*The Annals of Probability*
2003, Vol. 31, No. 3, 1565–1582


# VAPNIK–CHERVONENKIS TYPE CONDITIONS AND UNIFORM DONSKER CLASSES OF FUNCTIONS[1]

BY MICHEL TALAGRAND

*Université Paris VI and Ohio State University*

Given a bounded class of functions, we introduce a combinatorial quantity (related to the idea of Vapnik–Chervonenkis classes) that is much more explicit than the Koltchinskii–Pollard entropy, but is proved to be essentially of the same order.

**1. Introduction.** The theory of empirical processes studies the uniform behavior of a class $\mathcal{F}$ of functions (defined on a measurable space $\Omega$) with respect to the law of large numbers (Glivenko–Cantelli classes), the central limit theorem (Donsker classes) and so on. For the applications to statistics, where the underlying probability distribution is not known a priori, it is natural to require that these properties hold for each underlying distribution (universality) or even uniformly over the underlying distribution (uniformity). Thus a uniform Donsker class of functions is a class over which the central limit theorem holds uniformly over the underlying probability. Giving a precise definition is however a nontrivial technical tasks. (There exist several possible natural definitions that are not obviously equivalent.) The technicalities involved in these tasks are irrelevant to the present paper, so we refer the reader to [2, 6, 4] for definitions and a detailed background on these notions.

In the important case where $\mathcal{F}$ consists of the indicators of a class $\mathcal{C}$ of sets, the important structure is clearly identified. It is the so-called Vapnik–Chervonenkis (VC) classes of sets. Let us recall that a class $\mathcal{C}$ of sets *shatters* a finite set $F$ if given a subset $G$ of $F$ there is a set $C$ in $\mathcal{C}$ for which $G = F \cap C$. A class $\mathcal{C}$ of sets is called a VC class if for some integer $n$, $\mathcal{C}$ does not shatter any set of cardinality $n$. The largest cardinality of a set shattered by $\mathcal{C}$ is called its VC dimension.

The situation is much less satisfactory concerning classes of functions. Several definitions have been proposed (see [2] and Chapter 4 of [3]). These definitions are interesting from the point of view of combinatorics. Unfortunately they are all distinct and none is the obvious choice. Roughly speaking, from the point of view of probability, what matters is that the class of functions does not oscillate too much. Oscillations of small amplitude are much less dangerous than oscillations of

---

Received May 2001; revised August 2002.
[1]Supported in part by an NSF grant.
*AMS 2000 subject classifications.* Primary 60F17; secondary 60C05.
*Key words and phrases.* Shattering, combinatorial dimension, Donsker classes.

big amplitude, but the amplitude of the oscillations is not sufficiently distinguished by the definitions of VC classes.

A natural concept is as follows.

DEFINITION 1.1. We say that a class $\mathcal{F}$ of functions shatters a set $F$ at levels $\alpha, \beta$ ($\alpha < \beta$) if given a subset $G$ of $F$, there exists $f$ in $\mathcal{F}$ such that $f(x) \le \alpha$ if $x \in G$, $f(x) \ge \beta$ if $x \in F \setminus G$.

It is natural to define

$$n_{\mathcal{F}}(\varepsilon) = \sup\{n; \exists\, \alpha, \exists\, F,\ \operatorname{card} F = n, \mathcal{F} \text{ shatters } F \text{ at levels } \alpha, \alpha + \varepsilon\}.$$

It is proved in [10] that (under measurability), given a probability measure $P$ (that we will assume atomless for clarity), a uniformly bounded class $\mathcal{F}$ of functions is a Glivenko–Cantelli (GC) class for $P$ if there does *not* exist a set $A$ with $P(A) > 0$ and $\alpha < \beta$ such that for each $n$, almost every choice of $x_1, \dots, x_n$ in $A$ (drawn at random independently according to the restriction of $P$ to $A$), the set $\{x_1, \dots, x_n\}$ is shattered by $\mathcal{F}$ at levels $\alpha, \beta$. The proof is complicated. A modern proof seem to require both tools for empirical processes (such as put forward, e.g., in the foundational paper [5]) and combinatorial arguments. Suitable combinatorial tools are provided by the paper [7]. These are brought forward in the context of the GC problem in [12], to prove the main result of [10], result that was mentioned above. As an intermediate step, it is proved in [12] that if $S_n(\alpha, \beta)$ denotes the set of $n$-tuples $x_1, \dots, x_n$ such that the set $\{x_1, \dots, x_n\}$ is shattered at levels $\alpha, \beta$ by $\mathcal{F}$, then $\mathcal{F}$ is a GC class for $P$ if for each $\alpha < \beta$, the sequence $a_n(\alpha, \beta) = (P^{\otimes n}(S_n(\alpha, \beta)))^{1/n}$ converges to zero. The rate of convergence in the GC theorem is determined by the rate of convergence to zero of these sequences. When it is true that $n_{\mathcal{F}}(\varepsilon) < \infty$ for each $\varepsilon > 0$, the sets $S_n(\alpha, \beta)$ are empty for $\beta - \alpha > \varepsilon$, and therefore the convergence is uniform over $P$. Thus, a uniformly bounded class of functions is a uniform GC class if (and only if) $n_{\mathcal{F}}(\varepsilon)$ is finite for each $\varepsilon > 0$. Since this fact is not stated in [12], it belongs to the paper by Alon, Ben-David, Cesa-Bianchi and Haussler [1] to state it first. As in [12], [1] uses the Giné–Zinn methods, but the combinatorial ingredients are completely different and of independent interest.

The previous results do not help when one is interested in rates of convergence (and in central limit theorems, that resemble laws of large numbers with rate $n^{-1/2}$). As we will explain soon, "shattering at fixed levels" as in Definition 1.1 is no longer an adequate measure of the oscillations of $\mathcal{F}$. We need to use a different level for each point of $F$. The following definition is stated in [1].

DEFINITION 1.2. We say that a class of $\mathcal{F}$ of functions $\varepsilon$-shatters a set $F$ if for some point $x$ in $F$ there exists a level $\alpha(x) \in \mathbb{R}$, such that, given any subset $G$ of $F$, we can find a function $f$ in $\mathcal{F}$, such that $f(x) \le \alpha(x)$ for $x$ in $G$, while $f(x) \ge \alpha(x) + \varepsilon$ for $x$ in $F \setminus G$.

The very same property was used in [11] to construct $\ell_1$ sequences, even though it was not given a name there.

From the point of view of the behavior of $\mathcal{F}$ with respect to probability theory, it is just as bad that there exists a set $F$ that is $\varepsilon$-shattered by $\mathcal{F}$, whether or not the shattering levels $\alpha(x)$ depend on $x$. (If this statement is not obvious now it will become obvious at the beginning of the proof of Proposition 1.4.) On the other hand, if, say, $\mathcal{F}$ consists of functions $f$ with $|f| \le 1$, given a set $F$ that is $\varepsilon$-shattered by $\mathcal{F}$, we can construct fixed levels $\alpha$, $\alpha + \varepsilon/2$, and a subset $F'$ of $F$, such that $\mathcal{F}$ shatters $F'$ at levels $\alpha, \alpha + \varepsilon/2$; but we cannot in general achieve better than $\operatorname{card} F'$ about $\varepsilon \operatorname{card} F$, a loss of accuracy that is devastating when dealing with precise rates.

DEFINITION 1.3. The shattering function $S(\varepsilon) = S_{\mathcal{F}}(\varepsilon)$ of the class $\mathcal{F}$ is defined by $S_{\mathcal{F}}(\varepsilon) = \sup\{n;\ \mathcal{F}\ \varepsilon\text{-shatters a set of cardinality } n\}$.

We believe that the shattering function is a quantity of fundamental importance, but this will be better explained after the statement of the main result. Throughout the paper we will assume that $\mathcal{F}$ is uniformly bounded. Let us recall the definition of the Koltchinskii–Pollard capacity:

$$D(\varepsilon, \mathcal{F}) = \sup\sup\left\{n;\ \exists f_1, \dots, f_n \in \mathcal{F},\ i < j \le n \Rightarrow \int (f_i - f_j)^2\, d\gamma \ge \varepsilon^2\right\},$$

where the outer supremum is taken over all probability measures $\gamma$ supported by a finite subset of the underlying space $\Omega$. We then define the Koltchinskii–Pollard entropy as $\log D(\varepsilon, \mathcal{F})$.

PROPOSITION 1.4. *For some universal constant $K$, we have for each $\varepsilon > 0$,*

$$S_{\mathcal{F}}(\varepsilon) \le K \log D\left(\frac{\varepsilon}{2}, \mathcal{F}\right). \tag{1.1}$$

This (easy) fact can be expressed by saying that the shattering function is dominated by the Koltchinskii–Pollard entropy. Our main result is that, under minimum regularity the converse is nearly true. Arguably the interesting case is when $S_{\mathcal{F}}(\varepsilon)$ resembles a positive power of $1/\varepsilon$; thus the regularity condition (1.2) below is mild.

THEOREM 1.5. *Consider a continuous positive nonincreasing function $\xi$ on $(0, \infty)$. Assume that for some $\alpha \ge 1$ and all $0 < x < y$ we have*

$$\frac{1}{2}\left(\frac{x}{y}\right)^{\alpha} \le \frac{\xi(y)}{\xi(x)} \le 2\left(\frac{x}{y}\right)^{1/\alpha}. \tag{1.2}$$

*Assume that for each $\varepsilon > 0$ we have*

$$S_{\mathcal{F}}(\varepsilon) \le \xi(\varepsilon). \tag{1.3}$$

*Assume that $\mathcal{F}$ is uniformly bounded, and let $\Delta$ be its diameter for the uniform norm. Then*

$$\forall\, \varepsilon < \Delta, \qquad \log D(\varepsilon, \mathcal{F}) \le \xi(\varepsilon)\left(\log \frac{4\Delta}{\varepsilon}\right)^K \tag{1.4}$$

*where $K$ depends upon $\alpha$ only.*

Thus (within log terms) we essentially have $S_{\mathcal{F}}(\varepsilon) \approx \log D(\varepsilon, \mathcal{F})$. This result is of the type "concentration of pathology." Suppose we know that $L = \log D(\varepsilon, \mathcal{F})$ is large. This simply means that $\mathcal{F}$ contains many well separated functions; but we know very little about what kind of pattern they form. The content of the theorem is that it is possible (with a lot of work) to construct a set $F$ of size nearly $L$ and for $x \in F$ levels $\alpha(x)$ for which the phenomenon described in Definition 1.2 occurs. We now have a very precise structure that witnesses that $\mathcal{F}$ is large, and one can expect that this structure will be easy to identify. This result is exactly in the line of the previously mentioned author's characterization of Glivenko–Cantelli classes [11].

COROLLARY 1.6. *There exists a universal constant $M$, such that a countable uniformly bounded class $\mathcal{F}$ of Borel functions that satisfies*

$$\sup_{\varepsilon} \varepsilon^2\, S_{\mathcal{F}}(\varepsilon)\left(\log \frac{4\Delta}{\varepsilon}\right)^M < \infty$$

*is a uniform Donsker class.*

PROOF. Take $M > K(2) + 2$. Use Theorem 1.5 with a function $\xi(\varepsilon)$ equal to $\varepsilon^{-2} \log(2\Delta/\varepsilon)^{M-2}$ for $\varepsilon \le \Delta$ and use Proposition 3.1 of [6]. $\square$

It is certainly not necessary in Corollary 1.6 that $\mathcal{F}$ be countable or the functions of $\mathcal{F}$ be Borel, but some kind of measurability is certainly needed. A convenient condition, that seems sufficient for all practical purposes, is Dudley's notion of Souslin admissible classes; see [3].

One might like to know how small $M$ can be taken in Corollary 1.6. It is probably a rather difficult question to find a sharp value of $M$. This question however does not seem very important because the natural examples of uniform Donsker classes (such as those of [2]) satisfy $\sup \varepsilon^{\alpha} S_{\mathcal{F}}(\varepsilon) < \infty$ for some $\alpha < 2$. Thus, even though (1.5) does not provide a characterization of uniform Donsker classes, it seems that it does "for natural situations."

Theorem 1.5 is proved implicitly, in another language in [10], and the aim of the present paper is to make it known in the empirical process community. We present complete proofs. The key lemmas reproduce the proofs of [10] with notation more adapted to the present setting. The key argument is a (very powerful) iteration method that is commonly used in Banach space theory. This central part of the argument has been completely rewritten, in a less savagely technical way, now more adapted to the sharply diminished number of functioning neurons available to the author. In retrospect, it however appears obvious that the present paper should have either been written much earlier or not at all. Soon after it was submitted, Mendelson and Vershynin obtained optimal results in the same direction, with simpler and much nicer proofs [9].

**2. The iteration argument.** First, a word about Proposition 1.4. Consider a finite set $F$ that is $\varepsilon$-shattered by $\mathcal{F}$. Let $n = \operatorname{card} F$, and for each subset $G$ of $F$, consider a function $f_G$ in $\mathcal{F}$ with $f_G(x) \le \alpha(x)$ when $x \in G$, $f_G(x) \ge \alpha(x) + \varepsilon$ when $x \notin G$, where the numbers $\alpha(x)$ are as in Definition 1.3. Consider the probability $\gamma$ on $\Omega$ that gives mass $1/n$ to each point of $F$. Then, clearly, for two subsets $G, G'$ of $F$, we have

$$\int |f_G - f_{G'}|^2 \, d\gamma \ge \varepsilon^2 \gamma(G \Delta G'). \tag{2.1}$$

It is elementary that one can find a family $\mathcal{C}$ of subsets of $F$ with $\operatorname{card} \mathcal{C} \ge 2^{n/K}$ (for a certain number $K$) and $\operatorname{card}(G \Delta G') \ge n/4$ when $G \ne G'$, $G, G' \in \mathcal{C}$ so that if $G \ne G'$

$$\int |f_G - f_{G'}|^2 \, d\gamma \ge \left(\frac{\varepsilon}{2}\right)^2,$$

and thus $\log D(\varepsilon/2, \mathcal{F}) \ge n/K$. Taking the supremum over $n$ yields (1.1).

We can replace $\mathcal{F}$ by $a\mathcal{F} + h$ for a suitable number $a$ and a suitable function $h$ to see that there is no loss of generality to assume

$$f \in \mathcal{F} \quad \Longrightarrow \quad 0 \le f \le 1, \tag{2.2}$$

$$\exists\, f, g \in \mathcal{F},\ \exists\, \omega, \qquad |f(\omega) - g(\omega)| \ge 2/3. \tag{2.3}$$

These conditions are assumed through the rest of the paper. In particular, to prove (1.4) we can and do assume that $\varepsilon \le 1$.

Many arguments of the paper involve numerical constants (such as 2/3 above). The choice of these is never critical and is largely irrelevant. There is all the room in the world to choose them. We have made no efforts whatsoever to make choices anywhere close to optimal. In particular, the reader should not be disturbed if in certain inequalities below we happen to lose a few factors of 2 for no particular reason. These factors do not matter. It was easier while writing the proofs to allow room at the beginning for possible later changes, and useless afterwards to attempt to tighten the constants.

Given a function $h$ (that need not belong to $\mathcal{F}$) and a probability measure $\gamma$ on $\Omega$, supported by a finite set, we define $N(\varepsilon, h, \gamma)$ as the supremum of the cardinalities of the finite subsets $Z$ of $\mathcal{F}$ such that

$$f \in Z \quad \Longrightarrow \quad \int (f-h)^2\,d\gamma \le 4\varepsilon^2, \tag{2.4}$$

$$f, g \in Z, f \neq g \quad \Longrightarrow \quad \int (f-g)^2\,d\gamma \ge \frac{\varepsilon^2}{4}. \tag{2.5}$$

Of course $N(\varepsilon, h, \gamma)$ depends upon $\mathcal{F}$, but the dependence is kept implicit since we now fix $\mathcal{F}$ once and for all.

LEMMA 2.2. *To prove Theorem* 1.5 *it suffices to prove that for each* $\varepsilon \le 1$, *each* $h$ *and each* $\gamma$ *we have*

$$\log N(\varepsilon, h, \gamma) \le \xi(\varepsilon)\left(\log\left(\frac{4}{\varepsilon}\right)\right)^K. \tag{2.6}$$

There and below, $K$ denotes a constant depending upon $\alpha$ only. It need not be the same at each occurrence.

PROOF OF LEMMA 2.2. Denoting by $N(\varepsilon, \gamma)$ the supremum of $N(\varepsilon, h, \gamma)$ over $h$, the trace on $\mathcal{F}$ of a ball in $L_2(\gamma)$ of radius $2\varepsilon$ can be covered by $N(\varepsilon, \gamma)$ balls in $L_2(\gamma)$ of radius $\le \varepsilon$. [If $Z$ satisfies (2.4) and (2.5), and has the largest possible cardinality, the union of the balls of $L_2(\gamma)$ centered at $Z$, of radius $\varepsilon$, covers the trace on $\mathcal{F}$ of the ball of $L_2(\gamma)$ centered at $h$, of radius $2\varepsilon$.] Then, for each $j \ge 0$, we see that $\mathcal{F}$ can be covered by $N_\gamma(j) = \prod_{i=0}^{i=j} N(2^{-i}, \gamma)$ balls for $L_2(\gamma)$ of radius $2^{-j}$. By the pigeonhole principle, given a subset $Z$ of $\mathcal{F}$ of cardinal $> N_\gamma(j)$, there exist two elements of $Z$ within distance $2^{-j+1}$. So, if $2^{-j} < \varepsilon/2$, we see that, using (2.6)

$$\log D(\varepsilon, \mathcal{F}) \le \sup_\gamma \log N_\gamma(j) \le K \sum_{i=0}^{i=j} \xi(2^{-i})(\log 2^{i+2})^K.$$

Taking $j$ as small as possible implies (1.4), using (1.2) to see that the term for $i = j$ is of the same order as entire sum. $\square$

The basic idea to prove (2.6) is as follows. Starting with a family $Z$ that satisfies (2.4) and (2.5), we will try to realize these conditions when $\gamma$ gives mass $1/n$ to $n$ points $x_1, \dots, x_n$ of $\Omega$. We want to make $n$ as small as possible (possibly at the cost of slightly decreasing $Z$). Ideally, if $\eta$ satisfies

$$\xi(\eta) = \log \operatorname{card} Z,$$

we would like to succeed in taking $n$ small enough that

$$n\varepsilon^2 \text{ is about } \eta^2 \xi(\eta). \tag{2.7}$$

The space $L_2(\gamma)$ is of dimension $n$, so by looking at volumes, a ball of radius $2\varepsilon$ can contain at most $9^n$ points that are $\varepsilon/2$ separated. (Notice that the open balls of radius $\varepsilon/4$ centered at these points are disjoint and entirely contained in a ball of radius $9\varepsilon/4$. This type of argument will later be referred to as a "volume argument.") Thus

$$\xi(\eta) = \log \operatorname{card} Z \le n \log 9$$

and

$$\xi(\eta)\varepsilon^2 \le n\varepsilon^2 \log 9.$$

The right-hand side is about $\eta^2\xi(\eta)$, so that the ratio $\varepsilon/\eta$ is not large. Thus, by (1.2), $\log \operatorname{card} Z = \xi(\eta)$ is not much larger than $\xi(\varepsilon)$.

To achieve the previous program we will use the following reduction principle.

PROPOSITION 2.3. *Consider a family $Z$ of functions with* $\operatorname{card} Z \ge 16$ *that satisfies* (2.4), *and a number $M$ such that*

$$f, g \in Z, f \neq g \quad \Longrightarrow \quad \int (f-g)^2 \wedge M^2 \, d\gamma \ge \frac{\varepsilon^2}{8}. \tag{2.8}$$

*We set*

$$n = \lceil (64 M^2 \log \operatorname{card} Z)/\varepsilon^2 \rceil. \tag{2.9}$$

*Then we can find a number $\varepsilon_0 > 0$, a function $h_0$, a family $Z_0 \subset Z$ and $n$ points $x_1, \dots, x_n$ in $\Omega$ with the following property. If $\gamma_0$ denotes the probability measure $n^{-1}\sum_{i=1}^{i=n} \delta_{x_i}$, we have*

$$\frac{1}{8} \le \frac{\varepsilon_0}{\varepsilon} \le 4, \tag{2.10}$$

$$\operatorname{card} Z_0 \ge \left(\tfrac{1}{2} \operatorname{card} Z\right)^{1/6}, \tag{2.11}$$

$$\forall f \in Z_0, \qquad \int (f-h_0)^2 \, d\gamma_0 \le 4\varepsilon_0^2, \tag{2.12}$$

$$\forall f, \qquad g \in Z_0, f \neq g \quad \Longrightarrow \quad \int |f-g|^2 \, d\gamma_0 \ge \varepsilon_0^2/4. \tag{2.13}$$

In words, starting with a situation (2.4) and (2.5), we create a similar situation without changing $\varepsilon$ too much [as (2.10) witnesses] or without decreasing $Z$ too much [as shown by (2.11)], where the number $n$ of points that $\gamma$ charges satisfies (2.9). The points $x_1, \dots, x_n$ will simply be selected at random independently according to $\gamma$. In order to make the procedure efficient (i.e., to obtain $n$ small) we need to have a good control on $M$ (i.e., we need $M$ to be small). To achieve this, starting with a situation where we know (2.5), we have to achieve (2.8) with as small $M$ as possible, without reducing $Z$ too much.

PROPOSITION 2.4. *There exists a number $K_1$, depending upon $\alpha$ only, with the following property. Consider $n$ points $x_1, \dots, x_n$ (that need not be distinct) and the probability measure $\gamma = n^{-1}\sum_{1\le i\le n}\delta_{x_i}$. Assume $n \ge \xi(1)$. Consider a subset $Z$ of $\mathcal{F}$ that satisfies (2.4) and (2.5). Consider an integer $j$ (which might be negative) and a positive number $\theta \le 2^{-6}$ that satisfy*

$$\log\left(\frac{en}{\xi(1)}\right)\xi(\theta 2^{-j}) + n\varepsilon^2\theta 2^{2j} \le \frac{\log \operatorname{card} Z}{K_1}. \tag{2.14}$$

*Then we can find a subset $Z'$ of $Z$ such that*

$$\operatorname{card} Z' \ge (\operatorname{card} Z)^{1/2} \tag{2.15}$$

*and*

$$\forall f, \qquad g \in Z', f \ne g \quad \Longrightarrow \quad \int |f-g|^2 \wedge 2^{-2j+2}\, d\gamma \ge \frac{\varepsilon^2}{8}. \tag{2.16}$$

In words, under (2.14), we can achieve (2.6) with $M^2 = 2^{-2j+2}$, without reducing $Z$ too much. Of course in (2.14) we will choose $\theta$ suitably, and then $j$ as large as we can. But one gets a clearer argument if one resists the temptation to optimize over $\theta$ in (2.14). To achieve (2.14) with $2^{2j}$ large (i.e., $M^2$ small), it helps a lot that $n$ is not large. Thus, the situation is as follows: Proposition 2.2 lets us reduce $n$ (the smaller $M$, the better it works) while Proposition 2.3 lets us reduce $M$ (the smaller $n$, the better it works). The idea is then to use these propositions in turn a large number of times. The nice (and somewhat unexpected) feature is that this iteration brings us within logarithmic terms of the optimal result. We delay the proof of Propositions 2.3 and 2.4 until Section 3, and we perform this iteration now. We start with a subset $Z$ of $\mathcal{F}$ with $\operatorname{card} Z \ge 16$ that satisfies (2.4) and (2.5), and we consider an $\eta$ such that

$$\log \operatorname{card} Z = \xi(\eta). \tag{2.17}$$

As already pointed out, we can assume $\varepsilon \le 1$, and we can assume $\eta \le \varepsilon$ (for there is nothing to prove otherwise).

The first step of the construction is special. We apply Proposition 2.3 with $M = 1$. Since we assume $0 \le f \le 1$ for $f$ in $\mathcal{F}$ (2.8) follows from (2.5). We set

$$n_0 = \lceil (64 \log \operatorname{card} Z)/\varepsilon^2 \rceil.$$

Using (1.2), and since $\eta \le \varepsilon$ we have

$$\frac{n_0}{\xi(1)} \le \frac{K\xi(\eta)}{\varepsilon^2\xi(1)} \le \left(\frac{2}{\eta}\right)^K. \tag{2.18}$$

We then find $\varepsilon_0$, $Z_0$ as in (2.10) to (2.13). We observe that $n_0\varepsilon^2$ is about $\xi(\eta)$, while our goal is to reduce it to about $\eta^2\xi(\eta)$.

We now describe the general step of the iteration. At the beginning of this step, we start with a subset $Z_k$ of $\mathcal{F}$, a number $\varepsilon_k \geq \varepsilon 2^{-k-3}$, an integer $n_k$ and a probability measure $\gamma_k$ that is the sum of point masses $1/n_k$ at each of $n_k$ points (not necessarily distinct). The elements of $Z_k$ satisfy

$$f \in Z_k \quad \Longrightarrow \quad \int (f - h_k)^2 \, d\gamma_k \leq 4\varepsilon_k^2, \tag{2.19}$$

$$f, g \in Z_k, f \neq g \quad \Longrightarrow \quad \int (f - g)^2 \, d\gamma_k \geq \varepsilon_k^2/4, \tag{2.20}$$

and we have

$$\log \operatorname{card} Z_k \geq 2^{-6(k+1)} \xi(\eta). \tag{2.21}$$

For simplicity we write

$$a_k = 2^{-6(k+1)}. \tag{2.22}$$

The general step consists in successive applications of Propositions 2.4 and then 2.3, in that order. If $n_k \leq \xi(1)$ or if $\operatorname{card} Z_k < 16$, we stop the construction. We will show later that there is nothing to prove. So we assume that $n_k \geq \xi(1)$ and $\operatorname{card} Z_k \geq 16$. We will apply Proposition 2.4 to $Z = Z_k$ and $n = n_k$. We must choose $\theta$ and $j$ so that (2.14) holds.

We see from (2.21) that (2.14) will hold provided

$$(\log c_k)\xi(\theta 2^{-j}) + n_k \varepsilon_k^2 \theta 2^{2j} \leq \frac{a_k \xi(\eta)}{K_1}, \tag{2.23}$$

where for clarity we write $c_k = e n_k / \xi(1)$.

Since our goal is to reduce $n_k \varepsilon_k^2$ to $\eta^2 \xi(\eta)$, it clarifies matters to set

$$u_k = \frac{n_k \varepsilon_k^2}{\eta^2 \xi(\eta)}.$$

Consider $v$ and $j$ such that $2^{-j-1} \leq v\eta \leq 2^{-j}$. Then (2.23) will hold as soon as

$$(\log c_k) \frac{\xi(\eta\theta v)}{\xi(\eta)} + 4\frac{\theta u_k}{v^2} \leq \frac{a_k}{K_1}. \tag{2.24}$$

If we can now take $\theta$ such that $t := \theta v > 1$, then since $\xi(t\eta) \leq 2\xi(\eta) t^{-1/\alpha}$ by (1.2), to achieve (2.24) it will suffices to have

$$2t^{-1/\alpha} \log c_k + 4\frac{u_k t}{v^3} \leq \frac{a_k}{K_1}, \tag{2.25}$$

which will follow from

$$t = \left( \frac{4K_1 \log c_k}{a_k} \right)^{\alpha}; \qquad v = \left( \frac{8K_1 t u_k}{a_k} \right)^{1/3}. \tag{2.26}$$

We have to be careful however that Proposition 2.4 requires $\theta = t/v \le 2^{-6}$, that is, $t \le 2^{-6}v$. Simple algebra show that this is implied by the condition

$$u_k \ge K_2 \frac{(\log c_k)^{2\alpha}}{a_k^{2\alpha-1}}, \tag{2.27}$$

where $K_2 = K_2(\alpha)$ is now fixed once and for all.

If (2.27) fails, that is, if

$$u_k < \frac{K_2(\log c_k)^{2\alpha}}{a_k^{2\alpha-1}}, \tag{2.28}$$

we simply stop the construction. We also stop the construction if $a_k\xi(\eta) \le 4$. We will explain later why the construction has already succeeded. Thus, the construction continues only if (2.27) holds, if $n_k \ge \xi(1)$, if $\operatorname{card} Z_k \ge 16$ and if $a_k\xi(\eta) \ge 2$.

We now assume that these conditions hold; we take $t$ and $v$ as in (2.26); we take for $j$ the smallest integer with $2^{-j-1} \le v\eta$. Hence (2.14) holds. Proposition 2.4 gives us a subset $Z'$ of $Z_k$ such that, using (2.21), and since $\log\operatorname{card} Z' \ge (1/2)\log\operatorname{card} Z_k$,

$$\log\operatorname{card} Z' \ge \frac{a_k}{2}\xi(\eta) \tag{2.29}$$

and that

$$\forall f, \qquad g \in Z', f \ne g \quad \Longrightarrow \quad \int (f-g)^2 \wedge M^2\, d\gamma_k \ge \frac{\varepsilon_k^2}{8}$$

for $M = 2^{-j+2} \le 4\eta v$, so that by (2.26),

$$M^2 \le 16\eta^2 \biggl(\frac{8K_1 t u_k}{a_k}\biggr)^{2/3}. \tag{2.30}$$

We now apply Proposition 2.3 to $Z'$, with $\min(M, 1)$ rather than $M$ [see (2.2)] to find a number $\varepsilon_{k+1}$ satisfying

$$\varepsilon_{k+1} \ge \frac{\varepsilon_k}{8} \ge 2^{-3(k+1)}\varepsilon \tag{2.31}$$

an $h_{k+1}$ (the $h_0$ of Proposition 2.3) and a subset $Z_{k+1}$ of $Z'$ satisfying

$$\operatorname{card} Z_{k+1} \ge \bigl(\tfrac{1}{2}\operatorname{card} Z'\bigr)^{1/6} \tag{2.32}$$

such that (2.19) and (2.20) hold (for $k+1$ rather than $k$) with

$$n_{k+1} = \biggl\lceil \frac{64(\min(M,1))^2 \log\operatorname{card} Z'}{\varepsilon_k^2} \biggr\rceil.$$

We observe that $\operatorname{card} Z' \geq 4$ by (2.29) and since $a_k \xi(\eta) \geq 4$ and that, by the equation after (2.29), we have $M^2 \geq \varepsilon_k^2/8$. Thus, since $\lceil x \rceil \leq 2x$ for $x \geq 1$, we have

$$n_{k+1} \leq \frac{128(\min(M, 1))^2 \log \operatorname{card} Z'}{\varepsilon_k^2}.$$

Since $\log \operatorname{card} Z' \leq \log \operatorname{card} Z$, and using (2.17) to bound $\log \operatorname{card} Z$, we get

$$n_{k+1} \leq \frac{K(\min(M, 1))^2 \log \operatorname{card} Z'}{\varepsilon_k^2} \leq \frac{K}{\varepsilon_{k+1}^2}\left(\frac{t u_k}{a_k}\right)^{2/3} \eta^2 \xi(\eta). \tag{2.33}$$

Let us observe that (2.33) implies that

$$u_{k+1} = \frac{n_{k+1}\varepsilon_{k+1}^2}{\eta^2 \xi(\eta)} \leq K\left(\frac{t}{a_k}\right)^{2/3} u_k^{2/3}. \tag{2.34}$$

Since $\operatorname{card} Z' \geq 4$, we have

$$\tfrac{1}{2}\operatorname{card} Z' \geq (\operatorname{card} Z')^{1/2}, \tag{2.35}$$

and combining (2.29), (2.32) and (2.35), we get

$$\log \operatorname{card} Z_{k+1} \geq \frac{a_k}{32}\xi(\eta) = a_{k+1}\xi(\eta).$$

At this point we observe that $\varepsilon_k^2 \geq 2^{-6(k+1)}\varepsilon^2 = \varepsilon^2 a_k$ and that, since we assume $a_k \xi(\eta) \geq 4$, without loss of generality, we can reduce $Z'$ in (2.29) so that $\log \operatorname{card} Z'$ is not larger than $\xi(\eta) a_k$. Then the first inequality of (2.33) and the definition of $n_0$ show that $n_{k+1} \leq K n_0$. Thus $c_k \leq K c_0$. From (2.22) and (2.26), we see that $t \leq K^k (\log K c_0)^\alpha$ so that by (2.34), we have

$$u_{k+1} \leq K^{k+1}(\log K c_0)^{2\alpha/3} u_k^{2/3} \tag{2.36}$$

and thus

$$\frac{u_{k+1}}{K^{3(k+2)}(\log K c_0)^{2\alpha}} \leq \left(\frac{u_k}{K^{3(k+1)}(\log K c_0)^{2\alpha}}\right)^{2/3}. \tag{2.37}$$

Recall that $0 < \eta < \varepsilon < 1$. Consider the smallest integer $m$ such that $(2/3)^{m-1}\log(4/\eta) \leq 1$ and assume first that the construction has not stopped before $m$. Then combining the inequalities (2.37) for $k = m-1, \dots, k = 0$ shows that

$$\frac{u_m}{K^{3(m+1)}(\log K c_0)^{2\alpha}} \leq \left(\frac{u_0}{K^3(\log K c_0)^{2\alpha}}\right)^{(2/3)^m}.$$

We recall that $u_0 \leq K/\eta^2$ by (2.18) and (2.10). Thus we have

$$\left(\frac{u_0}{K^3(\log K c_0)^{2\alpha}}\right)^{(2/3)^m} \leq \left(\frac{K}{\eta^2}\right)^{(2/3)^m} \leq K$$

and thus

$$u_m \le K^m (\log K c_0)^{2\alpha} \le \left(\log \frac{4}{\eta}\right)^K, \tag{2.38}$$

where we have used in the second inequality that, by definition of $m$ we have $(3/2)^m \le \log(4/\eta)$ that, since $\eta \le 1$, we have $\log(4/\eta) \ge \log 4 > 1$, and also that by (2.18) we have $\log c_0 \le K \log(2/\eta)$.

Thus

$$n_m \varepsilon_m^2 \le \left(\log \frac{4}{\eta}\right)^K \eta^2 \xi(\eta)$$

and thus, since $\varepsilon_m \ge 2^{-3(m+1)}\varepsilon \ge (\log(4/\eta))^{-K}\varepsilon$, we have

$$n_m \le \left(\log \frac{4}{\eta}\right)^K \frac{\eta^2}{\varepsilon^2} \xi(\eta).$$

Now, using the "volume argument" in the first inequality below and (2.21) in the second, we have

$$n_m \ge \frac{1}{K} \log \operatorname{card} Z_m \ge \frac{1}{2^{6(m+1)}} \log \operatorname{card} Z \ge \left(\log \frac{4}{\eta}\right)^{-K} \xi(\eta)$$

and thus

$$\varepsilon \le \left(\log \frac{4}{\eta}\right)^K \eta.$$

It is simple to deduce that $\eta \ge \varepsilon/(\log(4/\varepsilon))^K$ and thus by (1.2),

$$\xi(\eta) \le \xi(\varepsilon)\left(\log \frac{4}{\varepsilon}\right)^K. \tag{2.39}$$

This completes the proof in the case where the construction does not stop before $m$.

Now we consider the case where the construction stopped before $m$. There are 4 cases. The first is that it happened for some $k \le m$ that $\log \operatorname{card} Z_k \le 4$. By (2.21) this is because

$$\log \operatorname{card} Z = \xi(\eta) \le \left(\log \frac{4}{\eta}\right)^K. \tag{2.40}$$

Since we assume $S_{\mathcal{F}}(\varepsilon) \le \xi(\varepsilon)$ and since $S_{\mathcal{F}}(2/3) \ge 1$ by (2.3), we see from (1.2) and (2.40) that $\eta \ge 1/K$. Since $\varepsilon \le 1$, we get $\log \operatorname{card} Z = \xi(\eta) \le K\xi(\varepsilon)$ and there is nothing more to prove.

The second case is that it happened for some $k \le m$ that $a_k \xi(\eta) \le 4$. It is entirely similar to the first case.

The third case is that it happened for some $k \le m$ that $n_k \le \xi(1)$. Since by volume arguments, we have $\log \operatorname{card} Z_k \le K n_k$, and since $\log \operatorname{card} Z \le 2^{6(k+1)} \times \log \operatorname{card} Z_k$, we have

$$\xi(\eta) = \log \operatorname{card} Z \le K 2^{6(k+1)} \xi(1) \le K \xi(1) \left( \log \frac{4}{\eta} \right)^K$$

and, as before, we get $\eta \ge 1/K$. The fourth case is that it happened that (2.28) occurs for $k \le m$. In that case one simply repeats the argument that deduces (2.39) from (2.38).

## 3. Proof of the reduction principles.

PROOF OF PROPOSITION 2.3. We start with the following well known elementary inequality (see, e.g., [11], Lemma 2.2 or [8], Theorem 2 for a sharper result with a harder proof). If $(X_i)_{i \le n}$ are i.i.d. r.v., $0 \le X_i \le A$, then

$$P\left( \sum_{i \le n} X_i \le \frac{n}{4} E X_1 \right) \le \exp\left( - \frac{n}{4A} E X_1 \right). \tag{3.1}$$

Consider an i.i.d. sequence of points $(x_i)_{i \le n}$ in $\Omega$, distributed like $\gamma$. Fixing $f, g$ in $Z$ with $f \ne g$, we consider $X_i = (f(x_i) - g(x_i))^2 \wedge M^2$, so that $0 \le X_i \le M^2$ and $E X_i \ge \varepsilon^2/8$ by (2.8). Thus by (3.1), we have

$$P\left( \sum_{i \le n} X_i \le \frac{n\varepsilon^2}{32} \right) \le \exp\left( - \frac{n\varepsilon^2}{32 M^2} \right).$$

By (2.9),

$$\frac{n\varepsilon^2}{32 M^2} \ge 2 \log \operatorname{card} Z \tag{3.2}$$

so that the event

$$\forall f, g \in Z, f \ne g, \qquad \frac{1}{n} \sum_{i \le n} \big( f(x_i) - g(x_i) \big)^2 \ge \frac{\varepsilon^2}{32} \tag{3.3}$$

has probability $\ge 1 - N(N-1)/(2N^2) > 1/2$ where $N = \operatorname{card} Z$. On the other hand, given $f$ in $Z$, the event

$$\frac{1}{n} \sum_{i \le n} \big( f(x_i) - h(x_i) \big)^2 \le 16 \varepsilon^2 \tag{3.4}$$

has a probability at least $3/4$ by (2.4) and the Markov inequality. By the Fubini theorem, with probability at least $1/2$, (3.4) occurs for at least half of the elements of $Z$. Thus we can find $x_1, \dots, x_n$ such that (3.3) occurs, and that (3.4) occurs for $f \in Z_1$, where $\operatorname{card} Z_1 \ge (1/2) \operatorname{card} Z$. We then choose for $\gamma_0$ the probability $n^{-1} \sum_{1 \le i \le n} \delta_{x_i}$.

Let us consider the $L^2(\gamma_0)$ metric $d(f,g)$. Thus $Z_1$ consists of elements of a ball of radius $4\varepsilon$, and they are $\varepsilon/4\sqrt{2}$ separated by (3.3). This is not what we need. We need $\varepsilon_0/2$ separated elements in a ball of radius $2\varepsilon_0$. To obtain this we repeat the first argument of Lemma 2.2. If for $j \geq 0$ we denote by $N(j)$ the maximum possible number of $2^{1-j}\varepsilon$ separated elements of $Z_1$ contained in a ball of radius $2^{2-j}\varepsilon$, we see that $Z_1$ can be covered by $\prod_{0\leq j\leq \ell} N(j)$ balls of radius $2^{1-\ell}\varepsilon$. For $\ell = 5$ these balls are of diameter $\leq \varepsilon/8$ and thereby can contain at most one element of $Z$. Thus

$$\operatorname{card} Z \leq \prod_{0\leq j\leq 5} N(j)$$

and thus there exists $0 \leq j_0 \leq 5$ such that $N(j_0) \geq (\operatorname{card} Z_1)^{1/6}$. To conclude we take $\varepsilon_0 = 2^{2-j_0}\varepsilon$, and $Z_0$ an $\varepsilon_0/2$ separated subset of $Z_1$ witnessing that $N(j_0) \geq (\operatorname{card} Z_1)^{1/6}$. $\square$

We now turn to the proof of Proposition 2.4. This proposition is itself the result of an iteration procedure, the basic step of which is given as follows, keeping the notation of Proposition 2.4. In particular the function $h$ is as in (2.4).

PROPOSITION 3.1. *Consider an integer $\ell \geq 0$ and a positive number $\theta \leq 2^{-6}$. Then we can find a subset $Z_1$ of $Z$ that satisfies*

$$\operatorname{card} Z_1 \geq \operatorname{card} Z \exp\left(-K\left(\log\left(\frac{2n}{\xi(1)}\right)\xi(\theta 2^{-\ell}) + \theta\varepsilon^2 n 2^{2\ell}\right)\right), \tag{3.5}$$

$$\begin{aligned} &\forall f \in Z_1,\ \forall g \in Z_1,\ \forall i \leq n,\\ &\quad f(x_i) - h(x_i) \in [2^{-\ell}, 2^{-\ell+1}]\\ &\qquad \Longrightarrow \quad |f(x_i) - g(x_i)| \leq 2^{-3}|f(x_i) - h(x_i)|. \end{aligned} \tag{3.6}$$

PROOF OF PROPOSITION 2.4. We can assume $j \geq 0$, for otherwise we simply take $Z' = Z$. We apply Proposition 3.1 recursively for all values of $0 \leq \ell \leq j$. We get a subset $Z_2$ of $Z$ with

$$\operatorname{card} Z_2 \geq \exp(-S) \operatorname{card} Z, \tag{3.7}$$

where

$$S = K\left(\log\frac{2n}{\xi(1)}\sum_{\ell\leq j}\xi(\theta 2^{-\ell}) + \theta\varepsilon^2 n \sum_{\ell\leq j} 2^{2\ell}\right) \tag{3.8}$$

and such that

$$\begin{aligned} &\forall f \in Z_2,\ \forall g \in Z_2,\ \forall i \leq n,\ \forall \ell \leq j,\\ &\quad f(x_i) - h(x_i) \in [2^{-\ell}, 2^{-\ell+1}]\\ &\qquad \Longrightarrow \quad |f(x_i) - g(x_i)| \leq 2^{-3}|f(x_i) - h(x_i)|. \end{aligned} \tag{3.9}$$

We use again the same construction for $-Z_2$ and $-h$, that is, we find a subset $Z'$ of $Z_2$ such that

$$\operatorname{card} Z' \ge \exp(-S) \operatorname{card} Z_2 \ge \exp(-2S) \operatorname{card} Z$$

and

$$\begin{aligned} &\forall f \in Z',\ \forall g \in Z',\ \forall i \le n,\ \forall \ell \le j, \\ &\quad -f(x_i) + h(x_i) \in [2^{-\ell}, 2^{-\ell+1}] \\ &\qquad \Longrightarrow \quad |f(x_i) - g(x_i)| \le 2^{-3}|f(x_i) - h(x_i)|. \end{aligned}$$

Thus in (3.9), if we replace $Z_2$ by $Z'$, we can replace the condition $f(x_i) - h(x_i) \in [2^{-\ell}, 2^{-\ell+1}]$ by the condition $|f(x_i) - h(x_i)| \in [2^{-\ell}, 2^{-\ell+1}]$, that is, we have

$$\begin{aligned} &\forall f \in Z',\ \forall g \in Z',\ \forall i \le n, \\ &\quad |f(x_i) - h(x_i)| \ge 2^{-j} \\ &\qquad \Longrightarrow \quad |f(x_i) - g(x_i)| \le 2^{-3}|f(x_i) - h(x_i)|. \end{aligned} \tag{3.10}$$

Consider then $f, g$ in $Z'$ and let $I = \{i \le n, |f(x_i) - h(x_i)| \ge 2^{-j}\}$. Then by (3.10) and (2.4), we have

$$\sum_{i\in I} \bigl(f(x_i) - g(x_i)\bigr)^2 \le 2^{-6} \sum_{i \in I} \bigl(f(x_i) - h(x_i)\bigr)^2 \le 2^{-4}\varepsilon^2 n. \tag{3.11}$$

Consider $J = \{i \le n; |g(x_i) - h(x_i)| \ge 2^{-j}\}$. We get, similarly,

$$\sum_{i \in J} \bigl(g(x_i) - f(x_i)\bigr)^2 \le 2^{-4}\varepsilon^2 n. \tag{3.12}$$

Set now $L = \{1, \dots, n\} \setminus (I \cup J)$. Combining (2.5) with (3.11) and (3.12), we have

$$\sum_{i \in L} \bigl(g(x_i) - f(x_i)\bigr)^2 \ge \frac{n\varepsilon^2}{8}.$$

Since

$$|g(x_i) - f(x_i)| \le |g(x_i) - h(x_i)| + |h(x_i) - f(x_i)| \le 2^{1-j}$$

for $i \in L$, we have shown that

$$\frac{1}{n} \sum_{i \le n} |g(x_i) - f(x_i)|^2 \wedge 2^{2-2j} \ge \frac{\varepsilon^2}{8}. \tag{3.13}$$

To conclude the proof, it suffices to observe that by (1.2),

$$2S \le K\biggl(\log\Bigl(\frac{2n}{\xi(1)}\Bigr)\xi(\theta 2^{-j}) + \theta\varepsilon^2 n 2^{2j}\biggr)$$

and to observe that together with (2.14) and the fact that $\operatorname{card} Z' \geq \exp(-2S) \times \operatorname{card} Z$ this implies (2.15). $\square$

The proof of Proposition 3.1 will require two steps. In the first step, for each $i \leq n$, we will construct 18 small intervals contained in the set $h(x_i) + [2^{-\ell-1}, 2^{-\ell+2}]$ and a large subset $Z''$ of $Z$ such that for $f$ in $Z''$, $f(x_i)$ never belongs to these small intervals.

We consider the set

$$H = \bigcup_{-1 \leq k \leq 2^4} [2^{-\ell}(1 + 2^{-4}k), 2^{-\ell}(1 + 2^{-4}k + \theta)],$$

that consists of 18 evenly spaced intervals of length $\theta 2^{-\ell}$. For $t \leq 2^{-4} - \theta$, we consider the sets

$$H_i(t) = H + h(x_i) + t2^{-\ell}.$$

LEMMA 3.2. *We can find a subset $Z''$ of $Z$ such that*

$$\operatorname{card} Z'' \geq \exp(-Kn\varepsilon^2\theta 2^{2\ell}) \operatorname{card} Z \tag{3.14}$$

*and numbers $(t_i)_{i \leq n}$, $0 \leq t_i \leq 2^{-4} - \theta$, such that*

$$\forall i \leq n, \qquad f \in Z'' \quad \Longrightarrow \quad f(x_i) \notin H_i(t_i). \tag{3.15}$$

PROOF. This is a counting argument. We consider the natural product probability $P$ on $\Omega' = [0, 2^{-4} - \theta]^n$. For $f$ in $Z$, we consider the event

$$A_i(f) = \{f(x_i) \notin H_i(t_i)\},$$

where $t = (t_i)_{i \leq n}$ is a generic point of $\Omega'$. If we denote by $\lambda$ the Lebesgue measure we have

$$(2^{-4} - \theta) P(A_i^c(f)) = \lambda\bigl(\{t_i \in [2^{-4} - \theta];\, f(x_i) \in H_i(t_i)\}\bigr).$$

Now,

$$\{t_i \in [2^{-4} - \theta];\, f(x_i) \in H_i(t_i)\} = \bigl\{[2^{-4} - \theta] \cap 2^{\ell}\bigl(-H + f(x_i) - h(x_i)\bigr)\bigr\}.$$

We observe that $2^{\ell}(-H + f(x_i) - h(x_i))$ is the union of intervals of length $\theta$ and that any two of these intervals are at a distance at least $2^{-4} - \theta$. Thus it should be obvious that the above set has a Lebesgue measure at most $\theta$. Thus

$$P(A_i(f)) \geq 1 - \frac{\theta}{2^{-4} - \theta} \geq 1 - 2^5\theta.$$

Moreover, $P(A_i(f)) = 1$ if $f(x_i) \leq h(x_i) + 2^{-\ell-1}$. We define

$$A(f) = \bigcap_{i \leq n} A_i(f) = \bigl\{t \in \Omega';\, \forall i \leq n,\, f(x_i) \notin H_i(t_i)\bigr\}.$$

By independence,

$$P(A(f)) \geq (1-2^5\theta)^m, \tag{3.16}$$

where $m = \operatorname{card}\{i \leq n;\ f(x_i) \geq h(x_i) + 2^{-\ell-1}\}$. By (2.4), we have $m \leq 16n\varepsilon^2 2^{2\ell}$. Since $\theta \leq 2^{-6}$, we have $1 - 2^5\theta \geq \exp(-2^6\theta)$ and thus by (3.16), we have

$$P(A(f)) \geq \exp(-2^{11}\theta n\varepsilon^2 2^{2\ell}). \tag{3.17}$$

Thus we can find $t$ such that the set

$$Z'' = \{f;\ A(f) \text{ occurs}\}$$

satisfies

$$\operatorname{card} Z'' \geq \operatorname{card} Z \exp(-2^{11}\theta n\varepsilon^2 2^{2\ell}). \qquad \square$$

PROOF OF PROPOSITION 3.1. We consider $Z''$ and the sets $H_i(t_i)$ as provided by Lemma 3.2. For a given $-1 \leq k \leq 2^4$ and $f$ in $Z''$, we consider the set

$$A(f;k) = \{i \leq n;\ f(x_i) \leq b_{i,k}\}, \tag{3.18}$$

where

$$b_{i,k} = h(x_i) + 2^{-\ell}(1 + 2^{-4}k + t_i)$$

and we consider the collection $\mathcal{A}_k$ of all the subsets of $\{1,\dots,n\}$ of the type (3.18).

The importance of (3.15) is that

$$i \notin A(f,k) \quad\Longrightarrow\quad f(x_i) \geq b_{i,k} + \theta 2^{-\ell} \tag{3.19}$$

because by (3.15), we have $f(x_i) > b_{i,k} \Rightarrow f(x_i) \geq b_{i,k} + \theta 2^{-\ell}$. Now if $\mathcal{A}_k$ shatters a subset $I$ of $\{1,\dots,n\}$, we see by (3.18) and (3.19) that $\mathcal{F}(\theta 2^{-\ell})$-shatters $I$ (by Definition 1.2); so that by (1.3), we have $\operatorname{card} I \leq m := \min(n, \lfloor \xi(\theta 2^{-\ell})\rfloor)$.

By the Sauer–Shelah theorem, we have

$$\operatorname{card} \mathcal{A}_k \leq \sum_{p\leq m} \binom{p}{n} \leq \left(\frac{en}{m}\right)^m \leq \exp\left(\log\left(\frac{en}{\xi(1)}\right)\xi(\theta 2^{-\ell})\right),$$

using that $\xi(\theta 2^{-\ell}) \geq \xi(1)$ and that the function $t \to (en/t)^t$ is increasing if $0 < t < n$. Thus there are at most $\exp(6\log(en/\xi(1))\xi(\theta 2^{-\ell}))$ possible 6-tuples $(A(f,k))_{-1\leq k\leq 4}$, $f$ in $Z$. Hence we can find a subset $Z_1$ of $Z''$ such that

$$\forall f, g \in Z_1,\ \forall k, -1 \leq k \leq 4, \qquad A(f,k) = A(g,k), \tag{3.20}$$

while

$$\operatorname{card} Z_1 \geq \exp\left(-6\log\left(\frac{en}{\xi(1)}\right)\xi(\theta 2^{-\ell})\right)\operatorname{card} Z''. \tag{3.21}$$

If

$$f(x_i) \in h(x_i) + [2^{-\ell}, 2^{-\ell+1}],$$

then there is $-1 \le k \le 2^4$ such that $b_{i,k} < f(x_i) \le b_{i,k+1}$ (there it helps to allow the value $k = -1$). Then $i \in A(f, k+1)$, $i \notin A(f, k)$, so that by (3.20), $i \in A(g, k+1)$, $i \notin A(g, k)$, that is, $b_{i,k} < g(x_i) \le b_{i,k+1}$, so that

$$|f(x_i) - g(x_i)| \le 2^{-4}2^{-\ell} \le 2^{-3}|f(x_i) - h(x_i)|,$$

because $f(x_i) - h(x_i) \ge 2^{-\ell}$. $\square$

**Acknowledgment.** I am very grateful to Richard Dudley for the gift of his book, a gift that triggered the long-delayed writing of this paper, and for thoughtful comments on the first version.

## REFERENCES


[1] ALON, N., BEN-DAVID, N., CESA-BIANCHI, D. and HAUSSLER, D. (1997). Scale-sensitive dimension, uniform convergence and learnability. *J. ACM* **44** 615–631.

[2] DUDLEY, R. M. (1987). Universal Donsker classes and metric entropy. *Ann Probab.* **12** 1306–1326.

[3] DUDLEY, R. M. (1999). *Uniform Central Limit Theorems.* Cambridge Univ. Press.

[4] DUDLEY, R. M., GINÉ, E. and ZINN, J. (1991). Uniform and universal Glivenko–Cantelli classes. *J. Theoret. Probab.* **4** 485–510.

[5] GINÉ, E. and ZINN, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.* **12** 929–989.

[6] GINÉ, E. and ZINN, J. (1991). Gaussian characterization of uniform Donsker classes of functions. *Ann. Probab.* **19** 758–782.

[7] ELTON, J. (1983). Sign embedding of $\ell_1^n$. *Trans. Amer. Math. Soc.* **175** 113–123.

[8] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–50.

[9] MENDELSON, S. and VERSHYNIN, R. (2003). Entropy and the combinatorial dimension. *Inv. Math.* **152** 37–55.

[10] TALAGRAND, M. (1987). The Glivenko–Cantelli problem. *Ann. Probab.* **15** 837–870.

[11] TALAGRAND, M. (1992). Type, infratype and the Elton–Pajor theorem. *Invent. Math.* **107** 41–59.

[12] TALAGRAND, M. (1996). The Glivenko–Cantelli theorem, 10 years later. *J. Theoret. Probab.* **9** 371–386.



EQUIPE D'ANALYSE FONCTIONNELLE
INSTITUT DE MATHÉMATIQUES
UMR 7586-CNRS
BOITE 186
4 PL. JUSSIEU 75230
PARIS 05
FRANCE
E-MAIL: mit@ccr.jussieu.fr

DEPARTMENT OF MATHEMATICS
OHIO STATE UNIVERSITY
COLUMBUS, OHIO 43210