mathcal{C}},L\{\overline{\mathbb{G}}_\mu\})$. According to (a), for every $n \in \mathbb{N}$ there exist $m_n \in \mathbb{N}$ and $C_{n,1},\ldots,C_{n,m_n} \in \mathcal{C}$ such that $\mathcal{C} = \bigcup_{i=1}^{m_n} B^c_{d_\mu}(C_{n,i},\frac{1}{n})$, where $B^c_{d_\mu}(C_{i,n},\frac{1}{n}) := \{C \in \mathcal{C}: d_\mu(C_{i,n},C) \leq \frac{1}{n}\}$; therefore, $\mathcal{D} := \bigcup_{n\in\mathbb{N}} \bigcup_{i=1}^{m_n} \{C_{n,i}\}$ is a countable and d_μ-dense subset of $\mathcal{C}$. Let $\mathcal{U}(\mathcal{D},d_\mu) := \{\varphi: \mathcal{D} \to \mathbb{R},\ \varphi \text{ uniformly } d_\mu\text{-continuous}\}$, and let $\overline{\mathbb{G}}_{\mu,\mathcal{D}} \equiv (\overline{G}_\mu(D))_{D\in\mathcal{D}}$, viewed as the coordinate process on $(\overline{\Omega},\overline{F},\overline{\mathbb{P}}_{\mathcal{D}}) = (\mathbb{R}^{\mathcal{D}},\mathcal{B}_{\mathcal{D}},L\{\overline{\mathbb{G}}_{\mu,\mathcal{D}}\})$. Then it suffices to show

(+) There exists a Gaussian process $\mathbb{G}_{\mu,\mathcal{D}} \equiv (G_\mu(D))_{D\in\mathcal{D}}$ on some p-space $(\Omega',F',\mathbb{P}')$ having all its sample paths in $\mathcal{U}(\mathcal{D},d_\mu)$ and such that $\mathbb{G}_{\mu,\mathcal{D}} \overset{L}{\underset{\text{f.d.}}{=}} \overline{\mathbb{G}}_{\mu,\mathcal{D}}$.

In fact, once (+) is shown, we can define for each $\omega' \in \Omega'$ $\mathbb{G}_\mu(\omega')$ as the uniquely determined uniformly d_μ-continuous extension on $\mathcal{C}$ of $\mathbb{G}_{\mu,\mathcal{D}}(\omega')$ (i.e., for each $C \in \mathcal{C}$ $G_\mu(C,\omega') = \lim_{n\to\infty} G_\mu(D_n,\omega')$, $(D_n)_{n\in\mathbb{N}} \subset \mathcal{D}$ being such that $d_\mu(C,D_n) \to 0$ as $n \to \infty$). It follows that

(*) $\mathbb{G}_\mu(\omega')$ is bounded for each ω', whence $\mathbb{G}_\mu(\omega') \in S_o$ for all ω',

and (**) $\mathbb{G}_\mu \overset{L}{\underset{\text{f.d.}}{=}} \overline{\mathbb{G}}_\mu$.

<u>ad (*):</u> By (a), for every $\varepsilon>0$ there exist an $n_o = n_o(\varepsilon) \in \mathbb{N}$ and $\mathcal{C}_j \subset \mathcal{C}$, $j=1,\ldots,n_o$, such that d_μ-diam $(\mathcal{C}_j) \leq 2\varepsilon$ and $\mathcal{C} = \bigcup_{j=1}^{n_o} \mathcal{C}_j$.

Let $\omega' \in \Omega'$ be arbitrary but fixed; since $\mathbb{G}_\mu(\omega')$ is uniformly continuous on $\mathcal{C}$, for each $\delta>0$ there exists an $\varepsilon = \varepsilon(\delta,\omega') > 0$ such that

$$|G_\mu(C_1,\omega') - G_\mu(C_2,\omega')| < \delta \text{ whenever } d_\mu(C_1,C_2) \leqq 2\varepsilon \text{ for } C_1,C_2 \in \mathcal{C}.$$

Now, given an arbitrary $C \in \mathcal{C}$, there exists a $j \in \{1,\dots,n_o\}$ and a $C_j \in \mathcal{C}_j$ such that $d_\mu(C,C_j) \leqq 2\varepsilon$, and therefore

$|G_\mu(C,\omega')| \leqq |G_\mu(C,\omega') - G_\mu(C_j,\omega')| + |G_\mu(C_j,\omega')| \leqq \delta + |G_\mu(C_j,\omega')|$, whence

$\sup_{C\in\mathcal{C}} |G_\mu(C,\omega')| \leqq \delta + \sup_{1\leqq j\leqq n_o} |G_\mu(C_j,\omega')| < \infty.$

<u>ad (**):</u> Let us confine here to show that $\mathcal{L}\{G_\mu(C)\} = \mathcal{L}\{\overline{G}_\mu(C)\}$ for each fixed $C \in \mathcal{C}$; concerning the higher-dimensional fidis the proof runs in a similar way.

Now, given any $C \in \mathcal{C}$, let $(D_n)_{n\in\mathbb{N}} \subset \mathcal{D}$ be such that $d_\mu(C,D_n) \to 0$ as $n \to \infty$, whence, by construction,

$$G_\mu(C,\omega') = \lim_{n\to\infty} G_\mu(D_n,\omega') \quad \text{for all } \omega' \in \Omega',$$

implying $G_\mu(D_n) \xrightarrow{\mathcal{L}} G_\mu(C)$. Now, by ⊕, $\mathcal{L}\{G_\mu(D_n)\} = \mathcal{L}\{\overline{G}_\mu(D_n)\} = \mathcal{N}(0,\mu(D_n)(1 - \mu(D_n)))$ (cf. (3) of Section 1) for each $n \in \mathbb{N}$, where $\mu(D_n) \to \mu(C)$ as $n \to \infty$, since $\lim_{n\to\infty} d_\mu(C,D_n) = 0$;

therefore $\mathcal{L}\{G_\mu(C)\} = \mathcal{N}(0,\mu(C)(1 - \mu(C))) = \mathcal{L}\{\overline{G}_\mu(C)\}$.

So it remains to show ⊕:

According to Lemma 7.2.31 and Satz 7.1.18 in Gaenssler-Stute (1977) ⊕ is equivalent with

$$(++) \qquad \mathbb{P}_{\mathcal{D}}(\{\varphi \in \mathbb{R}^{\mathcal{D}}: \varphi \in \mathcal{U}(\mathcal{D},d_\mu)\}) = 1,$$

where $\varphi \in \mathcal{U}(\mathcal{D},d_\mu)$ iff $\lim_{\delta\downarrow 0} w_\varphi^{\mathcal{D}}(\delta) = 0$ with

$$w_\varphi^{\mathcal{D}}(\delta) := \sup\{|\varphi(D_1) - \varphi(D_2)|: d_\mu(D_1,D_2) < \delta,\ D_1,D_2 \in \mathcal{D}\}$$

being $\mathcal{B}_{\mathcal{D}},\mathcal{B}$-measurable as a function in φ.

Note that for any $\varphi \in \mathbb{R}^{\mathcal{D}}$ and any $\delta>0$

$$w_\varphi^{\mathcal{D}_n}(\delta) \uparrow w_\varphi^{\mathcal{D}}(\delta) \quad \text{as } \mathcal{D}_n \uparrow \mathcal{D},$$

whence for any $\varepsilon>0$ we have

(c) $$\{\varphi \in \mathbb{R}^{\mathcal{D}} : \omega_{\varphi}^{\mathcal{D}}(\delta) > \varepsilon\} \subset \bigcup_{n\in\mathbb{N}} \{\varphi \in \mathbb{R}^{\mathcal{D}} : \omega_{\varphi}^{\mathcal{D}_n}(\delta) > \varepsilon\} \text{ as } \mathcal{D}_n \uparrow \mathcal{D}.$$

We are going to show next that (++) is implied by

(R_o): For any $\varepsilon,\eta > 0$ there exists a $\delta = \delta(\varepsilon,\eta)$, $0 < \delta < 1$, such that

$$\bar{\mathbb{P}}_{\mathcal{D}}(\{\varphi \in \mathbb{R}^{\mathcal{D}} : \omega_{\varphi}^{\mathcal{D}}(\delta) > \varepsilon\}) < \eta.$$

In fact, (R_o) implies that for each fixed $\varepsilon>0$

$$\sum_{n\in\mathbb{N}} \bar{\mathbb{P}}_{\mathcal{D}}(\{\varphi \in \mathbb{R}^{\mathcal{D}} : \omega_{\varphi}^{\mathcal{D}}(\delta_n) > \varepsilon\}) < \infty$$

for some suitable sequence $\delta_n \downarrow 0$, whence, by the Borel-Cantelli lemma, for $\bar{\mathbb{P}}_{\mathcal{D}}$-almost all $\varphi \in \mathbb{R}^{\mathcal{D}}$ there exists an $n(\varphi) \in \mathbb{N}$ such that for all $n \geqq n(\varphi)$ $\omega_{\varphi}^{\mathcal{D}}(\delta_n) \leqq \varepsilon$ which implies, by repeating the argument for a sequence of ε's tending to zero, that for $\bar{\mathbb{P}}_{\mathcal{D}}$-almost all $\varphi \in \mathbb{R}^{\mathcal{D}}$

$$\lim_{\delta\downarrow 0} \omega_{\varphi}^{\mathcal{D}}(\delta) = 0, \text{ i.e. } \varphi \in \mathcal{U}(\mathcal{D},d_{\mu}),$$

which proves (++).

So far we only made use of assumption (a); now, the proof of Proposition B_1 will be concluded by showing that the other assumption (b) implies (R_o): for this, remembering that $\mathcal{D}$ is countable, let $\mathcal{D}_n \subset \mathcal{D}$, $n \in \mathbb{N}$, with $|\mathcal{D}_n| < \infty$ $\mathcal{D}_n \uparrow \mathcal{D}$; then, according to (c) it suffices to show:

(d) For any $\varepsilon,\eta > 0$ there exists a $\delta = \delta(\varepsilon,\eta)$, $0 < \delta < 1$, such that for any $\mathcal{D}' \subset \mathcal{D}$ with $|\mathcal{D}'| < \infty$

$$\bar{\mathbb{P}}_{\mathcal{D}}(\{\varphi \in \mathbb{R}^{\mathcal{D}} : \omega_{\varphi}^{\mathcal{D}'}(\delta) > \varepsilon\}) < \eta,$$

where, for $\mathcal{D}' = \{D_1,\ldots,D_\ell\}$, $\omega_{\varphi}^{\mathcal{D}'}(\delta) > \varepsilon$ iff $(\varphi(D_1),\ldots,\varphi(D_\ell)) \in G$ with $G = G_{\varepsilon,\delta}$ being some <u>open</u> subset of $\mathbb{R}^\ell$.

Now, given an arbitrary $\varepsilon>0$ and an arbitrary $\eta>0$ choose $\delta = \delta(\varepsilon,\eta)$, $0 < \delta < 1$ according to (b) such that for all $n \geqq n_o(\varepsilon,\eta,\delta)$

(e) $$\mathbb{P}^*(\omega_{\beta_n}(\delta) > \varepsilon) < \eta.$$

Then it follows that for each $\mathcal{D}' = \{D_1,\ldots,D_\ell\} \subset \mathcal{D}$ $(\subset \mathcal{C})$

$$\bar{\mathbb{P}}_{\mathcal{D}}(\{\varphi \in \mathbb{R}^{\mathcal{D}}: w_\varphi^{\mathcal{D}'}(\delta) > \varepsilon\}) = \bar{\mathbb{P}}_{\mathcal{D}}(\{\varphi \in \mathbb{R}^{\mathcal{D}}: (\varphi(D_1),\ldots,\varphi(D_\ell)) \in G\})$$

$$= \bar{\mathbb{P}}_{\mathcal{D}} \circ \pi^{-1}_{D_1,\ldots,D_\ell}(G) = \mathcal{L}\{\mathbb{G}_\mu\} \circ \pi^{-1}_{D_1,\ldots,D_\ell}(G)$$

$$\leq \liminf_{n\to\infty} \mathcal{L}\{\beta_n\} \circ \pi^{-1}_{\{D_1,\ldots,D_\ell\}}(G) = \liminf_{n\to\infty} \mathbb{P} \circ (\pi_{\{D_1,\ldots,D_\ell\}} \circ \beta_n)^{-1}(G)$$

$$= \liminf_{n\to\infty} \mathbb{P}(w_{\beta_n}^{\mathcal{D}'}(\delta) > \varepsilon) \leq \liminf_{n\to\infty} \mathbb{P}^*(w_{\beta_n}(\delta) > \varepsilon) \underset{(e)}{\leq} \eta,$$

where for the first inequality above we made use of (28) and the fact that according to (52) and (**) $\beta_n \xrightarrow[\text{f.d.}]{\mathcal{L}} \mathbb{G}_\mu$.

This proves Proposition B_1. □

(56) REMARK. The proof just given of Proposition B_1 shows that in order to get a result like (53), it suffices to show that an entropy condition like (E_o) implies (R_o). This was nicely demonstrated by D. Pollard (1982) in one of his Seminar talks at Seattle using an analogue of the chaining argument of R.M. Dudley ((1978), pp. 915, 924); cf. also D. Pollard (1981), pp. 191-192.

PROPOSITION B_2. Suppose that (a) and (b) of Theorem B are fulfilled and also (M); then $(\mathcal{L}\{\beta_n\})_{n\in\mathbb{N}}$ is δ-tight w.r.t. $S_o \equiv \mathcal{U}^b(\mathcal{C},d_\mu)$.

(Note again that $\mathcal{L}\{\beta_n\} \in M_a^1(S)$, $S \equiv D_o(\mathcal{C},\mu)$.)

For the proof of Proposition B_2 we will make use of the Kirszbraun-McShane-Theorem (cf. M.D. Kirszbraun (1934) and McShane (1934)): let $S = (S,d)$ be a metric space, $A \subset S$, and let φ be a real-valued function defined on A such that

$$\sup\{|\varphi(x) - \varphi(y)|/d(x,y): x,y \in A,\ x \neq y\} =: K < \infty;$$

then φ can be extended to a function ψ on all of S with

$$\sup\{|\psi(x) - \psi(y)|/d(x,y): x,y \in S,\ x \neq y\} = K.$$

Proof of Proposition B_2 (cf. R.M. Dudley (1978), Lemma (1,3)).

For any $\varepsilon,\delta > 0$ let

$B_{\delta,\varepsilon} := \{\varphi \in D_o(\mathcal{C},\mu): \exists C_1,C_2 \in \mathcal{C}$ s.t. $d_\mu(C_1,C_2) < \delta$ and $|\varphi(C_1) - \varphi(C_2)| > \varepsilon\}$.

Note that $\varphi \in B_{\delta,\varepsilon}$ iff $\omega_\varphi(\delta) > \varepsilon$.

We have to show:

for any $0 < \varepsilon < 1$ there exists a compact set $K \subset \mathcal{U}^b(\mathcal{C},d_\mu)$ such that for each $\gamma>0$ $\mathcal{L}\{\beta_n\}(K^\gamma) > 1 - \varepsilon$ for n large enough.

(Note that $K^\gamma \in \mathcal{B}_b(S,\rho) \subset \mathcal{A}$ by (M).)

Let $0 < \varepsilon < 1$ be given; by (b) take $\delta = \delta(\varepsilon)$, $0 < \delta < 1$, such that

(a) $$\mathbb{P}^*(\beta_n \in B_{\delta,\varepsilon/2}) < \varepsilon/4 \text{ for all } n \geq n_o(\varepsilon,\delta(\varepsilon)).$$

According to (a) there exists a finite $\mathcal{C}_o = \mathcal{C}_o(\delta) \subset \mathcal{C}$ such that for all $C \in \mathcal{C}$, $d_\mu(C,C_o) < \delta$ for some $C_o \in \mathcal{C}_o$.

Let $k := |\mathcal{C}_o|$; then $k = k(\delta(\varepsilon)) \in \mathbb{N}$.

Take $M = M(\varepsilon)$ large enough so that $(M - 1)^{-2} < \varepsilon/k$; then

(b) $$\mathbb{P}(\sup_{C\in\mathcal{C}} |\beta_n(C)| > M) < \varepsilon/2 \quad \text{for all } n \geq n_o(\varepsilon) \equiv n_o(\varepsilon,\delta(\varepsilon)).$$

<u>ad (b):</u> Note that $\{\omega: \sup_{C\in\mathcal{C}} |\beta_n(C,\omega)| > M\} \in \mathcal{F}$ according to (M);

now, for each $C_o \in \mathcal{C}_o$, $\mathbb{P}(|\beta_n(C_o)| > M - 1) < \varepsilon/4k$ by Chebyshev's inequality (and the choice of M), whence

(b_o) $$\mathbb{P}(\sup_{C\in\mathcal{C}_o} |\beta_n(C_o)| > M - 1) < \varepsilon/4.$$

Next, $\sup_{C\in\mathcal{C}} |\beta_n(C,\omega)| > M$ and $|\beta_n(C_1,\omega) - \beta_n(C_2,\omega)| \leq \varepsilon/2$

for all $C_1,C_2 \in \mathcal{C}$ with $d_\mu(C_1,C_2) < \delta$ together imply (due to the choice of $\mathcal{C}_o$) that there exists a $C_o \in \mathcal{C}_o$ such that

$|\beta_n(C_o,\omega)| > M - \varepsilon/2 > M - 1$, whence

$$\{\sup_{C\in\mathcal{C}} |\beta_n(C)| > M\} \subset \{\beta_n \in B_{\delta,\varepsilon/2}\} \cup \{\sup_{C_o\in\mathcal{C}_o} |\beta_n(C_o)| > M - 1\}$$

which implies (b) according to (a) and (b_o).

Now, for any $j \in \mathbb{N}$, let $\varepsilon(j) := \varepsilon \cdot 2^{-j}$; then by (b) there exists a sequence $\delta(j) = \delta(j,\varepsilon) > 0$, $j \in \mathbb{N}$, such that

(i) $\delta(j+1) < \delta(j)/2$, and

(ii) $\mathbb{P}^*(\beta_n \in B_{\delta(j),\varepsilon(j)}) < \varepsilon(j)$ for all $n \geq n_o(\varepsilon,j)$.

Let $A_j := B_{\delta(j),\varepsilon(j)}$ and $\delta_j := \dfrac{\delta(j)\varepsilon}{2^{j+1}M} = \dfrac{\delta(j)\varepsilon(j)}{2M}$;

then, by (i), we have

(iii) $\delta_{j+1} < \delta_j/4$ and $\dfrac{\varepsilon(j)}{\delta_j}$ is increasing with j.

Furthermore, for $m \geq 2$, let

$$F_m := \{\varphi \in D_o(\mathcal{C},\mu) : \sup_{C\in\mathcal{C}} |\varphi(C)| \leq M \text{ and s.t. for all } C_1,C_2 \in \mathcal{C}$$

$$|\varphi(C_1) - \varphi(C_2)| \leq \varepsilon(j)\cdot\max\left(1, \frac{d_\mu(C_1,C_2)}{\delta_j}\right) \text{ for } j=2,\ldots,m\};$$

then

ⓒ $\sup_{C\in\mathcal{C}} |\varphi(C)| \leq M$ for some $\varphi \in D_o(\mathcal{C},\mu)$ and $\varphi \in \complement A_j$ for $j=2,\ldots,m$

together imply that $\varphi \in F_m$.

ad ⓒ: $\sup_{C\in\mathcal{C}} |\varphi(C)| \leq M$ implies that for all $C_1,C_2 \in \mathcal{C}$

$$|\varphi(C_1) - \varphi(C_2)| \leq 2M = \frac{\delta(j)\varepsilon(j)}{\delta_j} \leq \varepsilon(j)\,\frac{d_\mu(C_1,C_2)}{\delta_j}, \text{ if}$$

$d_\mu(C_1,C_2) \geq \delta(j)$ for all $C_1,C_2 \in \mathcal{C}$; on the other hand,

$d_\mu(C_1,C_2) < \delta(j)$ for some $C_1,C_2 \in \mathcal{C}$ together with $\varphi \in \complement A_j$

imply $|\varphi(C_1) - \varphi(C_2)| \leq \varepsilon(j)$, which proves ⓒ.

We will show next that (ii) together with ⓑ and ⓒ imply

ⓓ For each $m \geq 2$ there exists an $n_1 = n_1(\varepsilon,m) \in \mathbb{N}$ such that for all $n \geq n_1$ there exists an $E_{nm} \in \mathcal{F}$ with $\mathbb{P}(E_{nm}) > 1-\varepsilon$ and $\beta_n(\cdot,\omega) \in F_m$ for all $\omega \in E_{nm}$.

ad ⓓ: According to (ii), let $n_o(\varepsilon,m)$ be large enough such that for all $n \geq n_o(\varepsilon,m)$ and each $j=2,\ldots,m$ there exist $E'_{nj} \in \mathcal{F}$ with

$\{\beta_n \in A_j\} \subset E'_{nj}$ and $\mathbb{P}(E'_{nj}) < \varepsilon(j) = \varepsilon \cdot 2^{-j}$,

whence $\mathbb{P}(\complement \bigcup_{j=2}^{m} E'_{nj}) > 1 - \varepsilon/2$ and $\complement \bigcup_{j=2}^{m} E'_{nj} \subset \bigcap_{j=2}^{m} \{\beta_n \in \complement A_j\}$;

thus, for $E_{nm} := (\complement \bigcup_{j=2}^{m} E'_{nj}) \cap \{\sup_{C\in\mathcal{C}} |\beta_n(C)| \leq M\} \in \mathcal{F}$,

we obtain together with ⓑ and ⓒ that for $n \geq n_1 := \max(n_o(\varepsilon,m), n_o(\varepsilon))$

$\mathbb{P}(E_{nm}) > 1 - \varepsilon$ and $\beta_n(\cdot,\omega) \in F_m$ for all $\omega \in E_{nm}$.

This proves ⓓ.

Now let

$$K := \{\varphi \in \ell^{\infty}(\mathcal{C}): \sup_{C\in\mathcal{C}} |\varphi(C)| \leq M \text{ and s.t. for all } j \in \mathbb{N}$$

$$d_\mu(C_1,C_2) < \delta_j/2 \text{ implies } |\varphi(C_1) - \varphi(C_2)| \leq 3\varepsilon(j)\}.$$

Then $K \subset \mathcal{U}^b(\mathcal{C},d_\mu)$. Now, $(\mathcal{C},d_\mu)$ is totally bounded and K is a uniformly bounded and equicontinuous family of functions being ρ-closed in the Banach space $(\ell^{\infty}(\mathcal{C}),\rho)$ whence, by the Arzelà-Ascoli theorem (applied to the completion of $(\mathcal{C},d_\mu)$) it follows that K is compact.

So, it remains to show that for each $\gamma>0$

$$\mathcal{L}\{\beta_n\}(K^\gamma) > 1 - \varepsilon \quad \text{for n large enough.}$$

For this it suffices to prove

ⓔ For each $\gamma>0$ there exists an $m = m(\varepsilon,\gamma)$ such that $F_m \subset K^\gamma$.

In fact, ⓔ together with ⓓ imply $\mathcal{L}\{\beta_n\}(K^\gamma) \geq \mathbb{P}^*(\beta_n \in F_m) \geq \mathbb{P}(E_{nm}) > 1 - \varepsilon$ for $n \geq n_1(\varepsilon,m(\varepsilon,\gamma))$, which concludes the proof of Proposition B_2.

<u>ad ⓔ:</u> Given $\gamma>0$, choose $m = m(\varepsilon,\gamma)$ such that $\varepsilon(m) < \gamma/2$ and take a <u>maximal</u> set $\mathcal{C}_m \subset \mathcal{C}$ such that

$$d_\mu(C_1,C_2) \geq \delta_m \text{ for all } C_1 \neq C_2 \text{ in } \mathcal{C}_m.$$

Then $\mathcal{C}_m$ is finite by (a) and for all $C \in \mathcal{C}$, $d_\mu(C,C') < \delta_m$ for some $C' \in \mathcal{C}_m$ (by the maximality of $\mathcal{C}_m$).

Now, if $\varphi \in F_m$ and $C_1, C_2 \in \mathcal{C}_m$, then (since $\frac{d_\mu(C_1,C_2)}{\delta_m} \geq 1$ for $C_1 \neq C_2$) we obtain (cf. the definition of F_m)

$$|\varphi(C_1) - \varphi(C_2)| \leq \varepsilon(m) \frac{d_\mu(C_1,C_2)}{\delta_m}.$$

Applying the Kirszbraun-McShane Theorem, $\text{rest}_{\mathcal{C}_m} \varphi$ can be extended to a function ψ on $\mathcal{C}$ with

$$\text{(iv)} \qquad |\psi(C_1) - \psi(C_2)| \leq \varepsilon(m) \frac{d_\mu(C_1,C_2)}{\delta_m} \quad \text{for all } C_1, C_2 \in \mathcal{C}.$$

In addition, w.l.o.g. we may assume that $\sup_{C \in \mathcal{C}} |\psi(C)| \leq M$.

Let us show that $\psi \in K$, i.e.,

for all $j \in \mathbb{N}$ $\;d_\mu(C_1,C_2) < \delta_j/2$ implies $|\psi(C_1) - \psi(C_2)| \leq 3\varepsilon(j)$.

For $j \geq m$, since $\frac{\varepsilon(j)}{\delta_j} \geq \frac{\varepsilon(m)}{\delta_m}$ by (iii), we obtain from (iv)

$$|\psi(C_1) - \psi(C_2)| \leq \varepsilon(j) \text{ if } d_\mu(C_1,C_2) \leq \delta_j;$$

for $j < m$, given $C_i \in \mathcal{C}$, $i=1,2$, with $d_\mu(C_1,C_2) < \delta_j/2$, choose $C_i' \in \mathcal{C}_m$ such that $d_\mu(C_i,C_i') < \delta_m$, $i=1,2$;

then $d_\mu(C_1',C_2') < 2\delta_m + \delta_j/2 \underset{\text{(iii)}}{\leq} \delta_j$, and so by (iv)

(note that $\text{rest}_{\mathcal{C}_m} \psi = \text{rest}_{\mathcal{C}_m} \varphi$, $\varphi \in F_m$)

$$|\psi(C_1) - \psi(C_2)| \leq |\psi(C_1) - \psi(C_1')| + |\varphi(C_1') - \varphi(C_2')|$$
$$+ |\psi(C_2') - \psi(C_2)| \leq \varepsilon(m) + \varepsilon(j) + \varepsilon(m) \leq 3\varepsilon(j).$$

Thus $\psi \in K$.

Now, we have $\rho(\varphi,\psi) < \gamma$ since for any $C \in \mathcal{C}$ there exists a $C' \in \mathcal{C}_m$ such that $d_\mu(C,C') < \delta_m$, whence (since $\varphi \in F_m$ and by (iv))

$$|\varphi(C) - \psi(C)| \leq |\varphi(C) - \varphi(C')| + |\psi(C') - \psi(C)|$$
$$\leq 2\varepsilon(m) < \gamma.$$

So $F_m \subset K^\gamma$ which concludes the proof of ⓔ. □

We are now in a position to give the

<u>Proof of Theorem B.</u> First assume (a) and (b). Then, by Proposition B_1, we can view $\mathbb{G}_\mu \equiv (G_\mu(C))_{C\in\mathcal{C}}$ as a random element in $(S,\mathcal{B}_b(S,\rho))$ with $\mathcal{L}\{\mathbb{G}_\mu\}(S_o) = 1$, $S_o \equiv \mathcal{U}^b(\mathcal{C},d_\mu)$ being ρ-closed and separable; furthermore, as mentioned at the end of the proof of Proposition B_1, we have

(1) $$\beta_n \xrightarrow[\text{f.d.}]{\mathcal{L}} \mathbb{G}_\mu .$$

Now, by Proposition B_2, $(\mathcal{L}\{\beta_n\})_{n\in\mathbb{N}}$ is δ-tight w.r.t. S_o, whence it follows from Theorem 11* that

for every subsequence $(\mathcal{L}\{\beta_{n'}\})$ of $(\mathcal{L}\{\beta_n\})$ there exists a further subsequence $(\mathcal{L}\{\beta_{n''}\})$ of $(\mathcal{L}\{\beta_{n'}\})$ and a $\nu = \nu_{(n'),(n'')} \in M_b^1(S)$ with $\nu(S_o) = 1$ such that

(2) $$\mathcal{L}\{\beta_{n''}\} \xrightarrow[b]{} \nu .$$

Since each projection $\pi_{C_1,\ldots,C_k} : S \to \mathbb{R}^k$ is $\mathcal{A},\mathcal{B}_k$-measurable and ρ-continuous and since (M) is assumed, we obtain from (2) by Theorem 3 that

(3) $$\mathcal{L}\{\beta_{n''}\} \circ \pi^{-1}_{C_1,\ldots,C_k} \xrightarrow[b]{} \nu \circ \pi^{-1}_{C_1,\ldots,C_k}$$

for each $C_1,\ldots,C_k \in \mathcal{C}$.

Together with (1) this implies that ν and $\mathcal{L}\{\mathbb{G}_\mu\}$ must have the same fidis; thus $\nu = \mathcal{L}\{\mathbb{G}_\mu\}$ on $\mathcal{B}_b(S,\rho)$ (cf. Lemma 21) and therefore

$$\mathcal{L}\{\beta_n\} \xrightarrow[b]{} \mathcal{L}\{\mathbb{G}_\mu\}, \text{ i.e. } \beta_n \xrightarrow{\mathcal{L}_b} \mathbb{G}_\mu .$$

Conversely if $\mathcal{C}$ is a μ-Donsker class, then (a) holds (cf. Proposition 3.4 in R.M. Dudley (1967)). So it remains to prove (b) (where it suffices to prove the assertion there by taking $\eta = \varepsilon$).

Now, by Theorem 12 there exists a sequence $\hat{\beta}_n$, $n \in \mathbb{N}$, of random elements in $(S,\mathcal{A})$ and a random element $\hat{\mathbb{G}}_\mu$ in $(S,\mathcal{B}_b(S,\rho))$, all defined on an appropriate p-space $(\hat{\Omega},\hat{\mathcal{F}},\hat{\mathbb{P}})$, such that

④ $L\{\hat{\beta}_n\} = L\{\beta_n\}$ (on A) for all $n \in \mathbb{N}$, $L\{\hat{\mathbb{G}}_\mu\} = L\{\mathbb{G}_\mu\}$ (on $\mathcal{B}_b(S,\rho)$)

and

⑤ $\rho(\hat{\beta}_n(\hat{\omega}), \hat{\mathbb{G}}_\mu(\hat{\omega})) = \sup_{C \in \mathcal{C}} |\hat{\beta}_n(C,\hat{\omega}) - \hat{G}_\mu(C,\hat{\omega})| \to 0$ as $n \to \infty$

for all $\hat{\omega} \in \hat{\Omega}_o \in \hat{F}$ with $\hat{\mathbb{P}}(\hat{\Omega}_o) = 1$.

Since $L\{\hat{\mathbb{G}}_\mu\}(S_o) \underset{④}{=} L\{\mathbb{G}_\mu\}(S_o) = 1$, we may assume w.l.o.g. that $\hat{\mathbb{G}}_\mu(\hat{\omega}) \in S_o$ for all $\hat{\omega} \in \hat{\Omega}$, whence for any $\varepsilon>0$ there exists a $\delta = \delta(\varepsilon) > 0$ such that

⑥ $\hat{\mathbb{P}}(w_{\hat{\mathbb{G}}_\mu}(\delta) > \varepsilon/2) < \varepsilon/2$.

(Note that $\{\hat{\omega} \in \hat{\Omega}: w_{\hat{\mathbb{G}}_\mu(\hat{\omega})}(\delta) > \varepsilon/2\} \in \hat{F}$ if, as just assumed, $\hat{\mathbb{G}}_\mu(\hat{\omega}) \in S_o$ for all $\hat{\omega} \in \hat{\Omega}$.)

Now, since S_o is separable, take a sequence $\{\varphi_m: m \in \mathbb{N}\}$ dense in $S_o \cap \complement B_{\delta,\varepsilon/2}$ (with $B_{\delta,\varepsilon/2} := \{\varphi \in S: \exists C_1,C_2 \in \mathcal{C}$ s.t. $d_\mu(C_1,C_2) < \delta$ and $|\varphi(C_1)-\varphi(C_2)| > \varepsilon/2\}$).

Let

$T_o := \bigcup_{m\in\mathbb{N}} B_\rho(\varphi_m,\varepsilon/4)$ ($B_\rho(\varphi_m,\varepsilon/4)$ denoting the open ρ-ball with center φ_m and radius $\varepsilon/4$);

then $T_o \in \mathcal{B}_b(S,\rho)$ whence, by (M), $\{\beta_n \notin T_o\} \in F$ as well as $\{\hat{\beta}_n \notin T_o\} \in \hat{F}$ for each n.

Furthermore we have

$$T_o \cap B_{\delta,\varepsilon} = \emptyset:$$

in fact, $\varphi \in T_o$ implies that $\rho(\varphi_m,\varphi) < \varepsilon/4$ for some $m \in \mathbb{N}$, and since $\varphi_m \in \complement B_{\delta,\varepsilon/2}$, we have for any $C_1,C_2 \in \mathcal{C}$

$$\text{either } d_\mu(C_1,C_2) \geq \delta \quad \text{or} \quad |\varphi_m(C_1) - \varphi_m(C_2)| \leq \varepsilon/2,$$

implying in the latter case that $|\varphi(C_1) - \varphi(C_2)| \leq |\varphi(C_1) - \varphi_m(C_1)| + |\varphi_m(C_1) - \varphi_m(C_2)| + |\varphi_m(C_2) - \varphi(C_2)| < \varepsilon$, whence $\varphi \in \complement B_{\delta,\varepsilon}$.

We thus obtain for each $n \in \mathbb{N}$

$$\mathbb{P}^*(w_{\beta_n}(\delta) > \varepsilon) = \mathbb{P}^*(\beta_n \in B_{\delta,\varepsilon}) \leq \mathbb{P}^*(\beta_n \notin T_o) = \mathbb{P}(\beta_n \notin T_o) = \hat{\mathbb{P}}(\hat{\beta}_n \notin T_o),$$

and so it remains to show

⑦ $\hat{\mathbb{P}}(\hat{\beta}_n \notin T_o) < \varepsilon$ for n sufficiently large.

This will follow now easily from ⑤ together with ⑥:

At first ⑤ implies that there exists an $n_o = n_o(\varepsilon) \in \mathbb{N}$ such that

⑧ $\hat{\mathbb{P}}^*(\rho(\hat{\beta}_n, \hat{\mathbb{G}}_\mu) > \varepsilon/8) < \varepsilon/2$ for all $n \geq n_o$.

Next, if $\hat{\beta}_n(\hat{\omega}) \notin T_o$ then $\rho(\varphi_m, \hat{\beta}_n(\hat{\omega})) \geq \varepsilon/4$ for all $m \in \mathbb{N}$, whence

either $\rho(\hat{\beta}_n(\hat{\omega}), \hat{\mathbb{G}}_\mu(\hat{\omega})) > \varepsilon/8$

or $\rho(\varphi_m, \hat{\mathbb{G}}_\mu(\hat{\omega})) \geq \varepsilon/8$ for all $m \in \mathbb{N}$

(note that $\rho(\hat{\beta}_n(\hat{\omega}), \hat{\mathbb{G}}_\mu(\hat{\omega})) \leq \varepsilon/8$ implies $\rho(\varphi_m, \hat{\mathbb{G}}_\mu(\hat{\omega})) \geq \rho(\varphi_m, \hat{\beta}_n(\hat{\omega})) - \rho(\hat{\beta}_n(\hat{\omega}), \hat{\mathbb{G}}_\mu(\hat{\omega})) \geq \varepsilon/4 - \varepsilon/8 = \varepsilon/8$ for all $m \in \mathbb{N}$).

But since $\rho(\varphi_m, \hat{\mathbb{G}}_\mu(\hat{\omega})) \geq \varepsilon/8$ for all $m \in \mathbb{N}$ implies $w_{\hat{\mathbb{G}}_\mu(\hat{\omega})}(\delta) > \varepsilon/2$

(note that $w_{\hat{\mathbb{G}}_\mu(\hat{\omega})}(\delta) \leq \varepsilon/2$ would imply $\hat{\mathbb{G}}_\mu(\hat{\omega}) \in H := S_o \cap \complement B_{\delta,\varepsilon/2}$, whence

$\rho(\varphi_m, \hat{\mathbb{G}}_\mu(\hat{\omega})) < \varepsilon/8$ for some $m \in \mathbb{N}$ since $\{\varphi_m : m \in \mathbb{N}\}$ is dense in H),

it follows from ⑥ together with ⑧ that ⑦ holds true.

This concludes the proof of Theorem B. □

After having taken great care in proving the fundamental characterization theorem for μ-Donsker classes[1)], we can confine ourselves now to giving Dudley's

Proof of Theorem A.

In view of (E_1) and (54) we have $N(\varepsilon, C, \mu) < \infty$ for each $\varepsilon > 0$, i.e. C is totally bounded for d_μ, and therefore by Theorem B it suffices to prove

---

1) By the way, if instead of Theorem 12 the Portmanteau theorem (cf. (b) there) is used, the last part of the proof becomes much simpler.

(+): (E_1) implies that for any $0 < \varepsilon < 1$ there exists a $\delta_o = \delta_o(\varepsilon)$, $0 < \delta_o < 1$, and there exists an $n_o = n_o(\varepsilon,\delta_o)$ such that for each $n > n_o$

$$\mathbb{P}^*(\omega_{\beta_n}(\delta_o) > \varepsilon) < \varepsilon.$$

Let $0 < \varepsilon < 1$ be arbitrary but fixed and $N_I(x) \equiv N_I(x,\mathcal{C},\mu)$.
Suppose that δ_k, k=0,1,2,... is a sequence of nonnegative real numbers tending to zero (δ_k will be specified below).
According to the definition of $N_I(\delta_k,\mathcal{C},\mu)$ take sets

$$A_{k1},\ldots,A_{km(k)} \in \mathcal{B}, \quad m(k) := N_I(\delta_k),$$

such that for each $C \in \mathcal{C}$ and k=0,1,2,... there exist

$$i(k) = i(k,C) \text{ and } j(k) = j(k,C),\ i(k),j(k) \in \{1,\ldots,m(k)\},$$

with $A_{ki(k)} \subset C \subset A_{kj(k)}$ and $\mu(A_{kj(k)} \setminus A_{ki(k)}) < \delta_k$.

Since $\{\omega_{\beta_n}(\delta_o) > \varepsilon\} = \{\sup[|\beta_n(C) - \beta_n(D)|: C,D \in \mathcal{C},\ \mu(C \Delta D) < \delta_o] > \varepsilon\}$

$$\subset \{\sup_{C\in\mathcal{C}} |\beta_n(C) - \beta_n(A_{oj(o,C)})| > \varepsilon/2\}$$

$$\cup \{\sup [|\beta_n(A_{or}) - \beta_n(A_{os})|,\ \mu(A_{or} \Delta A_{os}) < 3\delta_o,\ r,s \in \{1,\ldots,m(0)\}] > \varepsilon/2\}$$

$$= E_1(\varepsilon,\delta_o,n) \cup E_2(\varepsilon,\delta_o,n), \text{ say},$$

it suffices to show that $\mathbb{P}^*(E_i(\varepsilon,\delta_o,n)) < \varepsilon/2$, i=1,2, for an appropriate $\delta_o = \delta_o(\varepsilon)$ and n sufficiently large.

STEP ①: Let us consider first E_2 replacing (in view of STEP ② below) ε by $\varepsilon/2$, i.e. we will show that

$P_2 := \mathbb{P}^*(E_2(\varepsilon/2,\delta_o,n)) = \mathbb{P}(E_2(\varepsilon/2,\delta_o,n)) < \varepsilon/4$ for a proper choice of $\delta_o = \delta_o(\varepsilon)$ and n sufficiently large.

Applying Lemma 4 (i) of Section 1 we get

$$P_2 \leq 2\,[m(0)]^2 \exp\left(- \frac{\varepsilon^2/16}{6\delta_o + \frac{4}{3} n^{-1/2} \frac{\varepsilon}{4}}\right) \leq 2\,[m(0)]^2 \exp\left(- \frac{\varepsilon^2}{192\,\delta_o}\right)$$

for $n > n_o := \varepsilon^2/(256\,\delta_o^2)$;

now, as to m(0), it follows from (E_1) together with $N_I(x) \uparrow$ as $x \downarrow 0$ that

$x\log N_I(x) \to 0$ as $x \to 0$, whence there is a $\gamma = \gamma(\varepsilon) > 0$ such that

(1) $$N_I(x) \leqq \exp(\varepsilon^2/(800\,x)) \quad \text{for all } 0 < x \leqq \gamma.$$

Thus, for $\delta_o \leqq \gamma$ and $n > n_o$, $P_2 \leqq 2\exp\left(\frac{\varepsilon^2}{400\delta_o} - \frac{\varepsilon^2}{192\delta_o}\right) = 2\exp\left(-\frac{\varepsilon^2}{1600\delta_o}\right)$.

But since

(2) $$\exp\left(-\frac{\varepsilon^2}{1600\alpha}\right) < \varepsilon/8 \text{ for } \alpha \text{ small enough,}$$

we obtain for $\delta_o \leqq \min(\gamma,\alpha)$ that $P_2 < \varepsilon/4$ for all $n > n_o$.

STEP (2): To cope with the other event E_1 a certain chaining argument will be used: for this we note first that the entropy condition (E_1) is equivalent to

$$\int_0^1 y^{-1/2}(\log N_I(y))^{1/2}\,dy < \infty \text{ and to } \sum_{i\in\mathbb{N}} (2^{-i}\log N_I(2^{-i}))^{1/2} < \infty;$$

therefore, there exists a $u = u(\varepsilon)$ so that

(3) $$\sum_{i\geqq u} (2^{-i}\log N_I(2^{-i}))^{1/2} < \varepsilon/96 \quad \text{and}$$

(4) $$\sum_{\ell\geqq 0} \exp(-2^{\ell+u}\,\varepsilon^2/(9000(\ell+1)^4)) < \varepsilon/32.$$

Now, let $\delta_o = \delta_o(\varepsilon) := 2^{-r}$ with $r \geqq u$ and r large enough so that also $\delta_o \leqq \min(\gamma,\alpha)$ (cf. STEP (1)).

For $k=1,2,\ldots$ let $\delta_k := \delta_o \cdot 2^{-k} = 2^{-(r+k)}$ and $b_k := (2^{-k}\log m(k))^{1/2}$,

i.e. $b_k\delta_o^{1/2} = (2^{-(r+k)}\log N_I(2^{-(r+k)}))^{1/2}$ so that by (3) we have

(3*) $$\sum_{k\geqq 0} b_k\delta_o^{1/2} < \varepsilon/96.$$

Next, let $B_k = B_k(C) := A_{kj(k,C)} \setminus A_{k+1,j(k+1,C)}$ and

$D_k = D_k(C) := A_{k+1,j(k+1,C)} \setminus A_{kj(k,C)}$; then $\mu(B_k) < \delta_k$

and $\mu(D_k) < \delta_{k+1} < \delta_k$ (cf. STEP (1)).

As in STEP (1) we choose $n_o := \varepsilon^2/(256\,\delta_o^2)$. (Note that $\delta_o \leqq \alpha \underset{(2)}{<} \varepsilon^2/1600$, so that $n_o > 10.000/\varepsilon^2 \to \infty$ as $\varepsilon \to 0$.)

Then, for each $n > n_o$ there is a unique $k = k(n)$ such that

(5) $$1/2 < 8\delta_k n^{1/2}/\varepsilon \leq 1.$$

Now, for each $n > n_o$ and each $C \in \mathcal{C}$ we obtain (with $k = k(n)$, $i(k) = i(k,C)$ and $j(k) = j(k,C)$)

(6) $$\beta_n(A_{ki(k)}) - \varepsilon/8 \leq \beta_n(A_{ki(k)}) - \delta_k n^{1/2} \leq \beta_n(C) \leq \beta_n(A_{kj(k)}) + \varepsilon/8.$$

Also

(7) $$|\beta_n(A_{kj(k)}) - \beta_n(A_{oj(o)})| \leq \sum_{0\leq \ell<k} |\beta_n(A_{\ell j(\ell)}) - \beta_n(A_{\ell+1,j(\ell+1)})| \leq \sum_{0\leq \ell<k} [|\beta_n(B_\ell)| + |\beta_n(D_\ell)|].$$

Let $\mathcal{S}_\ell$ be the collection of sets $B = A_{\ell j} \setminus A_{\ell+1,m}$ or $A_{\ell+1,m} \setminus A_{\ell j}$ with $j \in \{1,\ldots,m(\ell)\}$ and $m \in \{1,\ldots,m(\ell+1)\}$, respectively, and so that $\mu(B) < \delta_\ell$. Then, for each $C \in \mathcal{C}$, $B_\ell(C)$ and $D_\ell(C) \in \mathcal{S}_\ell$.

The number of sets in $\mathcal{S}_\ell$ is bounded by

(8) $$|\mathcal{S}_\ell| \leq 2m(\ell)m(\ell+1).$$

For later use, note that (by the definition of b_ℓ)

$$m(\ell) = \exp(2^\ell b_\ell^2).$$

Let $d_\ell := \max((\ell+1)^{-2}\varepsilon/32,\ 6\, b_{\ell+1}\delta_o^{1/2})$; then by (3*)

(9) $$\sum_{\ell\geq 0} d_\ell < \varepsilon/8.$$

For each $\ell \leq k = k(n)$, $n > n_o$, we have $n^{1/2}\delta_\ell \geq n^{1/2}\delta_k \underset{(5)}{>} \varepsilon/16$; thus by (9)

$$d_\ell \leq 2n^{1/2}\delta_\ell.$$

Now, by Lemma 4 (ii) of Section 1 we obtain for each $B \in \mathcal{S}_\ell$

$$P_{\ell nB} := \mathbb{P}(|\beta_n(B)| > d_\ell) \leq 2\exp\left(-\frac{d_\ell^2}{2\mu(B)(1-\mu(B))+d_\ell n^{-1/2}}\right).$$

Thus, since $\mu(B) < \delta_\ell$ and $d_\ell n^{-1/2} \leq 2\delta_\ell$, we have

(10) $$P_{\ell nB} \leq 2\ \exp\ (-\ \frac{d_\ell^2}{4\delta_\ell}\ ).$$

Let $M_\ell := 4\ m(\ell)m(\ell+1) \leq 4[m(\ell+1)]^2 = 4\ \exp(2^{\ell+2}b_{\ell+1}^2)$.

Then, using (8) and (10) we obtain

$$P_{\ell n} := \mathbb{P}(|\beta_n(B)| > d_\ell \text{ for some } B \in \mathcal{S}_\ell)$$

$$\leq M_\ell\ \exp(-\ \frac{d_\ell^2}{4\delta_\ell}\ ) = M_\ell\ \exp(-\ \frac{2^\ell d_\ell^2}{4\delta_o}\ )$$

$$\leq 4\ \exp[2^\ell(4\ b_{\ell+1}^2 - d_\ell^2/(4\ \delta_o))].$$

Now, by definition of d_ℓ, $4\ b_{\ell+1}^2 \leq d_\ell^2/(8\ \delta_o)$ and

$-2^{-\ell+r}\ d_\ell^2/8 \leq -2^{-\ell+r}\ \varepsilon^2/(8\cdot(32)^2(\ell+1)^4)$ and so

$$P_{\ell n} \leq 4\ \exp(-2^\ell d_\ell^2/(8\ \delta_o)) \leq 4\ \exp(-2^{\ell+r}\ \varepsilon^2/(8\cdot(32)^2(\ell+1)^4))$$

$$\leq 4\ \exp(-2^{\ell+r}\ \varepsilon^2/(9000(\ell+1)^4)).$$

Thus, by (4), for each $k = k(n)$ $(n > n_o)$ we have

(11) $$\sum_{0\leq\ell<k} P_{\ell n} < 4\cdot\varepsilon/32 = \varepsilon/8.$$

Next, again for $k = k(n)$, $n > n_o$, let

$V_n := \sup\{|\beta_n(A_{kj}) - \beta_n(A_{ki})|: A_{ki} \subset A_{kj},\ \mu(A_{kj}\setminus A_{ki}) < \delta_k,\ i,j=1,\dots,m(k)\}$

and $Q_n := \mathbb{P}(V_n > \varepsilon/8)$.

Then by Lemma 4 (i) of Section 1 and (5) (according to which $\frac{4}{3}\ n^{-1/2}\ \frac{\varepsilon}{8} \leq 3\ \delta_k$)

$$Q_n \leq [m(k)]^2\cdot 2\ \exp(-\ \frac{\varepsilon^2/64}{2\delta_k + \frac{4}{3}\ n^{-1/2}\ \frac{\varepsilon}{8}}\ ) \leq [m(k)]^2\cdot 2\ \exp(-\ \frac{\varepsilon^2}{64\cdot 5\ \delta_k}\ )$$

$$= \exp(2\cdot 2^k b_k^2)\cdot 2\ \exp(-\ \frac{\varepsilon^2 2^k}{320\ \delta_o}\ ) = 2\ \exp[2^k(2b_k^2 - \frac{\varepsilon^2 2^r}{320}\ )].$$

Now, for $s := k+r$,

$$2b_k^2 = 2^{1-k}\log m(k) = 2^{1-k}\log N_I(\delta_k) = 2^{1-k}\log N_I(2^{-(k+r)})$$

$$= 2^{r+1-s} \log N_I(2^{-s}) \underset{①}{\leq} 2^{r+1-s} \frac{\varepsilon^2}{800\ 2^{-s}} = 2^r \frac{\varepsilon^2}{400}.$$

Thus $Q_n \leqq 2 \exp [2^{k+r} (\frac{\varepsilon^2}{400} - \frac{\varepsilon^2}{320})] = 2 \exp (-2^{k+r} \frac{\varepsilon^2}{1600})$

$$\leqq 2 \exp(- \frac{\varepsilon^2}{1600\alpha}) \underset{②}{<} \varepsilon/4.$$

Now, if $V_n \leqq \varepsilon/8$ then by ⑥ $|\beta_n(C) - \beta_n(A_{kj(k,C)})| \leqq \varepsilon/4$ for all $C \in \mathcal{C}$, and therefore

$$E_1 \equiv E_1(\varepsilon,\delta_o,n) := \{\sup_{C \in \mathcal{C}} |\beta_n(C) - \beta_n(A_{oj(o,C)})| > \varepsilon/2\}$$

$$= (E_1 \cap \{V_n > \varepsilon/8\}) \cup (E_1 \cap \{V_n \leqq \varepsilon/8\}) \subset \{V_n > \varepsilon/8\} \cup W_n$$

with $W_n := \{\sup_{C \in \mathcal{C}} |\beta_n(A_{kj(k,C)}) - \beta_n(A_{oj(o,C)})| > \varepsilon/4\}$.

Now $W_n \underset{⑦}{\subset} W'_n := \{\sup_{C \in \mathcal{C}}[\sum_{0 \leq \ell < k} |\beta_n(B_\ell(C))|] > \varepsilon/8\} \cup \{\sup_{C \in \mathcal{C}}[\sum_{0 \leq \ell < k} |\beta_n(D_\ell(C))|] > \varepsilon/8\}$,

where according to ⑨ (note that $B_\ell(C), D_\ell(C) \in \mathcal{S}_\ell$)

$$\mathbb{P}(W'_n) \leqq \sum_{0 \leq \ell < k} P_{\ell n} + \sum_{0 \leq \ell < k} P_{\ell n} \underset{⑪}{<} \varepsilon/4;$$

thus, together with $\mathbb{P}(V_n > \varepsilon/8) = Q_n < \varepsilon/4$ it follows that

$$\mathbb{P}^*(E_1(\varepsilon,\delta_o,n)) < \varepsilon/2 \text{ for } n > n_o.$$

This proves (+) and concludes the proof of Theorem A. □

(57) REMARK. The above proof shows that the two conditions (a) and (b) of Theorem B are implied by (E_1) without imposing (M). I.S. Borisov (1981) has shown that (E_1) cannot be weakened, being necessary in case $\mathcal{C}$ is the collection of all subsets of a countable set X, where (E_1) is equivalent to $\sum_{x \in X} (\mu(\{x\}))^{1/2} < \infty$; cf. also M. Durst and R.M. Dudley (1980).

(58) EXAMPLE. As an illustration of the applicability of Theorem A we will show that in $(X,\mathcal{B}) = (\mathbb{R}^k,\mathcal{B}_k)$, $k \geqq 1$, the class $\mathcal{C} = \mathcal{J}_k$ of all lower left orthants is a μ-Donsker class for any p-measure μ on $\mathcal{B}_k$ (proved by M.D. Donsker (1952) for k = 1 and by R.M. Dudley (1966) for k ≧ 1).

As remarked in (48) (a), condition (M) holds true for $\mathcal{J}_k$; so, by Theorem A, we must show that (E_1) is fulfilled:

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

$$-\infty =: t_o < t_1 < \dots < t_{m-1} < t_m := \infty \text{ of } \mathbb{R}, \text{ where}$$

$$t_{i+1} := \sup \{t \in \mathbb{R}: \mu((t_i,t]) \leq \varepsilon/2\}.$$

Since $\mu((t_i,t_{i+1}]) \geq \varepsilon/2$ and $\mu(\mathbb{R}) = 1$, we have $m - 1 \leq 2/\varepsilon$.

Then, taking as A_i's in the definition of $N_I(\varepsilon,\mathcal{J}_1,\mu)$ all sets of the form $\emptyset$, $(-\infty,t_1),(-\infty,t_1],(-\infty,t_2),(-\infty,t_2],\dots,(-\infty,t_{m-1}),(-\infty,t_{m-1}]$, $\mathbb{R}$ we obtain

$$\min \{n \in \mathbb{N}: \exists A_1,\dots,A_n \in \mathcal{B} \text{ s.t. for all } C \in \mathcal{J}_1 \text{ there exist } i,j \text{ with}$$

$$A_i \subset C \subset A_j \text{ and } \mu(A_j \setminus A_i) < \varepsilon\}$$

$$\leq 2(m-1) + 2 = 2m \leq 4/\varepsilon + 2 \leq 6/\varepsilon.$$

This implies that $\log N_I(\varepsilon^2,\mathcal{J}_1,\mu) \leq \log 6/\varepsilon^2$ showing that (E_1) is fulfilled for $k = 1$.

b) For $k > 1$ the result is an immediate consequence of a) and the inequality (59) of the following lemma (formulated in greater generality as needed in the present case).

<u>LEMMA.</u> Let $(X,\mathcal{B})$ be a measurable space and let μ be a probability measure on the product σ-algebra $\bigotimes_1^k \mathcal{B}$ in X^k, $k \geq 1$, with marginal laws $\pi_i\mu$ on $\mathcal{B}$, $i=1,\dots,k$. Let $\mathcal{C}_i \subset \mathcal{B}$, $i=1,\dots,k$, be given classes of sets and

$$\mathcal{C} := \{ \mathop{\times}_{i=k}^{k} C_i: C_i \in \mathcal{C}_i,\ i=1,\dots,k\}.$$

Then

$$N_I(\varepsilon,\mathcal{C},\mu) \leq \prod_{i=1}^{k} N_I(\varepsilon/k,\mathcal{C}_i,\pi_i\mu). \tag{59}$$

<u>Proof.</u> We may and do assume that $n_i := N_I(\varepsilon/k,\mathcal{C}_i,\pi_i\mu) < \infty$ for each $i=1,\dots,k$. Then there exist $A_{i1},\dots,A_{in_i} \in \mathcal{B}$ such that for any $C_i \in \mathcal{C}_i$ there exist $r_i,s_i \in \{1,\dots,n_i\}$ with

$$A_{ir_i} \subset C_i \subset A_{is_i} \text{ and } \pi_i \mu(A_{is_i} \setminus A_{ir_i}) < \varepsilon/k,$$

$i=1,\dots,k$. This implies that

$$\mathop{\times}_{i=1}^{k} A_{ir_i} \subset \mathop{\times}_{i=1}^{k} C_i \subset \mathop{\times}_{i=1}^{k} A_{is_i} \quad \text{and} \quad \mu\Big(\mathop{\times}_{i=1}^{k} A_{is_i} \setminus \mathop{\times}_{i=1}^{k} A_{ir_i}\Big)$$

$$\leq \sum_{i=1}^{k} \mu(B_i) \text{ (with } B_i := X \times \dots \times X \times (A_{is_i} \setminus A_{ir_i}) \times X \times \dots \times X)$$

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

Since there are at most $n_1 \cdot n_2 \cdot \dots \cdot n_k$ approximating sets of the form $\mathop{\times}_{i=1}^{k} A_{it_i} \in \bigotimes_1^k \mathcal{B}$, (59) follows. □

SOME REMARKS ON OTHER MEASURABILITY ASSUMPTIONS AND FURTHER RESULTS:

Instead of (M) Dudley (1978) used the following measurability assumption (M_o) (again w.r.t. the canonical model $(\Omega,\mathcal{F},\mathbb{P}) = (X^{\mathbb{N}}, \mathcal{B}_{\mathbb{N}}, \times_{\mathbb{N}} \mu)$):

(M_o): $\beta_n: \Omega \to S \equiv D_o(\mathcal{C},\mu)$ is $\tilde{\mathcal{F}},\mathcal{B}_b(S,\rho)$-measurable,

where $\tilde{\mathcal{F}}$ denotes the measure-theoretic completion of $\mathcal{F}$ w.r.t. $\mathbb{P} = \times_{\mathbb{N}} \mu$.

Imposing (M), it follows that β_n is $\mathcal{F},\mathcal{B}_b(S,\rho)$-measurable, whence

(M) implies (M_o).

On the other hand, replacing $\mathcal{A} = \sigma(\{\pi_C: C \in \mathcal{C}\})$ by

$$\mathcal{A}_o := \sigma(\{\pi_C: C \in \mathcal{C};\ \rho(\cdot,\varphi): \varphi \in S\})$$

and imposing (M_o) instead of (M), it follows that

β_n is $\tilde{\mathcal{F}},\mathcal{A}_o$-measurable,

where $\mathcal{B}_b(S,\rho) \subset \mathcal{A}_o \subset \mathcal{B}(S,\rho)$ (cf. (47)), which means that also under (M_o) one meets the basic model of Section 3.

Thus, Theorem A and Theorem B hold as well (with the same proof) if (M) is replaced by (M_o).

Besides (M_o) Dudley (1978) introduced a second measurability assumption

(M_1) (called a $\mu\in$ Suslin property for $\mathcal{C}$),

stronger that (M_o), which turned out to be verifiable in cases of interest where (M) or (SE) fails to hold (cf. (48) (b)).
As shown in Gaenssler (1983), based on Theorem A (with (M) replaced by (M_o)) one obtains a functional central limit theorem for empirical $\mathcal{C}$-processes indexed by classes $\mathcal{C}$ allowing a finite-dimensional parametrization in the sense of the following theorem:

<u>THEOREM C.</u> Let X be a locally compact, separable metric space, $\mathcal{B} = \mathcal{B}(X)$ be the σ-algebra of Borel sets in X, and let K be a compact subset of $\mathbb{R}^\ell$, $\ell \geq 1$. Suppose that

$$f: X \times K \to \mathbb{R}$$

is a function satisfying the following conditions (i) - (iii) ((iii) with respect to a given probability measure μ on $\mathcal{B}$):

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

(ii) $f.(x): K \to \mathbb{R}$ is "uniformly Lipschitz", i.e.,

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

(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 $\mathcal{C} \subset \mathcal{B}$ be defined by $\mathcal{C} := \{\{f_z \geq 0\}: z \in K\}$.
Then $\mathcal{C}$ is a μ-Donsker class; furthermore, (M_1) (and therefore also (M_o)) is satisfied for $\mathcal{C}$ and μ.

<u>(60) EXAMPLES.</u> (a): Let $(X,\mathcal{B},\mu) = ([0,1]^k,[0,1]^k \cap \mathbb{B}_k,\lambda_k)$, $k \geq 1$, λ_k being the k-dimensional Lebesgue measure on $[0,1]^k \cap \mathbb{B}_k$, and let $\mathcal{C} \subset \mathcal{B}$ be the class of all closed Euclidean balls in $[0,1]^k$. Then $\mathcal{C}$ is a λ_k-Donsker class and (M_1) is satisfied for $\mathcal{C}$ and λ_k.

In fact, take

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

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

$$f(x,z) := e(x,\complement B^o(y,r)) - e(x,B^c(y,r)),\ z = (y,r) \in K$$
$$(= r - e(x,y))$$

where $B^o(y,r) := \{x \in \mathbb{R}^k: e(x,y) < r\}$.

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

(b) (cf. (48)(b)): consider the same p-space $(X,\mathcal{B},\mu)$ as in (a) and let

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

C being a fixed closed and convex subset of $X = [0,1]^k$, $k \geq 1$, (cf. R. Pyke (1979)). As in (a) let $f(x,z) := e(x,\complement C_z^o) - e(x,C_z)$, $x,z \in [0,1]^k$, with $C_z := C + z$ and C_z^o denoting the interior of C_z.

Then $\mathcal{C}$ is a λ_k-Donsker class and (M_1) is satisfied for $\mathcal{C}$ and λ_k. This follows again from Theorem C; for this we have to verify the conditions (i) - (iii) there and also that

$$\mathcal{C} = \{\{f_z \geq 0\}: z \in [0,1]^k\}, \text{ i.e., that}$$

(+) $$C_z = \{f_z \geq 0\} \text{ for each } z \in [0,1]^k.$$

ad (+): $x \in C_z$ implies that $e(x,C_z) = 0$ whence $f_z(x) = e(x,\complement C_z^o) \geq 0$; on the other hand, if $x \in \complement C_z$ then $e(x,C_z) > 0$, since C_z is closed, and $e(x,\complement C_z^o) = 0$ whence $f_z(x) = -e(x,C_z) < 0$; this shows (+).

ad (i): follows immediately from the fact that for any $\emptyset \neq A \subset X$

$$|e(x_1,A) - e(x_2,A)| \leq e(x_1,x_2) \text{ for each } x_1,x_2 \in X.$$

ad (ii): let $f_z'(x) := e(x,\complement C_z^o)$ and $f_z''(x) := e(x,C_z)$, i.e., $f_\cdot(x) = f_\cdot'(x) - f_\cdot''(x)$ for all $x \in [0,1]^k$. Then it suffices to show that both

$f_\cdot'(x)$ and $f_\cdot''(x)$ are uniformly Lipschitz:

as to $f_\cdot'(x)$ this follows from

$$① \quad \forall x \in [0,1]^k: \ |e(x,\complement C_z^o) - e(x,\complement C_{z'}^o)| \leq e(z,z') \ \forall z,z' \in [0,1]^k.$$

ad ①: we use the following fact which is easy to prove:

(+) For any closed $F \subset [0,1]^k$ and any $x \in F^o$ there exists a $w \in \partial F$ such that $e(x,w) = e(x,\complement F^o)$.

Now, given any $x \in [0,1]^k$ let w.l.o.g. z and z' be such that $x \in C_z^o \cap C_{z'}^o$; applying then (+) for $F = C_z$ and $F = C_{z'}$, respectively, we obtain

$e(x,\complement C_z^o) = e(x,w_{x,z})$ and $e(x,\complement C_{z'}^o) = e(x,w_{x,z'})$ for some $w_{x,z} \in \partial C_z$ and $w_{x,z'} \in \partial C_{z'}$, respectively.

Furthermore, since C_z and $C_{z'}$ are closed,

$$w_{x,z} = c_{x,z} + z \quad \text{and} \quad w_{x,z'} = c_{x,z'} + z'$$

for some $c_{x,z} \in C$ and $c_{x,z'} \in C$, respectively, and

$$(++) \quad e(x,c_{x,z} + z) \leq e(x,c_{x,z'} + z) \text{ and } e(x,c_{x,z'} + z') \leq e(x,c_{x,z} + z'), \text{ respectively.}$$

Thus

$$e(x,\complement C_z^o) - e(x,\complement C_{z'}^o) = e(x,c_{x,z} + z) - e(x,c_{x,z'} + z')$$

$$\underset{(++)}{\leq} e(x,c_{x,z'} + z) - e(x,c_{x,z'} + z') \leq e(c_{x,z'} + z,c_{x,z'} + z') = e(z,z').$$

This proves ①.

That also $f_\cdot''$ is uniformly Lipschitz follows from

$$② \qquad \forall x \in [0,1]^k: \ |e(x,C_z) - e(x,C_{z'})| \leq e(z,z') \ \forall z,z' \in [0,1]^k.$$

ad ②: Given any $x \in [0,1]^k$ and any $\varepsilon>0$ there exists a $c = c(x,\varepsilon) \in C$ such that for all $z,z' \in [0,1]^k$, $e(x,c + z) \leq e(x,C_z) + \varepsilon$ and thus

$$e(x,C_{z'}) \leq e(x,c + z') \leq e(x,c + z) + e(c + z,c + z')$$
$$= e(x,c + z) + e(z,z') \leq e(x,C_z) + \varepsilon + e(z,z') \text{ for any } \varepsilon>0,$$

whence $e(x,C_{z'}) - e(x,C_z) \leq e(z,z')$ yielding ② by symmetry.

Before proving (iii), let us remark that so far we have only used that C is a closed subset of $[0,1]^k$; for proving (iii), in addition, some smoothness of the boundary of C is needed. So we will now use that C is convex.

ad (iii): We must show that

$$\lambda_k(\{f_z \in [-\varepsilon,\varepsilon)\}) = \mathcal{O}(\varepsilon) \text{ uniformly in } z \in K.$$

For this it suffices to prove

③ $$\{f_z \in [-\varepsilon,\varepsilon)\} \subset C_z^\varepsilon \setminus {}_\varepsilon C_z \quad \text{for all } z \in [0,1]^k, \text{ and}$$

④ $$\sup_{z\in[0,1]^k} \lambda_k(C_z^\varepsilon \setminus {}_\varepsilon C_z) \leq c_k\varepsilon \quad \text{for } \varepsilon \downarrow 0 \text{ with some}$$

constant c_k depending only on k.

(Here $A^\varepsilon := \{x\colon e(x,A) \leq \varepsilon\}$ and ${}_\varepsilon A := \{x\colon e(x,\complement A) > \varepsilon\}$.)

ad ③: Suppose that $-\varepsilon \leq f_z(x) < \varepsilon$, where $f_z(x) = e(x,\complement C_z^o) - e(x,C_z)$, $x \in X$, whence

(a) $f_z(x) = -e(x,C_z)$ iff $x \in \complement C_z^o$,

(b) $f_z(x) = e(x,\complement C_z^o)$ iff $x \in C_z$, and

(c) $f_z(x) = 0$ iff $x \in \partial C_z$.

Thus (note that $X = [(\complement C_z^o)\setminus\partial C_z] + (C_z\setminus\partial C_z) + \partial C_z$)

$$\left\{\begin{matrix} -\varepsilon \leq f_z(x) < \varepsilon \\ x \in (\complement C_z^o)\setminus\partial C_z \end{matrix}\right\} \overset{\Leftrightarrow}{\text{(a)}} \left\{\begin{matrix} -\varepsilon < e(x,C_z) \leq \varepsilon \\ x \in (\complement C_z^o)\setminus\partial C_z \end{matrix}\right\} \Rightarrow \quad x \in C_z^\varepsilon \setminus {}_\varepsilon C_z,$$

$$\left\{\begin{matrix} -\varepsilon \leq f_z(x) < \varepsilon \\ x \in C_z\setminus\partial C_z \end{matrix}\right\} \overset{\Leftrightarrow}{\text{(b)}} \left\{\begin{matrix} -\varepsilon \leq e(x,\complement C_z^o) < \varepsilon \\ x \in C_z\setminus\partial C_z \end{matrix}\right\} \Rightarrow \quad x \in C_z^\varepsilon \setminus {}_\varepsilon C_z, \text{ and}$$

$$\left\{\begin{matrix} -\varepsilon \leq f_z(x) < \varepsilon \\ x \in \partial C_z \end{matrix}\right\} \overset{\Leftrightarrow}{\text{(c)}} \left\{\begin{matrix} f_z(x) = 0 \\ x \in \partial C_z \end{matrix}\right\} \Rightarrow \quad x \in C_z^\varepsilon \setminus {}_\varepsilon C_z.$$

This proves ③.

Ad ④: Due to the translation invariance of λ_k ④ is equivalent to

$$\lambda_k(C^\varepsilon \setminus {}_\varepsilon C) \leq c_k\varepsilon \text{ as } \varepsilon \downarrow 0.$$

Now, as shown in Gaenssler (1981) one has

(61) $\sup_{C\in\mathfrak{C}_k} \lambda_k(C^{\varepsilon}\setminus_{\varepsilon}C) \leqq c_k\varepsilon$ as $\varepsilon \downarrow 0$,

where $\mathfrak{C}_k$ denotes the class of all convex Borel sets in $[0,1]^k$, $k \geqq 1$.

This proves the assertion of Example (b). □

(62) ADDITIONAL REMARKS. (a): the above considerations show that the set of all translates of a fixed closed but not necessarily convex set C is a λ_k-Donsker class provided that C has a smooth boundary in the sense that $\lambda_k(C^{\varepsilon}\setminus_{\varepsilon}C) = \mathcal{O}(\varepsilon)$. Based on a result of E.M. Bronštein (1976) it was shown by Dudley (1981a) that for the class $\mathfrak{C}_k^c$ of all closed convex Borel sets in $[0,1]^k$, $k \geqq 2$, the following inequality holds true:

(63) $N_I(\varepsilon,\mathfrak{C}_k^c,\lambda_k) \leqq \exp(M/\varepsilon^{(k-1)/2})$ for $0 < \varepsilon \leqq 1$ and some constant $M < \infty$ depending only on k.

For k = 2 this yields that (E_1) is fulfilled for $\mathfrak{C}_2^c$ and λ_2; in fact,

$$\int_0^1 (\log N_I(x^2,\mathfrak{C}_2^c,\lambda_2))^{1/2}\, dx \leqq \int_0^1 M^{1/2}\, x^{-1/2}\, dx < \infty,$$

implying a result of Bolthausen (1978) according to which $\mathfrak{C}_2^c$ is a λ_2-Donsker class.

But, for $k \geqq 3$, (63) does not yield (E_1) for $\mathfrak{C}_k^c$ and λ_k which is in accordance with a result of Dudley (1979a) showing that $\mathfrak{C}_k^c$ is not a λ_k-Donsker class for $k \geqq 3$.

(b): let us reconsider the example in (60)(b) according to which for any fixed closed and convex set C in $X = [0,1]^k$, $k \geqq 1$,

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

is a λ_k-Donsker class and also (M_1) (and therefore also (M_o)) is satisfied for $\mathcal{C}$ and λ_k. The way we derived this result from Theorem C shows that λ_k can be replaced by any p-measure μ on $[0,1]^k \cap \mathcal{B}_k$ having a bounded density w.r.t.

λ_k, whence, by Theorem 3 (using (M_o)),

$$n^{1/2} \sup_{C\in\mathcal{C}} |\mu_n(C) - \mu(C)| \xrightarrow{\mathcal{L}} \sup_{C\in\mathcal{C}} |G_\mu(C)|$$

implying (note that w.l.o.g. $G_\mu(\cdot,\omega) \in S_o$ and therefore $\sup_{C\in\mathcal{C}} |G_\mu(C,\omega)| < \infty$ for each ω)

$$\sup_{C\in\mathcal{C}} |\mu_n(C) - \mu(C)| = D_n(\mathcal{C},\mu) \xrightarrow{\mathbb{P}} 0$$

being equivalent with

$$D_n(\mathcal{C},\mu) \to 0 \quad \mathbb{P}\text{-a.s.}$$

according to Lemma 6 in Section 1. (Note that the necessary measurability for $D_n(\mathcal{C},\mu)$ is implied by (M_o).)

Thus

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

is also a Glivenko-Cantelli class (compare this with our conjecture at the end of Section 2 stating that $\mathcal{C}$ is <u>not</u> a Vapnik-Chervonenkis class). Of course the above reasoning works in general, i.e. one gets

(64) Any μ-Donsker class $\mathcal{C}$ satisfying (M_o) is also a Glivenko-Cantelli class.

(c): Let $(X,\mathcal{B})$ be the Euclidean space $\mathbb{R}^k$, $k \geq 1$, with its Borel σ-algebra $\mathcal{B} = \mathbb{B}_k$ and let μ be any p-measure on $\mathbb{B}_k$. Let $\mathbb{B}_k$ be the class of all closed Euclidean balls in $\mathbb{R}^k$. As shown at the end of Section 2 $\mathbb{B}_k$ is a Vapnik-Chervonenkis class (VCC); we also know from (48)(a) that (M) holds true for $\mathbb{B}_k$. Furthermore, as pointed out in Gaenssler (1983), also (M_1) is satisfied for $\mathcal{C} = \mathbb{B}_k$ and any μ on $\mathbb{B}_k$. Thus, for any μ, $\mathbb{B}_k$ is a μ-Donsker class according to the following general result of R.M. Dudley ((1978), Theorem 7.1) stated here without proof (cf. also D. Pollard (1981):

<u>THEOREM D.</u> Let $(X,\mathcal{B},\mu)$ be an arbitrary sample space and $\mathcal{C} \subset \mathcal{B}$ be a VCC such that (M_1) is satisfied for $\mathcal{C}$ and μ; then $\mathcal{C}$ is a μ-Donsker class.

(d): A condition like (61) was also basic for the results of R. Pyke (1977 and 1982) on the Haar function construction of Brownian motion indexed by sets and on functional limit theorems for partial-sum processes indexed by sets (1982a). In fact, Pyke considers classes $\mathcal{C}$ of closed sets in $X = [0,1]^k$, $k \geq 1$, fulfilling (besides an entropy condition) the following two conditions:

A1. There is a constant $c > 0$ such that for all $\varepsilon > 0$ and $C \in \mathcal{C}$

$$\lambda_k(C^{\varepsilon} \setminus {}_{\varepsilon}C) \leq c\varepsilon.$$

A2. $\mathcal{C}$ is totally bounded with respect to the Hausdorff metric d_H defined by

$d_H(C,D) := \inf\{\varepsilon > 0: C \subset D^{\varepsilon} \text{ and } D \subset C^{\varepsilon}\}$ for $C,D \in \mathcal{C}$

(and $C^{\varepsilon} := \{x: e(x,C) \leq \varepsilon\}$).

In another very important and original contribution of T.G. Sun and R. Pyke (1982) on weak convergence of empirical processes, a certain index family $\mathcal{C}$ of closed sets in $[0,1]^k$, $k \geq 1$, closely related to one introduced by Dudley (1974) is studied and it is shown in particular that this class fulfills A1.

In contrary to Dudley's (1978) approach to functional central limit theorems for empirical measures (i.e., empirical $\mathcal{C}$-processes) the paper of Sun and Pyke (based on results of Sun's thesis (1977)) involves first the study of a SMOOTHED VERSION of the empirical processes obtained by replacing the unit point masses assigned to each observation by a uniform distribution of equal mass on a small ball (in the sample space $(X,\mathcal{B},\mu) = ([0,1]^k, [0,1]^k \cap \mathcal{B}_k, \lambda_k))$ of radius r centered at the observations (i.e. $\beta_n(C,\omega)$ is replaced by

$\beta_n^r(C,\omega) := n^{-1/2} \sum_{i=1}^{n} \zeta_i^r(C,\omega)$ with $\zeta_i^r(C,\omega) := \lambda_k(C \cap B^c(\xi_i(\omega),r))/\lambda_k(B^c(\xi_i(\omega),r))$ $-\lambda_k(C)$, where $B^c(\xi_i(\omega),r)$ denotes the closed ball of radius r centered at the observation $\xi_i(\omega)$).

This approach has the advantage that the smoothed version has continuous sample paths in the space of all d_H-continuous functions on $\mathcal{C}$. The remaining steps in the Sun-Pyke approach are then to show the uniform (w.r.t. $\mathcal{C}$) closeness of the smoothed and unsmoothed versions and to establish weak sequential compactness which amounts to verify a conditions like (b) in Theorem B on the uniform (w.r.t. n) behaviour of the modulus of continuity.

In this context the following mode of weak convergence is used (cf. R. Pyke and G. Shorack (1968)):

If η_n, $n \in \mathbb{N}$, and η are defined on some p-space $(\Omega,\mathcal{F},\mathbb{P})$ with values in a metric space $S = (S,d)$ (like e.g., $D_o(\mathcal{C},d_H)$), the η_n's and η being not assumed $\mathcal{F},\mathcal{A}$-measurable for some σ-algebra $\mathcal{A}$ in S (with $\mathcal{B}_b(S,d) \subset \mathcal{A} \subset \mathcal{B}(S)$), then η_n is said to converge weakly to η iff $\lim_{n\to\infty} \mathbb{E}(f(\eta_n)) = \mathbb{E}(f(\eta))$ for all $f \in C^b(S)$ which are, in addition, such that each $f(\eta_n)$, $n \in \mathbb{N}$, and $f(\eta)$ is a random variable, i.e., $\mathcal{F},\mathcal{B}$-measurable.

This concludes our remarks on the other measurability assumptions and further results. For other extensions the reader is referred to our concluding remarks at the end of Section 4.

At this place we prefer to present some of the interesting results obtained by G. Shorack (1979).

FUNCTIONAL CENTRAL LIMIT THEOREMS FOR WEIGHTED EMPIRICAL PROCESSES:

This part is concerned with some results on weak convergence of so-called weighted empirical processes supplementing in another way our earlier remarks in Section 2 on the a.s. behaviour of weighted discrepancies and giving at the same time a further illustration of the special results concerning the D[0,1]-case summarized at the end of Section 3. We will follow closely the presentation in Shorack's (1979) paper using some modifications due to W. Schneemeier in a first draft of his Diploma-Thesis, University of Munich, 1981/82.

Let $(\xi_{ni})_{1\le i\le n}$, $n \in \mathbb{N}$, be an array of row-wise independent random variables defined on some p-space $(\Omega,\mathcal{F},\mathbb{P})$ with distribution functions F_{ni}, $1\le i\le n$, $n \in \mathbb{N}$, being concentrated on $[0,1]$ (i.e. $F_{ni}(0) = 0$ and $F_{ni}(1) = 1$ for all $1\le i\le n$, $n \in \mathbb{N}$).

Before introducing some weight functions q as in Section 2, let us start with the consideration of the following form of a WEIGHTED EMPIRICAL PROCESS W_n based on (ξ_{ni}) and on a given array of so-called scores $(c_{ni})_{1\leq i\leq n}$, $n \in \mathbb{N}$:

(65) $$W_n(t) := n^{-1/2} \sum_{i=1}^{n} c_{ni}[1_{[0,t]}(\xi_{ni}) - F_{ni}(t)], \quad t \in [0,1],$$

where the constant scores c_{ni} are assumed to satisfy

(66) $$n^{-1} \sum_{i=1}^{n} c_{ni}^2 = 1 \quad \text{for each } n.$$

Note that for $c_{ni} \equiv 1$ and for ξ_{ni} being uniformly distributed on $[0,1]$ W_n reduces to the uniform empirical process α_n considered at the end of Section 3. In the same way as α_n there, also W_n will be considered as a random element in $(D,\mathcal{B}_b(D,\rho))$ as well as in $(D,\mathcal{B}(D,s))$.

Generalizing Donsker's functional central limit theorem for α_n we are going to give sufficient conditions under which there exists a certain Gaussian stochastic process W being a random element in $(D,\mathcal{B}_b(D,\rho))$ with $\mathcal{L}\{W\}(C) = 1$ ($C \equiv C[0,1]$ being again the space of continuous functions on $[0,1]$) and such that $W_n \xrightarrow{\mathcal{L}_b} W$.

Before proving one of the main results of Shorack (1979), Theorem 1.1, we will mention some basic facts and preliminary results.

<u>(67) REMARKS.</u> (a) It follows from (66) that ν_n, defined by

$$\nu_n(t) := n^{-1} \sum_{i=1}^{n} c_{ni}^2 F_{ni}(t), \quad t \in [0,1],$$

is a distribution function on $[0,1]$ (with $\nu_n(0) = 0$ and $\nu_n(1) = 1$).

(b) For each $0 \leq s, t \leq 1$ we have $\mathbb{E}(W_n(t)) = 0$ and

$$K_n(s,t) := \operatorname{cov}(W_n(s),W_n(t)) = n^{-1} \sum_{i=1}^{n} c_{ni}^2[F_{ni}(s\wedge t) - F_{ni}(s)F_{ni}(t)],$$

whence

(68) $$\mathbb{E}(W_n^2(t)) \leq \nu_n(t) \quad \text{for all} \quad t \in [0,1].$$

In the following, let $F_{ni}(s,t] := F_{ni}(t) - F_{ni}(s)$, $\nu_n(s,t] := \nu_n(t) - \nu_n(s)$, and $W_n(s,t] := W_n(t) - W_n(s)$ for $0 \leq s \leq t \leq 1$; then we have

LEMMA 22. (i) $\mathbb{E}(W_n^2(r,s] \cdot W_n^2(s,t]) \leq 3\nu_n(r,s] \cdot \nu_n(s,t]$, $0 \leq r \leq s \leq t \leq 1$;

(ii) $\mathbb{E}(W_n^4(s,t]) \leq 3\nu_n^2(s,t] + (\max_{1 \leq i \leq n} \frac{c_{ni}^2}{n}) \cdot \nu_n(s,t]$, $0 \leq s \leq t \leq 1$.

Proof (cf. G. Shorack (1979), INEQUALITY 1.1). Writing c_i for $n^{-1/2}c_{ni}$ we have for $0 \leq r \leq s \leq 1$

(a) $W_n(r,s] = \sum_{i=1}^n c_i X_i(r,s)$ with $X_i(r,s) := 1_{(r,s]}(\xi_{ni}) - F_{ni}(r,s]$;

furthermore, for $0 \leq r \leq s \leq t \leq 1$,

(b1) $\mathbb{E}(X_i(r,s)) = 0$

(b2) $\mathbb{E}(X_i^2(r,s)) = F_{ni}(r,s](1 - F_{ni}(r,s]) \leq F_{ni}(r,s]$

(b3) $\mathbb{E}(X_i^4(r,s)) \leq F_{ni}(r,s]$

(b4) $\mathbb{E}(X_i(r,s)X_i(s,t)) = -F_{ni}(r,s] \cdot F_{ni}(s,t]$

(b5) $\mathbb{E}(X_i^2(r,s)\, X_i^2(s,t)) \leq F_{ni}(r,s] \cdot F_{ni}(s,t]$

(b6) If $\{i,j,k,\ell\} \subset \{1,\ldots,n\}$ such that $|\{i,j,k,\ell\}| \geq 3$, then assuming w.l.o.g. that $i \notin \{j,k,\ell\}$, we have (by independence)
$\mathbb{E}(X_i(r,s)X_j(r,s)X_k(s,t)X_\ell(s,t))$
$= \mathbb{E}(X_i(r,s))\,\mathbb{E}(X_j(r,s)X_k(s,t)X_\ell(s,t)) = 0.$

Therefore,

$$\mathbb{E}(W_n^2(r,s]W_n^2(s,t]) = \mathbb{E}((\sum_{i=1}^n c_i X_i(r,s))^2 (\sum_{i=1}^n c_i X_i(s,t))^2)$$

$$= \sum_{i,j,k,\ell=1}^n c_i c_j c_k c_\ell\, \mathbb{E}(X_i(r,s)X_j(r,s)X_k(s,t)X_\ell(s,t))$$

$$\underset{(b6)}{=} \sum_{\substack{i,j,k,\ell=1 \\ |\{i,j,k,\ell\}| \leq 2}}^n c_i c_j c_k c_\ell\, \mathbb{E}(X_i(r,s)X_j(r,s)X_k(s,t)X_\ell(s,t))$$

$$= \sum_{i=1}^n c_i^4\, \mathbb{E}(X_i^2(r,s)X_i^2(s,t)) + \sum_{\substack{i,j,k,\ell=1 \\ i=j\neq k=\ell}}^n c_i^2 c_k^2\, \mathbb{E}(X_i^2(r,s))\,\mathbb{E}(X_k^2(s,t))$$

$$+ \sum_{\substack{i,j,k,\ell=1 \\ i=k\neq j=\ell}}^{n} c_i^2 c_j^2 \, \mathbb{E}(X_i(r,s)X_i(s,t)) \, \mathbb{E}(X_j(r,s)X_j(s,t))$$

$$+ \sum_{\substack{i,j,k,\ell=1 \\ i=\ell\neq k=j}}^{n} c_i^2 c_k^2 \, \mathbb{E}(X_i(r,s)X_i(s,t)) \, \mathbb{E}(X_k(r,s)X_k(s,t))$$

$$\underset{(b5),(b2),(b4)}{\leq} \sum_{i=1}^{n} c_i^4 \, F_{ni}(r,s] \, F_{ni}(s,t] + 3 \sum_{\substack{i,k=1 \\ i\neq k}}^{n} c_i^2 c_k^2 \, F_{ni}(r,s] \, F_{nk}(s,t]$$

$$\leq 3\Big(\sum_{i=1}^{n} c_i^2 \, F_{ni}(r,s]\Big)\Big(\sum_{k=1}^{n} c_k^2 \, F_{nk}(s,t]\Big) = 3\nu_n(r,s]\nu_n(s,t] \text{ proving (i).}$$

As to (ii) we have $\mathbb{E}(W_n^4(s,t]) = \mathbb{E}((\sum_{i=1}^{n} c_i X_i(s,t))^4)$

$$= \sum_{i=1}^{n} c_i^4 \, \mathbb{E}(X_i^4(s,t)) + 3 \sum_{\substack{i,k=1 \\ i\neq k}}^{n} c_i^2 c_k^2 \, \mathbb{E}(X_i^2(s,t)X_k^2(s,t))$$

$$= 3\Big(\sum_{i=1}^{n} c_i^2 \, \mathbb{E}(X_i^2(s,t))\Big)^2 - 3 \sum_{i=1}^{n} c_i^4 (\mathbb{E}(X_i^2(s,t)))^2 + \sum_{i=1}^{n} c_i^4 \, \mathbb{E}(X_i^4(s,t))$$

$$\underset{(b2),(b3)}{\leq} 3\nu_n^2(s,t] + (\max_{1\leq i\leq n} c_i^2)\cdot\nu_n(s,t] \text{ proving (ii). } \square$$

<u>THEOREM 18 (G. Shorack (1979), Theorem 1.1).</u>

(i) If there exists a monotone increasing and continuous function

$G: [0,1] \to \mathbb{R}_+$ for which

either ⓐ $\nu_n(r,s] \leq G(r,s] := G(s) - G(r)$ for all $0 \leq r \leq s \leq 1$ and all $n \in \mathbb{N}$

or ⓑ $\lim_{n\to\infty} \nu_n(t) = G(t)$ for all $t \in [0,1]$,

then $(W_n)_{n\in\mathbb{N}}$ is relatively L-sequentially compact.

(ii) If further

$$\max_{1\leq i\leq n} \frac{c_{ni}^2}{n} \to 0 \quad \text{as} \quad n \to \infty,$$

then any possible limiting process, i.e., any random element W in $(D,\mathcal{B}(D,s)) = (D,\mathcal{B}_b(D,\rho))$ such that $W_{n'} \xrightarrow{L} W$ for some subsequence (n') of $\mathbb{N}$, satisfies $L\{W\}(C) = 1$;

thus (cf. Lemma 18) $(W_n)_{n\in\mathbb{N}}$ is relatively L_b-sequentially compact.

(iii) Suppose the hypotheses of (i) and (ii) hold. Then

there exists a random element W in $(D,\mathcal{B}_b(D,\rho))$ being a mean-zero Gaussian stochastic process with $\mathrm{cov}(W(s),W(t)) = K(s,t)$, $L\{W\}(C) = 1$ and such that $W_n \xrightarrow{L_b} W$ in (D,ρ)

if and only if

$\lim_{n\to\infty} K_n(s,t) = K(s,t)$ for all $0 \leqq s \leqq t \leqq 1$.

Proof. ad (i): We shall apply Theorem 14 in connection with our remarks (41) and (42). In view of this we have to verify the conditions (A), (B'), (C') and (D) for $\xi_n = W_n$ choosing $F_n := \sqrt{3}\,\nu_n$ and $F := \sqrt{3}\,G$.

ad (D): Follows immediately from the above assumptions (a) or (b) according to the choice of F_n and F in the present situation.

ad (C'): Follows from Lemma 22 (i) according to the choice of F_n.

ad (B'): Given any $\varepsilon > 0$ we have to show that

(+) $\limsup_{n\to\infty} \mathbb{P}(|W_n(\delta) - W_n(0)| \geqq \varepsilon) \to 0$ as $\delta \to 0$

and

(++) $\limsup_{n\to\infty} \mathbb{P}(|W_n(1) - W_n(\delta)| \geqq \varepsilon) \to 0$ as $\delta \to 1$.

ad (+): $\mathbb{P}(|W_n(\delta) - W_n(0)| \geqq \varepsilon) \leqq \varepsilon^{-2}\,\mathbb{E}(W_n^2(0,\delta])$

$$\underset{(b1),(b2),L.22}{=} \varepsilon^{-2} n^{-1} \sum_{i=1}^{n} c_{ni}^2 [F_{ni}(0,\delta] - F_{ni}^2(0,\delta]] \leqq \varepsilon^{-2} n^{-1} \sum_{i=1}^{n} c_{ni}^2\, F_{ni}(0,\delta]$$

$= \varepsilon^{-2}\,\nu_n(0,\delta]$, whence by (a) or (b)

$\limsup_{n\to\infty} \mathbb{P}(|W_n(\delta) - W_n(0)| \geqq \varepsilon) \leqq \varepsilon^{-2}(G(\delta) - G(0)) \to 0$ as $\delta \to 0$

since G was assumed to be continuous.

ad (++): Follows in the same way as (+).

ad (A): By (42) (C') implies (C) which together with (D) implies (C") by (41). According to (C"), given any $\eta > 0$ there exist $\delta = \delta(\eta) > 0$ and

$n_o = n_o(\delta(\eta)) \in \mathbb{N}$ such that for all $n \geq n_o$ $\mathbb{P}(\omega''_{W_n}(\delta) \geq 1) \leq \eta/2$.

Now, choose $k \in \mathbb{N}$ and $a > 0$ such that

$$k^{-1} < \delta \quad \text{and} \quad a^{-2} \leq \eta/2;$$

then

$$\mathbb{P}(\bigcup_{i=1}^{k} \{|W_n(\tfrac{i-1}{k}, \tfrac{i}{k}]| \geq a\}) \leq a^{-2} \sum_{i=1}^{k} \mathbb{E}(W_n^2(\tfrac{i-1}{k}, \tfrac{i}{k}])$$

$$\leq a^{-2} \sum_{i=1}^{k} \nu_n(\tfrac{i-1}{k}, \tfrac{i}{k}] \underset{(67)(a)}{\leq} a^{-2} \leq \eta/2.$$

Furthermore, for $n \geq n_o$, let

$$A_n := \bigcap_{i=1}^{k} \{|W_n(\tfrac{i-1}{k}, \tfrac{i}{k}]| < a\} \cap \{\omega''_{W_n}(\delta) < 1\} \cap \{W_n(0) = 0\};$$

then $\mathbb{P}(A_n) \geq 1 - \eta$ and for each $\omega \in A_n$ we have for any $t \in [0,1]$ and $1 \leq i \leq k$ so that $t \in [\frac{i-1}{k},\frac{i}{k})$

$$|W_n(t,\omega) - W_n(\tfrac{i-1}{k},\omega)| \leq |W_n(\tfrac{i}{k},\omega) - W_n(\tfrac{i-1}{k},\omega)|$$

$$+ \min\{|W_n(\tfrac{i}{k},\omega) - W_n(t,\omega)|, |W_n(t,\omega) - W_n(\tfrac{i-1}{k},\omega)|\}$$

$$< a + \omega''_{W_n(\omega)}(k^{-1}) \leq a + \omega''_{W_n(\omega)}(\delta) < a + 1, \text{ whence}$$

$|W_n(t,\omega)| < ka + 1$, and therefore

$\mathbb{P}(\|W_n\| \leq ka + 1) \geq 1 - \eta$ for all $n \geq n_o$ which proves Ⓐ.

This concludes the proof of (i).

<u>ad (ii):</u> Suppose that for some random element W in $(D,\mathcal{B}(D,s)) = (D,\mathcal{B}_b(D,\rho))$

$W_{n'} \xrightarrow{L} W$ for some subsequence (n') of $\mathbb{N}$;

then for any $0 \leq s \leq t \leq 1$ such that

$$s,t \in T_W := \{r \in [0,1]: \pi_r \text{ is } L\{W\}\text{-a.e. s-continuous}\}$$

it follows from Theorem 5.1 in Billingsley (1968) that

$$(+) \qquad |W_{n'}(t) - W_{n'}(s)|^3 \xrightarrow{L} |W(t) - W(s)|^3;$$

on the other hand it follows from Lemma 22 (ii) that

(++) $\quad \mathbb{E}(W^4_{n'}(s,t]) \leqq 3\nu^2_{n'}(s,t] + (\max_{1\leqq i\leqq n'} \frac{c^2_{n'i}}{n'}) \nu_{n'}(s,t] \leqq 4$ for all n' (cf. (66) and (67)(a)).

But (+) together with (++) imply (cf. Gaenssler-Stute (1977), Exercise 1.14.4, p. 114) that

$$\mathbb{E}(|W(s,t]|^3) = \lim_{n'\to\infty} \mathbb{E}(|W_{n'}(s,t]|^3) \underset{(\text{Hölder})}{\leqq} \limsup_{n'\to\infty} (\mathbb{E}(W^4_{n'}(s,t]))^{3/4}$$

$$\underset{(++)\text{ and ⓐ or ⓑ}}{\leqq} (3G^2(s,t])^{3/4} \leqq 3(G(s,t])^{3/2}.$$

Since T_W contains 0 and 1 and is dense in [0,1] (cf. Billingsley (1968)), it follows that

$$\mathbb{E}(|W(t) - W(s)|^3) \leqq 3|G(t) - G(s)|^{3/2}$$

for <u>all</u> $0 \leqq s \leqq t \leqq 1$,

whence by Lemma 19 $\mathbb{P}(W \in C) = 1$.

<u>ad (iii):</u> Assume first that

$$\lim_{n\to\infty} K_n(s,t) = K(s,t) \quad \text{for all } 0 \leqq s \leqq t \leqq 1.$$

We are going to show that for any $\alpha_1,\ldots,\alpha_k \in \mathbb{R}$ and any $t_1,\ldots,t_k \in [0,1]$

(+++) $$\sum_{j=1}^k \alpha_j W_n(t_j) \xrightarrow{L} N(0,V) \text{ with } V := \sum_{r,s=1}^k \alpha_r\alpha_s K(t_r,t_s).$$

<u>ad (+++):</u> If V = 0, then for any $\varepsilon > 0$

$$\mathbb{P}(|\sum_{j=1}^k \alpha_j W_n(t_j)| \geqq \varepsilon) \leqq \varepsilon^{-2}\,\mathbb{E}(|\sum_{j=1}^k \alpha_j W_n(t_j)|^2)$$

$$\underset{(\text{cf.}(67)(\text{b}))}{=} \varepsilon^{-2} \sum_{r,s=1}^k \alpha_r\alpha_s K_n(t_r,t_s) \longrightarrow \varepsilon^{-2} V = 0$$ as $n \to \infty$ which proves (+++) in case V = 0.

If V > 0, consider

$$V^{-1/2}\sum_{j=1}^k \alpha_j W_n(t_j) = V^{-1/2}\sum_{j=1}^k \alpha_j [n^{-1/2}\sum_{i=1}^n c_{ni}(1_{[0,t_j]}(\xi_{ni}) - F_{ni}(t_j))]$$

$$= \sum_{i=1}^n (nV)^{-1/2} c_{ni}[\sum_{j=1}^k \alpha_j(1_{[0,t_j]}(\xi_{ni}) - F_{ni}(t_j))] =: \sum_{i=1}^n \zeta_{ni}, \text{ say.}$$

Then the ζ_{ni}'s form a triangular array of row-wise independent random variables with $\mathbb{E}(\zeta_{ni}) = 0$ and such that

$$\sum_{i=1}^{n} \mathbb{E}(\zeta_{ni}^2) = \mathbb{E}((\sum_{i=1}^{n} \zeta_{ni})^2) = V^{-1}\,\mathbb{E}((\sum_{j=1}^{k} \alpha_j W_n(t_j))^2)$$

$$= V^{-1} \sum_{r,s=1}^{k} \alpha_r \alpha_s K_n(t_r,t_s) \to 1 \quad \text{as } n \to \infty;$$

furthermore, for any $\delta > 0$ we have

$$\sum_{i=1}^{n} \mathbb{E}(\zeta_{ni}^2\, 1_{\{|\zeta_{ni}| > \delta\}}) \to 0 \quad \text{as} \quad n \to \infty:$$

in fact, given any $\varepsilon > 0$ let $p := 1 + \varepsilon/2$ and $q > 0$ be such that $1/p + 1/q = 1$; then, by Hölder's and Markov's inequality we obtain for any $\delta > 0$

$$\sum_{i=1}^{n} \mathbb{E}(\zeta_{ni}^2\, 1_{\{|\zeta_{ni}| > \delta\}}) \leq \sum_{i=1}^{n} \mathbb{E}(|\zeta_{ni}|^{2+\varepsilon})^{1/p}\, \mathbb{P}(|\zeta_{ni}| > \delta)^{1/q}$$

$$\leq \sum_{i=1}^{n} \mathbb{E}(|\zeta_{ni}|^{2+\varepsilon})^{1/p}\, (\delta^{-(2+\varepsilon)})^{1/q}\, \mathbb{E}(|\zeta_{ni}|^{2+\varepsilon})^{1/q} = \delta^{-2p/q} \sum_{i=1}^{n} \mathbb{E}(|\zeta_{ni}|^{2+\varepsilon})$$

$$\leq \delta^{-2p/q} \sum_{i=1}^{n} \left(\frac{|c_{ni}|}{(Vn)^{1/2}}\right)^{2+\varepsilon} \left(\sum_{j=1}^{k} |\alpha_j|\right)^{2+\varepsilon}$$

$$\leq \delta^{-2p/q}\, V^{-1-\varepsilon/2} \underbrace{\left(\max_{1\leq i\leq n} \frac{|c_{ni}|}{n^{1/2}}\right)^{\varepsilon}}_{\to 0 \text{ as } n\to\infty} \underbrace{\left(\sum_{i=1}^{n} \frac{c_{ni}^2}{n}\right)}_{=1} \left(\sum_{j=1}^{k} |\alpha_j|\right)^{2+\varepsilon} \to 0 \text{ as } n \to \infty.$$

Thus, it follows from the Central Limit Theorem (cf. Gaenssler-Stute (1977), 9.2.9) that

$$\sum_{i=1}^{n} \zeta_{ni} \xrightarrow{\mathcal{L}} N(0,1) \quad \text{which proves (+++).}$$

Next, let $\bar{W}$ be a mean-zero Gaussian process with covariance structure given by K. Then, again for any $\alpha_1,\ldots,\alpha_k \in \mathbb{R}$ and any $t_1,\ldots,t_k \in [0,1]$

$$\mathcal{L}\{\sum_{j=1}^{k} \alpha_j \bar{W}(t_j)\} = N(0,V) \quad \text{with V defined as above.}$$

Therefore, by the Cramér-Wold Device (cf. Gaenssler-Stute (1977), 8.7.6) it follows together with (+++) that

$$W_n \xrightarrow[\text{f.d.}]{L} \overline{W}.$$

Now, by (ii), for any subsequence $(W_{n'})$ of (W_n) there exists a further subsequence $(W_{n''})$ and a random element $W_{(n')(n'')}$ in $(D,\mathcal{B}_b(D,\rho))$ such that $L\{W_{(n')(n'')}\}(C) = 1$ and

$$W_{n''} \xrightarrow{L_b} W_{(n')(n'')} \text{ in } (D,\rho).$$

Applying Theorem 3 and using the fact that each $W_{(n')(n'')}$ is uniquely determined by its fidis, it follows that

$$W_{(n')(n'')} \overset{L}{\equiv}: W \underset{\text{f.d.}}{\overset{L}{=}} \overline{W} \text{ and therefore}$$

$$W_n \xrightarrow{L_b} W \text{ in } (D,\rho).$$

To prove the other direction, suppose that $W_n \xrightarrow{L_b} W$ in (D,ρ) where W is a r.e. in $(D,\mathcal{B}_b(D,\rho))$ and is a mean-zero Gaussian process with covariance structure given by K (and such that $L\{W\}(C) = 1$). Then, by Theorem 3, for any $0 \leqq s \leqq t \leqq 1$

$$W_n(s)\cdot W_n(t) \xrightarrow{L} W(s)\cdot W(t) \quad \text{as} \quad n \to \infty,$$

and $\quad \mathbb{E}(W_n^2(s)\cdot W_n^2(t)) \leqq \mathbb{E}(W_n^4(s))^{1/2}\, \mathbb{E}(W_n^4(t))^{1/2} \leqq 4$

for all $n \in \mathbb{N}$ (cf. (+++) above), whence (by the same reasoning as in the proof of (ii))

$$\mathbb{E}(W(s)\cdot W(t)) = \lim_{n\to\infty} \mathbb{E}(W_n(s)\cdot W_n(t)), \text{ i.e.,}$$

$$\lim_{n\to\infty} K_n(s,t) = K(s,t) \quad \text{for all } 0 \leqq s \leqq t \leqq 1. \quad \square$$

SOME GENERAL REMARKS ON WEAK CONVERGENCE OF RANDOM ELEMENTS IN $D \equiv D[0,1]$ w.r.t. ρ_q-METRICS:

Let q be any weight function belonging to the set

$\mathcal{Q}_2 := \{q\colon [0,1] \to \mathbb{R}$, q continuous, $q(0) = q(1) = 0$ and $q(t) > 0$ for all $t \in (0,1)$, having the following additional properties (i_q) - $(iii_q)\}$:

There exists a $\delta^* = \delta^*(q)$, $0 < \delta^* \leqq 1/2$, such that

(i_q) $q(t)$ and $q(1-t)$ are monotone increasing on $[0,\delta^*]$;

(ii_q) $\frac{q(t)}{\sqrt{t}}$ and $\frac{q(1-t)}{\sqrt{t}}$ are monotone decreasing on $[0,\delta^*]$;

(iii_q) $\int_0^{\delta^*} q^{-2}(u)du < \infty$ and $\int_{1-\delta^*}^1 q^{-2}(u)du < \infty$

(i.e. q^{-1} square integrable).

Here and in the following we make use of the convention $\frac{0}{0} := 0$.

REMARK. Let $q \in \mathcal{Q}_2$; then for any $t \leq \delta^*$

$\frac{\sqrt{t}}{q(t)} = (\frac{t}{q^2(t)})^{1/2} = (\int_0^t q^{-2}(t)du)^{1/2} \leq (\int_0^t q^{-2}(u)du)^{1/2}$, whence by (iii_q) one has

(iv_q) $\frac{\sqrt{t}}{q(t)} \to 0$ as $t \to 0$; similarly, by symmetry,

$\frac{\sqrt{1-t}}{q(t)} \to 0$ as $t \to 1$.

Now, let ξ_n, $n \geq 0$, be random elements in $(D,\mathcal{B}_b(D,\rho))$, defined on a common p-space $(\Omega,\mathcal{F},\mathbb{P})$, such that $\mathbb{P}(\xi_n(0) = 0,\ \xi_n(1) = 0) = 1$ and $\mathbb{P}(\xi_n/q \in D) = 1$ for all $n \geq 0$; in this case we shall assume w.l.o.g. that

$$\frac{\xi_n(\cdot,\omega)}{q(\cdot)} \in D \text{ for each } \omega \in \Omega \text{ and all } n \geq 0.$$

Then the ξ_n/q, $n \geq 0$, are also random elements in $(D,\mathcal{B}_b(D,\rho))$ (cf. (37)).

Let $q \in \mathcal{Q}_2$ and define

$$D_q := \{y = qx\colon x \in D\} \equiv qD, \text{ and } C_q := qC$$

with $C \equiv C[0,1]$, and define the ρ_q-METRIC on D_q by

$\rho_q(y_1,y_2) := \rho(x_1,x_2)$ if $y_i = qx_i \in D_q$, $i=1,2$, where we tacitly assume that $x(0) = x(1) = 0$ whenever $x \in D$ occurs.

Let $\mathcal{B}_b(D_q,\rho_q)$ be the σ-algebra in D_q generated by the open ρ_q-balls and consider the map

$T_1\colon D \to D_q$, defined by $T_1(x) := qx$ for $x \in D$;

then T_1 is $\mathcal{B}_b(D,\rho)$, $\mathcal{B}_b(D_q,\rho_q)$-measurable:

in fact, let $B_{\rho_q}(y,\varepsilon)$ be the open ρ_q-ball with center $y \in D_q$ and radius ε; then

$$T_1^{-1}(B_{\rho_q}(y,\varepsilon)) = \{x \in D: \rho_q(y,qx) < \varepsilon\} = \{x \in D: \rho(y/q,x) < \varepsilon\} \in \mathcal{B}_b(D,\rho).$$

In the same way

$T_2: D_q \to D$, defined by $T_2(y) := y/q$, is $\mathcal{B}_b(D_q,\rho_q)$, $\mathcal{B}_b(D,\rho)$-measurable.

This implies that

$$\xi/q \text{ is a random element in } (D,\mathcal{B}_b(D,\rho)) \text{ iff}$$
$$\xi \text{ is a random element in } (D_q,\mathcal{B}_b(D_q,\rho_q)).$$

Note also (cf. (39)) that $C_q \in \mathcal{B}_b(D_q,\rho_q)$ and that (C_q,ρ_q) is a separable and closed subspace of (D_q,ρ_q).

Furthermore, one has the following

<u>LEMMA 23.</u> In the just described setting, the following two statements are equivalent:

(i) $\xi_n/q \xrightarrow{L_b} \xi_o/q$ in (D,ρ) and $L\{\xi_o/q\}(C) = 1$

(ii) $\xi_n \xrightarrow{L_b} \xi_o$ in (D_q,ρ_q) and $L\{\xi_o\}(C_q) = 1.$

<u>Proof.</u> (i) $\Rightarrow$ (ii): Note first that $L\{\xi_o\}(C_q) = \mathbb{P}(\xi_o \in C_q = qC)$ $= \mathbb{P}(\xi_o/q \in C) = L\{\xi_o/q\}(C)$.

Now, according to (28) (cf. (h') there) it remains to show

(+) $\mathbb{E}(f(\xi_n)) \to \mathbb{E}(f(\xi_o))$ for every $f: D_q \to \mathbb{R}$ which is bounded, uniformly ρ_q-continuous and $\mathcal{B}_b(D_q,\rho_q)$, $\mathcal{B}$-measurable.

So, let $f: D_q \to \mathbb{R}$ be bounded, uniformly ρ_q-continuous and $\mathcal{B}_b(D_q,\rho_q)$, $\mathcal{B}$-measurable, and let $g: D \to \mathbb{R}$ be defined by $g(x) := f(qx)$, $x \in D$; then g is bounded, $\mathcal{B}_b(D,\rho)$, $\mathcal{B}$-measurable (since $g = f \circ T_1$) and uniformly ρ-continuous (since $\rho(x_1,x_2) = \rho_q(qx_1,qx_2)$ and $|g(x_1) - g(x_2)| = |f(qx_1) - f(qx_2)|$ for any $x_1,x_2 \in D$, i.e. $qx_1,qx_2 \in D_q$). Therefore, by (i) and (28) $\mathbb{E}(g(\xi_n/q)) \to \mathbb{E}(g(\xi_o/q))$ which implies (+) since $\mathbb{E}(g(\xi_n/q)) = \mathbb{E}(f(\xi_n))$ for all

$n \geqq 0$.

(ii) ⇒ (i): follows in the same way. □

FUNCTIONAL CENTRAL LIMIT THEOREMS FOR WEIGHTED EMPIRICAL PROCESSES w.r.t. ρ_q-METRICS:

As before let W_n be a weighted empirical process based on an array (ξ_{ni}) of row-wise independent random variables ξ_{ni}, $1 \leqq i \leqq n$, $n \in \mathbb{N}$, defined on some p-space $(\Omega, F, \mathbb{P})$, and on an array (c_{ni}) of given scores (cf. (65)).
We assume again that the distribution functions F_{ni} of the ξ_{ni}'s are concentrated on $[0,1]$; here, in addition, we suppose that

(69) for each $n \in \mathbb{N}$: $n^{-1} \sum_{i=1}^{n} F_{ni}(t) = t$ for all $t \in [0,1]$.

Then, for any $q \in \mathcal{Q}_2$, we have

LEMMA 24. $\mathbb{P}(W_n/q \in D) = 1$ for all n, whence we may and do assume w.l.o.g. that $W_n(\cdot,\omega)/q(\cdot) \in D$ for each $\omega \in \Omega$ and all $n \in \mathbb{N}$.

Proof. According to the definition of W_n, for $\mathbb{P}$-a.a. $\omega \in \Omega$ there exists a $t = t(\omega) \leqq \delta^*$ such that by (69) and (iv_q)

$$\frac{|W_n(t,\omega)|}{q(t)} \leqq n^{-1/2} \sum_{i=1}^{n} |c_{ni}| \frac{F_{ni}(t)}{q(t)} \leqq \left(\max_{1\leqq i\leqq n} |c_{ni}|\right) n^{1/2} \frac{t}{q(t)} \to 0$$

as $t \to 0$; similarly, $\frac{|W_n(t,\omega)|}{q(t)} \to 0$ as $t \to 1$ for $\mathbb{P}$-a.a. ω which implies the assertion (imposing the convention $\frac{0}{0} := 0$). □

Now, for uniformly bounded scores, Shorack (1979) has shown:

THEOREM 19 (Shorack (1979), Theorem 1.2).

Suppose that

(70) $$\sup_{n\in\mathbb{N}} \left(\max_{1\leqq i\leqq n} |c_{ni}|\right) \leqq M < \infty.$$

Then for all $q \in \mathcal{Q}_2$ we have

(i) $(W_n/q)_{n\in\mathbb{N}}$ is relatively L-sequentially compact (in (D,s)).

(ii) Any limiting process, i.e. any random element W in $(D,\mathcal{B}(D,s)) = (D,\mathcal{B}_b(D,\rho))$ such that $W_{n'}/q \xrightarrow{L} W$ for some subsequence (n') of $\mathbb{N}$, satisfies $L\{W\}(C) = 1$, whence, by Lemma 18, $(W_n/q)_{n\in\mathbb{N}}$ is relatively L_b-sequentially compact in (D,ρ) such that for any limiting process W $L\{W\}(C) = 1$,

and therefore, by Lemma 23,

$(W_n)_{n\in\mathbb{N}}$ is relatively L_b-sequentially compact in (D_q,ρ_q) such that for any limiting process W_o $L\{W_o\}(C_q) = 1$.

(iii) There exists a random element W_o in $(D_q,\mathcal{B}_b(D_q,\rho_q))$ being a mean-zero Gaussian stochastic process with

$\mathrm{cov}(W_o(s), W_o(t)) = K_o(s,t)$, $L\{W_o\}(C_q) = 1$ and such that

$W_n \xrightarrow{L_b} W_o$ in (D_q,ρ_q)

if and only if (for $K_n(s,t) = \mathrm{cov}(W_n(s), W_n(t))$)

$\lim_{n\to\infty} K_n(s,t) = K_o(s,t)$ for all $0 \leqq s \leqq t \leqq 1$.

The proof of Theorem 19 (being based on Theorem 18 and Theorem 17) can be carried through along the lines presented in Shorack's (1979) paper with some slight modifications being necessary due to our choice of $\mathcal{Q}_2$; by the way, instead of (15) on p. 171 it suffices to impose (iv_q) and instead of $\mathbb{P}(A_N) \leqq \exp(-1/a_N)$ one shows $\mathbb{P}(A_N) \geqq 1 - 1/a_N$ to get (v) on p. 181. We are not going to give further details here. Instead, since the proof of Theorem 1.2 in Shorack (1979) seems not suited to carry over to give a proof of his Theorem 1, 3 as mentioned there on p. 182 (note that in the case of not uniformly bounded scores it is not possible to estimate

$$(\max_{1\leqq i\leqq n} \frac{c_{ni}^2}{n})(t - s) \text{ by } M^2(t - s)^2 \text{ for } t - s > n^{-1},$$ which was essentially used

to get (c) on p. 179) we want to present here a completely different proof of the following result:

THEOREM 20 (Shorack (1979), Theorem 1.3).

If all ξ_{ni} are uniformly distributed on [0,1] and if instead of (70)

(71) $$\max_{1\leq i\leq n} \frac{c_{ni}^2}{n} \to 0 \quad \text{as} \quad n \to \infty,$$

then, for any $q \in \mathcal{Q}_2$, $W_n \xrightarrow{\mathcal{L}_b} B^o$ in (D_q, ρ_q) as $n \to \infty$ and $\mathcal{L}\{B^o\}(C_q) = 1$, where B^o denotes the Brownian bridge.

The proof of Theorem 20 is based on the following lemmata which may be of independent interest.

The first lemma is concerned with a martingale property of the weighted empirical process W_n based on ξ_{ni} which are uniformly distributed on [0,1].

LEMMA 25. Let $n \in \mathbb{N}$ be arbitrary but fixed and write ξ_i and c_i instead of ξ_{ni} and c_{ni}, respectively. Suppose that $F_{ni}(t) = t$ for all $t \in [0,1]$. Then for any $(c_1,\ldots,c_n) \in \mathbb{R}^n$

$$\left(\frac{n^{1/2}W_n(t)}{1-t}\right)_{0\leq t<1} = \left(\frac{\sum_{i=1}^n c_i[1_{[0,t]}(\xi_i) - t]}{1-t}\right)_{0\leq t<1} \text{ is a}$$

martingale w.r.t. $\mathcal{F}_t := \sigma(\{\frac{n^{1/2}W_n(s)}{1-s} : s \leq t\})$, $0 \leq t < 1$.

Proof. We use the following auxiliary result which is easy to prove:

(72) Let $(\zeta_t)_{0\leq t<T}$ and $(\eta_t)_{0\leq t<T}$, $T \leq \infty$, be martingales w.r.t.

$(\mathcal{F}_t := \sigma(\{\zeta_s : s \leq t\}))_{0\leq t<T}$ and $(\mathcal{G}_t := \sigma(\{\eta_s : s \leq t\}))_{0\leq t<T}$,

respectively. Assume that $(\zeta_t)_{0\leq t<T}$ and $(\eta_t)_{0\leq t<T}$ are independent.

Then $(\zeta_t + \eta_t)_{0\leq t<T}$ is a martingale w.r.t.

$(\mathcal{H}_t := \sigma(\{\zeta_s, \eta_s : s \leq t\}))_{0\leq t<T}$ and therefore also w.r.t.

$(\sigma(\{\zeta_s + \eta_s : s \leq t\}))_{0\leq t<T}$.

Now, given any $(c_1,\ldots,c_{n+1}) \in \mathbb{R}^{n+1}$, put

$$\zeta_t = \zeta_t^{(n)} := \sum_{i=1}^n c_i[1_{[0,t]}(\xi_i) - t]/(1-t),$$

$$\eta_t := \zeta_t^{(n+1)} - \zeta_t^{(n)},\ 0 \leq t < 1,$$

and apply (72) to get the assertion of Lemma 25 by induction on n; for n = 1 cf. Lemma 3 in Section 1 (choosing $(X,\mathcal{B},\mu) = ([0,1],[0,1] \cap \mathbb{B},\lambda)$, λ = Lebesgue measure, and $\mathcal{C} := \{[0,t]: 0 \leq t < 1\}$ there).

LEMMA 26. Suppose that $F_{ni}(t) = t$ for all $t \in [0,1]$; let $q \in \mathcal{Q}_2$ and $\delta^* = \delta^*(q)$, $0 < \delta^* \leq 1/2$, be as in the definition of $\mathcal{Q}_2$.
Then, for any $n \in \mathbb{N}$, each $\varepsilon > 0$, and any $\delta \leq \delta^*$, one has

(i) $$\mathbb{P}(\sup_{t\in[0,\delta]} |\frac{W_n(t)}{q(t)}| > \varepsilon) \leq \varepsilon^{-2} \cdot 8 \int_0^{\delta} q^{-2}(u)du, \text{ and}$$

(ii) $$\mathbb{P}(\sup_{t\in[1-\delta,1]} |\frac{W_n(t)}{q(t)}| > \varepsilon) \leq \varepsilon^{-2} \cdot 8 \int_{1-\delta}^{1} q^{-2}(u)du.$$

Proof. ad (i): let $n \in \mathbb{N}$ be arbitrary but fixed and for each $k \in \mathbb{N}$ and $i \in \{0,\ldots,2^k\}$ let

$$t_{ki} := i \cdot \delta/2^k.$$

Then, due to the path properties of W_n/q (cf. Lemma 24), it suffices to show that for each $\varepsilon > 0$ and any $k \in \mathbb{N}$ one has

(+) $$\mathbb{P}(\sup_{1\leq i\leq 2^k} |\frac{W_n(t_{ki})}{q(t_{ki})}| > \varepsilon) \leq \varepsilon^{-2} \cdot 8 \int_0^{\delta} q^{-2}(u)du.$$

For later use note that

(++) $$1 \leq (1 - t_{ki})^{-2} \leq 4 \quad \text{for each } k \text{ and } i.$$

ad (+): let $\varepsilon > 0$ and $k \in \mathbb{N}$ be arbitrary but fixed.

Since, by Lemma 25, for any fixed $n \in \mathbb{N}$

$(\frac{W_n(t)}{1-t})_{0\leq t<1}$ is a martingale, we can apply Chow's inequality (cf. Gaenssler-Stute (1977), (6.6.2)) on the submartingale

$((\frac{W_n(t)}{1-t})^2)_{0\leq t<1}$ to obtain

$$\varepsilon^2\, \mathbb{P}(\max_{1\leq i\leq 2^k} |\frac{W_n(t_{ki})}{q(t_{ki})}| > \varepsilon) \underset{(++)}{\leq} \varepsilon^2\, \mathbb{P}(\max_{1\leq i\leq 2^k} [q^{-2}(t_{ki})(\frac{W_n(t_{ki})}{1-t_{ki}})^2] > \varepsilon^2)$$

$$\leq q^{-2}(t_{k1})\ \mathbb{E}\Big(\Big(\frac{W_n(t_{k1})}{1-t_{k1}}\Big)^2\Big) + \sum_{i=2}^{2^k} q^{-2}(t_{ki})\ \mathbb{E}\Big(\Big(\frac{W_n(t_{ki})}{1-t_{ki}}\Big)^2 - \Big(\frac{W_n(t_{k,i-1})}{1-t_{k,i-1}}\Big)^2\Big).$$

Now, since $\mathbb{E}(W_n^2(t)) = n^{-1}\sum_{i=1}^{n} c_{ni}^2[F_{ni}(t) - F_{ni}^2(t)] = n^{-1}\sum_{i=1}^{n} c_{ni}^2(t - t^2)$

$\underset{(66)}{=}$ $t(1 - t) \leq t$, we get by the second inequality in (++)

$$q^{-2}(t_{k1})\ \mathbb{E}\Big(\Big(\frac{W_n(t_{k1})}{1-t_{k1}}\Big)^2\Big) \leq 4\ q^{-2}(t_{k1})t_{k1} \underset{(i_q)}{\leq} 4 \int_0^{t_{k1}} q^{-2}(u)du \leq 4 \int_0^{\delta} q^{-2}(u)du;$$

on the other hand

$$\sum_{i=2}^{2^k} q^{-2}(t_{ki})\ \mathbb{E}\Big(\Big(\frac{W_n(t_{ki})}{1-t_{ki}}\Big)^2 - \Big(\frac{W_n(t_{k,i-1})}{1-t_{k,i-1}}\Big)^2\Big)$$

$$= \sum_{i=2}^{2^k} q^{-2}(t_{ki}) \Big[\frac{t_{ki}}{1-t_{ki}} - \frac{t_{k,i-1}}{1-t_{k,i-1}}\Big] \underset{(\text{cf.}(++))}{\leq} 4 \sum_{i=2}^{2^k} q^{-2}(t_{ki})(t_{ki} - t_{k,i-1})$$

$$\leq 4 \int_{t_{k1}}^{\delta} q^{-2}(u)du \leq 4 \int_0^{\delta} q^{-2}(u)du.$$

So, in summary we have

$$\mathbb{P}\Big(\max_{1\leq i\leq 2^k}\Big|\frac{W_n(t_{ki})}{q(t_{ki})}\Big| > \varepsilon\Big) \leq \varepsilon^{-2}\cdot 8 \int_0^{\delta} q^{-2}(u)du$$

which proves (+).

ad (ii): by symmetry this follows in the same way. □

LEMMA 27. For any $q \in \mathcal{Q}_2$ we have $\mathbb{P}(B^o/q \in C) = 1$, where B^o denotes the Brownian bridge and C is the space of all continuous functions on [0,1].

Proof. We have to show that B^o/q is $\mathbb{P}$-a.s. continuous at 0 (and also at 1 which is shown similarly). For this, according to Lemma 19, it suffices to show that for some constants $a > 1$, $b > 0$ and some continuous function $F: [0,1] \to \mathbb{R}$

(+) $\qquad \mathbb{E}(|\overline{B}^o(t) - \overline{B}^o(s)|^b) \leq |F(t) - F(s)|^a$

for all $0 \leq s \leq t \leq 1$, where (using again the convention $\frac{0}{0} := 0$)

$$\overline{B}^o(t) := \frac{B^o(t)}{q(t)}\, 1_{[0,\delta^*]}(t) + \frac{B^o(\delta^*)}{q(\delta^*)}\, 1_{(\delta^*,1]}(t),\ t \in [0,1]$$

($\delta^* = \delta^*(q)$ as in the definition of $\mathcal{Q}_2$).

ad (+): since, for any $0 \leqq s \leqq t \leqq 1$, $B^o(t) - B^o(s)$ is normally distributed with mean zero and variance $(t - s)(1 - (t - s))$, we have

(a) $$\mathbb{E}\left(\frac{(B^o(t) - B^o(s))^4}{q^4(t)} \right) = \frac{3(t - s)^2(1 - (t - s))^2}{q^4(t)}, \quad 0 \leqq s \leqq t \leqq 1.$$

On the other hand, for any $0 < s \leqq t \leqq \delta^*$, we have

$$\mathbb{E}((B^o(s))^4[\frac{1}{q(s)} - \frac{1}{q(t)}]^4) = [\frac{1}{q(s)} - \frac{1}{q(t)}]^4 \cdot 3\, s^2(1 - s)^2,$$

where

$$s^2[\frac{1}{q(s)} - \frac{1}{q(t)}]^4 = [\frac{\sqrt{s}}{q(s)}(1 - \frac{q(s)}{q(t)})]^4$$

$$\leqq [\frac{\sqrt{t}}{q(t)}(1 - \frac{\sqrt{s}}{t})]^4 \quad (\text{since } \frac{\sqrt{t}}{q(t)} \uparrow \text{ on } [0,\delta^*] \text{ by } (ii_q))$$

$$= [\frac{\sqrt{t} - \sqrt{s}}{q(t)}]^4 \leqq (\frac{t - s}{q^2(t)})^2, \text{ whence}$$

(b) $$\mathbb{E}((B^o(s))^4[\frac{1}{q(s)} - \frac{1}{q(t)}]^4) \leqq \frac{3(t-s)^2(1-s)^2}{q^4(t)}, \quad 0 < s \leqq t \leqq \delta^*.$$

Now, it follows from (a) and (b) that for $0 < s \leqq t \leqq \delta^*$

$$\mathbb{E}(|\frac{B^o(t)}{q(t)} - \frac{B^o(s)}{q(s)}|^4) = \mathbb{E}((\frac{B^o(t) - B^o(s)}{q(t)} + B^o(s)[\frac{1}{q(t)} - \frac{1}{q(s)}])^4)$$

$$\leqq 2^4[\mathbb{E}(\frac{(B^o(t) - B^o(s))^4}{q^4(t)}) + \mathbb{E}((B^o(s))^4[\frac{1}{q(t)} - \frac{1}{q(s)}]^4)]$$

$$\leqq 2^4[\frac{3(t-s)^2(1-(t-s))^2 + 3(t-s)^2(1-s)^2}{q^4(t)}]$$

$$\leqq 2^4 \cdot 6(q^{-2}(t)(t - s))^2 \underset{(i_q)}{\leqq} (4\sqrt{6}\int_s^t q^{-2}(u)du)^2.$$

Thus, taking $F(t) := 4\sqrt{6}\int_0^t q^{-2}(u)du$, we get (+) (with $b = 4$, $a = 2$) for all $0 < s \leqq t \leqq 1$.

It remains to consider the case $s = 0$ and $0 < t \leqq \delta^*$; but, by (a),

$$\mathbb{E}(|\frac{B^o(t)}{q(t)}|^4) \leqq 3\, q^{-4}(t)t^2 \underset{(i_q)}{\leqq} (\sqrt{3}\int_0^t q^{-2}(u)du)^2$$

$$\leqq F^2(t) = |F(t) - F(0)|^2. \text{ This proves (+).} \quad \square$$

Proof of Theorem 20. In the setting of Theorem 18 and its preceding remarks (67) we have in the present situation (where $F_{ni}(t) = t$ for all $t \in [0,1]$) that (cf. (66))

$$\nu_n(t) = t \quad (=: G(t)) \text{ for all } t \in [0,1]$$

and $$K_n(s,t) \equiv K(s,t) = s\wedge t - st = \operatorname{cov}(B^o(s), B^o(t)).$$

Therefore, by Theorem 18 (iii), we have

(a) $$W_n \xrightarrow{\mathcal{L}_b} B^o \quad \text{in } (D,\rho).$$

Furthermore, by Lemma 24 and Lemma 27, for any $q \in \mathcal{Q}_2$ we have

(b) $$\mathbb{P}(W_n/q \in D) = 1 \quad \text{for all } n \in \mathbb{N} \text{ and } \mathbb{P}(B^o/q \in C) = 1.$$

We are thus in a situation where our general remarks on weak convergence of random elements in $D = D[0,1]$ w.r.t. ρ_q-metrics can be applied. So, by Lemma 23, it remains to show

(c) $$W_n/q \xrightarrow{\mathcal{L}_b} B^o/q \quad \text{in } (D,\rho).$$

ad (c): let $q_o :\equiv 1$ and for $m \geq 1$ let

$$q_m := q\, 1_{(\frac{1}{m}, 1-\frac{1}{m})} + q(\tfrac{1}{m})\, 1_{[0,\frac{1}{m}]} + q(1-\tfrac{1}{m})\, 1_{[1-\frac{1}{m},1]}.$$

Since q_m is continuous and $q_m > 0$ on $[0,1]$

(d) $$W_n/q_m \xrightarrow{\mathcal{L}_b} B^o/q_m \quad \text{in } (D,\rho) \text{ as } n \to \infty.$$

Now, according to (28), (c) holds if we show that

$$\lim_{n\to\infty} \mathbb{E}(f(W_n/q)) = \mathbb{E}(f(B^o/q)) \quad \text{for all } f \in U_b^b(D,\rho).$$

But, again by (28) and (d), we have for each m that

$$\lim_{n\to\infty} \mathbb{E}(f(W_n/q_m)) = \mathbb{E}(f(B^o/q_m)) \text{ for all } f \in U_b^b(D,\rho);$$

furthermore, by Lebesgue's theorem

$$\lim_{m\to\infty} \mathbb{E}(f(B^o/q_m)) = \mathbb{E}(f(B^o/q)) \text{ for all } f \in U_b^b(D,\rho),$$

since, by Lemma 27, $\mathbb{P}(B^o/q \in C) = 1$ and therefore $\lim_{m\to\infty} \rho(B^o/q_m, B^o/q) = 0$ $\mathbb{P}$-a.s.

Thus, given any $f \in U_b^b(D,\rho)$, choosing for each $m \in \mathbb{N}$ $k_m > k_{m-1}$ (with $k_o := 0$) such that

$$|\mathbb{E}(f(W_n/q_m)) - \mathbb{E}(f(B^o/q_m))| \leq \frac{1}{m} \text{ for } n \geq k_m,$$

and putting for each $n \in \mathbb{N}$ $i_n := m$ if $n \in \{k_m,\ldots,k_{m+1}-1\}$, we obtain

$$\lim_{n\to\infty} \mathbb{E}(f(W_n/q_{i_n})) = \mathbb{E}(f(B^o/q)).$$

So, it remains to show

(e) $$\lim_{n\to\infty} |\mathbb{E}(f(W_n/q_{i_n})) - \mathbb{E}(f(W_n/q))| = 0.$$

For this, let $\varepsilon > 0$ be arbitrary and $\delta = \delta(\varepsilon) > 0$ be such that $\rho(x,y) \leq \delta$ implies $|f(x) - f(y)| \leq \varepsilon$. (Note also that $||f|| := \sup_{x\in D} |f(x)| < \infty$.) Then for n sufficiently large (i.e. such that $i_n^{-1} \leq \delta^*$)

$$|\mathbb{E}(f(W_n/q_{i_n})) - \mathbb{E}(f(W_n/q))| \leq \mathbb{E}(|f(W_n/q_{i_n}) - f(W_n/q)|)$$

$$\leq \varepsilon + 2\,||f|| \cdot \mathbb{P}(\rho(W_n/q_{i_n}, W_n/q) > \delta)$$

$$\leq \varepsilon + 2\,||f|| \cdot [\mathbb{P}(\sup_{t\in[0,\frac{1}{i_n}]} |\frac{W_n(t)}{q(t)}| > \delta/2) + \mathbb{P}(\sup_{t\in[1-\frac{1}{i_n},1]} |\frac{W_n(t)}{q(t)}| > \delta/2)],$$

whence it follows from Lemma 26 that

$$|\mathbb{E}(f(W_n/q_{i_n})) - \mathbb{E}(f(W_n/q))|$$

$$\leq \varepsilon + 2\,||f|| \cdot [(\delta/2)^{-2} \cdot 8(\int_0^{\frac{1}{i_n}} q^{-2}(u)du + \int_{1-\frac{1}{i_n}}^{1} q^{-2}(u)du)],$$

and therefore, by (iii_q),

$$\limsup_{n\to\infty} |\mathbb{E}(f(W_n/q_{i_n})) - \mathbb{E}(f(W_n/q))| \leq \varepsilon .$$

Since $\varepsilon > 0$ was arbitrary, this proves (e) and therefore (c) is shown. □

<u>(73) REMARKS.</u> (a) W. Schneemeier (1982) has given an example showing that Theorem 19 fails to hold if the uniform boundedness condition (70) on the scores is replaced by the condition

$$\max_{1\leq i\leq n} \frac{c_{ni}^2}{n} \to 0 \quad \text{as} \quad n \to \infty$$

which was imposed in Theorem 18. Thus, the assumption in Theorem 20 of ξ_{ni} being uniformly distributed on $[0,1]$ cannot be weakened to the assumption that for every $n \in \mathbb{N}$ $n^{-1} \sum_{i=1}^{n} F_{ni}(t) = t$ for all $t \in [0,1]$ (cf. (69)) without strengthening the condition on the scores.

(b) As to the L_b-statements in Theorem 18 and Theorem 19 it is possible by making use of Theorem 11 a) (or Theorem 11^*) to modify the given proofs such that they operate totally within our theory of L_b-convergence.

Note, for example, that along the same lines as in the proof of Proposition B_2 together with an application of Theorem 11^* one obtains (within the theory of L_b-convergence) that any sequence $(\xi_n)_{n\in\mathbb{N}}$ of random elements in $(D,\mathcal{B}_b(D,\rho))$ which satisfies the following two conditions

(i) $\lim_{\delta\downarrow 0} \limsup_{n\to\infty} \mathbb{P}(w_{\xi_n}(\delta) > \varepsilon) = 0$ for each $\varepsilon > 0$

and

(ii) $\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}(\|\xi_n\| > M) = 0$

is relatively L_b-sequentially compact and such that for any limiting random element ξ_o one has $L\{\xi_o\}(C) = 1$.

Further results in this direction will be contained in a forthcoming paper by P. Gaenssler, E. Haeusler and W. Schneemeier (1983).

CONCLUDING REMARKS ON FURTHER RESULTS FOR EMPIRICAL PROCESSES INDEXED BY CLASSES OF SETS OR CLASSES OF FUNCTIONS:

(a) FUNCTIONAL LAWS OF THE ITERATED LOGARITHM

(cf. Gaenssler-Stute (1979), Section 1.3, concerning results for the uniform empirical process α_n).

One of the main theorems in Kuelbs and Dudley (1980) states that for any p-space $(X,\mathcal{B},\mu)$ the following holds true:

(74) If (M_1) is satisfied for a class $\mathcal{C} \subset \mathcal{B}$ and μ, and if $\mathcal{C}$ is a μ-Donsker class, then $\mathcal{C}$ is a STRASSEN LOG·LOG CLASS for μ, i.e., with probability one the set

$$\{ \left(\frac{\beta_n(C)}{(2\log\log n)^{1/2}} \right)_{C\in\mathcal{C}} : n \geq n_o \}$$ is relatively compact

(w.r.t. the supremum metric ρ in $D_o(\mathcal{C},\mu)$) with limit set

$B_{\mathcal{C}} := \{\varphi: C \mapsto \int_C f d\mu,\ C \in \mathcal{C};\ f \in B\}$, where

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

(Note that $B_{\mathcal{C}} \subset U^b(\mathcal{C},d_\mu) \subset D_o(\mathcal{C},\mu)$.)

Now, as pointed out in Gaenssler (1983), since for $(X,\mathcal{B},\mu) = (\mathbb{R}^k,\mathbb{B}_k,\mu)$, $k \geq 1$, the class $\mathcal{C} = \mathcal{J}_k$ of all lower left orthants satisfies (M_1) and is a μ-Donsker class for any μ by (58), one obtains by (74) the results of Finkelstein (1971) and Richter (1974), namely

(75) $\mathcal{J}_k$ is a Strassen log log class for every p-measure μ on $\mathbb{B}_k$, $k \geq 1$.

That the same holds true for $\mathcal{C} = \mathbb{B}_k$ (the class of all closed Euclidean balls in $\mathbb{R}^k$, $k \geq 1$) is a consequence of our remarks preceding Theorem D and of Corollary 2.4 in Kuelbs and Dudley (1980) according to which one has

(76) If (M_1) is satisfied for μ and a Vapnik-Chervonenkis class $\mathcal{C}$, then $\mathcal{C}$ is a Strassen log log class for μ.

(b) DONSKER CLASSES OF FUNCTIONS.

Let α_n be the uniform empirical process (cf. the end of Section 3) and let q be some weight function considered above in connection with weak convergence of random elements in $D \equiv D[0,1]$ w.r.t. ρ_q-metrics. For any $q \in \mathcal{Q}_2$ we know from Theorem 20 (with $c_{ni} \equiv 1$) that

$$\alpha_n \xrightarrow{L_b} B^o \quad \text{in} \quad (D_q,\rho_q)$$

or, equivalently by Lemma 23, that

(77) $$\alpha_n/q \xrightarrow{L_b} B^o/q \quad (\text{in } (D,\rho)).$$

Now, from a different point of view, taking for each $t \in [0,1]$ the functions $f_t\colon [0,1] \to \mathbb{R}$ defined by

$$f_t(s) := q^{-1}(t)\, 1_{[0,t]}(s),\ s \in [0,1],$$

α_n/q can be considered as an empirical process indexed by a class of functions; in fact, let

$$\mathfrak{F}_o := \{f_t\colon t \in [0,1]\},$$

then for each $t \in [0,1]$

$$\alpha_n(t)/q(t) = \int_0^1 f_t(s)\, d\alpha_n(s) =: \alpha_n(f_t).$$

Also the limiting process in (77) can be viewed as a mean-zero Gaussian process $\mathbb{G}_\mu$ (μ being here Lebesgue measure on $X = [0,1]$) indexed by $\mathfrak{F}_o$, i.e., $\mathbb{G}_\mu \equiv (G_\mu(f))_{f\in\mathfrak{F}_o}$, with covariance structure

(78) $$\mathrm{cov}(G_\mu(f_{t_1}), G_\mu(f_{t_2})) = \int_0^1 f_{t_1} f_{t_2}\, d\mu - \int_0^1 f_{t_1}\, d\mu \cdot \int_0^1 f_{t_2}\, d\mu;$$

note that $$\mathrm{cov}(q^{-1}(t_1)\, B^o(t_1),\ q^{-1}(t_2)\, B^o(t_2))$$

$$= q^{-1}(t_1)q^{-1}(t_2)[t_1 \wedge t_2 - t_1t_2] = \int_0^1 f_{t_1} f_{t_2}\, d\mu - \int_0^1 f_{t_1}\, d\mu \cdot \int_0^1 f_{t_2}\, d\mu.$$

Hence (77) is equivalent to

(79) $$(\alpha_n(f))_{f\in\mathfrak{F}_o} \xrightarrow{L_b} \mathbb{G}_\mu \equiv (G_\mu(f))_{f\in\mathfrak{F}_o}.$$

This leads to the problem of generalizing Dudley's central limit theory from empirical $\mathcal{C}$-processes to the case of

$$\text{EMPIRICAL } \mathfrak{F}\text{-PROCESSES} \quad \beta_n \equiv (\beta_n(f))_{f\in\mathfrak{F}},$$

defined by

$$\beta_n(f) := n^{1/2}\, (\mu_n(f) - \mu(f)),\ f \in \mathfrak{F},$$

where $\mathfrak{F}$ is a given class of measurable functions defined on an arbitrary sample space $(X,\mathcal{B},\mu)$, and where

$$\mu_n(f) := \int f d\mu_n,\quad \mu(f) := \int f d\mu,\ f \in \mathfrak{F},$$

μ_n being the empirical measure based on i.i.d. ξ_i's with values in X and distribution μ on $\mathcal{B}$.

For uniformly bounded classes of functions such an extension is more or less straightforward, but this does of course not meet the special case mentioned before (note that q^{-1} is approaching ∞ at the endpoints of [0,1]). For possibly unbounded classes $\mathfrak{F}$ the present knowledge is by recent work of R.M. Dudley (1981a), R.M. Dudley (1981b) and D. Pollard (1981a) as follows: let $(X,\mathcal{B},\mu)$ be an arbitrary p-space and $\beta_n \equiv (\beta_n(f))_{f\in\mathfrak{F}}$ be an empirical $\mathfrak{F}$-process with $\mathfrak{F} \subset L^2(X,\mathcal{B},\mu)$. It turns out that there are proper extensions of the spaces $S_o \equiv \mathcal{U}^b(\mathcal{C},d_\mu)$ and $S \equiv D_o(\mathcal{C},\mu)$ considered at the beginning of Section 4 (corresponding to the special situation $\mathfrak{F} = 1_{\mathcal{C}} := \{1_C\colon C \in \mathcal{C}\}$) to the present case with d_μ on $\mathcal{C}$ being replaced by

$$e_\mu(f_1,f_2) := (\int_X (f_1 - f_2)^2 d\mu)^{1/2},\ f_1,f_2 \in \mathfrak{F},$$

$$\text{(or better } \rho_\mu(f_1,f_2) := (\int_X (f_1 - f_2 - \int_X (f_1 - f_2)d\mu)^2 d\mu)^{1/2},\ f_1,f_2 \in \mathfrak{F}),$$

leading to certain spaces $S_o = S_o(\mathfrak{F},e_\mu)$ and $S = S(\mathfrak{F},\mu)$ of functions $\varphi\colon \mathfrak{F} \to \mathbb{R}$ which can be chosen in such a way that under certain conditions on $\mathfrak{F}$ $S_o(\mathfrak{F},e_\mu)$ becomes a separable subspace of $(S(\mathfrak{F},\mu),\rho)$ and such that $(\beta_n(f))_{f\in\mathfrak{F}}$ has all its sample paths in $S(\mathfrak{F},\mu)$; here as before ρ denotes the supremum metric, i.e.,

$$\rho(\varphi_1,\varphi_2) := \sup_{f\in\mathfrak{F}} |\varphi_1(f) - \varphi_2(f)| \quad \text{for } \varphi_1,\varphi_2 \in S(\mathfrak{F},\mu).$$

Now, again under certain measurability assumptions (like (M) or (M_o) imposed in the case of empirical $\mathcal{C}$-processes) the setting of a functional limit theorem for empirical $\mathfrak{F}$-processes $\beta_n \equiv (\beta_n(f))_{f\in\mathfrak{F}}$ in the sense of L_b-convergence for random elements in $(S(\mathfrak{F},\mu),\ \mathcal{B}_b(S(\mathfrak{F},\mu),\rho))$ applies, i.e., one can speak of

(80) $\beta_n \xrightarrow{L_b} \mathbb{G}_\mu$, where $\mathbb{G}_\mu \equiv (G_\mu(f))_{f\in\mathfrak{F}}$ is a mean-zero Gaussian process with covariance structure (cf. (78))

$$\operatorname{cov}(G_\mu(f_1),\ G_\mu(f_2)) = \int_X f_1 f_2 d\mu - \int_X f_1 d\mu \cdot \int_X f_2 d\mu.$$

If (80) holds true, $\mathcal{F}$ is called a μ-DONSKER CLASS OF FUNCTIONS.

Generalizing Theorem A from classes of sets to classes of functions the main result of R.M. Dudley (1981a) is:

(81) Suppose (M_o) (which means here that $\beta_n: X^{\mathbb{N}} \to S(\mathcal{F},\mu)$ is measurable from the measure-theoretic completion of $(X^{\mathbb{N}}, \mathcal{B}_{\mathbb{N}}, \underset{\mathbb{N}}{\times}\, \mu)$ to $(S(\mathcal{F},\mu), \mathcal{B}_b(S(\mathcal{F},\mu),\rho)))$ and suppose that $F := \sup\{|f|: f \in \mathcal{F}\} \in L^p(X,\mathcal{B},\mu)$ for some $p > 2$; assume further that for γ with $0 < \gamma < 1 - 2/p$ and some $M < \infty$

(E_2) $N_I(\varepsilon,\mathcal{F},\mu) \leq \exp(M\,\varepsilon^{-\gamma})$ for ε small enough.

Then $\mathcal{F}$ is a μ-Donsker class.

In this connection, $N_I(\varepsilon,\mathcal{F},\mu)$, a natural extension of $N_I(\varepsilon,\mathcal{C},\mu)$, is defined as the smallest $m \in \mathbb{N}$ such that for some $f_1,\dots,f_m \in L^2(X,\mathcal{B},\mu)$ (not necessarily in $\mathcal{F}$), for every $f \in \mathcal{F}$ there exist $j,k \leq m$ with $f_j(x) \leq f(x) \leq f_k(x)$ for all $x \in X$ and such that $\int_X (f_k - f_j)d\mu < \varepsilon$.

Note that for $\mathcal{F} = 1_{\mathcal{C}}$ with $\mathcal{C} \subset \mathcal{B}$ one has for any μ

(82) $$N_I(\varepsilon,1_{\mathcal{C}},\mu) \leq N_I(\varepsilon,\mathcal{C},\mu) \leq 2\, N_I(\varepsilon,1_{\mathcal{C}},\mu).$$

In fact, as to the first inequality, suppose that $n := N_I(\varepsilon,\mathcal{C},\mu) < \infty$; then there exist $A_1,\dots,A_n \in \mathcal{B}$ such that for every $C \in \mathcal{C}$ there exist i,j with $A_i \subset C \subset A_j$ and $\mu(A_j \setminus A_i) < \varepsilon$. Take $f_i := 1_{A_i}$, $i=1,\dots,n$, to obtain $f_i \in L^2(X,\mathcal{B},\mu)$ so that for every $f = 1_C$ $f_i \leq f \leq f_j$ and $\int(f_j - f_i)d\mu = \mu(A_j \setminus A_i) < \varepsilon$. To verify the second inequality, let $m := N_I(\varepsilon,1_{\mathcal{C}},\mu) < \infty$; then there exist $f_1,\dots,f_m \in L^2(X,\mathcal{B},\mu)$ such that for every $f = 1_C \in 1_{\mathcal{C}}$ there exist $j,k \leq m$ with $f_j \leq 1_C \leq f_k$ and $\int(f_k - f_j)d\mu < \varepsilon$. Taking as $A_1,\dots,A_{2m}$ all sets of the form $\{f_i > 0\}$ and $\{f_i \geq 1\}$, $i=1,\dots,m$, we obtain that for every $C \in \mathcal{C}$ there exist $j,k \leq 2m$ such that $A_j \subset C \subset A_k$ and $\mu(A_k \setminus A_j) < \varepsilon$; in fact, $A_j := \{f_j > 0\}$ and $A_k := \{f_k \geq 1\}$ serves for this.

(83) (R.M. Dudley (1981a)): as $p \to \infty$ the condition on γ in (81) approaches $\gamma < 1$; if $\mathcal{F}$ is a collection of indicator functions of sets, i.e.,

$\mathcal{F} = 1_C$ for some $C \subset \mathcal{B}$, then (E_2) does imply (E_1) for C (cf. (82)). For $\gamma = 1$ it appears that (81) fails, specifically it fails when $\mathcal{F}$ is the collection of indicator functions of convex sets in $\mathbb{R}^3$ and μ is Lebesgue measure on the unit cube (cf. (63) and its consequences).

Now, if one would try to infer (79) from (81), there is the problem of verifying (E_2); on the other hand the condition on the envelope function F imposed in (81) is rather restrictive since this forces q^{-1} to be in $L^p(X,\mathcal{B},\mu)$ for some $p > 2$ (cf. instead the condition (iii_q) imposed in the definition of $\mathcal{Q}_2$). But, from another point of view, the class $\mathcal{F}_o = \{q^{-1}(t)\ 1_{[0,t]}: t \in [0,1]\}$ considered in (79) is of the following special form:

$\mathcal{F}_o = \{f_o \cdot g_t: t \in [0,1]\}$ with $f_o = q^{-1}$ and $g_t(s) := \frac{q(s)}{q(t)} 1_{[0,t]}(s),\ s \in X = [0,1]$,

where $\frac{q(s)}{q(t)} \to 0$ as $s \to 0$.

Thus, restricting our attention at this place to weight functions q for which q^{-1} is continuous, monotone decreasing on $(0,1/2)$, symmetric around $s = 1/2$ and such that $q^{-1}(s) \geq \delta > 0$ for all $s \in [0,1]$, then there exists some $M < \infty$ such that for each $t \in [0,1]$ $\sup_{s\in[0,1]} |g_t(s)| \leq M$ and such that

$$\{g_t^{-1}((a,b]): a < b\} \text{ forms a Vapnik-Chervonenkis class,}$$

since for each $a < b$ $g_t^{-1}((a,b])$ consists of one or at most two disjoint intervals $(c,d]$ in $[0,1]$ (cf. FIGURE 5).

Thus, the following result of R.M. Dudley (1981b) gives another way to obtain (79) (for proper weight functions q):

(84) Suppose $\mathcal{F} = \{f_o \cdot g: g \in \mathcal{G}\}$ where for some constant $M < \infty$ and some (suitably measurable) Vapnik-Chervonenkis class C

a) $\mathcal{G} = \{g: X \to [-M,M],\ g^{-1}((a,b]) \in C\ \forall a < b\}$ and

b) $f_o \geq 0$, f_o is measurable and $\mu(\{f_o > t\}) = o(t^{-2}(\log t)^{-\beta})$ as $t \to \infty$ for some $\beta > 4$,

then $\mathcal{F}$ is a μ-Donsker class.

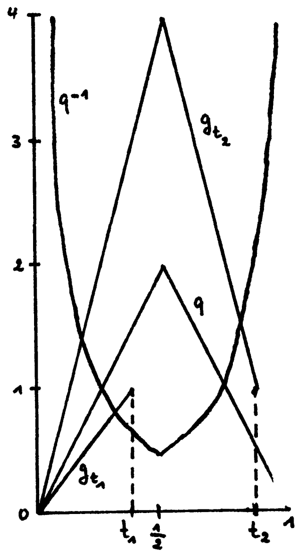

FIGURE 5

Note that b), even for $\beta > 1$, implies $f_o \in L^2(X,\mathcal{B},\mu)$. Conversely, the condition on $\mu(\{f_o > t\})$ is implied by $f_o \in L^p(X,\mathcal{B},\mu)$ for some $p > 2$.

Note also that by taking $f_o \equiv 1$ one obtains Theorem D as a corollary of (84). In this context the following result of D. Pollard (1981a) extends a special case of (84) to the case where only $f_o \in L^2(X,\mathcal{B},\mu)$ is assumed:

(85) If $f \in L^2(X,\mathcal{B},\mu)$ and if $\mathcal{C}$ is a Vapnik-Chervonenkis class of sets, then (for a separable version of $\beta_n \equiv (\beta_n(f))_{f\in\mathcal{F}}$), $\mathcal{F} := \{f_o \cdot 1_C: C \in \mathcal{C}\}$ is a μ-Donsker class.

(c) STRONG APPROXIMATIONS (cf. Gaenssler-Stute (1979), Section 3, concerning results for the uniform empirical process α_n).

In a recent paper by R.M. Dudley and W. Philipp (1983) almost sure and probability invariance principles are established for sums of independent not necessarily (Borel-)measurable random elements with values in a not necessarily separable Banach space like the closure of $D_o(\mathcal{C},\mu)$ in $(\ell^\infty(\mathcal{C}),\rho)$ fitting readily into the theory of empirical $\mathcal{C}$-processes $\beta_n \equiv (\beta_n(C))_{C\in\mathcal{C}}$ being now viewed as partial sum processes

$$\beta_n = n^{-1/2} \sum_{i=1}^{n} \zeta_i$$

with $\zeta_i \equiv (\zeta_i(C))_{C\in\mathcal{C}}$ defined by $\zeta_i(C) := 1_C(\xi_i) - \mu(C)$

(for a given sequence $(\xi_i)_{i\in\mathbb{N}}$ of random elements in $(X,\mathcal{B})$ with distribution μ on $\mathcal{B}$) having its values in $D_o(\mathcal{C},\mu)$.

In an analogous way the same viewpoint applies for empirical $\mathfrak{F}$-processes.

This approach of getting strong resp. weak invariance principles has the advantage that one can bypass most of the problems of measurability and topological characteristics which occurred in our theory of L_b-convergence where it was essential to choose proper sample spaces S and S_o for the processes β_n and $\mathbb{G}_\mu$, respectively, together with suitable σ-algebras in S and S_o on which the laws of β_n and $\mathbb{G}_\mu$ could be defined.

On the other hand, we think that the availability of the presented theory of weak convergence of empirical processes is at the least necessary to support Dudley's and Philipp's claim that strong approximation results are strengthened versions of functional central limit theorems.