Working Paper 93-23
Statistics and Econometrics Series 18
October, 1993

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

THE CENTRAL LIMIT THEOREM FOR EMPIRICAL PROCESSES ON V-Č CLASSES: A MAJORIZING MEASURE APPROACH

Juan Romo*

Abstract

Alexander (1987) gave necessary and sufficient conditions for the central limit theorem for empirical processes on Vapnik-Červonenkis classes of functions. In this paper we present a different version of his result using Talagrand's analytic characterization of pregaussianness (the majorizing measure condition). Our proof can be directly extended to give the corresponding result in the non-gaussian stable case.

Key Words
Central limit theorem, majorizing measure, Vapnik-Červonenkis classes.

*Universidad Carlos III de Madrid. Research partially supported by DGICYT PB90-0266 (Spain).

THE CENTRAL LIMIT THEOREM FOR EMPIRICAL PROCESSES ON V-Č CLASSES: A MAJORIZING MEASURE APPROACH

JUAN ROMO[1]
Universidad Carlos III de Madrid

Abstract

Alexander (1987) gave necessary and sufficient conditions for the central limit theorem for empirical processes on Vapnik-Červonenkis classes of functions. In this paper we present a different version of his result using Talagrand's analytic characterization of pregaussianness (the majorizing measure condition). Our proof can be directly extended to give the corresponding result in the non-gaussian stable case.

Key words and phrases: central limit theorem; majorizing measure; Vapnik-Červonenkis classes.

[1]Research partially supported by DGICYT PB90-0266 (Spain)

1 Introduction and preliminaries

Alexander (1987) completely characterized the central limit theorem for empirical processes on Vapnik-Červonenkis classes of functions. His proof relies on comparison with Gaussian processes. In this paper we present a different version of this result using Talagrand's analytic characterization of pregaussianness (the majorizing measure condition). Our proof can be directly extended to give the corresponding result in the non-gaussian stable case.

Let $(S, \mathcal{S}, P)$ be a complete probability space where $\mathcal{S}$ is countably generated and let $\{X_n : n \in N\}$ be a sequence of independent and identically distributed random variables with law P. By $\{\varepsilon_n : n \in N\}$ we represent a sequence of independent random variables such that

$$P\{\varepsilon_n = +1\} = P\{\varepsilon_n = -1\} = \frac{1}{2}, \qquad n \in N,$$

(a Rademacher sequence). In what follows, we use the sample space $(\Omega, \mathcal{A}, Pr) = (S^N \times [0,1], \mathcal{S}^N \times \mathcal{B}_{[0,1]}, P^N \times \lambda)$, where $\mathcal{B}_{[0,1]}$ is the Borel σ-algebra for the usual topology in [0,1] and λ is the Lebesgue measure. This space is rich enough to support the sequence $\{X_n : n \in N\}$ and a Rademacher sequence $\{\varepsilon_n : n \in N\}$ independent from each other.

$\mathcal{L}_p(S, \mathcal{S}, P)$ (or simply $\mathcal{L}_p$), $1 \leq p \leq 2$, is the class of real measurable functions on S such that $\int_S | f |^p \, dP < \infty$. On $\mathcal{L}_p(S, \mathcal{S}, P)$ we define the pseudometric

$$\rho_p(f,g) = \rho_p^P(f,g) = (\int_S | f - g |^p \, dP)^{\frac{1}{p}}, \quad f, g \in \mathcal{L}_p(S, \mathcal{S}, P),$$

We will write Pf instead of $\int_S f dP$. Let $\mathcal{F}$ be a class of real measurable functions on S with envelope $F(s) = \sup_{f \in \mathcal{F}} | f(s) |$ finite for all $s \in S$. Let $F^* : S \longrightarrow \mathbf{R}$ be the measurable envelope of $\mathcal{F}$ (see Dudley (1984), Theorem 3.1.1.).

For a probability space $(\mathcal{X}, \mathcal{E}, \mu)$, μ^* is the outer probability measure for μ. From Proposition 2.2 in Andersen (1985) it follows that

$$P^*\{F > t\} = P\{F^* > t\}, \quad for \quad t \in \mathbf{R}.$$

It also holds that

$$Pr^*\{F(X_n) > t\} = P\{F^* > t\}, \quad for \quad all \quad t \in \mathbf{R}.$$

Let $\{P_n : n \in N\}$ be the sequence of empirical probability measures corresponding to $\{X_n : n \in N\}$, $P_n = n^{-1}\sum_{i=1}^{n}\delta_{x_i}$. If we write

$$\nu_n = n^{\frac{1}{2}}(P_n - P), \quad n \in N,$$

since F is finite for all $s \in S$, we have that for all $n \in N$,

$$\{\nu_n(f) : f \in \mathcal{F}\} = \{n^{-\frac{1}{2}}\sum_{i=1}^{n}(f(X_i) - Pf) : f \in \mathcal{F}\}$$

is a stochastic process with sample paths in the space $\ell^\infty(\mathcal{F})$ of real bounded functions defined on $\mathcal{F}$. We endow $\ell^\infty(\mathcal{F})$ with the sup norm, $\| H \|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} | H(f) |$, $H \in \ell^\infty(\mathcal{F})$. If ρ is a pseudometric on $\mathcal{F}$, $C_u(\mathcal{F}, \rho)$ is the space of uniformly ρ-continuous bounded real functions.

Some of the calculations to be carried out need certain measurability conditions on the class $\mathcal{F}$; we will assume both that $\mathcal{F}$ is admissibly measurable for P and that $\mathcal{F}' = \{f - g : f, g \in \mathcal{F}\}$ and $\mathcal{F}'_{\rho_2}(\beta) = \{f - g : f, g \in \mathcal{F}, \rho_2(f, g) \leq \beta\}$ are deviation measurable for P (see Alexander (1987) for these definitions).

In part (a) of the following lemma, we recall the subgaussian inequality, which is a particular case of Hoeffding's inequality (1963); part (b) of the lemma gives Bernstein's inequality (see Bennet (1962)).

LEMMA 1.1. (a) If $\{\varepsilon_n : n \in N\}$ is a Rademacher sequence and $\{a_n : n \in N\} \subset \mathbf{R}$ then, for all $n \in N$,

$$P\{\sum_{i=1}^{n} a_i\varepsilon_i > t\} \leq \exp\{-\frac{t^2}{2}\sum_{i=1}^{n} a_i^2\}, \qquad t \in \mathbf{R}.$$

(b) Let $\{Y_n : n \in N\}$ be a sequence of independent real random variables such that $| Y_n | \leq M$ and $EY_n = 0$, for all $n \in N$. Then

$$P\{\sum_{i=1}^{n} Y_i > t\} \leq \exp\left\{-\frac{t^2}{2\sum_{i=1}^{n} EY_i^2 + \frac{2}{3}Mt}\right\}, \quad t \in \mathbf{R}.$$

Since $(\ell^\infty(\mathcal{F}), \|\cdot\|_{\mathcal{F}})$ is a nonseparable (except for finite $\mathcal{F}$) Banach space and since $\nu_n : \Omega \to \ell^\infty(\mathcal{F})$ is not necessarily measurable with respect to $\mathcal{A}$ and the Borel σ-algebra corresponding to the topology induced by $\|\cdot\|_{\mathcal{F}}$, we have to consider a convenient definition of weak convergence. Following Hoffmann-Jϕrgensen (1984), we say that $\xi : \Omega \longrightarrow \ell^\infty(\mathcal{F})$ is a random element in $\ell^\infty(\mathcal{F})$ if $\lim_{M\to\infty} Pr^*\{\| \xi \|_{\mathcal{F}} > M\} = 0$. We say that the sequence $\{\xi_n : n \in N\}$ of random elements in $\ell^\infty(\mathcal{F})$ *converges in law to a Radon limit* γ if there exists a probability Radon measure γ in $\ell^\infty(\mathcal{F})$ such that for all bounded and $\| \cdot \|_{\mathcal{F}}$-continuons $H : \ell^\infty \mathcal{F} \longrightarrow \mathbf{R}$ we have

$$\lim_{n\to\infty} E^* H(\xi_n) = \int H d\gamma,$$

where E^* is the upper integral.

We introduce now two useful concepts in the study of empirical processes: metric entropy and majorizing measure. Both give information about the "size" or "shape" of the pseudometric space which is the index set of the process.

DEFINITION 1.2. Let $(\mathcal{F}, \rho)$ be a pseudometric space and let $\varepsilon > 0$. The *covering number* of $\mathcal{F}$ using balls of ρ-radius ε is

$$N(\varepsilon, \rho, \mathcal{F}) \equiv \min\{n \in N : \ \ there \ \ exist \ \ f_1, f_2, \ldots, f_n \in \mathcal{F} \\ such \ \ that \ \ \sup_{f\in\mathcal{F}} \min_{i\le n} \rho(f, f_i) \le \varepsilon\}.$$

The function $H(\varepsilon, \rho, \mathcal{F}) = \log N(\varepsilon, \rho, \mathcal{F})$ is called *metric entropy* and it was introduced by Kolmogorov and Tikhomirov (1959). The idea of majorizing measure first appeared in Preston (1972) and was used by Fernique (1974) and Talagrand (1987) in their study of sample paths continuity for Gaussian processes. Let $(\mathcal{F}, \rho)$ be a pseudometric space and let $B_\rho(f, \varepsilon) = \{g \in \mathcal{F} : \rho(f, g) \le \varepsilon\}$. Let $\Phi : \mathbf{R}_+ \longrightarrow \mathbf{R}_+$ be a nonincreasing monotone function such that $\Phi(x + y) \le \Phi(x) + \Phi(y)$ and $\int_o^a \Phi(x)dx < \infty$ for some $a > 0$. Under these conditions we have.

DEFINITION 1.3. A Borel probability measure on $(\mathcal{F}, \rho)$ is a *majorizing measure* (for $(\mathcal{F}, \rho)$ with respect to Φ) if $\sup_{f\in\mathcal{F}} \int_0^\infty \Phi(\mu\{B_\rho(f, \varepsilon)\})d\varepsilon < \infty$. A probability measure μ is a *discrete majorizing measure* if there exists a sequence $\{\pi_q : q \in N\}$ such that $\pi_q : \mathcal{F} \to \mathcal{F}$ with $\sharp(\pi_q\mathcal{F}) < \infty$, μ is supported on $\bigcup_{q\in N} \pi_q\mathcal{F}$ and

$$(i) \qquad \rho(f, \pi_q f) \leq 2^{-q} \quad , for \;\; all \;\; f \in \mathcal{F}, \;\; and \tag{1.1}$$

$$(ii) \qquad \sup_{f\in\mathcal{F}} \sum_{q=1}^{\infty} 2^{-q}\Phi(\mu\{\pi_q f\}) < \infty.$$

The proof of the next result can be seen in Andersen, Giné, Ossiander and Zinn (1988) (Lemmas 2.1 and 2.4).

LEMMA 1.4. (a) If $(\mathcal{F}, \rho)$ admits a majorizing measure then it also admits a discrete majorizing measure. If the majorizing measure τ verifies that

$$\lim_{\delta\to 0} \sup_{f\in\mathcal{F}} \int_0^\delta \Phi(\tau\{B_\rho(f,\varepsilon)\})d\varepsilon = 0 \tag{1.2}$$

then the discrete majorizing measure μ can be chosen such that

$$\lim_{q_0\to\infty} \sup_{f\in\mathcal{F}} \sum_{q=q_0}^{\infty} 2^{-q}\Phi(\mu\{\pi_q f\}) = 0. \tag{1.3}$$

(b) If μ is a discrete majorizing measure for $(\mathcal{F}, \rho)$ supported on $\bigcup_{q\in N} \pi_q\mathcal{F}$ then it also verifies that

$$\sup_{f\in\mathcal{F}} \sum_{q=1}^{\infty} 2^{-q}\Phi(2^{-q}\mu\{\pi_1 f\} \ldots \mu\{\pi_q f\}) < \infty.$$

If, moveover, Φ satisfies (1.3) then

$$\lim_{q_0\to\infty} \sup_{f\in\mathcal{F}} \sum_{q=q_0}^{\infty} 2^{-q}\Phi(2^{-q}\mu\{\pi_1 f\} \ldots \mu\{\pi_q f\}) = 0.$$

The lemma we state next allows to replace entropy by majorizing measure in "chaining arguments"; its proof can be seen in Andersen, Giné, Ossiander and Zinn (1988), where this observation is attributed to Talagrand.

LEMMA 1.5. If μ is a discrete majorizing measure for $(\mathcal{F},\rho)$ supported on $\bigcup_{q\in N}\pi_q\mathcal{F}$ then

$$\lim_{r\to\infty}\sum_{q=1}^{\infty}\sum_{t\in T_q}\exp\{-r\varphi^{-1}(s\gamma_q(t)\}=0$$

for all $s>0$, with finite sum for all $r>1$, where $T_q=\{(\pi_1 f,\ldots,\pi_q f): f\in\mathcal{F}\}$ and

$$\gamma_q(t)=\gamma_q(f)=\varphi\left(\log\frac{2^q}{\mu\{\pi_1 f\},\ldots,\mu\{\pi_q f\}}\right),$$

for $t=(\pi_1 f,\ldots,\pi_q f)\in T_q$.

The function Φ that we will use in next section will have the form

$$\Phi_2(x)=(\log\frac{1}{x})^{\frac{1}{2}},\quad x>0.$$

Next we present in some detail the relationship between entropy and majorizing measure in the case of Gaussian processes; this will make clear the connection between the central limit theorem for empirical processes of Alexander (1987) and the reformulation of it that we give in Theorem 3.1.

Let $Z=\{Z(t): t\in T\}$ be a Gaussian process. It induces the pseudometric $\sigma(s,t)=(E(Z(s)-Z(t))^2)^{\frac{1}{2}}$, $s,t\in T$, and let $|T|_\sigma=\sup_{s,t\in T}\sigma(s,t)$. Dudley (1967, 1973) proved that if

$$\int_0^{|T|_\sigma}(\log N(\varepsilon,\sigma,T))^{\frac{1}{2}}d\varepsilon<\infty \tag{1.4}$$

then the process Z has a version with almost surely bounded and σ-uniformly continuous sample paths. In Fernique (1974) it is shown that if (1.4) holds then there exists a majorizing measure τ for (T,σ) such that

$$\lim_{\delta\to 0}\sup_{t\in T}\int_0^{\delta}\left(\log\frac{1}{\tau\{B_\sigma(t,\varepsilon)\}}\right)^{\frac{1}{2}}d\varepsilon=0, \tag{1.5}$$

i.e., (1.2) holds with $\Phi_2(x)=(\log\frac{1}{x})^{\frac{1}{2}}$. In the same paper, Fernique stated that if there exists a majorizing measure for (T,σ) satisfying (1.5) and $\tilde{Z}$ is a measurable and separable version of Z, then $\tilde{Z}$ has almost surely bounded and

σ-uniformly continuous sample paths. In the strongly stationary Gaussian case, Fernique proved that (1.5) is necessary and sufficient for the continuity of the sample paths. Talagrand (1987) finally completed the solution of the problem by showing that (1.5) is necessary for the continuity and boundedness of the sample paths of any Gaussian process.

The following result gives sufficient conditions for the convergence of $\{\nu_n : n \in N\}$ (see Andersen, Giné, Ossiander and Zinn, 1988). This proposition expresses the idea of finite-dimensional approximation for the central limit theorem in infinite dimensions in a way which is very convenient to combine with discrete majorizing measures, as we will do in the proof of Theorem 3.1.

PROPOSITION 1.6. Let $\mathcal{F}$ be a class of functions in $\mathcal{L}_2(S, \mathcal{S}, P)$ with envelope F finite everywhere. Assume that there exist applications $\pi_q : \mathcal{F} \longrightarrow \mathcal{F}$, $q \in N$, such that $\sharp(\pi_q \mathcal{F}) < \infty$ for all $q \in N$ and

$$(i) \quad \lim_{q_0 \to \infty} \limsup_n \sup_{f \in \mathcal{F}} Pr^*\{| n^{-\frac{1}{2}} \sum_{i=1}^{n} (f - \pi_{q_0} f) - P(f - \pi_{q_0} f)(X_i) | > \eta\} = 0,$$

and

$$(ii) \quad \lim_{q_0 \to \infty} \limsup_n Pr^*\{\| n^{-\frac{1}{2}} \sum_{i=1}^{n} \varepsilon_i (f - \pi_{q_0} f)(X_i) \|_{\mathcal{F}} > \eta\} = 0,$$

for all $\eta > 0$. Then $\{\{n^{-\frac{1}{2}} \sum_{i=1}^{n} (f(X_i - Pf) : f \in \mathcal{F}\} : n \in N\}$ converges in law in $\ell^\infty(\mathcal{F})$ to a centered Gaussian process G_P with sample paths in $C_u(\mathcal{F}, \rho_2)$ and covariance structure determined by P.

Note that hypothesis (i) is a centering condition which allows for desymmetrization in (ii) (using Lemma 2.7 (b) in Giné and Zinn (1986)) transforming it into an "asymptotic equicontinuity condition" which is the usual one to prove the central limit theorem for stochastic processes.

2 VAPNIK-ČERVONENKIS CLASSES OF FUNCTIONS

The property we describe now was introduced by Vapnik and Červonenkis (1971) in their study of Glivenko-Cantelli's theorem for classes of sets more general than semiintervals in **R**.

Let S be a nonempty set and let $A \subset S$. We will write

$$\Delta^{\mathcal{C}}(A) = \sharp\{C \cap A : C \in \mathcal{C}\}$$

and

$$m^{\mathcal{C}}(n) = \max\{\Delta^{\mathcal{C}}(A) : A \subset S, \sharp A = n\}, \quad n \in N.$$

$m^{\mathcal{C}}(n)$ is the *growth function of* $\mathcal{C}$. We define the (Vapnik-Červonenkis) *index* of $\mathcal{C}$ as $v(\mathcal{C}) = \min\{n \in N : m^{\mathcal{C}}(n) < 2^n\}$ or $v(\mathcal{C}) = +\infty$ if $m^{\mathcal{C}}(n) = 2^n$ for all $n \in N$. Say that $\mathcal{C}$ is a *Vapnik-Červonenkis class of sets* if $v(\mathcal{C}) < \infty$, i.e., if these exists an integer n such that for every $A \subset S$ with n elements, $\{C \cap A : C \in \mathcal{C}\}$ does not contain all the subsets of A.

If f is a real function on S, its graph is $G(f) = \{(s,t) \in S \times \mathbf{R} : 0 \leq t \leq f(s)$ *or* $f(s) \leq t \leq 0\}$. The class $\mathcal{F} \subset \mathbf{R}^S$ is a *Vapnik-Červonenkis class of functions* if $\{G(f) : f \in \mathcal{F}\}$ is a Vapnik-Červonenkis class of sets in $S \times R$. The basic property of the entropy of Vapnik-Cervonenkis classes of functions is given in the following lemma which can be seen in Pollard (1984).

LEMMA 2.1. Let $\mathcal{F}$ be a class of measurable real functions defined on a measurable space $(S, \mathcal{S})$ with envelope F and let P be a probability measure on $(S, \mathcal{S})$ such that $0 < PF < \infty$. If $\mathcal{F}$ is a Vapnik-Červonenkis class of functions then there exist constants A and v, independent of F and P, such that

$$N(\varepsilon \parallel F \parallel_1, \rho_1^P, \mathcal{F}) \leq A\varepsilon^{-v}, \quad for \quad 0 < \varepsilon < 1.$$

The exponential inequality we prove next is similar to Proposition 4.8 in Alexander (1987); although this one is weaker than Alexander's, it need less hypotheses. The proof below which is more elementary than Alexander's, is based on Lemma 5.2 in Giné and Zinn (1984) where a technique by Le Cam (the "square root trick") is adapted. First, we state Lemma 5.2 in Giné and Zinn (1984). If $\mathcal{G}$ is a class of functions, let $\mathcal{G}^2 = \{g^2 : g \in \mathcal{G}\}$.

LEMMA 2.2. Let $\mathcal{G}$ be a class of real functions on S such that $\mathcal{G}^2$ is uniformly bounded by d and deviation measurable for P. Let $M_n = n^{\frac{1}{2}} \sup_{g\in\mathcal{G}} Pg^2$, $n \in N$. Let $t > 0$ and $r > 0$ be such that $\lambda = (\frac{t}{2})^{\frac{1}{2}} - (2M_n)^{\frac{1}{2}} - 2r > 0$. Then for every $n \in N$ and $A > 0$,

$$Pr^*\left\{\sup_{g\in\mathcal{G}} \sum_{i=1}^{n} g^2(X_i) > tn^{\frac{1}{2}}\right\} \leq 4Pr^*\{N\left(\frac{r}{n^{\frac{1}{4}}}, \rho_2^{P_n}, \mathcal{G}\right) > A\} + 8A\exp\left\{-\frac{\lambda^2 n^{\frac{1}{2}}}{2d}\right\}.$$

Before stating the inequality, we give two properties of covering numbers: if $\mathcal{F}$ is a class of functions with envelope F and $\mathcal{F}' = \{f - g : f \in \mathcal{F}, g \in \mathcal{F}\}$ then

$$N(2\varepsilon PF, \rho_i, \mathcal{F}') \leq N^2(\varepsilon PF, \rho_i, \mathcal{F}) \quad , \quad i = 1,2$$

and if $\mathcal{F}$ is uniformly bounded by d then

$$N(\varepsilon, \rho_1, \mathcal{F}) \leq N(\varepsilon, \rho_2, \mathcal{F}) \leq N\left(\frac{\varepsilon^2}{2d}, \rho_1, \mathcal{F}\right).$$

PROPOSITION 2.3. Let $\mathcal{F}$ be a Vapnik-Červonenkis class of functions uniformly bounded by $a > 0$ and such that, for all $\beta > 0$,$(\mathcal{F}'_{\rho_2}(\beta))^2$ is deviation measurable for P. There exist constants K_1 and K_2, depending only on the Vapnik-Červonenkis index of $\mathcal{F}$, such that if M and m satisfy

$$aM \geq 28m^{\frac{1}{2}}\beta^2 \tag{2.6}$$

and

$$\frac{Mm^{\frac{1}{2}}}{a} \geq K_1 \log\frac{a^2}{\beta^2} + K_2 \tag{2.7}$$

then

$$Pr^*\left\{\| \sum_{i=1}^{m} \varepsilon_i f(X_i) \|_{\mathcal{F}'_{\rho_2}(\beta)} > Mm^{\frac{1}{2}}\right\} \leq 2\exp\left\{-\frac{Mm^{\frac{1}{2}}}{33a}\right\}.$$

PROOF. We have that

$$Pr^*\left\{\| \sum_{i=1}^{m} \varepsilon_i f(X_i) \|_{\mathcal{F}'_{\rho_2}(\beta)} > Mm^{\frac{1}{2}}\right\} \leq Pr^*\left\{N\left(\frac{M}{2m^{\frac{1}{2}}}, \rho_1^{P_m}, \mathcal{F}'_{\rho_2}(\beta)\right) > H\right\}$$

$$+Pr^*\left\{\| \sum_{i=1}^{m} f^2(X_i) \|_{\mathcal{F}'_{\rho_2}(\beta)} > 4aMm^{\frac{1}{2}}\right\}$$

$$+ Pr^*\left\{\| \sum_{i=1}^{m} f^2(X_i) \|_{\mathcal{F}'_{\rho_2}(\beta)} \leq 4aMm^{\frac{1}{2}}, N\left(\frac{M}{2m^{\frac{1}{2}}}, \rho_1^{P_m}, \mathcal{F}'_{\rho_2}(\beta)\right) \leq H\ , \quad (2.8)$$

$$\| \sum_{i=1}^{m} \varepsilon_i f(X_i) \|_{\mathcal{F}_{\rho'_2}(\beta)} > Mm^{\frac{1}{2}}\right\}.$$

By Lemma 2.1 and the remark above,

$$N\left(\frac{M}{2m^{\frac{1}{2}}}, \rho_1^{P_m}, \mathcal{F}'_{\rho_2}(\beta)\right) \leq K_0\left(\frac{M}{4am^{\frac{1}{2}}}\right)^{-2v}.$$

Taking $H = K_0(\frac{M}{4am^{\frac{1}{2}}})^{-2v}$, (2.8) is less than or equal to

$$Pr^*\left\{\| \sum_{i=1}^{m} f^2(X_i) \|_{\mathcal{F}'_{\rho_2}(\beta)} > 4aMm^{\frac{1}{2}}\right\}$$

$$+ Pr^*\left\{\| \sum_{i=1}^{m} f^2(X_i) \|_{\mathcal{F}'_{\rho_2}(\beta)} \leq 4aMm^{\frac{1}{2}}, N\left(\frac{M}{2m^{\frac{1}{2}}}, \rho_1^{P_m}, \mathcal{F}'_{\rho_2}(\beta)\right) \leq H\ , \quad (2.9)$$

$$\| \sum_{i=1}^{m} \varepsilon_i f(X_i) \|_{\mathcal{F}'_{\rho_2}(\beta)} > Mm^{\frac{1}{2}}\ \right\}.$$

Adding and subtracting to each $f \in \mathcal{F}'_{\rho_2}(\beta)$ the corresponding f^* from a minimal net of size $N(\frac{M}{2m^{\frac{1}{2}}}, \rho_1^{P_m}, \mathcal{F}'_{\rho_2}(\beta))$ and using Lemma 1.(a), the second term in (2.9) is less than or equal to

$$2K_0\left(\frac{M}{4am^{\frac{1}{2}}}\right)^{-2v} \exp\left\{-\frac{Mm^{\frac{1}{2}}}{32a}\right\}. \quad (2.10)$$

Now, Lemma 2.2 gives a bound for the first term in (2.9): let $\mathcal{G} = \mathcal{F}'_{\rho_2}(\beta)$, $d = 4a^2$, $M_m = m^{\frac{1}{2}}\beta^2$, $t = 4aM$ and $r = \frac{a^{\frac{1}{2}}M^{\frac{1}{2}}}{9}$. From (2.6) it follows that

$$\lambda = 2^{\frac{1}{2}}(a^{\frac{1}{2}}M^{\frac{1}{2}} - m^{\frac{1}{4}}\beta) - \frac{2a^{\frac{1}{2}}M^{\frac{1}{2}}}{9} > a^{\frac{1}{2}}M^{\frac{1}{2}} > 0$$

and so

$$Pr^*\left\{\| \sum_{i=1}^{m} f^2(X_i) \|_{\mathcal{F}'_{\rho_2}(\beta)} > 4aMm^{\frac{1}{2}}\right\}$$

$$\leq 4Pr^*\left\{N\left(\frac{2^{\frac{1}{2}}M^{\frac{1}{2}}}{9m^{\frac{1}{4}}}, \rho_2^{P_m}, \mathcal{F}'_{\rho_2}(\beta)\right) > A\right\} \tag{2.11}$$

$$+8A\exp\left\{-\frac{Mm^{\frac{1}{2}}}{8a}\right\}.$$

Taking $A = K_0'(\frac{M}{18^2 \cdot am^{\frac{1}{2}}})^{-v}$, with K_0' conveniently chosen (depending only on the Vapnik-Červonenkis index of $\mathcal{F}$), from (2.10) and (2.11) we obtain that (2.8) is less than or equal to

$$\left(8K_0'\left(\frac{M}{18^2 am^{\frac{1}{2}}}\right)^{-v} + 2K_0\left(\frac{M}{4am^{\frac{1}{2}}}\right)^{-2v}\right)\exp\left\{-\frac{Mm^{\frac{1}{2}}}{32a}\right\}$$

$$\leq \left(8K_0'\left(\frac{\beta^2}{18^2 a^2}\right)^{-v} + 2K_0\left(\frac{\beta^2}{4a^2}\right)^{-2v}\right)\exp\left\{-\frac{Mm^{\frac{1}{2}}}{32a}\right\},$$

applying (2.6) again. Choosing K_1 and K_2 such that

$$\max\left\{8 \cdot 18^{2v}K_0'\left(\frac{a^2}{\beta^2}\right)^{v}, 2 \cdot 16^{v}K_0\left(\frac{a^2}{\beta^2}\right)^{2v}\right\} \leq e^{(\frac{1}{32}-\frac{1}{33})\frac{Mm^{\frac{1}{2}}}{a}},$$

the result follows.

We end this section by recalling a theorem by Chang (1964). Let G be a probability distribution function on $\mathbf{R}$. The same proof given by Chang, using now the function $1 - G(-x)$, allows us to state the following reformulation of the theorem which is convenient to use it in our proof of Theorem 3.1

THEOREM 2.4. Let G be a probability distribution on R and let G_n be the empirical distribution function obtained from the corresponding random sample. If $\{b_n : n \in N\}$ is a sequence of positive numbers that tends to infinity then, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} Pr\left\{\sup_{\frac{b_n}{n}\leq 1-G(x)} \mid \frac{1-G_n(x)}{1-G(x)} - 1 \mid > \varepsilon\right\} = 0.$$

3 Results

The definitive central limit theorem for empirical processes on Vapnik-Červonenkis classes of functions was obtained by Alexander (1987). Here we present a different proof of his result which does not make use of Gaussian comparison; instead, we combine results of Talagrand (1987) about majorizing measures and a modification of the classic "chaining argument" that appears in Andersen, Giné, Ossiander and Zinn (1988) and Andersen, Giné and Zinn (1988). The characterization of Gaussian processes with continuous sample paths allows for the following reformulation of Alexander's theorem (compare with the result in Alexander (1987)).

THEOREM 3.1. Let $\mathcal{F} \subset \mathcal{L}_2(S, \mathcal{S}, P)$ be an admissibly measurable Vapnik-Červonenkis class of functions with envelope F finite everywhere. Assume that $\sup_{f\in\mathcal{F}} \mid Pf \mid < \infty$. Then

$$(i) \qquad t^2 P\{F^* > t\} \to 0 \quad \textit{when} \quad t \to \infty, \qquad \textit{and} \qquad (3.12)$$

(ii) there exists a Borel probability measure μ on $(\mathcal{F}, \rho_2)$ such that

$$\lim_{\delta\to 0}\sup_{f\in\mathcal{F}} \int_0^\delta (\log \frac{1}{\mu\{B_{\rho_2}(f,\varepsilon)\}})^{\frac{1}{2}} d\varepsilon = 0$$

and

$$\sup_{f\in\mathcal{F}} \int_0^\infty (\log \frac{1}{\mu\{B_{\rho_2}(f,\varepsilon)\}})^{\frac{1}{2}} d\varepsilon < \infty$$

if, and only if,

$$\{\{n^{-\frac{1}{2}} \sum_{i=1}^{n} (f(X_i) - Pf) : f \in \mathcal{F}\} : n \in N\}$$

converges in law in $\ell^\infty(\mathcal{F})$ to a centered Gaussian process G_P with sample paths in $C_u(\mathcal{F}, \rho_2)$ and the covariance structure determined by P.

PROOF. Necessity follows in the same way as in Corollary 2.3 in Alexander (1987). Next we prove sufficiency. By Lemma 1.4, we can assume that there exists a sequence $\{\pi_q : q \in N\}$, with $\pi_q : \mathcal{F} \to \mathcal{F}$ such that $\sharp(\pi_q \mathcal{F}) < \infty$ for each $q \in N$, and that μ is a discrete probability measure supported on $\bigcup_{q \in N} \pi_q \mathcal{F}$ verifying

$$(a) \qquad \rho_2(f, \pi_q f) \leq 2^{-q}, \quad for \;\; every \;\; f \in \mathcal{F}$$

$$(b) \qquad \lim_{q_0 \to \infty} \sup_{f \in \mathcal{F}} \sum_{q=q_0}^{\infty} 2^{-q} \gamma_q(f) = 0 \quad and$$

$$\sup_{f \in \mathcal{F}} \sum_{q=1}^{\infty} 2^{-q} \gamma_q(f) < \infty$$

where $\gamma_q(f) = (\log \frac{2^q}{\mu\{\pi_1 f\} \ldots \mu\{\pi_q f\}})^{\frac{1}{2}}$. The function γ_q is defined on

$$T_q = \{(\pi_1 f, \ldots, \pi_q f) : f \in \mathcal{F}\}$$

and for each fixed $f \in \mathcal{F}$, the sequence $\{\gamma_q(f) : q \in N\}$ is increasing. Let $\beta_{q_0} = \sup_{f \in \mathcal{F}} \sum_{q=q_0}^{\infty} 2^{-q} \gamma_q(f)$.

The centering condition in Proposition 1.6 follows from (i), hence it is enough to show that

$$\lim_{q_0 \to \infty} \limsup_{n \to \infty} Pr^* \{ \| n^{-\frac{1}{2}} \sum_{i=1}^{n} \varepsilon_i (f - \pi_{q_0} f)(X_i) \|_{\mathcal{F}} > \eta \} = 0,$$

for all $\eta > 0$.

The first step is truncation. From (i), there exists a sequence $\{\delta_n : n \in N\}$, $\delta_n \to 0$, such that $nP\{F^* > \delta_n n^{\frac{1}{2}}\} \to 0$ as $n \to \infty$. Fix $\eta > 0$, $\phi > 0$ and define

$$\tau_n = \inf\{\tau > 0 : nP\{F^* > \tau n^{\frac{1}{2}}\} < \phi \eta \delta_n^{-1}\}.$$

For large enough n, $\tau_n \leq \delta_n$ and so $\tau_n \to 0$. On the one hand, we have that

$$Pr^* \{ \| n^{-\frac{1}{2}} \sum_{i=1}^{n} \varepsilon_i (f - \pi_q f) I_{\{F^* > \delta_n n^{\frac{1}{2}}\}} (X_i) \|_{\mathcal{F}} > \eta \} \leq$$

$$\leq nP\{2F^* > \delta_n n^{\frac{1}{2}}\} \to 0 \quad , \quad as \quad n \to \infty.$$

On the other hand, for big enough n,

$$Pr^*\{\|n^{-\frac{1}{2}}\sum_{i=1}^{n}\varepsilon_i(f-\pi_q f)I_{\{\tau_n n^{\frac{1}{2}}<F^*\leq\delta_n n^{\frac{1}{2}}\}}(X_i)\|_{\mathcal{F}} > \eta\}$$

$$\leq \frac{E\sum_{i=1}^{n} \mid \varepsilon_i(f-\pi_q f)I_{\{\tau_n n^{\frac{1}{2}}<F^*\leq\delta_n n^{\frac{1}{2}}\}}(X_i) \mid}{\eta n^{\frac{1}{2}}}$$

$$\leq \frac{2\delta_n n^{\frac{1}{2}} nP\{F^* > \tau_n n^{\frac{1}{2}}\}}{\eta n^{\frac{1}{2}}} \leq \frac{2\delta_n n^{\frac{1}{2}}\eta\phi\delta_n^{-1}}{\eta n^{\frac{1}{2}}} = 2\phi,$$

where the last inequality is true since the distribution function is right continuous.

Therefore, since

$$Pr^*\{\|n^{-\frac{1}{2}}\sum_{i=1}^{n}\varepsilon_i(f-\pi_q f)(X_i)\|_{\mathcal{F}} > 3\eta\}$$

$$\leq Pr^*\{\|n^{-\frac{1}{2}}\sum_{i=1}^{n}\varepsilon_i(f-\pi_q f)I_{\{F^*>\delta_n n^{\frac{1}{2}}\}}(X_i)\|_{\mathcal{F}} > \eta\}$$

$$+Pr^*\{\|n^{-\frac{1}{2}}\sum_{i=1}^{n}\varepsilon_i(f-\pi_q f)I_{\{\tau_n n^{\frac{1}{2}}<F^*\leq\delta_n n^{\frac{1}{2}}\}}(X_i)\|_{\mathcal{F}} > \eta\}$$

$$+Pr^*\{\|n^{-\frac{1}{2}}\sum_{i=1}^{n}\varepsilon_i(f-\pi_q f)I_{\{F^*\leq\tau_n n^{\frac{1}{2}}\}}(X_i)\|_{\mathcal{F}} > \eta\},$$

it is enough to show that

$$\lim_{q_0\to\infty}\limsup_{n\to\infty} Pr^*\{\|n^{-\frac{1}{2}}\sum_{i=1}^{n}\varepsilon_i(f-\pi_{q_0} f)I_{\{F^*\leq\tau_n n^{\frac{1}{2}}\}}(X_i)\|_{\mathcal{F}} > \eta\} = 0,$$

for all $\eta > 0$. Hereafter, we will denote $fI_{\{F^*\leq\tau_n n^{\frac{1}{p}}\}}$ by $\bar{f}$ for any $f \in \mathcal{F}$.

Now, we utilize a decomposition of $\bar{f}$ using "summation by parts" (see Andersen, Giné, Ossiander and Zinn (1988); the idea of stratifying comes from Bass (1984) through Ossiander (1987); Andersen *et al.* (1988) consider strata whose levels depend on f). For $q_0 \in N$ and $q_1 > q_0 + 1$,

$$\overline{f} - \overline{\pi_{q_0} f} = (\overline{f} - \overline{\pi_{q_1} f}) + \sum_{q=q_0+1}^{q_1} (\overline{\pi_q f} - \overline{\pi_{q-1} f})$$

$$= (\bar{f} - \overline{\pi_{q_1} f}) + \sum_{q=q_0+1}^{q_1} (\overline{\pi_q f} - \overline{\pi_{q-1} f}) I_{\{F^* \leq c_{n,q}(f)\}}$$

$$+ \sum_{q=q_0+1}^{q_1} (\overline{\pi_q f} - \overline{\pi_{q-1} f}) I_{\{F^* > c_{n,q}(f)\}}$$

$$= (\bar{f} - \overline{\pi_{q_1} f}) I_{\{F^* \leq c_{n,q_1}(f)\}} + (\bar{f} - \overline{\pi_{q_0} f}) I_{\{F^* > c_{n,q_0+1}(f)\}}$$

$$+ \sum_{q=q_0+1}^{q_1} (\overline{\pi_q f} - \overline{\pi_{q-1} f}) I_{\{F^* \leq c_{n,q}(f)\}}$$

$$+ \sum_{q=q_0+2}^{q_1} (\bar{f} - \overline{\pi_{q-1} f}) I_{\{c_{n,q}(f) < F^* \leq c_{n,q-1}(f)\}},$$

where $c_{n,q}(f) = \frac{2^{-q} n^{\frac{1}{2}}}{\gamma_q(f)}$ which, for fixed $n \in N$ and $f \in \mathcal{F}$, is a decreasing sequence in q.
Hence, we have that

$$Pr^*\{\|n^{-\frac{1}{2}} \sum_{i=1}^{n} \varepsilon_i (f - \pi_{q_0} f) I_{\{F^* \leq \tau_n n^{\frac{1}{2}}\}}(X_i)\|_{\mathcal{F}} > 4\eta\}$$

$$\leq Pr^*\{\|n^{-\frac{1}{2}} \sum_{i=1}^{n} \varepsilon_i (\bar{f} - \overline{\pi_{q_1} f}) I_{\{F^* \leq c_{n,q_1}(f)\}}(X_i)\|_{\mathcal{F}} > \eta\}$$

$$+ Pr^*\{\|n^{-\frac{1}{2}} \sum_{i=1}^{n} \varepsilon_i (\bar{f} - \overline{\pi_{q_0} f}) I_{\{F^* > c_{n,q_0+1}(f)\}}(X_i)\|_{\mathcal{F}} > \eta\} \tag{3.13}$$

$$+ Pr^*\{\|n^{-\frac{1}{2}} \sum_{q=q_0+1}^{q_1} \sum_{i=1}^{n} \varepsilon_i (\overline{\pi_q f} - \overline{\pi_{q-1} f}) I_{\{F^* \leq c_{n,q}(f)\}}(X_i)\|_{\mathcal{F}} > \eta\}$$

$$+ Pr^*\{\|n^{-\frac{1}{2}} \sum_{q=q_0+2}^{q_1} \sum_{i=1}^{n} \varepsilon_i (\bar{f} - \overline{\pi_{q-1} f}) I_{\{c_{n,q}(f) < F^* \leq c_{n,q-1}(f)\}}(X_i)\|_{\mathcal{F}} > \eta\}$$

and we have to show that each of these terms tends to zero when n and q_0 go to infinity.

The first term

$$Pr^* \left\{ \| \sum_{i=1}^{n} \varepsilon_i (\bar{f} - \overline{\pi_{q_1} f}) I_{\{F^* \leq \frac{n^{\frac{1}{2}} 2^{-q_1}}{\gamma_{q_1}(f)}\}}(X_i)\|_{\mathcal{F}} > \eta n^{\frac{1}{2}} \right\}$$

vanishes if

$$2n \sup_{f\in\mathcal{F}} \frac{n^{\frac{1}{2}} 2^{-q_1}}{\gamma_{q_1}(f)} \leq \eta n^{\frac{1}{2}},$$

i.e., if

$$2n 2^{-q_1} \leq \eta \inf_{f\in\mathcal{F}} \gamma_{q_1}(f).$$

Since $q_1^{\frac{1}{2}} \leq \gamma_{q_1}(f)$ for all $f \in \mathcal{F}$, it is enough to have

$$\frac{2}{\eta} n \leq q_1^{\frac{1}{2}} 2^{q_1}$$

and this holds taking $q_1(n) \approx \log n$ for all $n > N_0(\eta)$ large enough.

For the second term, we get that

$$Pr^* \left\{ \| \sum_{i=1}^{n} \varepsilon_i (\overline{f} - \overline{\pi_{q_0} f}) I_{\{F^* > c_{n,q_0+1}(f)\}}(X_i) \|_{\mathcal{F}} > \eta n^{\frac{1}{2}} \right\}$$
$$\leq nP \left\{ 2F^* > \inf_{f\in\mathcal{F}} \frac{2^{-(q_0+1)} n^{\frac{1}{2}}}{\gamma_{q_0+1}(f)} \right\}$$
$$\leq nP \left\{ 2F^* > n^{\frac{1}{2}} \frac{2^{-(q_0+1)}}{2^{q_0}} \right\}$$

which tends to zero as n tends to infinity by (i).

To obtain the convergence of the third term, we combine Bernstein's inequality (Lemma 1.1(b)) with properties of the majorizing measure (Lemma 1.5). Let $\triangle_q f = \overline{\pi_q f} - \overline{\pi_{q-1} f}$. For positive numbers $a_1, \ldots, a_n, b_1, \ldots, b_n$,

$$\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \leq \max_{1\leq i\leq n} \frac{a_i}{b_i}.$$

Using this and the definition of β_{q_0}, we can write

$$Pr^* \left\{ \left\| \sum_{i=1}^{n} \sum_{q=q_0+1}^{q_1(n)} \varepsilon_i \triangle_q f I_{\{F^* \leq c_{n,q}(f)\}}(X_i) \right\|_{\mathcal{F}} > \eta n^{\frac{1}{2}} \right\} \tag{3.14}$$
$$\leq Pr^* \left\{ \left\| \frac{\sum_{q=q_0+1}^{q_1(n9)} \sum_{i=1}^{n} \varepsilon_i \triangle_q f I_{\{F^* \leq c_{n,q}(f)\}}(X_i)}{\sum_{q=q_0+1}^{q_1(n)} 2^{-q} \gamma_q(f)} \right\|_{\mathcal{F}} > \frac{\eta n^{\frac{1}{2}}}{\beta_{q_0}} \right\}$$

$$\leq Pr^*\left\{\exists q \in (q_0, q_1(n)] \cap N \;\; \exists \Delta_q f : \frac{\sum_{i=1}^n \varepsilon_i \Delta_q f I_{\{F^* \leq c_{n,q}(f)\}}(X_i)}{2^{-q}\gamma_q(f)} > \frac{\eta n^{\frac{1}{2}}}{\beta_{q_0}}\right\}$$

$$\leq \sum_{q=q_0+1}^{q_1(n)} \sum_{t \in T_q} Pr^*\left\{\sum_{i=1}^n \varepsilon_i \,\Delta_q\, t I_{\{F^* \leq c_{n,q}(t)\}}(X_i) > \frac{\eta n^{\frac{1}{2}} 2^{-q}\gamma_q(t)}{\beta_{q_0}}\right\} \qquad (3.15)$$

where we abuse notation to write $\Delta_q f = \Delta_q t, c_{n,q}(f) = c_{n,q}(t)$ and $\gamma_q(f) = \gamma_q(t)$ if $t = (\pi_1 f, \ldots, \pi_q f)$.

Now Bernstein's inequality implies that (3.15) is less than or equal to

$$\sum_{q=q_0+1}^{q_1(n)} \sum_{t \in T_q} \exp\left\{-\frac{\frac{\eta^2 n (2^{-q})^2 \gamma_q^2(t)}{(\beta_{q_0})^2}}{4(2^{-q})^2 + \frac{2}{3}\frac{\eta n^{\frac{1}{2}} 2^{-q}\gamma_q(t) c_{n,q}(t)}{\beta_{q_0}}}\right\}.$$

Replacing $c_{n,q}(t)$ by its value $\frac{n^{\frac{1}{2}} 2^{-q}}{\gamma_q(t)}$, this expression becomes

$$\sum_{q=q_0+1}^{q_1(n)} \sum_{t \in T_q} \exp\left\{\frac{-\eta\beta_{q_0}^{-2}}{4 + \frac{2}{3}\eta\beta_{q_0}^{-1}}(\gamma_q(t))^2\right\}$$

$$= \sum_{q=q_0+1}^{q_1(n)} \sum_{t \in T_q} \exp\{-r_{q_0}(\eta)\gamma_q^2(t)\} \qquad (3.16)$$

where $r_{q_0}(\eta) \to \infty$ when $q_0 \to \infty$ since β_{q_0} goes to zero by (b). Lemma 1.5 implies that (3.16) tends to zero when q_0 tends to infinity; this shows that (3.14) goes to zero when q_0 tends to infinity.

Finally, to get the convergence to zero of the last term in (3.13), let us fix $n \in N$ and let us define

$$C_{n,q} = \sup_{f \in \mathcal{F}} c_{n,q}(f)$$

$$c_{n,q} = \inf_{f \in \mathcal{F}} c_{n,q}(f) \quad , \quad q \geq q_0$$

and

$$q(n) = \max\{q \leq q_1(n) : q \in N, c_{n,q} > \tau_n n^{\frac{1}{2}}\}.$$

For fixed $k \in N \backslash \{0\}$ and $q \in N$ such that $q(n) < q \leq q_1(n)$, let $P^{(n,q)}$ be the probability measure defined on $(S, \mathcal{S})$ by

$$P^{(n,q)}\{\cdot\} = P\{\cdot \mid F^* > c_{n,q}\}$$

and let $Pr_{n,q} = (P^{(n,q)})^N \times \lambda$ be the product measure defined on $(S^N, \mathcal{S}^N)$ $\times$ $([0,1], \mathcal{B}_{[0,1]})$, where λ is Lebesgue measure on $([0,1], \mathcal{B}_{[0,1]})$. Define $\tilde{\nu}_{n,q}$: $S^N \times [0,1] \to \ell^\infty(\mathcal{F})$, $n \in N$, by

$$\{\tilde{\nu}_{n,q}(f) : f \in \mathcal{F}\} = \left\{ n^{-\frac{1}{2}} \sum_{i=1}^{n} \varepsilon_i(\bar{f} - \overline{\pi_{q-1} f}) I_{\{c_{n,q}(f) < F^* \leq c_{n,q-1}(f)\}}(X_i) : f \in \mathcal{F} \right\}.$$

We want to use now Proposition 2.3 and it is convenient to restrict ourselves to sets of the form

$$A_{n,q,k} = \{nP_n\{F^* > c_{n,q}\} = k\} = \left\{ \sum_{i=1}^{n} I_{\{F^* > c_{n,q}\}}(X_i) = k \right\}$$

because, in this case, we have, for any $\eta > 0$,

$$Pr\{\{\|\tilde{\nu}_{n,q}\|_{\mathcal{F}} > \eta\} \cap A_{n,q,k}\} = Pr^*\{\{\|\tilde{\nu}_{n,q}\|_{\mathcal{F}} > \eta\} \mid A_{n,q,k}\} Pr\{A_{n,q,k}\} \quad (3.17)$$

$$= Pr^*\{\{\|n^{\frac{-1}{2}} \sum_{i=1}^{n} \varepsilon_i(\overline{f} - \overline{\pi_{q-1} f}) I_{\{c_{n,q}(f) < F^* \leq c_{n,q-1}(f)\}}(X_i)\|_{\mathcal{F}} > \eta\} \mid \{\sum_{i=1}^{n} I_{\{F^* > c_{n,q}\}}(X_i)$$

$$= k\}\} Pr\{A_{n,q,k}\}$$

$$\leq (\frac{n}{k}) Pr^*_{n,q}\{\|(\frac{k}{n})^{\frac{1}{2}} \frac{\sum_{i=1}^{k} \varepsilon_i(\overline{f} - \overline{\pi_{q-1} f}) I_{\{c_{n,q} < F^* \leq C_{n,q-1}\}}(X_i)}{k^{\frac{1}{2}}} \|_{\mathcal{F}} > \eta\} \frac{1}{(\frac{n}{k})} Pr\{A_{n,q,k}\}$$

$$= Pr^*_{n,q}\{\{\|(\frac{k}{n})^{\frac{1}{2}} \frac{\sum_{i=1}^{k} \varepsilon_i(\overline{f} - \overline{\pi_{q-1} f}) I_{\{c_{n,q} < F^* \leq C_{n,q-1}\}}(X_i)}{k^{\frac{1}{2}}} \|_{\mathcal{F}} > \eta\} \cap A_{n,q,k}\}$$

(we have used that for any measurable set A, $Pr^*\{\cdot \mid A\} = \frac{Pr^*\{\cdot \cap A\}}{P(A)}$, which is easy to prove).

Chang's (1964) result described in Theorem 2.4 allows us to restrict the values of k that we have to consider. Let

$$R_n = \{1 \leq nP_n\{F^* > c_{n,q}\} \leq 2nP\{F^* > c_{n,q}\} \quad for \quad all \quad q(n) < q \leq q_1(n)\}.$$

We want to show that $Pr^*\{R_n^c\}$ tends to zero when n goes to infinity. We have that

$$Pr\{R_n^c\} \leq Pr\{nP_n\{F^* > c_{n,q}\} < 1 \quad for \quad some \quad q(n) < q \leq q_1(n)\}$$

$$+ Pr\{2P\{F^* > c_{n,q}\} < P_n\{F^* > c_{n,q}\} \quad for \quad some \quad q(n) < q \le q_1(n)\}. \tag{3.18}$$

On the one hand, since $c_{n,q(n)+1} < \tau_n n^{\frac{1}{2}}$ implies (by the definition of τ_n) that

$$nP\{F^* > c_{n,q(n)+1}\} > \phi\eta\delta_n^{-1} \to \infty \quad , \quad n \to \infty, \tag{3.19}$$

we have

$$Pr\{nP_n\{F^* > c_{n,q}\} < 1 \quad for \quad some \quad q(n) < q \le q_1(n)\}$$

$$\le Pr\{nP_n\{F^* > c_{n,q(n)+1}\} < 1\} \le Pr\left\{\sum_{i=1}^{n} I_{\{F^* > c_{n,q(n)+1}\}}(X_i) < 1\right\}$$

$$= (1 - P\{F^* > c_{n,q(n)+1}\})^n \to 0 \quad as \quad n \to \infty.$$

On the other hand, also due to (3.19), we can take

$$b_n = \min_{q(n) < q \le q_1(n)} nP\{F^* > c_{n,q}\}$$

and $\varepsilon = 1$ in Theorem 2.4 to get

$$\lim_{n\to\infty} Pr\{ \sup_{\frac{b_n}{n} \le P\{F^* > x\}} \mid \frac{P_n\{F^* > x\}}{P\{F^* > x\}} - 1 \mid \ge 1\} = 0.$$

Therefore, the second term in (3.18) goes to zero when n tends to infinity.

These simplifications allow us to apply now Proposition 2.3. Let

$$p_{n,q} = P\{F^* > c_{n,q}\}$$

and

$$\rho_{n,q} = \rho_2^{P(n,q)}(f, g), \quad f, g \in \mathcal{F}.$$

For functions f, g supported on $\{F^* > c_{n,q}\}$, we have that

$$\rho_{n,q}(f, g) = p_{n,q}^{-\frac{1}{2}} \rho(f, g)$$

and hence

$$\tilde{\mathcal{F}}'_\rho(2^{-(q-1)}) = \tilde{\mathcal{F}}'_{\rho_{n,q}}(2^{-(q-1)} p_{n,q}^{-\frac{1}{2}})$$

with

$$\mathcal{F}'_{\rho_{n,q}}(\alpha) = \{fI_{\{F^* > c_{n,q}\}} - gI_{\{F^* > c_{n,q}\}} : f, g \in \mathcal{F}, \rho_{n,q}(fI_{\{F^* > c_{n,q}\}}, gI_{\{F^* > c_{n,q}\}}) \le \alpha\}$$

and analogously for $\mathcal{F}'_\rho(\alpha)$.

Finally, we get that the last term in (3.13) is less than or equal to

$$Pr^*\left\{\left\|n^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{q=q_0+2}^{q_1(n)}\varepsilon_i(\overline{f}-\overline{\pi_{q-1}f})I_{\{c_{n,q}(f)<F^*\le c_{n,q-1}(f)\}}(X_i)\right\|_{\mathcal{F}}>\eta\right\}$$

$$=Pr^*\left\{\left\|n^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{q=q(n)+1}^{q_1(n)}\varepsilon_i(\overline{f}-\overline{\pi_{q-1}f})I_{\{c_{n,q}(f)<F^*\le c_{n,q-1}(f)\}}(X_i)\right\|_{\mathcal{F}}>\eta\right\}$$

$$\le Pr^*\left\{\left\|\frac{n^{-\frac{1}{2}}\sum_{i=1}^{n}\sum_{q=q(n)+1}^{q_1(n)}\varepsilon_i(\overline{f}-\overline{\pi_{q-1}f})I_{\{c_{n,q}(f)<F^*\le c_{n,q-1}(f)\}}(X_i)}{\sum_{q=q(n)+1}^{q_1(n)}2^{-q}\gamma_q(f)}\right\|_{\mathcal{F}}>\frac{\eta}{\beta_{q_0}}\right\}$$

$$\le Pr\{R_n^c\}+\sum_{q=q(n)+1}^{q_1(n)}Pr^*\left\{\left\{\left\|n^{-\frac{1}{2}}\sum_{i=1}^{n}\varepsilon_i(\overline{f}-\overline{\pi_{q-1}f})I_{\{c_{n,q}(f)<F^*\le c_{n,q-1}(f)\}}(X_i)\right\|_{\mathcal{F}}>\frac{\eta 2^{-(q-1)}\inf_{f\in\mathcal{F}}\gamma_{q-1}(f)}{\beta_{q_0}}\right\}\cap R_n\right\}$$

$$\le Pr\{R_n^c\}+\sum_{q=q(n)+1}^{q_1}\sum_{k=1}^{2np_{n,q}}Pr^*_{n,q}\left\{\left\|k^{-\frac{1}{2}}\sum_{i=1}^{k}\varepsilon_i g I_{\{F^*\le C_{n,q-1}\}}(X_i)\right\|_{\mathcal{F}_{\rho'_{n,q}}(2^{-(q-1)}p_{n,q}^{-\frac{1}{2}})}\right.$$

$$\left.>\left(\frac{n}{k}\right)^{\frac{1}{2}}\frac{\eta 2^{-(q-1)}\inf_{f\in\mathcal{F}}\gamma_{q-1}(f)}{\beta_{q_0}}\right\}P\{nP_n\{F^*>c_{n,q}\}=k\},\qquad(3.20)$$

where (3.17) was used in the last inequality.

Hence, by Proposition 2.3, choosing

$$a=C_{n,q-1}=\frac{n^{\frac{1}{2}}2^{-(q-1)}}{\inf_{f\in\mathcal{F}}\gamma_{q-1}(f)},\qquad(3.21)$$

$$M = \left(\frac{n}{k}\right)^{\frac{1}{2}} \frac{\eta 2^{-(q-1)} \inf_{f\in\mathcal{F}} \gamma_{q-1}(f)}{\beta_{q_0}}, \tag{3.22}$$

$$\beta = 2^{-(q-1)} p_{n,q}^{-\frac{1}{2}} \tag{3.23}$$

and

$$m = k, \tag{3.24}$$

we obtain that (3.19) is less than or equal to

$$o(1)+2 \sum_{q=q(n)+1}^{q_1(n)} \exp\left\{-\left(\frac{n}{k}\right)^{\frac{1}{2}} \frac{\eta 2^{-(q-1)} \inf_{f\in\mathcal{F}} \gamma_{q-1}(f)}{33\beta_{q_0}} k^{\frac{1}{2}} \left(\frac{n^{\frac{1}{2}} 2^{-(q-1)}}{\inf_{f\in\mathcal{F}} \gamma_{q-1}(f)}\right)^{-1}\right\}$$

$$\leq o(1) + 2 \sum_{q=q(n)+1}^{q_1(n)} \exp\left\{-\frac{\eta}{33\beta_{q_0}} (\inf_{f\in\mathcal{F}} \gamma_{q-1}(f))^2\right\}. \tag{3.25}$$

Since, for every $f \in \mathcal{F}$, $\gamma_{q-1}(f) \geq (q-1)^{\frac{1}{2}}$, (3.25) is less than or equal to

$$o(1) + \sum_{q=q(n)+1}^{q_1(n)} \exp\left\{-\frac{n}{33\beta_{q_0}}(q-1)\right\}$$

which tends to zero when q_0 tends to infinity. It only remains to check that values (3.19) to (3.24) fulfill conditions (2.6) and (2.7) in Proposition 2.3. The left hand side in (2.6) is

$$aM = \frac{n^{\frac{1}{2}} 2^{-(q-1)}}{\inf_{f\in\mathcal{F}} \gamma_{q-1}(f)} \left(\frac{n}{k}\right)^{\frac{1}{2}} \frac{\eta 2^{-(q-1)} \inf_{f\in\mathcal{F}} \gamma_{q-1}(f)}{\beta_{q_0}} = \frac{n(2^{-(q-1)})^2 \eta}{k^{\frac{1}{2}} \beta_{q_0}}$$

and the right hand side is equal to $28k^{\frac{1}{2}}(2^{-(q-1)})^2 p_{n,q}^{-1}$. So it is enough to have

$$\frac{n\eta}{\beta_{q_0}} \geq 28k p_{n,q}^{-1}$$

and this holds for large q_0 since $k \leq 2np_{n,q}$ and β_{q_0} tends to zero when q_0 tends to infinity.

From (2.7) we get

$$\frac{Mm^{\frac{1}{2}}}{a} = \left(\frac{n}{k}\right) \frac{\eta 2^{-(q-1)} \inf_{f\in\mathcal{F}} \gamma_{q-1}(f)}{\beta_{q_0}} k^{\frac{1}{2}} \left(\frac{n^{\frac{1}{2}} 2^{-(q-1)}}{\inf_{f\in\mathcal{F}} \gamma_{q-1}(f)}\right)^{-1} \geq \frac{\eta}{\beta_{q_0}}(q-1) \tag{3.26}$$

since $\inf_{f\in\mathcal{F}} \gamma_{q-1}(f) \geq (q-1)^{\frac{1}{2}}$. The right hand side becomes

$$K_1 \log \frac{a^2}{\beta^2} + K_2 = K_1 \log C^2_{n,q-1} p_{n,q} 2^{2(q-1)} + K_2. \tag{3.27}$$

From hypothesis (i) in the theorem, it follows that

$$\sup_{t>0} t^2 P\{F^* > t\} \leq D < \infty \quad , \quad D \in R, \tag{3.28}$$

and hence,

$$c^2_{n,q} p_{n,q} \leq D. \tag{3.29}$$

Multiplying and dividing by $c^2_{n,q}$ in (3.27) and recalling that $\sup_{f\in\mathcal{F}} \gamma_q(f) \leq 2^q$ and that $\inf_{f\in\mathcal{F}} \gamma_{q-1}(f) \geq (q-1)^{\frac{1}{2}}$, we get that (3.27) is less than or equal to

$$K_1' \log 2^{4(q-1)} + K_2'' = K'''(q-1) + K_2''. \tag{3.30}$$

Comparing (3.26) and (3.30), and recalling that β_{q_0} tends to zero when q_0 tends to infinity, the condition (2.7) in Proposition 2.3 holds. This proves the theorem.

REMARKS. (a) The condition $\sup_{f\in\mathcal{F}} \mid Pf \mid < \infty$ (which is nothing but a centering condition) allows us to affirm that (3.12) is necessary. If this condition is not assumed, we can only say that

$$t^2 P\{F_p^* > t\} \to 0 \quad as \quad t \to \infty,$$

where $F_p(s) = \sup_{f\in\mathcal{F}} \mid f(s) - Pf \mid$. An alternative hypothesis would be to assume that $\{f - Pf : f \in \mathcal{F}\}$ be an admissibly measurable Vapnik-Červónenkis class of functions (see Alexander (1987)).

(b) The stratification used by Alexander (already applied by Alexander and Pyke (1986) in the study of partial sum processes) is based on intervals of the form $(4^k, 4^{k+1}]$ which do not depend on $f \in \mathcal{F}$; on the contrary, in the proof of Theorem 3.1, the strata depend on $c_{n,q}(f)$. This allows us to take advantage of the "lack of uniformity" of the majorizing measure.

(c) This proof can directly be generalized to give the corresponding stable central limit theorems for $1 \leq p < 2$.

REFERENCES

ALEXANDER, K. S. (1987). The central limit theorem for empirical processes on Vapnik-Červonenkis classes of functions. *Ann. Probab.* **15**, 178-203.

ALEXANDER, K. S. and PYKE, R. (1986). A uniform central limit theorem for set-indexed partial sum processes with finite variance. *Ann. Prob.* **14**, 582-597.

ANDERSEN, N. T. (1985). *The calculus of non-measurable functions and sets.* Aarhus Univ, Math. Inst. Various Publ. Seies No. **36**.

ANDERSEN, N. T., GINÉ, E., OSSIANDER, M. and ZINN, J. (1988). The central limit theorem and the law of iterated logarithm for empirical processes under local conditions. *Probab. Theory Related Fields* **77**, 271-305.

ANDERSEN, N. T., GINÉ, E. and ZINN, J. (1988). The central limit theorem for empirical processes under local conditions: the case of Radon infinitely divisible limits without Gaussian component. *Trans. Amer. Math. Soc.* **308**. 603-635.

BASS, R. F. (1984). Law of the iterated logarithm for set-indexed partial-sum processes with finite variance. *Preprint. Univ. Washington.*

BENNETT, G.W. (1962). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **57**, 33-45.

CHANG, L. C. (1964). On the ratio of the empirical distribution function to the theoretical distribution function. *Sel. Transl. Math. Statist. Prob.* **4**, 17-38.

DUDLEY, R. M. (1967). The sizes of compact subsets of Hilbert spaces and continuity of Gaussian processes. *J. Func. Anal.* **1**, 290-330.

DUDLEY, R. M. (1973). Sample functions of the Gaussian process. *Ann.*

Prob. **1**, 66-103.

DUDLEY, R. M. (1984). A course on empirical processes. *Lect. Notes in Math.* **1097**, 1-142. Springer-Verlag, Berlin.

FERNIQUE, X. (1974). Regularité des trajectoires des fonctions aléatoires gaussienes. *Lect. Notes in Math.* **480**, 1-96. Springer-Verlag, Berlin.

GINÉ, E. and ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Prob.* **12**, 929-989.

GINÉ, E. and ZINN, J. (1986). Lectures on the central limit theorem for empirical processes. *Lect. Notes in Math.* **1221**, 50-113.

HOEFFDING, W. (1963). Probability inequalites for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58**, 13-30

HOFFMANN-JΦRGENSEN, J. (1984). *Stochastic Processes on Polish Spaces.* Preprint, Aarhus University.

KOLMOGOROV, A. N. and TIKHOMIROV, V. M. (1959). ε-entropy and ε-capacity of sets in function spaces. *Amer. Math. Soc. Transls.* (ser. 2) **17** (1961), 277-364.

OSSIANDER, M. (1987). A central limit theorem under metric entropy with L_2bracketing. *Ann. Prob.* **15**, 897-919.

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer-Verlag, Berlin.

PRESTON, C. (1972). Continuity properties of some Gaussian processes. *Ann. Math. Statis.* **43**, 285-292.

TALAGRAND, M. (1987). Regularity of Gaussian processes. *Acta Math.*. **159**, 99-149.

VAPNIK. V. N. and ČERVONENKIS, A. Ya. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theor. Prob. Applic.* **16**, 264-280.