On the Prokhorov distance between the empirical process and the associated Gaussian bridge

By L. Le Cam*

Technical Report No. 170
September 1988

*Research Supported by National Science Foundation Grant DMS87-01426.

Department of Statistics
University of California
Berkeley, California

On the Prokhorov distance between the empirical process and the associated Gaussian bridge

By L. Le Cam*
University of California, Berkeley

1) Introduction.

Let P_j $j = 1, \ldots, n$ be probability measures on a σ-field A. Let $\mathbf{F} \subset \mathbf{A}$ be a Vapnik-Červonenkis class. If $\xi_1, \ldots, \xi_n$ are independent observations with $P_j = \mathbf{L}(\xi_j)$ the empirical process Z_n is defined on $\mathbf{A}$ by

$$Z_n(A) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} [I_A(\xi_j) - P_j(A)].$$

Let G_n be the Gaussian process that has mean zero and the same covariance structure as Z_n. Consider the space $\mathbf{B}$ of bounded functions on the V.C. class $\mathbf{F}$ with its uniform norm. It is well known that G_n defines a probability measure, say M_n, on $\mathbf{B}$. Similarly, under standard measurability conditions Z_n will define a measure, say L_n, on a reasonably big σ-field of subsets of $\mathbf{B}$.

We are interested in the Prokhorov distance $\pi(L_n, M_n)$ between these two measures. Our purpose is to prove the following:

There exists a universal function $(v, n) \rightarrow \phi(v, n)$ of the V.C. exponent v of $\mathbf{F}$ and the integer n such that

1) $\phi(v, n) \rightarrow 0$ as $n \rightarrow \infty$

2) $\pi(L_n, M_n) \leq \phi(v, n)$.

Actually this result is known. It has been proved by Pascal Massart in [5] and the proof given below differs little from that of Massart. The reason for rewriting it is to emphasize the universality of the function ϕ. It does not depend on the P_j at all. It depends on the class $\mathbf{F}$ only through the exponent v. The proof will supply a particular ϕ. It is not necessarily the optimal one. The same technique can be used to obtain other results, with bounds that depend on P and on the class $\mathbf{F}$. The proof uses three known results; a) The entropy bound of R.M. Dudley [2] with the chain argument; b) a bound on the norm of empirical processes on V.C. classes consisting of sets with small probability and c) a theorem of Yurinskii on the Central Limit theorem in $\mathbf{R}^k$.

---

* Research Supported by National Science Foundation Grant DMS87-01426.

Measurability requirements are essentially ignored at first. They are discussed in the last section.

One could let the class **F** depend on n. The measures P_j, $j = 1, \ldots, n$ can also depend on n in an arbitrary manner. One can also obtain results where the Vapniks exponent v of **F** (and **F**) varies with n provided that one puts restrictions on the entropy integrals of the form $\int_0^{\varepsilon} L_n(x^2)\,dx$ with $L_n^2(x) = \log \frac{K_n^2(x)}{2x}$ that occurs in the application of the chaining argument. For further results, see Massart [1986].

For applications and v fixed $\phi(v,n)$ can be taken equal to $C(v)\frac{1}{n^{\gamma(v)}}$ where $\gamma(v) = \frac{1}{9+20v}$ and where C is a certain universal function of v.

Improvements in the exponent $\gamma(v)$ seem possible but they depend on improvements in the Central Limit theorem in $\mathbf{R}^k$ for the uniform norm and arbitrary covariance matrices. The literature contains several such improvements. In Section ~~8~~ we discuss briefly the possible use of a theorem of Zaitsev (1987).

2) A theorem of Yurinskii.

Consider the k-dimensional space $\mathbf{R}^k$ and provide it with a norm $\|\cdot\|$. That norm will be called "Hilbertian" if it satisfies the median equality $\|\frac{x+y}{2}\|^2 + \|\frac{x-y}{2}\|^2 = \frac{1}{2}[\|x\|^2 + \|y\|^2]$.

Theorem 1. *Let* $X_1, X_2, \ldots, X_n$ *be independent random variables with values in* $\mathbf{R}^k$. *Assume* $EX_j = 0$ *and* $E\|X_j\|^3 < \infty$. *Let* F *be the distribution of the sum* $S = X_1 + X_2 + \ldots + X_n$ *and let* G *be the Gaussian measure with the same expectation and covariance structure as* F. *Then:*

1) If $\|\cdot|$ *is Hilbertian then*

$$\pi(F,G) \;\le\; 8\{(S_3k)[1 + \frac{5}{8}\,\frac{1}{k}\,|\log S_3k|]\}^{1/4}$$

2) *If* $\|\cdot\|$ *is arbitrary, then*

$$\pi(F,G) \;\le\; 8\{(S_3k^4)[1 + \frac{5}{8}\,\frac{1}{k}\,|\log(S_3k^4)|]\}^{1/4}$$

3) *If* $\|\cdot\|$ *is the maximum coordinate norm of* $\mathbf{R}^k$, *then*

$$\pi(F,G) \;\le\; 8\{S_3k^{5/2})[1 + \frac{5}{8}\,\frac{1}{k}\,|\log(S_3k^{5/2})|]\}^{1/4}$$

where $S_3 = \frac{5}{3} \Sigma E \| X_j \|^3$

Remark. Statement (1) is imitated from Yurinskii (1977). It is not quite the same as Yurinskii's because we had difficulties following some of his arguments and may have redone them differently.

Statements (2) and (3) are obtained from statement (1), following a technique used by Dehling (1983).

Proof. Take a function f defined on $\mathbb{R}^k$. Assume that f is twice differentiable. This can be taken to mean that for each $x \in \mathbb{R}^k$ there is a vector $A(x)$ and a matrix $B(x)$ such that $\frac{1}{\|y\|^2} |f(x+y) - f(x) - A(x)y - \frac{1}{2} y' B(x) y|$ tends to zero as $\|y\| \to 0$. Assume in addition that B satisfies a Lipschitz condition

$$|B(x) - B(z)| \leq C \|x - z\|$$

so that $|z'[B(x_1) - B(x_2)]y| \leq C \|x_1 - x_2\| \|z\| \|y\|$.

Let $Y_1, \ldots, Y_n$ be independent, independent of the X_j with $EY_j = EX_j$ and $EY_j Y_j' = EX_j X_j'$. Let $S = \Sigma X_j$, $T = \Sigma Y_j$. Then Lindeberg's argument shows that

$$|Ef(S) - Ef(T)| \leq \frac{C}{6} \{\Sigma E \|X_j\|^3 + \Sigma E \|Y_j\|^3\}.$$

Now assume that $\|\cdot\|$ is Hilbertian and Y_j Gaussian. Then

$$\begin{aligned} E\|Y_j\|^3 &\leq 4[E\|Y_j\|^2]^{3/2} = 4[E\|X_j\|^2]^{3/2} \\ &\leq 4E\|X_j\|^3. \end{aligned}$$

Thus

$$|Ef(S) - Ef(T)| \leq \frac{5}{6} C \sum_j E \|X_j\|^3$$

Now, assuming again that $\|\cdot\|$ is Hilbertian, let us create some function f that satisfy the above differentiability conditions.

Let F_i, $i = 1,2$ be two disjoint closed subsets of $\mathbb{R}^k$. Let $\rho(x, F_i) = \inf\{\|x - y\|; y \in F_i\}$ and $\rho = \rho(F_1, F_2) = \inf_x \{\rho(x, F_2) : x \in F_1\}$. Then the function g defined by $g(x) = \frac{\rho(x, F_2)}{\rho(x, F_1) + \rho(x, F_2)}$ satisfies the Lipschitz condition $|g(x_1) - g(x_2)| \leq \frac{1}{\rho} \|x_1 - x_2\|$. Let H_α be the Gaussian density whose exponential term is $\exp\{-\frac{1}{2\alpha^2} \|x\|^2\}$.

Consider the function $f(x) = Eg(x+y)$ where y has distribution H_α. This may also be written

$$K \int g(x+y) \exp\{-\frac{1}{2\alpha^2} \|y\|^2\} \, dy$$
$$= K \int g(y) \exp\{-\frac{1}{2\alpha^2} \|y-x\|^2\} \, dy.$$

Look at $f(x+\lambda z)$ and take derivatives with respect to λ. The second derivative of $\exp\{-\frac{1}{2\alpha^2}\|y-x-\lambda z\|^2\}$ is given by the same exponential multiplied by

$$\frac{1}{\alpha^4}[(y-x)'z - \lambda z'z]^2 - \frac{1}{\alpha^2}\|z\|^2$$
$$= \frac{1}{\alpha^4}[(y-(x+\lambda z)')z]^2 - \frac{1}{\alpha^2}\|z\|^2.$$

Taking as a new variable of integration $y-(x+\lambda z)$ one obtains

$$\frac{\partial^2}{\partial\lambda^2} f(x+\lambda z) = \frac{1}{\alpha^4} E g[y+(x+\lambda z)] \times [(y'z)^2 - \alpha^2\|z\|^2].$$

Therefore

$$\left|\frac{\partial^2}{\partial\lambda^2} f(x+\lambda z) - \frac{\partial^2}{\partial\lambda^2} f(x+\lambda z)|_{\lambda=0}\right|$$
$$\leq \frac{1}{\alpha^4} E|g(y+x+\lambda z) - g(y+x)|\,|[z'y]^2 - \alpha^2\|z\|^2|$$
$$\leq \frac{1}{\rho\alpha^4}\lambda\|z\| E|(z'y)^2 - \alpha^2\|z\|^2|$$
$$\leq \frac{2}{\rho\alpha^2}\lambda\|z\|\,\|z\|^2 \leq \frac{2}{\rho\alpha^2}\|z\|^3.$$

Thus the derivative $B(x)$ of the function f will satisfy the Lipschitz condition

$$|y'[B(x_1) - B(x_2)]y| \leq \frac{2}{\rho\alpha^2}\|x_1 - x_2\|\,\|y\|^2$$

Now take any arbitrary closed set A. Let $A^\beta = \{x; \rho(x,A) < \beta\}$. Apply the foregoing to $F_1 = \overline{A^\beta}$ and $F_2 = (A^{\beta+\rho})^c$. This will give a certain function f whose second derivative satisfies a Lipschitz condition with coefficient $C \leq \frac{2}{\rho\alpha^2}$. In addition for $x \in A$ one will have $f(x) \geq 1 - \Pr[\alpha^2\chi_k^2 > \beta^2]$. For $x \in (A^{2\beta+\rho})^c$ one will have $f(x) \leq \Pr[\alpha^2\chi_k^2 > \beta^2]$. If we let $\varepsilon = \Pr[\alpha^2\chi_k^2 > \beta^2]$ this will yield

$$P(S \in A) \leq Ef(S) + \varepsilon$$

$$\leq \; Ef(T) \; + \; \frac{5}{3}\,\frac{1}{\rho\alpha^2}\sum_j E\|X_j\|^3 \; + \; \varepsilon$$

$$\leq \; \Pr[\,T \in A^{2\beta+\rho}\,] \; + \; \varepsilon + \frac{5}{3}\,\frac{1}{\rho\alpha^2}\sum_j E\|X_j\|^3 \; + \; \varepsilon$$

$$\leq \; \Pr[\,T \in A^{2\beta+\rho}\,] \; + \; 2\varepsilon + \frac{5}{3}\,\frac{1}{\rho\alpha^2}\sum_j E\|X_j\|^3.$$

To obtain a bound it will be sufficient to select α, β and ρ in an appropriate manner. To do this note that $\Pr[\,\alpha^2\chi_k^2 > \beta^2\,]$ is the probability that a gamma variable, say V, with density $\dfrac{1}{\Gamma(\frac{k}{2})}\, e^{-x}\, x^{\frac{k}{2}-1}$ be larger than $\dfrac{\beta^2}{2\alpha^2}$. To bound this probability note that $E\,e^{tV} = (1-t)^{-k/2}$ for $t \in (0,1)$. Thus $\Pr[\,V > v\,] \leq (1-t)^{-k/2}\,e^{-tv}$.

Write $v = \dfrac{k}{2}\,z$ and minimize with respect to t. This gives

$$\Pr[\,V > v\,] \;\leq\; \exp\{-k/2\,[\,z - 1 - \log z\,]\}.$$

Treating $z > 1$ as fixed for the time being we get an inequality of the type

$$\Pr[\,S \in A\,] \;\leq\; \Pr[\,T \in A^{2\beta+\rho}\,] + 2\varepsilon \;+\; \frac{S_3}{\rho\alpha^2}$$

where $\dfrac{\beta^2}{\alpha^2} = kz$ and $\varepsilon = \exp\{-k/2\,[\,z - 1 - \log z\,]\}$ and $S_3 = \dfrac{5}{3}\,\Sigma E\|X_j\|^3$. This suggests taking

$$2\beta \;+\; \rho \;=\; \frac{S_3}{\rho\alpha^2} \;=\; \frac{S_3\,kz}{\rho\beta^2}.$$

This will give a value of ρ such that $(\rho + \beta)^2 = \dfrac{S_3\,kz}{\beta^2} + \beta^2$ and

$$2\beta + \rho \;=\; \beta + \{\frac{S_3kz}{\beta^2} + \beta^2\}^{1/2} \;\leq\; 2\beta + \frac{1}{\beta}\,(S_3\,kz)^{1/2}.$$

Minimizing with respect to β we obtain

$$\beta \;=\; \frac{1}{2}\,[\,S_3\,kz\,]^{1/4}$$

and

$$2\beta + \rho \;\leq\; 3[\,S_3\,kz\,]^{1/4}.$$

Finally we get

$$\pi(F,G) \;\leq\; 3\,[\,S_3\,kz\,]^{1/4} \;+\; 2\exp\{-k/2\,[\,z - 1 - \log z\,]\}$$

Since the Prokhorov distance never exceeds unity one can restrict all considerations to the case where $3(S_3 k)^{1/4} < 1$.

To select a value of z one can try to equate $(S_3 kz)^{1/4}$ and $\exp\{-k/2[z - 1 - \log z]\}$.

This yields the equality

$$z - 1 - (1 - \frac{1}{2k}) \log z = \frac{1}{2k} |\log S_3 k|.$$

This equality can also be written $x[1 - \frac{2k-1}{2k} \frac{\log(1+x)}{x}] = \frac{1}{2k} |\log S_3 k|$ with $x = z - 1$. Since $\frac{1}{x} \log(1 + x)$ is decreasing, it follows that if x is substantial then it will be about equal to $\frac{1}{2k} |\log S_3 k|$. We can assume, arbitrarily, $x \geq 15$. For such values, $x \leq \frac{5}{8k} |\log S_3 k|$. Then we shall have

$$\pi(F, G) \leq 8\{(S_3 k)[1 + \frac{5}{8k} |\log S_3 k|]\}^{1/4}$$

For the second statement in the theorem one can proceed as follows. See [4] pages 16-17.

Let $\|\cdot\|$ be an arbitrary norm on $\mathbf{R}^k$. Let B denote the unit ball of $(\mathbf{R}^k, \|\cdot\|)$ and let B^* be the unit ball dual to B.

Take an arbitrary $x_1 \in B$ with $\|x_1\| = 1$. There is a $y_1 \in B^*$ such that $< y_1, x_1 > = 1$. Let $B_1 = \{x : x \in B, < y_1, x > = 0\}$. Take a $x_2 \in B_1$ with $\|x_2\| = 1$ and a $y_2 \in B^*$ with $< y_2, x_2 > = 1$. Continue with $B_2 = \{x : x \in B, < y_i, x > = 0, i = 1,2\}$ and so forth. This gives a set of k pairs (x_i, y_i) with $< y_i, x_j > = \delta_{i,j}$, the Kronecker δ.

Now consider the norm $|\cdot|$ defined by $|x|^2 = \frac{1}{k} \sum_{j=1}^{k} |< y_j, x >|^2$.

Clearly $|x| \leq \max_j |< y_j, x >| \leq \|x\|$. Also, by convexity $\|x\| \leq \sum_j |< y_j, x >| \leq k\{\frac{1}{k} \Sigma |< y_j, x >|^2\}^{1/2} = k|x|$. Thus $|x| \leq \|x\| \leq k|x|$ and $|\cdot|$ is a Hilbertian norm. The foregoing argument applies to $|\cdot|$. Consider $A_\gamma = \{x : \inf_{y \in A} \|y - x\| \leq \gamma\}$. If $\gamma \geq (2\beta + \rho)k$ then $A_\gamma \supset A^{2\beta+\rho}$. Thus we have $P(S \in A) \leq P[T \in A_{(2\beta+\rho)k}] + \frac{\overline{S}_3 kz}{\rho\beta^2} + 2\exp\{-\frac{1}{2}k[z - 1 - \log z]\}$ as before. Here

$$\overline{S}_3 = \frac{5}{3} \Sigma E|X_j|^3 \leq \frac{5}{3} \Sigma E\|X_j\|^3 \leq S_3.$$

This suggest taking

$$(2\beta + \rho)k = \frac{S_3 kz}{\rho\beta^2} \text{ or equivalently}$$

$$(2\beta + \rho)\rho = \frac{S_3 z}{\beta^2}, \text{ which yields}$$

$$2\beta + \rho \le 2\beta + \frac{1}{\beta}(S_3 z)^{1/2}$$

and, minimizing with respect to β

$$(2\beta + \rho)_{\min} \le 3(S_3 z)^{1/4}.$$

Thus we have

$$\pi(F,G) \le 3k(S_3 z)^{1/4} + 2\exp\{-k/2[z - 1 - \log z]\}.$$

Selection z by the same procedure as before we get

$$\pi(F,G) \le 8\{S_3 k^4 [1 + \frac{5}{8}\frac{1}{k}|\log(S_3 k^4)|]\}^{1/4}.$$

The same argument applies to statement (3) of the theorem except that now the sup norm $\|x\| = \max_j |< y_j, x >|$ satisfies the inequalities $|x| \le \|x\| \le |x|\sqrt{k}$ for the Hilbertian norm defined by $|x|^2 = \frac{1}{k}\sum_j |< y_j, x >|^2$.

Remark. In the sequel we shall use statement (3) of the theorem under conditions where the actual dimension of the space is not known but where it is known to be bounded by a given k. The bounds are still applicable since the bound in statement (3) is monotone increasing as function of k.

3) **Reduction to the finite dimensional case.**

In this section **F** will be a V.C. class of sets and Z_n will be the empirical process defined in the introduction. For each $x > 0$, $\mathbf{F}_x$ will be a minimal subset of **F** with the property that $\sup_F \inf_S \{P[F\Delta S]; S \in \mathbf{F}_x, F \in \mathbf{F}\} \le x$. The cardinality of $\mathbf{F}_x$ will be $K(x)$ and $L(x)$ will be $\left[\log \frac{K^2(x)}{2x}\right]^{1/2}$.

If v is the exponent of the V.C. class it is known (see Dudley (1978), Le Cam (1986)) that

$$K(x) \le \exp\{2v[\log v + \log 2/x]\}.$$

The following lemma is a version of a chain argument used by many authors

Lemma 1. *Let* m *be an integer and let* $\alpha \in [0, 1/2]$ *be such that* $4L^2(\alpha) \le n\alpha$. *Then there is a map* τ *from* $\mathbf{F}_\alpha$ *to* $\mathbf{F}_{\alpha 4^m}$ *such that*

$$\sup\{|Z_n(A) - Z_n[\tau(A)]|;\ A \in \mathbf{F}_\alpha\}$$
$$\le\ 32 \int_{\sqrt{\frac{\alpha}{2}}}^{(2^m-1)\sqrt{\alpha}} L(x^2)\,dx$$

except for cases whose total probability does not exceed $4(2^m - 1)\sqrt{2\alpha}$.

This will allow us to approximate the class $\mathbf{F}_\alpha$ by the class $\mathbf{F}_{\alpha 4^m}$ whose cardinality is or can be very much smaller than that of $\mathbf{F}_\alpha$.

It will remain to approximate $\mathbf{F}$ itself by $\mathbf{F}_\alpha$. Later on the variable α will be made to depend on n.

To pass from $\mathbf{F}_\alpha$ to $\mathbf{F}$, introduce the class $\mathbf{D}_\alpha$ of sets of differences $S \setminus \xi(S)$ or $\xi(S) \setminus S$ where $\xi(S)$ is selected in $\mathbf{F}_\alpha$ so that $P[S \Delta \xi(S)] \le \alpha$.

It is known (see Dudley (1978)) that $\mathbf{D}_\alpha$ is a V.C. class of exponent at most 2v.

Consider a set of pairs $W = \{(x_j, y_j) : j = 1,2, \dots, n\}$ of points in the space $\mathbf{X}$ that carries $\mathbf{F}$. If $S \in \mathbf{D}_\alpha$ it determines a "pattern" on W as follows: The pattern of S on W is a sequence $\{u_j; j = 1, \dots, n\}$ with $u_j = 1$ if $x_j \in S$, $y_j \notin S$. It is $u_j = -1$ if $x_j \notin S$, $y_j \in S$. Otherwise $u_j = 0$.

The number of different patterns carved out on W by the elements S of $\mathbf{D}_\alpha$ will be called $M(\mathbf{D}_\alpha, W)$.

It is known (see Dudley (1978), Le Cam (1986)), that $M(\mathbf{D}_\alpha, W) \le (2n)^{2v}$ where v is the V.C. exponent of $\mathbf{F}$.

Consider also the sum $\nu(S,W) = \Sigma |u_j(S)|$ for the pattern carved by S on W and let $N(W) = \sup\{\sum_j |u_j(S)|;\ S \in \mathbf{D}_\alpha\}$.

According to Le Cam (1986), Lemma 6 page 546

Lemma 2. *Let conditions (A) and (B) of Le Cam (1986) page 545 hold. Let* $x_n^2(\varepsilon) = 2\log \frac{2(2n)^{2v}}{\varepsilon}$. *Then*

1) $\Pr^*\{N(W) \ge 2[x_n(\varepsilon) + \sqrt{n\alpha+1}\,]^2\} \le 8\varepsilon$

2) $\Pr^*\{\sup|Z_n(S)|;\ S \in \mathbf{D}_\alpha$
$$\ge\ \frac{1}{\sqrt{n}}[1 + \sqrt{2}\,x_n(\varepsilon)[x_n(\varepsilon) + \sqrt{n\alpha+1}\,]]\} \le 20\varepsilon.$$

For this to be true one needs some conditions on $\mathbf{D}_\alpha$ but they are of the nature of measurability restrictions. They are satisfied in most usual applications.

To combine Lemma 1 and Lemma 2 we need to select a value α_n of α. According to Lemma 1 one should take α_n so that $4L^2(\alpha_n) \le n\alpha_n$.

Recall that $L^2(x) = \log \frac{K^2(x)}{2x} = 2\log K(x) + \log \frac{1}{2x} \le (2v+1)\log \frac{1}{x} + 2v \log 2v$. One can verify that

$$\alpha_n = 4(2v+1)\frac{\log n}{n}$$

will satisfy the desired inequality. If so the bound in Lemma 2, for the probability (20)ε will become

$$\frac{1}{\sqrt{n}} + \frac{\sqrt{2}}{\sqrt{n}}[4v\log 2n + 2\log\frac{2}{\varepsilon}]$$
$$+ \frac{\sqrt{2}}{\sqrt{n}}[\sqrt{4(2v+1)\log n + 1}\,][4v\log 2n + 2\log\frac{2}{\varepsilon}]^{1/2}.$$

Keeping ε fixed for the time being, assume that $\varepsilon^2 > \alpha_n$ and let m be the smallest integer such that $4^m\alpha_n \ge \varepsilon^2$. That is $m = \mathrm{int}\left[\frac{1}{\log 4}\log\frac{\varepsilon^2}{\alpha_n}\right] + 1$. The class selected in Lemma 2 is a class $\mathbf{F}_{\alpha_n}$. Lemma 1 can be used to pass to the class $\mathbf{F}_{\alpha_n 4^m}$ selecting sets $\xi_1(S) \in \mathbf{F}_{\alpha_n 4^m}$ so that $\sup_S \{|Z_n(S) - Z_n[\xi_1(S)]|,\ S \in \mathbf{F}\} \le 32\int_0^{\varepsilon\sqrt{2}} L(x^2)\,dx$. Since $L^2(x) \le (2v+1)\log\frac{1}{x} + 2v\log 2v$ one has

$$L(x) \le \{(2v+1)\log\frac{1}{x}\}^{1/2} + [2v\log 2v]^{1/2}$$

and

$$\int_0^{\varepsilon\sqrt{2}} L(x^2)\,dx \le (4v+2)^{1/2}\int_0^{\varepsilon\sqrt{2}}(\log\frac{1}{x})^{1/2}dx + \varepsilon[4v\log 2v]^{1/2}$$

taking $y = \log\frac{1}{x}$ the integral $\int_0^{\varepsilon\sqrt{2}}(\log\frac{1}{x})^{1/2}dx$ can be written

$$\int_{\log\frac{1}{\varepsilon\sqrt{2}}}^{\infty} e^{-y}\sqrt{y}\,dy = e^{-y}\sqrt{y}\Big|_{\log\frac{1}{\varepsilon\sqrt{2}}} + \int_{\log\frac{1}{\varepsilon\sqrt{2}}}^{\infty}\frac{1}{2\sqrt{y}}e^{-y}\,dy$$
$$\le (\varepsilon\sqrt{2})\{[\log\frac{1}{\varepsilon\sqrt{2}}]^{1/2} + \frac{1}{2}\}$$

assuming $\varepsilon\sqrt{2} \leq 1$.

Combining the two maps ξ and ξ_1 one obtains a map τ from **F** to $\mathbf{F}_{\alpha_n 4^m}$ such that except for total outer probability at most $28\,\varepsilon$ one has

$$\sup_S \{|Z_n(S) - Z_n[\tau(S)]|;\ S \in \mathbf{F}\}$$

$$\leq\ 64\sqrt{2v+1}\,\varepsilon\{[4v\log 2v]^{1/2} + \frac{1}{2} + (\log\frac{1}{\varepsilon\sqrt{2}})^{1/2}\} + \frac{1}{\sqrt{n}}$$

$$+ \sqrt{\frac{2}{n}}[4v\log 2n + 2\log\frac{2}{\varepsilon}]$$

$$+ \sqrt{\frac{2}{n}}\{[4(2v+1)\log n + 1][4v\log 2n + 2\log\frac{2}{\varepsilon}]\}^{1/2}.$$

A similar argument can be applied to the Gaussian process, say Z_∞, that has the same covariance structure as Z_n. However, here one can pass directly from **F** to $\mathbf{F}_{\alpha_n 4^m}$ with a map τ such that, except for probability at most 8ε will satisfy

$$\sup_S \{|Z_\infty(S) - Z_\infty[\tau'(S)]|:\ S \in \mathbf{F}\}$$

$$\leq\ (64)(\sqrt{2v+1})\ \varepsilon\{[4v\log 2v]^{1/2} + \frac{1}{2} + (\log\frac{1}{\varepsilon\sqrt{2}})^{1/2}\}.$$

Note that since $\alpha_n 4^m \geq \varepsilon^2$ the cardinality of $\mathbf{F}_{\alpha_n 4^m}$ is at most

$$K(\varepsilon^2)\ \leq\ \exp\{2v\log 2v\ +\ 4v\log\frac{1}{\varepsilon}\}$$

$$=\ (2v)^{2v}\frac{1}{\varepsilon^{4v}}.$$

If one compares the bounds obtained above with the bound of theorem 1 with $k = K(\varepsilon^2)$ this suggest the choice of an ε of the type $\varepsilon = \frac{1}{n^\gamma}$. We shall make a particular choice of γ in the next section.

4) **A bound on the Prokhorov distance.**

We are now in a position to prove the following result:

Theorem 2. *Let* **F** *be a V.C. class of subsets of* **X** *with exponent* v. *Assume that suitable measurability restrictions are satisfied. Let* Z_n *be the empirical process for* n *independent observations* $\xi_1, \xi_2, \ldots, \xi_n$ *with arbitrary distributions* $L(\xi_j) = P_j$.

Let Z_∞ *be the Gaussian process defined on* **F** *with the same covariance structure as* Z_n.

Then one can construct a probability space Ω and processes Z_n^* and Z_∞^* defined on Ω and such that

1) $\mathbf{L}(Z_n) = \mathbf{L}(Z_n^*)$ *and* $\mathbf{L}(Z_\infty^*) = \mathbf{L}(Z_\infty)$,

and

2) $\Pr^*\{\sup_s |Z_n^*(S) - Z_\infty^*(S)| \ge \phi(v,n)\} \le \phi(v,n)$ *where for* $\gamma = [8 + 20v]^{-1}$ *one has*

$$n^\gamma \phi(v,n) \le C_3(v) + C_2(v)\sqrt{\log n}$$

with $C_2(v) = (128)\sqrt{\gamma(2v+1)}$ *and*

$$C_3(v) = C_1'(v) + \max(36, 9(2v)^{5v/4}),$$

$$C_1'(v) = (128)[(2v+1)4v\log 2v]^{1/2} + (\sqrt{2})(4+\sqrt{2})(2v+1)\frac{(8+20v)}{(3+10v)}.$$

Proof. For the time being we shall not bother about the measurability restrictions. They will be debated in the next section.

Choose α_n as described in Section 3. Let $\varepsilon_n = \frac{1}{n^\gamma}$.

Consider the empirical processes $Z_{n,r}$ equal to Z_n restricted to the class $\mathbf{F}_{\alpha_n 4^n}$ described in Section 3. On this class the processes can be considered as random vectors in $\mathbf{R}^{k(n)}$ with a dimension $k(n) \le (2v)^{2v} n^{4v\gamma}$. According to Theorem 1, and the remark that follows its proof, one can match $Z_{n,r}$ with $Z_{\infty,r}$ in such a way that, for the uniform norm $\|\cdot\|$ on functions on $\mathbf{F}_{\alpha_n 4^n}$,

$$\Pr\{\|Z_{n,r} - Z_{\infty,r}\| \ge \psi(v,n)\} \le \psi(v,n)$$

where $\psi(v,n)$ is the function

$$\psi(v,n) = 5(2v)^{5v/4} n^{5v\gamma/2 - 1/8} \times \left[1 + \frac{5}{8k(n)}|\log\frac{[k(n)]^{5/2}}{\sqrt{n}}|\right]^{1/4}.$$

To obtain a Z_n^* reconstruct all the $Z_n(S)$ from the conditional probabilities of $\{Z_n(S); S \in \mathbf{F}\}$ given $\{Z_n(S); S \in \mathbf{F}_{\alpha_n 4^n}\}$.

Proceed similarly for Z_∞^*. According to Section 3 one will have

$$\Pr\{\sup_S |Z_n^*(S) - Z_\infty^*(S)| \ge \psi(v,n) + \psi_1(v,n)\}$$

$$\le 36\varepsilon_n + \psi(v,n)$$

for a function $\psi_1(v,n)$ equal to

$$(128)\sqrt{2v+1}\,\frac{1}{n^\gamma}\{[4v\log 2v]^{1/2} + \frac{1}{2} + (\gamma\log n + \log\frac{1}{\sqrt{2}})^{1/2}\} + \psi_2(v,n),$$

$$\psi_2(v,n) = \frac{1}{\sqrt{n}} + \sqrt{\frac{2}{n}}\,[\,(4v+2)\log 2 + (4v+2\gamma)\log n\,]$$

$$+ \sqrt{\frac{2}{n}}\,\{[\,1 + 4(2v+1)\log n\,]\,[\,(4v+2)\log 2 + (4v+2\gamma)\log n\,]\}^{1/2}.$$

Equating the powers of n in front of $\psi(v,n)$ and $\psi_1(v,n)$ suggests taking $\gamma = \frac{1}{8+20v}$. Note that then $\psi(v,n)$ and the first part of $\psi_1(v,n)$ start with powers $n^{-\gamma}$. By contrast $\psi_2(v,n)$ starts with powers $n^{-1/2} = n^{-\gamma-\gamma_1}$ where $\gamma_1 = \frac{1}{2} - \gamma = \frac{3+10v}{8+20v}$. Using repeatedly the argument that for $A < B$ and $x \in (0,1)$ the maximum of $x[\,A - B\log x\,]$ is not more than B one obtains that

$$\frac{1}{n^{\gamma_1}}[\,1 + \sqrt{2}\,(4v+2)\log 2 + \sqrt{2}\,(4v+2\gamma)\log n\,] \;\le\; \frac{1}{\gamma_1}\sqrt{2}\,(4v+2\gamma) \text{ and}$$

$$\frac{1}{n^{\gamma_1}}[\,1 + 4(2v+1)\log n\,] \;\le\; \frac{1}{\gamma_1}4(2v+1).$$

Thus

$$\begin{aligned} n^{\gamma}\psi_2(v,n) &\le \frac{\sqrt{2}}{\gamma_1}\{(4v+2\gamma) \;+\; [\,[\,4(2v+1)\,]\sqrt{2}\,[\,4v+2\gamma\,]\,]^{1/2} \\ &\le \frac{\sqrt{2}}{\gamma_1}[\,(2+\sqrt{2})(2v+\gamma) + 2(2v+1\,] \\ &\le \frac{\sqrt{2}}{\gamma_1}(4+\sqrt{2})(2v+1). \end{aligned}$$

This yields

$$n^{\gamma}\psi_1(v,n) \;\le\; C_1'(v) \;+\; C_2(v)\sqrt{\log n}$$

where

$$C_1'(v) \;=\; (128)\sqrt{2v+1}\,(4v\log 2v)^{1/2} \;+\; \frac{\sqrt{2}}{\gamma_1}(4+\sqrt{2})(2v+1)$$

and

$$C_2(v) \;=\; 128\sqrt{\gamma(2v+1)}.$$

Finally

$$n^{\gamma}\psi(v,n) \;\le\; 8(2v)^{5v/4}[\,1 + \frac{5}{8k(n)}\,|\log\frac{k(n)^{5/2}}{\sqrt{n}}\,|\,]^{1/4}$$

with $k(n) \le (2v)^{2v}n^{4v\gamma}$. By the remark made after the proof of Theorem 1 one can replace $k(n)$ by that upper bound. Using the same argument about function of the type x $[A - B \log x]$ and noting that $(2+5v)\gamma = 1/4$ it can be seen that, whenever $(2v)^{5v} \le n^{4\gamma}$, one will have $1 + \frac{5}{8}\frac{1}{k(n)}|\log \frac{(k(n))^{5/2}}{\sqrt{n}}|$ smaller than $1 + \frac{5}{8}\frac{1}{v(2v)^{2v}}$. This will yield

$$n^{\gamma}\psi(v,n) \le 8(2v)^{5v/4}[1 + \frac{5}{8}\frac{1}{v(2v)^{2v}}]^{1/4}.$$

Then

$$n^{\gamma}[\psi(v,n) + \psi_1(v,n)] \le C_1(v) + C_2(v)\sqrt{\log n}$$

with $C_2(v) = (128)\gamma\sqrt{2v+1}$ as above and

$$C_1(v) = C_1'(v) + 9(2v)^{5v/4}$$

where we have replaced $1 + \frac{5}{8}\frac{1}{v(2v)^{2v}}$ by 41/36 for $v \ge 1$.

This yields a function $\phi(v,n)$ of the type described in the Theorem

(Note. The exponent $\gamma = \frac{1}{8+20v}$ obtained here seems smaller than the exponent in Massart [5]. However Massart 2d is the same as our 4v. Thus the rates of convergence are about the same). (See Section 6 for modifications)

5) **Measurability conditions.**

The arguments of Section 4 require certain measurability restrictions for their validity. The conditions can be stated as follows:

A) Let **D** be the class of differences $S_1 \setminus S_2$, $S_i \in \mathbf{F}$. Let $\{\xi_1, \xi_2, \ldots, \xi_n; \eta_1, \eta_2, \ldots, \eta_n\} = W$ be a sample of independent variables with $\mathbf{L}(\xi_j) = \mathbf{L}(\eta_j)$. Let ε_j; $j = 1,2,\ldots,n$ be independent variables, independent of W, with $\Pr[\varepsilon_j = 1] = \Pr[\varepsilon_j = -1] = 1/2$. Then

$$\{\sum_j \varepsilon_j(\delta_{\xi_j} - \delta_{\eta_j})(S);\ S \in \mathbf{D}\}$$

has the same distribution as $\{\sum_j(\delta_{\xi_j} - \delta_{\eta_j})(S);\ S \in \mathbf{D}\}$ and that distribution can be obtained by first conditioning on W.

B) Let m(S) be a median $\frac{1}{\sqrt{n}}\sum_j[\delta_{\xi_j}(S) - P_j(S)] = Z_n(S)$. Then

$$\Pr\{\sup_S[|Z_n(S) - m(S)|;\ S \in \mathbf{S}] \ge x\} \le 2\Pr\{[\sup|Z_n(S) - Z_n'(S)|;$$

$S \in \mathbf{S}] \ge x\}$ for Z_n' a copy of Z_n independent of it and for any subclass **S** of **D**.

Dudley (1978) has given conditions that are sufficient for the validity of (A) (B).

We have stated (A) (B) in terms of probabilities. However it is sufficient that they be valid with outer probabilities.

Here one may note that Dudley's conditions are meant to apply to the empirical processes as *naturally defined*. (That is taking for measurable sets in the space $\mathbf{B}(\mathbf{D})$ those whose inverse images are measurable in the space of the $\{\xi_1, \ldots, \xi_n\}$) They are automatically satisfied if $\mathbf{F}$ is countable. If $\mathbf{F}$ is uncountable one may be tempted to look at Z_n as a process with trajectories in the space $\mathbf{B}(\mathbf{D})$ of bounded functions on $\mathbf{D}$ and work with various "versions" of the process. For instance one may note that for any system P_j; $j = 1, \ldots, n$ of distributions $P_j = \mathbf{L}(\xi_j)$ the V.C. class $\mathbf{D}$ admits a countable dense subset for the distance $\overline{P}(S_1 \Delta S_2) = \frac{1}{n} \Sigma P_j(S_1 \Delta S_2)$. Indeed $\mathbf{D}$ is precompact for that pseudometric. Thus one can use "separable" versions.

Noting that if $n\alpha < 1/4$ the median of a binomial $\mathbf{B}(n, \alpha)$ variable is zero and applying Lemma 6 page 546 of Le Cam (1986) one sees that if $\mathbf{D}(\alpha)$ is the class of sets $\{D \in \mathbf{D};\ \overline{P}(D) \le \alpha\}$ then

$$\Pr{}^* \{\sup_D |Z_n(D)|;\ D \in \mathbf{D}(\alpha) \ge \sqrt{\frac{2}{n}[x(\varepsilon)]^2}\} \le 20\varepsilon$$

where $[x(\varepsilon)]^2 = 4v \log 2n + 4 \log \frac{2}{\varepsilon}$

Thus the validity of conditions (A) (B) already implies that the asymptotic behavior of Z_n in $\mathbf{F}$ can be deduced, within terms of order $\log 2n / \sqrt{n}$, from its behavior in the countable subclass $\mathbf{F}_0 = \bigcup_m \mathbf{F}_{\frac{1}{m}}$ union of finite classes $\mathbf{F}_{1/m}$ that approximate $\mathbf{F}$ within $\frac{1}{m}$.

Kakutani defined a "distribution" for Z_n in $\mathbf{B}(\mathbf{F})$ as follows. Consider the line $\overline{\mathbf{R}}$ compactified by adjunction of points at ∞. Then Z_n defines a unique Radon measure on $(\overline{\mathbf{R}})^{\mathbf{F}}$. Since Z_n has bounded paths the measure in question is already a Radon measure on $\mathbf{B}(\mathbf{F})$ topologized by pointwise convergence on $\mathbf{F}$. Since that Radon measure is already well defined by its projection on subspaces of the type $(\overline{\mathbf{R}})^{\mathbf{S}}$ with $\mathbf{S} \subset \mathbf{F}$ and countable, it will automatically satisfy (A) and (B) but need not coincide with the natural version of $\mathbf{L}(Z_n)$ for sets where both are defined.

The argument used in Section 4 conditioning on projections of the type $\{Z_n(S);\ S \in \mathbf{F}_\beta\}$ seems to require additional measurability restrictions, but it does not.

Since $\mathbf{F}$ is a V.C. class, the Gaussian process Z_∞ defined on $\mathbf{F}$ admits a version with continuous paths for the pseudometric $\overline{P}(S_1 \Delta S_2)$. It is well defined. For a class $\mathbf{F}_\beta$ that is finite one can find a joint distribution Q in $\mathbf{R}^{\mathbf{F}_\beta} \times \mathbf{R}^{\mathbf{F}_\beta}$ that has marginal $\mathbf{L}[Z_n(S);\ S \in \mathbf{F}_\beta]$ on the first $\mathbf{R}^{\mathbf{F}_\beta}$ and $\mathbf{L}[Z_\infty(S);\ S \in \mathbf{F}_\beta]$ on the second $\mathbf{R}^{\mathbf{F}_\beta}$. It can be selected so that $Q\{\|Z_n - Z_\infty\|_{F_\beta} \geq \pi\} \leq \pi$ for the Prokhorov distance between the two marginals. This joint distribution yields a Markov kernel $Q(x, B)$ that maps the first $\mathbf{R}^{F_\beta}$ to measures on the second. For the Gaussian process Z_∞ one can select another Markov kernel, say $H(y, C)$, regular conditional distribution of $\{Z_\infty(S) : S \in \mathbf{F}\}$ given $\{Z_\infty(S) : S \in \mathbf{F}_\beta\}$. To pair Z_n with Z_∞, keep for Z_n the original process whether it is in its natural form or a modified version. Pass from $\{Z_n(S);\ S \in \mathbf{F}\}$ to $x = \{Z_n(S);\ S \in \mathbf{F}_\beta\}$. Now apply $Q(x, B)$ to get a $y \in \mathbf{R}^{F_\beta}$ of the form $y = \{y(S);\ S \in \mathbf{F}_\beta\}$. Then apply $H(y, \cdot)$ to get a point $z = \{z(S) : S \in \mathbf{F}\}$. This will yield the pairing as needed.

In Section 2 we have used a Prokhorov distance $\pi(P,Q)$ obtained from inequalities $P(A) \leq Q(A^\varepsilon) + \varepsilon$ for closed sets A. In Section 4 we have used the fact, due to Strassen, that such inequalities imply the possibility of a coupling with $\Pr\{\|Z_{n,r} - Z_{\infty,r}\| > \varepsilon\} \leq \varepsilon$. There the $Z_{n,r}$ and $Z_{\infty,r}$ are vectors in a finite dimensional space. Thus Strassen's theorem is certainly applicable. However we have stated Theorem 2 in the coupling form and not in the $P(A) \leq Q(A^\varepsilon) + \varepsilon$ form because that would necessitate specifying for which (closed ?) sets A the probabilities $P[Z_n \in A]$ are defined. The coupling form, with outer probabilities, avoids this specification.

6) Application of a theorem of Zaitsev.

In this section we discuss briefly some possible improvements on the rates of convergence obtained in Section 4. One small improvement can be obtained by replacing the bound on $K(\varepsilon^2)$ derived from Lemma 3 page 543 of Le Cam (1986). However it seems that major improvements will depend on the use of better finite dimensional results to replace Yurinskii's theorem (Theorem 1, here). There are several possibilities. One of them is a theorem of Zaitsev (1987). Unfortunately, as we shall see, Zaitsev's theorem, as published, does not quite fulfill its promise. It does replace the $n^{1/8}$ of Theorem 1 by $n^{1/2}$ but the power of the dimension k is increased. This may just be a feature that could be changed by redoing Zaitsev's proof. However the proof is rather complex and at the time of this writing we have not yet succeeded in carrying out the necessary modifications.

Let us start with the improvement on $K(\varepsilon^2)$. Here we have used the bound

$$K(x) \leq \exp\{2v[\log v + \log 2/x]$$

as given in Le Cam (1986) page 543. This is obtained there by writing $\xi = \frac{1}{v}\log K$ and noting that ξ is at most equal to that solution y of the equation $y = \log y + a$, (with $a = \log(2v/x)$) that is larger than unity. In fact that solution satisfies the inequality $y \leq a + \frac{a}{a-1}\log a = a[1 + (a-1)^{-1}\log a]$. This was replaced by 2a in Le Cam

(1986). For a large $(a-1)^{-1}\log a$ becomes small.

This means that, for small ε, the bound $K(\varepsilon^2) \le (2v)^{2v}\varepsilon^{-4v}$ given at the end of Section 3 can be replaced by $(2v)^v\varepsilon^{-2v}\mu(\varepsilon, v)$ where $\mu(\varepsilon, v) = (a)^{\frac{a}{a-1}v}$, $a = \log\frac{2v}{\varepsilon^2}$. This extra factor $\mu(\varepsilon, v)$ is logarithmic in ε.

If this is taken into account the exponent $\gamma = \frac{1}{8+20v}$ in the rate n^γ can be replaced by $1/(8+10v)$.

Now let us pass to Zaitsev's theorem. Zaitsev considers a random vector X with distribution given by a measure F on $\mathbb{R}^k$. It is assumed that $EX = 0$ and that for $v \in \mathbb{R}^k$ the variance $(Dv, v) = E(v'x)^2$ exists. Let G be the Gaussian measure with expectation zero and the same covariance system as F.

It is assumed that $\mathbb{R}^k$ is provided with a Hilbertian norm denoted $|\cdot|$. This norm is extended to the product $\mathbb{C}^k$ of k complex planes as usual. The corresponding inner product will be denoted $\langle z, x\rangle$, $z \in \mathbb{C}^k$, $x \in \mathbb{R}^k$ or $\mathbb{C}^k$. Define ϕ on $\mathbb{C}^k$ by

$$\phi(z) = \log E e^{\langle z, X\rangle}.$$

The function ϕ is subject to the following restrictions: There is a $\tau > 0$ such that

A_1) ϕ is defined and analytic for

$$\tau|z| \le 1, \quad z \in \mathbb{C}^k.$$

A_2) For all u and v in $\mathbb{R}^k$ the mixed third derivative satisfy the condition

$$\left|\frac{\partial}{\partial u}\frac{\partial^2}{\partial v^2}\phi(z)\right| \le |u|(Dv, v)\tau.$$

The derivatives are taken as usual so that $\frac{\partial}{\partial u}\phi(z)$ means $\lim_{\varepsilon\to 0}\frac{\phi(z+\varepsilon u)-\phi(z)}{\varepsilon}$

One of Zaitsev's results is as follows.

Theorem 3. *Let* F *satisfy the conditions* A_1 *and* A_2 *for a* $\tau > 0$. *Then for the norm* $|\cdot|$ *the Prokhorov distance* $\pi_2(F, G)$ *between* F *and the Gaussian with the same first and second moments satisfies the inequality*

$$\pi_2(F, G) \le Ck^2\tau[1 + |\log\tau|]$$

where C *is a universal constant.*

To apply this result here consider the case where X is the empirical process $Z_n = \{Z_n(A);\ A \in \mathbf{A}\}$ where $\mathbf{A}$ is a class of k subsets of the sample space. For complex vectors $z = \{z_1, z_2, \ldots, z_k\}$ let $\langle z, z_n\rangle = \sum_{j=1}^{k} z_j Z_n(A_j)$.

The function $\phi(z)$ is easily obtainable even though a bit messy looking. The third mixed derivative can be computed. Its main term contains an expectation of the type $H(u,v) = E|<u, Z_n>| [<v, z_n>]^2$. This can be bounded by

$$H(u,v) \le \frac{1}{\sqrt{n}} |u|_1 (Dv, v)$$

where $|u|_1$ is the L_1-norm $|u|_1 = \sum_{j=1}^{k} |u_j| \le |u| \sqrt{k}$ for the Hilbertian norm $(\Sigma |u_j|^2)^{1/2}$.

Thus Zaitsev's theorem applies with a number τ of the type $\tau = b\sqrt{\frac{k}{n}}$ for a certain constant b.

To obtain a theorem similar to Theorem 2 of Section 4 we need to use the Prokhorov distance $\pi(F, G)$ computed for the uniform norm $\|\cdot\|$ instead of the $\pi_2(F, G)$ computed for the Hilbertian norm. Since $|x| \le \|x\| \le |x| \sqrt{k}$ one will have $\pi(F, G) \le \pi_2(F, G)\sqrt{k}$. Finally this gives the following result

Theorem 4. *For the distribution* F *of the empirical process on* k *subsets of the sample space and for the uniform norm there are constants* C_1 *and* b *such that*

$$\pi(F, G) \le C_1 \frac{k^3}{\sqrt{n}} [1 + |\log(b\sqrt{\frac{k}{n}})|].$$

Note the term $k^3/\sqrt{n}$. It corresponds here to the $[k^{5/2}/\sqrt{n}]^{1/4}$ of Theorem 1. This shows that Zaitsev's result is a considerable improvement on Yurinskii's as far as powers of n are concerned. Unfortunately we were unable to beat down the k^3 to a $k^{5/2}$. Whether this is possible by rewriting Zaitsev's proof or by a better evaluation of the term τ is not known to this writer at this time.

In any event application of the above result to the computations carried out in Section 4 will give a bound where the Prokhorov distance between Z_n and the corresponding Gaussian process will tend to zero as $n^{-\gamma}$, ignoring some logarithmic terms, but here γ can be taken equal to $\gamma = \frac{1}{2 + 12v}$.

This is a definite improvement over the $\gamma = \frac{1}{8+20v}$ of Section 4. It is unfortunate that this particular γ was replaced by $(8 + 10v)^{-1}$ at the beginning of the present section. Since $8 + 10v$ and $2 + 12v$ are not comparable it suggests that a better argument should lead to a value of γ larger than $(2 + 10v)^{-1}$.

We shall return to this question and to other possible approaches in a later report.

References

1) **Dehling, H.** (1983). ''Limit theorems for sum of weakly dependent Banach space valued random variables'' *Z. Wahrscheinlichkeitstheorie u.v. Gebiete* pp. 391-432.

2) **Dudley, R.M.** (1978). ''Central limit theorems for empirical measures''. *Ann. Prob.* vol **6**, pp. 899-929. See also *Lecture Notes in Mathematics* no. 1097 (1982). Springer Verlag.

(3) **Le Cam, L.** (1986). *Asymptotic methods in statistical Decision theory.* Springer Verlag.

4) **Lindenstrauss, J. and Tzafriri, L.** (1977). *Classical Banach Spaces I. Sequence Spaces.* Springer Verlag, Ergebnisse #92.

5) **Massart, P.** (1986). ''Rates of convergence in the Central limit theorem for empirical processes''. *Ann. Inst. Henri Poincaré* vol **22**, pp. 381-423.

6) **Yurinskii, V.V.** (1977). ''On the error of the Gaussian approximation for convolutions''. *Th. Probability applications* vol **22**, #2 pp. 236-247.

7) **Zaitsev, A. Yu** (1987). ''Estimates of the Lévy-Prokhorov distance in the multivariate central limit theorem for random variables with finite exponential moments''. *Theory Probab. Applications* vol **31**, pp. 203-220.