Ratio Limit Theorems for Empirical Processes

Evarist Giné[1], Vladimir Koltchinskii[2], and Jon A. Wellner[3]

Abstract. Concentration inequalities are used to derive some new inequalities for ratio-type suprema of empirical processes. These general inequalities are used to prove several new limit theorems for ratio-type suprema and to recover a number of the results from [1] and [2]. As a statistical application, an oracle inequality for nonparametric regression is obtained via ratio bounds.

1. Introduction

Let $\mathcal{F}$ be a uniformly bounded class of real valued measurable functions on a probability space $(S, \mathcal{A}, P)$. To be specific, we assume most often that $\mathcal{F}$ takes values in $[0, 1]$ (although, in some places below, the class will be scaled differently). Let $X, X_i,\ i \in \mathbb{N}$, be i.i.d. random variables in $(S, \mathcal{A})$ with distribution P. We denote by P_n the empirical measure based on the sample $(X_1, \dots, X_n)$, $P_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$. Suppose that $\sigma_P(f)$ is defined in such a way that

$$\sigma_P^2(f) \geq \mathrm{Var}_P(f) := Pf^2 - (Pf)^2,\ f \in \mathcal{F}.$$

In particular, $\sigma_P(f)$ can be the standard deviation itself, or in can be equal to $\sqrt{Pf}$ (recall that f takes values in $[0, 1]$). In this note we present a simple technique to study the asymptotic behavior of the supremum of the standardized empirical process,

$$\sup_{f \in \mathcal{F}, \sigma_P(f) > \delta_n} \frac{\sqrt{n}|P_n f - Pf|}{\sigma_P(f)},$$

as well as some of its variations such as

$$\sup_{f \in \mathcal{F}, Pf > \delta_n} \left| \frac{P_n f}{Pf} - 1 \right| \quad \text{and} \quad \sup_{f \in \mathcal{F}, \sigma_P(f) > \delta_n} \frac{\sqrt{n}|P_n f - Pf|}{\omega(\sigma_P(f))}$$

for suitable 'moduli' ω and properly chosen 'cutoffs' δ_n depending on the complexity of the class $\mathcal{F}$. These questions for Vapnik-Červonenkis (VC) classes of sets

Key words and phrases. empirical processes, concentration inequalities, ratio limit theorems, nonparametric regression, oracle inequalities.
[1]Research partially supported by NSF Grant No. DMS-0070382.
[2]Research partially supported by NSA Grant No. MDA904-02-1-0075.
[3]Research partially supported by NSF Grants DMS-9532039 and DMS-0203320.

were studied by Alexander in [2], and his proofs were technically rather sophisticated. Our results apply to very general classes of functions and, particularly in the form they take when we specialize them to VC classes of functions, they may be considered as analogues of some of Alexander's results for VC classes of sets. The need for this kind of results, in the generality given here, is illustrated by an example. Indeed, as an application of our general theorems we obtain an 'oracle inequality' in a simple but quite general non-parametric regression setting (cf., [16], [19]). So, the type of inequalities proved in this article may turn out to be useful for bounding errors of prediction in Statistics and in Machine Learning.

The main advance on empirical process theory since 1987, when Alexander proved his results, has been Talagrand's discovery of concentration inequalities ([30],[31]). This tool allows us to handle ratios very easily by proving several simple exponential bounds expressed in terms of expectations of localized sup norms of empirical processes. These bounds are obtained by stratifying the class $\mathcal{F}$ according to variance size, applying Talagrand's inequality to each stratum and then collecting terms. This approach, as carried out here, originated in the more specialized setting of statistical learning theory and was developed by several authors (see, e.g., Koltchinskii and Panchenko [20], Koltchinskii [21], Panchenko [26, 27] Bousquet, Koltchinskii and Panchenko [10], Bartlett, Bousquet and Mendelson [5] and, especially, the Ph. D. dissertations of Panchenko [25] and Bousquet [7]). A very close approach has been developed in some other statistical applications even earlier (see [23] and references therein), and in one form or other is also present in [2]. The exponential bounds for ratios together with some new bounds on expectations of suprema of empirical processes over VC classes of functions ([30], [12], [11], [24]) allow one to obtain Alexander type theorems without any effort. The present approach may open a possibility to understand much better, and for much more general classes than VC, this important class of limit theorems for empirical processes, and in particular, to widen the scope of their applicability.

There is an extensive literature on ratio limit theorems for classical empirical processes (see e.g. Wellner [35]). For general empirical processes indexed by sets or functions some important references are [1], [2], [3], [4], [15], [19], [28], [32], [34].

In order to avoid measurability problems, in what follows, we will assume that the supremum over the class $\mathcal{F}$ or over any of the subclasses we consider is in fact a countable supremum. In this case we say that the class $\mathcal{F}$ is measurable.

With some abuse of notation we will write $\log m$ for $1 \vee \log m$ and $\log\log m$ for $1 \vee \log\log m$.

2. Ratio limit theorems: normalization with $\sigma_P(f)$

We introduce some notations used in what follows. We set

$$\mathcal{F}(r) := \{f \in \mathcal{F} : \sigma_P(f) \le r\}$$

and, for $r < s$,

$$\mathcal{F}(r, s] := \mathcal{F}(s) \setminus \mathcal{F}(r).$$

We denote

$$\xi_n(r,s] := \sup_{f\in\mathcal{F}(r,s]} \frac{|P_n f - Pf|}{\sigma_P(f)}.$$

Let now (given $r, s,\ r < s$) $q > 1$ and suppose that $s := rq^l$, for some $l \in \mathbb{N}$, so that

$$l = \log_q \frac{s}{r}.$$

[This will not be a loss of generality since the choice of q will be in our hands.] Let

$$\rho_j := rq^j,\ j = 0, \dots, l$$

(with $\rho_0 = r,\ \rho_l = s$). Then we define a function $\psi_{n,q}$ from $(r, s]$ into the real line by setting

$$\psi_{n,q}(u) := \mathbb{E}\|P_n - P\|_{\mathcal{F}(\rho_{j-1},\rho_j]},\ u \in (\rho_{j-1}, \rho_j],\ j = 1, \dots, l,$$

and we also set

$$\beta_{n,q}(r,s] := \sup_{u\in(r,s]} \frac{\psi_{n,q}(u)}{u}.$$

Given two sequences $\{r_n\},\ \{s_n\}$ of positive numbers such that $r_n < s_n$ we set

$$\xi_n := \xi_n(r_n, s_n],$$

and if $q_n \downarrow 1$ (so that $s_n = r_n q_n^{l_n}$ for an integer l_n) we define

$$\beta_n := \beta_{n,q_n}(r_n, s_n].$$

Our first goal is to prove the following general theorems. The only assumption on $\mathcal{F}$ is that it is a measurable class of functions taking values on $[0, 1]$.

Theorem 1. *Suppose that*

$$\sqrt{\frac{\log\log_{q_n}\frac{s_n}{r_n}}{n}} \bigvee \frac{1}{nr_n} = o(\beta_n).$$

Then

$$\frac{\xi_n}{\beta_n} \to 1 \ \text{ in Pr } \ \text{as } n \to \infty.$$

The following a.s. version holds under slightly stronger assumptions. For simplicity, we consider only the case of $s_n \equiv 1$.

Theorem 2. *Suppose that*

$$\sqrt{\frac{\log\log_{q_n}\frac{1}{r_n} + \log\log n}{n}} \bigvee \frac{\log\log n}{nr_n} = o(\beta_n).$$

In addition, suppose that

$$r_n \searrow \text{ and } \frac{\beta_n}{\sqrt{n}} \searrow .$$

Then

$$\limsup_{n\to\infty} \frac{\xi_n}{\beta_n} = 1 \text{ a.s.}$$

Somewhat stronger assumptions lead to a.s. convergence to 1:

Theorem 3. *Suppose that*

$$\sqrt{\frac{\log\log_{q_n}\frac{s_n}{r_n}+\log n}{n}}\bigvee\frac{\log n}{nr_n}=o(\beta_n).$$

Then

$$\lim_{n\to\infty}\frac{\xi_n}{\beta_n}=1\text{ a.s.}$$

Remarks. 1. It is easy to see that the conditions of Theorem 1 are satisfied if

$$\sqrt{\frac{\log\log n}{n}}=o(\beta_n)$$

and

$$r_n\sqrt{n\log\log n}\to\infty.$$

2. Only formal modifications in the proof given below for Theorem 1 also show that if

$$\sqrt{\frac{\log\log_{q_n}\frac{s_n}{r_n}}{n}}=O(\beta_n)$$

and the sequence $\{nr_n\beta_n\}$ is bounded away from 0, then

$$\frac{\xi_n}{\beta_n}=O_p(1).$$

3. Likewise, it can also be shown that if

$$\limsup\frac{1}{\beta_n}\sqrt{\frac{\log\log_{q_n}\frac{s_n}{r_n}}{n}}<1/2$$

and

$$\frac{1}{nr_n}=o(\beta_n),$$

then the sequence $\left\{\frac{\xi_n}{\beta_n}\right\}$ is both stochastically bounded and stochastically bounded away from 0.

4. If in Theorem 2 we replace $q_n\downarrow 1$ by $q>1$, and take $\beta_n\geq\beta_{n,q}(r_n,1]$, then the conditions

$$\sqrt{\frac{\log\log_{q_n}\frac{1}{r_n}+\log\log n}{n}}\bigvee\frac{\log\log n}{nr_n}=O(\beta_n),$$

$$r_n\searrow\text{ and }\frac{\beta_n}{\sqrt{n}}\searrow$$

imply

$$\limsup_{n\to\infty}\frac{\xi_n}{\beta_n}=R\text{ a.s.}$$

for some $R < \infty$. (This follows from obvious modifications in the proof of Theorem 2 given below together with the Kolmogorov 0-1 law as $nr_n\beta_n \to \infty$.)

5. For a general $\{s_n\}$, the extra conditions needed in Theorem 2 are $s_n \searrow$, $ns_n \nearrow$ and, most importantly,

$$\lim_{\varepsilon\to 0} \frac{\beta_{n,q_n}(r_n, s_n(1+\varepsilon)]}{\beta_n} = 1.$$

These assumptions (as well as the additional monotonicity assumptions in Theorem 2) come from a version of Lemma 7.2 in [2] that we use in the current proof. These assumptions might just be of a technical nature, and thus, perhaps superfluous.

The proofs are based on the following lemma.

Lemma 1. *For $t > 0$, define*

$$\Delta^{q,+}_{(r,s]}(t) := \sqrt{2q^2 \frac{t + 2\log\log_q \frac{s}{r}}{n} + 4\frac{t + c_q}{nr}\beta_{n,q}(r,s]} + \frac{t + c_q}{3nr}$$

and

$$\Delta^{q,-}_{(r,s]}(t) := \sqrt{2q^2 \frac{t + 2\log\log_q \frac{s}{r}}{n} + 4\frac{t + c_q}{nr}\beta_{n,q}(r,s]} + \frac{8}{3}\frac{t + c_q}{nr},$$

where $c_q := 2q\sup_{1\le j\le l} q^{-j}\log j$. Then

$$\mathbb{P}\left\{\sup_{f\in\mathcal{F}(r,s]} \frac{|P_nf - Pf|}{\sigma_P(f)} \ge \beta_{n,q}(r,s] + \Delta^{q,+}_{(r,s]}(t)\right\} \le 2e^{-t}$$

and

$$\mathbb{P}\left\{\sup_{f\in\mathcal{F}(r,s]} \frac{|P_nf - Pf|}{\sigma_P(f)} \le \frac{1}{q}\left[\beta_{n,q}(r,s] - \Delta^{q,-}_{(r,s]}(t)\right]\right\} \le 2e^{-t}.$$

Proof. For $q > 1$ and $0 < u \le 1$ define

$$\mathcal{F}_q(u) := \mathcal{F}(u/q, u]$$

and consider the following events

$$E^+_{q,u}(t) := \left\{\|P_n - P\|_{\mathcal{F}_q(u)} \le \mathbb{E}\|P_n - P\|_{\mathcal{F}_q(u)} + \sqrt{2\frac{t}{n}(u^2 + 2\psi_{n,q}(u))} + \frac{t}{3n}\right\}$$

and

$$E^-_{q,u}(t) := \left\{\|P_n - P\|_{\mathcal{F}_q(u)} \ge \mathbb{E}\|P_n - P\|_{\mathcal{F}_q(u)} - \sqrt{2\frac{t}{n}(u^2 + 2\psi_{n,q}(u))} - \frac{8t}{3n}\right\}.$$

By Talagrand's concentration inequalities for empirical processes (see [17], Theorem 1.1 and [9], Theorem 7.4), we have

$$\mathbb{P}(E^+_{q,u}(t)) \ge 1 - e^{-t} \text{ and } \mathbb{P}(E^-_{q,u}(t)) \ge 1 - e^{-t}$$

Let $\mathcal{F}_j := \mathcal{F}_{(\rho_{j-1},\rho_j]} = \mathcal{F}_q(\rho_j)$ and

$$E^+_j := E^+_{q,\rho_j}(t + 2\log j),\ E^-_j := E^-_{q,\rho_j}(t + 2\log j).$$

Then

$$\mathbb{P}\Big(\bigcap_{j=1}^{\infty} E_j^+\Big) \geq 1 - \sum_{j=1}^{\infty} e^{-t-2\log j} \geq 1 - e^{-t}\sum_{j=1}^{\infty} j^{-2} \geq 1 - 2e^{-t}$$

and similarly

$$\mathbb{P}\Big(\bigcap_{j=1}^{\infty} E_j^-\Big) \geq 1 - 2e^{-t}.$$

On the event $\bigcap E_j^+$, we have

$$\forall j\ \forall f \in \mathcal{F}_j, \quad \frac{|P_n f - Pf|}{\sigma_P(f)}$$

$$\leq \frac{\psi_{n,q}(\sigma_P(f))}{\sigma_P(f)} + \sqrt{2\frac{t+2\log j}{n}\left(q^2 + 2\frac{\psi_{n,q}(\sigma_P(f))}{\sigma_P^2(f)}\right)} + \frac{t+2\log j}{3n\sigma_P(f)}.$$

Note that, for $f \in \mathcal{F}_j$,

$$\frac{t+2\log j}{n\sigma_P(f)} \leq \frac{t+2\log j}{n\rho_{j-1}} \leq \frac{t}{nr} + 2\sup_j \frac{2\log j}{q^{j-1}}\frac{1}{nr} = \frac{t+c_q}{nr}.$$

Since also

$$j \leq l = \log_q \frac{s}{r},$$

we get

$$\begin{aligned}
&\frac{\psi_{n,q}(\sigma_P(f))}{\sigma_P(f)} + \sqrt{2\frac{t+2\log j}{n}\left(q^2 + 2\frac{\psi_{n,q}(\sigma_P(f))}{\sigma_P^2(f)}\right)} + \frac{t+2\log j}{3n\sigma_P(f)} \\
&\leq \quad \beta_{n,q}(r,s] + \sqrt{2q^2\frac{t+2\log\log_q \frac{s}{r}}{n} + 4\frac{t+c_q}{nr}\beta_{n,q}(r,s]} + \frac{t+c_q}{3nr} \\
&= \quad \beta_{n,q}(r,s] + \Delta^{q,+}_{(r,s]}(t).
\end{aligned}$$

Thus, on the event $\bigcap E_j^+$,

$$\forall\ f \in \mathcal{F}(r,s], \quad \frac{|P_n f - Pf|}{\sigma_P(f)} \leq \beta_{n,q}(r,s] + \Delta^{q,+}_{(r,s]}(t),$$

and the first bound follows.

Similarly, on the event $\bigcap_j E_j^-$, we have

$$\sup_{f\in\mathcal{F}(r,s]} \frac{|P_n f - Pf|}{\sigma_P(f)} = \sup_j \sup_{f\in\mathcal{F}_j} \frac{|P_n f - Pf|}{\sigma_P(f)} \geq \sup_j \frac{\|P_n - P\|_{\mathcal{F}_j}}{\rho_j}$$

$$\geq \sup_j \frac{\psi_{n,q}(\rho_j)}{\rho_j} - \sqrt{2\frac{t+2\log j}{n}\left(1 + 2\frac{\psi_{n,q}(\rho_j)}{\rho_j^2}\right)} - \frac{8}{3}\frac{t+2\log j}{n\rho_j}$$

$$\geq \frac{1}{q}\sup_j \sup_{\rho\in(\rho_{j-1},\rho_j]}\left[\frac{\psi_{n,q}(\rho)}{\rho} - \sqrt{2\frac{t+2\log j}{n}\left(1 + 2\frac{\psi_{n,q}(\rho)}{\rho^2}\right)} - \frac{8}{3}\frac{t+2\log j}{n\rho}\right]$$

and, exactly as in the case of the upper bound, this can be shown to be

$$\geq \frac{1}{q}\left[\beta_{n,q}(r,s] - \Delta^{q,-}_{(r,s]}(t)\right],$$

which yields the second inequality. □

Proof of Theorem 1. The condition of the theorem means that

$$\log\log_{q_n}\frac{s_n}{r_n} = o(n\beta_n^2)$$

and

$$n\beta_n r_n \to \infty \text{ as } n \to \infty.$$

It follows from Lemma 1 that with probability at least $1-4e^{-t}$,

$$q_n^{-1}\left(1 - \frac{1}{\beta_n}\Delta^{q_n,-}_{(r_n,s_n]}(t)\right) \leq \frac{\xi_n}{\beta_n} \leq 1 + \frac{1}{\beta_n}\Delta^{q_n,+}_{(r_n,s_n]}(t).$$

Then, the conditions of the theorem immediately imply that for all $t > 0$

$$\frac{1}{\beta_n}\Delta^{q_n,+}_{(r_n,s_n]}(t) \to 0 \text{ as } n \to \infty,$$

and the same is true for $\Delta^{q_n,-}_{(r_n,s_n]}(t)$, which is enough to complete the proof. □

Proof of Theorem 2. Passing from a probability upper bound to an a.s. upper bound, given Lemma 1, is standard, and the details within the present framework are worked out in [2], Lemmas 7.1 and 7.2. Alexander's Lemma 7.2 is stated for classes of sets, but its proof applies as well for classes of functions taking values in $[0,1]$. His lemma, adapted to our case, is as follows:

If $\beta_n/n^{1/2} \searrow$, $r_n \searrow$, then the condition

$$\mathbb{P}\left\{\frac{\xi_n}{\beta_n} > 1+\varepsilon\right\} = O\left((\log n)^{-1-\theta}\right) \tag{A}$$

for some $\varepsilon, \theta > 0$ implies

$$\mathbb{P}\left\{\frac{\xi_n}{\beta_n} > 1+2\varepsilon \quad \text{i.o.}\right\} = 0. \tag{B}$$

To establish (A), we take in Lemma 1 $t = 2\log\log n$, $q = q_n$, $s = 1$, $r = r_n$. The first bound of Lemma 1 then gives

$$\mathbb{P}\left\{\frac{\xi_n}{\beta_n} \geq 1 + \frac{1}{\beta_n}\Delta^{q_n,+}_{(r_n,1]}(2\log\log n)\right\} \leq 2\exp\{-2\log\log n\} = 2(\log n)^{-2}.$$

It follows from the conditions of the theorem that

$$\log\log_{q_n}\frac{1}{r_n} + \log\log n = o(n\beta_n^2)$$

and

$$\frac{n\beta_n r_n}{\log\log n} \to \infty \text{ as } n \to \infty.$$

This yields, by a straightforward computation, that

$$\frac{1}{\beta_n}\Delta^{q_n,+}_{(r_n,1]}(2\log\log n) \to 0 \text{ as } n \to \infty,$$

which establishes (A). Alexander's lemma now gives (B) for any $\varepsilon > 0$, so that

$$\limsup_{n\to\infty}\frac{\xi_n}{\beta_n} \le 1.$$

Hence, it remains to prove that the lim sup is ≥ 1. The second bound of Lemma 1 (with the same t, q, r, s as before) gives

$$\mathbb{P}\left\{\frac{\xi_n}{\beta_n} \le q_n^{-1}\left(1 - \frac{1}{\beta_n}\Delta^{q_n,-}_{(r_n,1]}(2\log\log n)\right)\right\} \le 2(\log n)^{-2}.$$

Take $n = n_k = e^k$. Since again, for all large enough n,

$$\frac{1}{\beta_n}\Delta^{q_n,-}_{(r_n,1]}(2\log\log n) < \varepsilon$$

we get

$$\mathbb{P}\left\{\frac{\xi_{n_k}}{\beta_{n_k}} \le 1 - \varepsilon\right\} = O(k^{-2}),$$

which by Borel–Cantelli Lemma implies that for all $\varepsilon > 0$

$$\mathbb{P}\left\{\frac{\xi_{n_k}}{\beta_{n_k}} \le 1 - \varepsilon \text{ i.o.}\right\} = 0,$$

and the result follows. □

Proof of Theorem 3. It is a straightforward application of Lemma 1 along with Borel–Cantelli Lemma. One should take $t := 2\log n$ this time. □

3. Continuity moduli of empirical processes

Our goal in this section is to study the asymptotic behavior of

$$\sup_{f\in\mathcal{F}(r_n,s_n]}\frac{n^{1/2}|P_nf - Pf|}{\omega(\sigma_P(f))}$$

with a properly chosen "continuity modulus" ω. This will provide a piece of information about the local continuity modulus of the empirical process $n^{1/2}(P_n - P)$ at $f = 0$. The global continuity modulus can be studied quite similarly. (See [2] for definitions and motivation.) Also, we concentrate on "in probability" results (their almost sure versions can be also obtained with a little extra work).

We use the notations of the previous section and define

$$\omega_n(u) := n^{1/2}\psi_{n,q_n}(u),\ u \in (r_n, s_n].$$

Theorem 4. *Let ω be a nonnegative nondecreasing bounded function on $[0,1]$, satisfying the conditions $\frac{\omega(u)}{u} \searrow$ and*

$$\sup_{u\in(0,1]} \frac{u}{\omega(u)}\sqrt{\log\log\frac{1}{u}} < \infty.$$

Suppose that

$$\omega_n(u) \le \omega(u),\ \ u \in [r_n, s_n].$$

If

$$\sup_n \frac{\log\log_{q_n}\frac{s_n}{r_n}}{\omega(r_n)\sqrt{n}} < \infty,$$

then

$$\sup_{f\in\mathcal{F}(r_n,s_n]} \frac{n^{1/2}|P_nf - Pf|}{\omega(\sigma_P(f))}$$

is stochastically bounded and uniformly (in n) bounded in L_1:

$$\sup_n \mathbb{E}\left\{\sup_{f\in\mathcal{F}(r_n,s_n]} \frac{n^{1/2}|P_nf - Pf|}{\omega(\sigma_P(f))}\right\} < \infty\,. \tag{3.1}$$

In Theorem 4, $q_n \downarrow 1$ can be replaced by $q > 1$ or by $1 < q_n < C < \infty$.

Theorem 5. *Let ω be a nonnegative nondecreasing bounded function on $[0,1]$, satisfying the conditions $\frac{\omega(u)}{u} \searrow$ and*

$$\frac{u}{\omega(u)}\sqrt{\log\log\frac{1}{u}} \to 0 \text{ as } u \to 0.$$

Suppose that

$$\sup_{u\in(r_n,s_n]}\left|\frac{\omega_n(u)}{\omega(u)} - 1\right| \to 0 \text{ as } n \to \infty.$$

If also

$$\frac{\omega(r_n)\sqrt{n}}{\log\log_{q_n}\frac{s_n}{r_n}} \to \infty,$$

then

$$\sup_{f\in\mathcal{F}(r_n,s_n]} \frac{n^{1/2}|P_nf - Pf|}{\omega(\sigma_P(f))} \to 1 \text{ as } n \to \infty \text{ in Pr.}$$

Proof of Theorem 4. We follow the proof of Lemma 1 with $r = r_n, s = s_n, q = q_n, l = l_n$. The definition of the events E_j^+, E_j^- is slightly different:

$$E_j^+ := E_{q,\rho_j}^+(t + 2\log(l-j+1)), E_j^- := E_{q,\rho_j}^-(t + 2\log(l-j+1)),$$

but we still have

$$\mathbb{P}\Big(\bigcap_{j=1}^{l} E_j^+\Big) \ge 1 - 2e^{-t},\ \mathbb{P}\Big(\bigcap_{j=1}^{l} E_j^-\Big) \ge 1 - 2e^{-t}.$$

On the event $\bigcap_{j=1}^{l} E_j^+$, we have

$$\forall j \ \forall\ f \in \mathcal{F}_j, \quad n^{1/2}|P_n f - Pf| \le \omega_n(\sigma_P(f))$$

$$+\sqrt{2\,(t + 2\log(l-j+1))\left(\sigma_P^2(f) + \frac{2\omega_n(\sigma_P(f))}{\sqrt{n}}\right)} + \frac{t + 2\log(l-j+1)}{3\sqrt{n}},$$

which under the assumptions about ω implies that

$$\forall j \ \forall\ f \in \mathcal{F}_j, \quad \frac{n^{1/2}|P_n f - Pf|}{\omega(\sigma_P(f))}$$

$$\le 1 + \sqrt{2(t + 2\log(l_n - j + 1))\left(\frac{\rho_j^2}{\omega^2(\rho_j)} + \frac{2}{\sqrt{n}\omega(\rho_{j-1})}\right)} + \frac{t + 2\log(l-j+1)}{3\sqrt{n}\omega(\sigma_P(f))}.$$

We have

$$\max_{1\le j\le l} \frac{\rho_j}{\omega(\rho_j)}\sqrt{\log(l-j+1)} \le \max_{1\le j\le l} \frac{\rho_j}{\omega(\rho_j)}\sqrt{\log\log_q \frac{s}{\rho_{j-1}}}$$

$$\le C \sup_{u\in(0,1]} \frac{u}{\omega(u)}\sqrt{\log\log\frac{1}{u}} =: K < +\infty$$

(for some constants C, K). Also, for all $j = 1, \dots, l$

$$\log(l-j+1)\frac{1}{\sqrt{n}\omega(\rho_{j-1})} \le \log\log_{q_n}\frac{s_n}{r_n}\frac{1}{\sqrt{n}\omega(r_n)},$$

which is bounded by the conditions. This allows us to easily conclude that on the event $\bigcap_{j=1}^{l} E_j^+$,

$$\forall j \ \forall\ f \in \mathcal{F}_j \ \frac{n^{1/2}|P_n f - Pf|}{\omega(\sigma_P(f))} \le K_1 t + K_2$$

with some constants K_1, K_2. Thus

$$\mathbb{P}\left\{\sup_{f\in\mathcal{F}(r_n,s_n]} \frac{n^{1/2}|P_n f - Pf|}{\omega(\sigma_P(f))} \ge K_1 t + K_2\right\} \le 2e^{-t},$$

implying the stochastic boundedness of the sequence in question. This bound also implies the boundedness in L_1 by simply integrating the tail bound. □

The proof of Theorem 5 requires to work out just several more details (including the lower bounds on the supremum) and it will not be given here.

4. Ratios $\frac{P_n f}{Pf}$: uniform LLN

We now turn to the study of

$$\sup_{Pf > r_n^2} \left| \frac{P_n f}{Pf} - 1 \right|.$$

Assuming that $nr_n^2 \to \infty$, we concentrate on determining necessary and sufficient conditions for the above suprema to converge to 0 in probability. Other types of ratio limit theorems can be studied as well using the methods of the previous sections.

In this section we set $\sigma_P(f) := \sqrt{Pf}$ and use all the notations of Section 2 (such as $\mathcal{F}(r,s]$, for instance). In particular, we need the functions $\psi_{n,q}$ to define the quantity

$$E_{n,q}(r,s] := \sup_{u \in (r,s]} \frac{\psi_{n,q}(u)}{u^2}.$$

Let the sequences r_n, s_n, q_n be as in Section 2 and let

$$E_n := E_{n,q_n}(r_n, s_n].$$

Theorem 6. *Suppose that $nr_n^2 \to \infty$ as $n \to \infty$. Choose $q_n \downarrow 1$ so that*

$$\log \frac{1}{q_n - 1} = o(nr_n^2) \text{ as } n \to \infty.$$

Then the condition $E_n \to 0$ as $n \to \infty$ is necessary and sufficient for

$$\sup_{f \in \mathcal{F}(r_n,s_n]} \left| \frac{P_n f}{Pf} - 1 \right| \to 0 \text{ in Pr.}$$

The proof is based on the following lemma.

Lemma 2. *For $t > 0$,*

$$\mathbb{P}\left\{ \sup_{f \in \mathcal{F}(r,s]} \left| \frac{P_n f}{Pf} - 1 \right| \geq E_{n,q}(r,s] + \sqrt{2\frac{t}{nr^2}(q^2 + 2E_{n,q}(r,s])} + \frac{t}{3nr^2} \right\} \leq \frac{q^2}{q^2-1} \frac{q}{t} e^{-t/q}$$

and

$$\mathbb{P}\left\{ \sup_{f \in \mathcal{F}(r,s]} \left| \frac{P_n f}{Pf} - 1 \right| \leq q^{-1}\left(E_{n,q}(r,s] - \sqrt{2\frac{t}{nr^2}(1 + 2E_{n,q}(r,s])} - \frac{8t}{3nr^2} \right) \right\} \leq \frac{q^2}{q^2-1} \frac{q}{t} e^{-t/q}.$$

Proof. It is similar to that of Lemma 1 and we use the notations introduced in that proof. The sets E_j^+, E_j^- are now defined as follows:

$$E_j^+ := E_{q,\rho_j}^+(tq^{2(j-1)}),\ E_j^- := E_{q,\rho_j}^-(tq^{2(j-1)}).$$

With this definition, we have

$$\mathbb{P}\Big(\bigcup_{j=1}^{\infty}(E_j^+)^c\Big) \le \sum_{j=0}^{\infty} e^{-tq^{2j}} = \frac{q^2}{q^2-1}\sum_{j=0}^{\infty} q^{-2j}\exp\{-tq^{2j}\}(q^{2j}-q^{2(j-1)})$$

$$\le \frac{q^2}{q^2-1}\int_{1/q}^{\infty} x^{-1}\exp\{-tx\}dx \le \frac{q^2}{q^2-1}\int_{t/q}^{\infty} y^{-1}e^{-y}dy \le \frac{q^2}{q^2-1}\frac{q}{t}e^{-t/q},$$

and similarly

$$\mathbb{P}\Big(\bigcup_{j=1}^{\infty}(E_j^-)^c\Big) \le \frac{q^2}{q^2-1}\frac{q}{t}e^{-t/q}.$$

On the event $\bigcap_{j=1}^{\infty} E_j^+$, we have

$$\forall j\ \forall f\in\mathcal{F}_j,\quad \frac{|P_nf-Pf|}{Pf}$$

$$\le \frac{\psi_{n,q}(\sqrt{Pf})}{Pf} + \sqrt{2\frac{tq^{2(j-1)}}{nPf}\Big(q^2+2\frac{\psi_{n,q}(\sqrt{Pf})}{Pf}\Big)} + \frac{tq^{2(j-1)}}{3nPf}.$$

Since for $f\in\mathcal{F}_j$,

$$q^{2(j-1)}r^2 < Pf \le q^{2j}r^2,$$

we get

$$\forall j\ \forall f\in\mathcal{F}_j,\quad \frac{|P_nf-Pf|}{Pf}$$

$$\le E_{n,q}(r,s] + \sqrt{2\frac{t}{nr^2}\Big(q^2+2E_{n,q}(r,s]\Big)} + \frac{t}{3nr^2},$$

which proves the first bound. The second one can be proved similarly. □

Proof of Theorem 6. Choose $t_n\to\infty$ so that $t_n = o(nr_n^2)$ and $\log(q_n-1)^{-1} = o(t_n)$. Then

$$\frac{q_n^2}{q_n^2-1}\frac{1}{t_n}e^{-t_n}\to 0$$

and we have

$$\begin{aligned} q_n^{-1}\Big(E_n - \sqrt{2\frac{t_n}{nr_n^2}(1+2E_n)} - \frac{8t_n}{3nr_n^2}\Big) \quad &\le \quad \sup_{f\in\mathcal{F}(r_n,s_n]}\Big|\frac{P_nf}{Pf}-1\Big| \\ &\le \quad E_n + \sqrt{2\frac{t_n}{nr_n^2}(q_n^2+2E_n)} + \frac{t_n}{3nr_n^2} \end{aligned}$$

with probability $1-o(1)$. This immediately implies the result. □

Here are two useful corollaries of Lemma 2 which we will use in section 7.

Corollary 1. *Suppose that $\mathcal{F}$ is a measurable class of functions with values in $[0,1]$ satisfying a.s.*

$$\log N(\mathcal{F}, L_2(P_n), \tau) \leq A\tau^{-\alpha}$$

for all $\tau > 0$ and some finite A and $\alpha \in (0,2)$. Then for $n \in \mathbb{N}$, $0 < \varepsilon \leq 1$, $q > 1$, $1 \geq \delta \geq \delta_n = (n\varepsilon)^{-2/(\alpha+2)}$, and a constant $C = C(A, \alpha, q)$ depending only on A, α and q,

$$\mathbb{P}\left\{ \sup_{f\in\mathcal{F}:\, Pf>\delta} \left| \frac{P_n f}{Pf} - 1 \right| \geq C\sqrt{\varepsilon} \right\} \leq \frac{q^2}{q^2-1} \frac{2q}{n\varepsilon\delta} e^{-n\varepsilon\delta/(2q)} = \frac{D}{n\varepsilon\delta} e^{-n\varepsilon\delta/(2q)} .$$

Proof. Choosing $r^2 = \delta$ and $t = n\varepsilon\delta/2$ in the first inequality of Lemma 2 gives

$$\mathbb{P}\left\{ \sup_{f\in\mathcal{F}: Pf>\delta} \left| \frac{P_n f}{Pf} - 1 \right| \geq E_{n,q}(\sqrt{\delta}, 1] + \sqrt{\varepsilon(q^2 + 2E_{n,q}(\sqrt{\delta}, 1])} + \varepsilon/6 \right\} \leq \frac{q^2}{q^2-1} \frac{2q}{n\varepsilon\delta} e^{-n\varepsilon\delta/(2q)} .$$

To bound $E_{n,q}(\sqrt{\delta}, 1]$ we argue as follows: By a standard symmetrization inequality, Dudley's entropy bound for Rademacher processes and our random entropy hypothesis

$$\begin{aligned}
\mathbb{E}\|P_n - P\|_{\mathcal{F}(r)} &\leq \frac{K}{\sqrt{n}} \mathbb{E}\left\{ \int_0^{\sqrt{\sup_{f\in\mathcal{F}(r)} P_n(f^2)}} \sqrt{\log N(\mathcal{F}(r), L_2(P_n), \tau)}\, d\tau \right\} \\
&\leq \frac{K}{\sqrt{n}} \int_0^{\sqrt{\mathbb{E}(\sup_{f\in\mathcal{F}(r)} P_n(f^2))}} \sqrt{A\tau^{-\alpha}}\, d\tau \\
&\leq \frac{K}{\sqrt{n}} \int_0^{\sqrt{\mathbb{E}(\sup_{f\in\mathcal{F}(r)} P_n(f))}} \sqrt{A\tau^{-\alpha}}\, d\tau \\
&\leq \frac{K}{\sqrt{n}} \int_0^{\sqrt{\mathbb{E}(\sup_{f\in\mathcal{F}(r)} |(P_n-P)(f)|)+r^2}} \sqrt{A\tau^{-\alpha}} d\tau \\
&\leq \frac{2}{2-\alpha} \frac{K\sqrt{A}}{\sqrt{n}} \left(E\|P_n - P\|_{\mathcal{F}(r)} + r^2 \right)^{1/2-\alpha/4} \\
&\leq \frac{B}{\sqrt{n}} \left\{ (E\|P_n - P\|_{\mathcal{F}(r)})^{1/2-\alpha/4} \vee (r^2)^{1/2-\alpha/4} \right\}
\end{aligned}$$

with $B = 2^{3/2-\alpha/4} K\sqrt{A}/(2-\alpha)$, where the second inequality follows from concavity of the integral $\int_0^x h(t)dt$ when h is non-increasing. Thus if $\mathbb{E}\|P_n - P\|_{\mathcal{F}(r)} \leq r^2$, then

$$\mathbb{E}\|P_n - P\|_{\mathcal{F}(r)} \leq \frac{B}{\sqrt{n}} r^{1-\alpha/2} ,$$

while if $\mathbb{E}\|P_n - P\|_{\mathcal{F}(r)} > r^2$, then

$$\mathbb{E}\|P_n - P\|_{\mathcal{F}(r)} \leq \left(\frac{B}{\sqrt{n}} \right)^{4/(2+\alpha)} = \tilde{B} n^{-2/(\alpha+2)} .$$

Combining these bounds yields

$$\mathbb{E}\|P_n - P\|_{\mathcal{F}(r)} \le \frac{B}{\sqrt{n}} r^{1-\alpha/2} \vee \tilde{B} n^{-2/(\alpha+2)}\,.$$

By taking $r = \rho_j = \sqrt{\delta_n} q^{2j}$ it follows that

$$\begin{aligned} \mathbb{E}\|P_n - P\|_{\mathcal{F}(\rho_j)} &\le \frac{B}{\sqrt{n}} \rho_j^{1-\alpha/2} \vee \tilde{B} n^{-2/(\alpha+2)} \\ &= \left\{ B\sqrt{\varepsilon}(n\varepsilon)^{-2/(\alpha+2)} q^{j(2-\alpha)} \right\} \bigvee \left\{ \tilde{B}\varepsilon^{2/(\alpha+2)} (n\varepsilon)^{-2/(\alpha+2)} \right\} \\ &\le M\sqrt{\varepsilon}\delta_n q^{j(2-\alpha)} \end{aligned}$$

with $M = B \vee \tilde{B}$. Hence it follows that

$$\begin{aligned} E_{n,q}(\sqrt{\delta}, 1] &= \sup_{\sqrt{\delta} < u \le 1} u^{-2} \psi_{n,q}(u) \le \max_{j \ge 1} \rho_{j-1}^{-2} \mathbb{E}\|P_n - P\|_{\mathcal{F}(\rho_j)} \\ &\le \max_{j\ge 1} \left\{ M\sqrt{\varepsilon} q^4 q^{-j(2+\alpha)} \right\} = M\sqrt{\varepsilon} q^{2-\alpha} \le M q^{2-\alpha}\,. \end{aligned}$$

Combining this bound with the first display of the proof yields the claimed inequality with

$$C \equiv C(A, \alpha, q) = M q^{2-\alpha} + \sqrt{q^2 + 2Mq^{2-\alpha}} + 1/6\,.$$

□

Similarly we can prove the following, using Lemma 2 and either direct computation or the bound in Corollary 3 below (or in Proposition 2.1 of [12]).

Corollary 2. *Suppose that $\mathcal{F}$ is a measurable class of functions with values in $[0,1]$ satisfying a.s.*

$$N(\mathcal{F}, L_2(P_n), \tau) \le \left(\frac{A}{\tau}\right)^v, \quad 0 < \tau \le 1,$$

for some $v \ge 1$ and $A \ge 2e\sqrt{v}$ (in particular, this holds if $\mathcal{F}$ is a VC class). Let $q > 1$ and let n, ε, δ and δ_n satisfy

$$\frac{1}{n} \le \varepsilon \le 1, \quad \text{and} \quad 1 \ge \delta \ge \delta_n := \frac{v \log \frac{A\sqrt{n\varepsilon}}{\sqrt{v}}}{n\varepsilon}.$$

Then, there exists a universal constant C such that

$$\mathbb{P}\left\{ \sup_{f \in \mathcal{F}:\, Pf > \delta} \left| \frac{P_n f}{Pf} - 1 \right| \ge C q^4 \sqrt{\varepsilon} \right\} \le \frac{q^2}{q^2 - 1} \frac{2q}{n\varepsilon\delta} e^{-n\varepsilon\delta/(2q)} = \frac{D}{n\varepsilon\delta} e^{-n\varepsilon\delta/(2q)}\,.$$

The constant C can be taken to be five times the constant in Corollary 3 below.

5. An inequality for expected values of empirical processes indexed by VC classes of functions

In the previous sections, either the results themselves or the conditions for their application are in terms of $\psi_{n,q}(u)$, that is, of $\mathbb{E}\|P_n - P\|_{\mathcal{F}(r,s]}$, and therefore require, for their application, of good estimates of the expectation of suprema of empirical processes indexed by general classes of functions. These are often available if the L_2 covering numbers of the classes are under control, as in the case of Vapnik-Červonenkis classes or the classes considered in Corollary 1. Next we give estimates for VC classes that improve on some of the estimates in the literature in that, instead of being in terms of σ and $\|F\|_\infty$, where F is a measurable envelope of the class, they are in terms of σ and $\|F\|_{L_2(P)}$ (when the functions in the class take values between -1 and 1). These estimates will be used in Section 6, and we think they can be useful elsewhere as well.

Let $\mathcal{F}$ be a uniformly bounded class of real valued measurable functions on a probability space $(S, \mathcal{A}, P)$. To be specific, assume the functions in $\mathcal{F}$ take values in $[-1, 1]$ and are centered. Assume also that the class $\mathcal{F}$ is adequately measurable (as described in the introduction) and VC, in particular,

$$N(\mathcal{F}, L_2(Q), \tau) \le \left(\frac{A\|F\|_{L_2(Q)}}{\tau}\right)^v \tag{5.1}$$

for all $0 < \tau < \|F\|_{L_2(Q)}$ and some finite A and v, that we assume $A \ge 2$ and $v \ge 1$ without loss of generality. Here, $1 \ge F \ge \sup_{f\in\mathcal{F}} |f|$ is a measurable envelope of the class $\mathcal{F}$. Let X, X_i, $i \in \mathbb{N}$, be i.i.d. (P) random variables (coordinates on a product probability space), and let P_n be the empirical measure corresponding to the variables X_i, as in previous sections. Let σ^2 be any number such that $\sup_f \mathbb{E}f^2(X) \le \sigma^2 \le \mathbb{E}F^2(X)$. The norm signs without specification will denote sup over the class $\mathcal{F}$. Here is the bound:

Theorem 7. *Under the assumptions in the above paragraph we have that for all $n \in \mathbb{N}$,*

$$\mathbb{E}\left\|\sum_{i=1}^n f(X_i)\right\|_{\mathcal{F}} \le C\left[\sqrt{v}\sqrt{n}\sigma\sqrt{\log\frac{A\|F\|_{L_2(P)}}{\sigma}} \vee v\log\frac{A\|F\|_{L_2(P)}}{\sigma} \vee \sqrt{v}\sqrt{n}A\|F\|_{L_2(P)}\exp\left(-\frac{9}{8}n\|F\|_{L_2(P)}^2\right)\right].$$

Proof. The square root trick for probabilities in [14], Lemma 3.3 and its remark -that misses a factor of 8-), give that for all $t \ge 47n\sigma^2$,

$$\mathbb{P}\left\{\left\|\sum_{i=1}^n f^2(X_i)\right\| \ge t\right\} \le \mathbb{E}\left[1 \wedge \left(8\left(\frac{A\|F\|_{L_2(P_n)}}{\sigma}\right)^v e^{-t/16}\right)\right].$$

By concavity of the function $1 \wedge x$ on $[0, \infty)$ and Hölder, we have

$$\mathbb{E}\left[1 \wedge \left(8\left(\frac{A\|F\|_{L_2(P_n)}}{\sigma}\right)^v e^{-t/16}\right)\right]$$

$$= \mathbb{E}\left[1 \wedge \left(8^{1/v}\left(\frac{A\|F\|_{L_2(P_n)}}{\sigma}\right)e^{-t/(16v)}\right)\right]^v$$
$$\le \mathbb{E}\left[1 \wedge \left(8^{1/v}\left(\frac{A\|F\|_{L_2(P_n)}}{\sigma}\right)e^{-t/(16v)}\right)\right]$$
$$\le 1 \wedge \left(\frac{8^{1/v}A\|F\|_{L_2(P)}}{\sigma}e^{-t/(16v)}\right).$$

Integrating this tail estimate one readily obtains:

Lemma 3. *Let $\mathcal{F}$ be a measurable VC class of P-centered functions taking values between -1 and 1, with $A \ge 2$ and $v \ge 1$ in (5.1). Let $F \ge \sup_{f\in\mathcal{F}}|f|$ be a measurable envelope of the class $\mathcal{F}$ and let $\sup_f \mathbb{E}f^2(X) \le \sigma^2 \le \mathbb{E}F^2(X)$. Then, for all $n \in \mathbb{N}$,*

$$\mathbb{E}\left\|\sum_{i=1}^n f^2(X_i)\right\| \le 135\left[n\sigma^2 \vee v\log\left(\frac{A\|F\|_{L_2(P)}}{\sigma}\right)\right].$$

Let ε_i, $i \in \mathbb{N}$, be independent Rademacher variables independent from the variables X_j, and let $\mathbb{E}_\varepsilon$ denote conditional expectation given the sequence $\{X_i\}$. The subgaussian entropy bound gives that

$$\mathbb{E}_\varepsilon\left\|\sum_{i=1}^n \varepsilon_i f(X_i)/\sqrt{n}\right\|_{\mathcal{F}} \le C\int_0^{\left(\|\sum_{i=1}^n f^2(X_i)\|/n\right)^{1/2}}\sqrt{v\log\left(\frac{A\|F\|_{L_2(P_n)}}{\varepsilon}\right)}d\varepsilon$$

for some universal constant C. Since

$$\int_0^{\left(\|\sum_{i=1}^n f^2(X_i)\|/n\right)^{1/2}}\sqrt{v\log\left(\frac{A\|F\|_{L_2(P_n)}}{\varepsilon}\right)}d\varepsilon \le D\sqrt{v}A\|F\|_{L_2(P_n)},$$

where $D = \int_0^1 \sqrt{\log u^{-1}}du$, the above integral is dominated by

$$\int_0^{\left(\|\sum_{i=1}^n f^2(X_i)\|/n\right)^{1/2}}\sqrt{v\log\left(\frac{2A\|F\|_{L_2(P)}}{\varepsilon}\right)}d\varepsilon$$
$$+ D\sqrt{v}A\|F\|_{L_2(P_n)}I_{\left[\|F\|_{L_2(P_n)}>2\|F\|_{L_2(P)}\right]}.$$

Regarding the second summand, Hölder's inequality followed by Bernstein's exponential inequality give

$$\mathbb{E}\left(\|F\|_{L_2(P_n)}I_{\left[\|F\|_{L_2(P_n)}>2\|F\|_{L_2(P)}\right]}\right) \le \|F\|_{L_2(P)}\exp\left(-\frac{9}{8}n\|F\|^2_{L_2(P)}\right).$$

For the first summand, we note that, by concavity of the integral of a decreasing function (as in Corollary 1), we have

$$\mathbb{E}\left[\int_0^{\left(\|\sum_{i=1}^n f^2(X_i)\|/n\right)^{1/2}}\sqrt{v\log\left(\frac{2A\|F\|_{L_2(P)}}{\varepsilon}\right)}d\varepsilon\right]$$

$$\leq \int_0^{\left(\mathbb{E}\|\sum_{i=1}^n f^2(X_i)\|/n\right)^{1/2}} \sqrt{v \log\left(\frac{2A\|F\|_{L_2(P)}}{\varepsilon}\right)} d\varepsilon.$$

Now, by regular variation, this integral is dominated by a constant times

$$\sqrt{v}\frac{1}{\sqrt{n}}\left(\mathbb{E}\left\|\sum_{i=1}^n f^2(X_i)\right\|\right)^{1/2}\left(\log\frac{2A\|F\|_{L_2(P)}}{\left(\mathbb{E}\left\|\sum_{i=1}^n f^2(X_i)\right\|/n\right)^{1/2}}\right)^{1/2},$$

which, by the lemma, is in turn dominated by a constant times

$$\frac{1}{\sqrt{n}}\left[\sqrt{n}\sigma\sqrt{v}\sqrt{\log\frac{A\|F\|_{L_2(P)}}{\sigma}} \vee v\log\frac{A\|F\|_{L_2(P)}}{\sigma}\right].$$

Collecting the above bounds and applying a desymmetrization inequality we obtain the desired bound. □

Corollary 3. *If in the previous theorem we also have $n\sigma^2 \geq A$, then there exists a universal constant C such that, for all $n \in \mathbb{N}$,*

$$\mathbb{E}\left\|\sum_{i=1}^n f(X_i)\right\|_{\mathcal{F}} \leq C\left[\sqrt{v}\sqrt{n}\sigma\sqrt{\log\frac{A\|F\|_{L_2(P)}}{\sigma}} \vee v\log\frac{A\|F\|_{L_2(P)}}{\sigma}\right].$$

Proof. It follows from the previous theorem and the inequality

$$\sqrt{\log x} \geq x\exp\left(-\frac{9}{8}x^2\right), \quad x \geq 2,$$

where we take $x = A\|F\|_{L_2(P)}/\sigma \geq A \geq 2$. □

The proof of Theorem 7 substantially modifies the proof of a similar bound (simpler, but with $U = \|F\|_\infty$ instead of $\|F\|_{L_2(P)}$) in [12]. In that proof, an abstract version of the square root trick (due to Ledoux and Talagrand [22]) was used, whereas here, as in [11], we use the Giné and Zinn [13] version of Le Cam's square root trick.

Remark. Bounds on expectations of empirical processes that take into account the norm of the envelope of the class can be obtained also under different assumptions on the entropy (in particular, for instance, in the setting of Corollary 1 of the previous section). We are not presenting these bounds here.

6. Ratio limit theorems for VC classes of functions

In this section we combine the main results from Sections 2-4 with the moment bound in section 5 in order to obtain analogues for VC classes of functions of some of the results in [2] for classes of sets. In what follows the class $\mathcal{F}$ is assumed to be a measurable VC class of functions (as defined in Section 5) taking values between 0 and 1, and otherwise, we resume the notation set up in Sections 1-4.

Let us fix $q > 1$. For $0 < r < 1$ we define

$$\mathcal{F}(r) := \{f \in \mathcal{F} : \sigma_P(f) \leq r\}, \quad \mathcal{F}_q(r) := \mathcal{F}(r) \setminus \mathcal{F}(r/q),$$

and let $F_{q,r} \le 1$ ($F_r \ge 1$) be a measurable envelope of the class $\mathcal{F}_q(r)$ (resp. $\mathcal{F}(r)$). These localized envelopes will play an important role in ratio limit theorems. The analogue for functions of the 'capacity function' in [2] is precisely

$$g(r) := \frac{q\|F_r\|_{L_2(P)}}{r} \vee 1,$$

however it is more convenient to localize a little more and define $g_q(r)$ as any function on $(0,1]$ such that

$$\frac{q\|F_{q,r}\|_{L_2(P)}}{r} \le g_q(r) \le \frac{q}{r}.$$

The following result is a version of Theorem 3.1, case (ii), in [2] for functions.

Theorem 8. *Let $\mathcal{F}$ be a measurable VC class of functions taking values on $[0,1]$ and let q be any number larger than 1. Define*

$$r_n := \sup\left\{r > 0 : \ r \le \sqrt{\frac{\log g_q(r) \vee \log\log n}{n}}\right\}$$

and

$$b_n := \sqrt{\log g_q(r_n) \vee \log\log n}.$$

Then, the sequence

$$\sup_{f\in\mathcal{F},\sigma_P(f)>r_n} \frac{n^{1/2}|P_n f - Pf|}{b_n \sigma_P(f)}, \quad n \in \mathbb{N},$$

is stochastically bounded. If moreover the sequence b_n/n is nonincreasing, then there is $R < \infty$ such that

$$\limsup_{n\to\infty} \sup_{f\in\mathcal{F},\sigma_P(f)>r_n} \frac{n^{1/2}|P_n f - Pf|}{b_n \sigma_P(f)} = R \quad a.s.$$

Proof. It suffices to show that r_n and $\beta_n := K b_n/\sqrt{n}$ for $K < \infty$ to be specified later, satisfy the conditions of the second remark following Theorem 3 for stochastic boundedness, and of the fourth for a.s. boundedness. The definitions readily imply that

$$r_n \ge \sqrt{\frac{\log\log n}{n}},$$

which immediately gives

$$\sqrt{\frac{\log\log r_n^{-1} \vee \log\log n}{n}} \bigvee \frac{\log\log n}{n r_n} = O\left(b_n/\sqrt{n}\right).$$

So, it remains to verify that

$$K b_n/\sqrt{n} \ge \beta_{n,q}(r_n, 1]$$

for some $K < \infty$ and from some n on. If $r > r_n$, then for some $K < \infty$ and for all $n > \exp(e^{q^2A^2})$ we have

$$r > \sqrt{\frac{\log g_q(r) \vee \log\log n}{n}} \geq \frac{1}{\sqrt{2}}\sqrt{\frac{\log(Ag_q(r)) \vee \log\log n}{n}} \geq \frac{1}{\sqrt{2}}\sqrt{\frac{\log(Ag_q(r))}{n}},$$

and also $n\sigma^2 \geq nr^2/q^2 > A^2$. Thus, the corollary to Theorem 7 shows that there is a constant K depending only on A and v such that

$$\mathbb{E}\,\|P_n - P\|_{\mathcal{F}_q(r)} \leq Kr\sqrt{\frac{\log(g_q(r))}{n}} \leq \frac{Krb_n}{\sqrt{n}},$$

which gives

$$Kb_n/\sqrt{n} \geq \beta_{n,q}(r_n, 1].$$

□

The next result is an analogue for VC classes of functions of Theorem 4.4 in [2].

Theorem 9. *Let $\mathcal{F}$ be a measurable VC class of functions taking values on $[0,1]$ and let q be any number larger than 1. Define*

$$r_n := \sup\left\{r > 0 :\ r \leq \sqrt{\frac{\log g_q(r) \vee \log\log r^{-1}}{n}}\right\}$$

and

$$\omega(r) := r\sqrt{\log g_q(r) \vee \log\log_q r^{-1}}.$$

Assume $\omega(r) \nearrow$ and $\omega(r)/r \searrow$. Then, the sequence

$$\sup_{f\in\mathcal{F},\sigma_P(f)>r_n} \frac{n^{1/2}|P_nf - Pf|}{\omega(\sigma_P(f))}, \quad n \in \mathbb{N},$$

is stochastically bounded.

Proof. We will check the conditions of Theorem 4 for $\bar{\omega} := K\omega$, for $K < \infty$ conveniently chosen, and for $q_n = q > 1$. If $r > r_n$ then, by the definition of r_n,

$$r > \sqrt{\frac{\log g_q(r) \vee \log\log r^{-1}}{n}}$$

and in particular, $r > \sqrt{(\log\log r^{-1})/n}$, which implies that $nr^2 > A^2$ from some n on, n depending only on A. These two observations imply, by the corollary to Theorem 7, that

$$\|P_n - P\|_{\mathcal{F}_q(r)} \leq \frac{K\omega(r)}{\sqrt{n}},$$

that is,

$$\omega_n(r) \leq \bar{\omega}(r), \quad r \in (r_n, 1]$$

if we take $\bar{\omega} = K\omega$. The definition of ω already implies that

$$\frac{u\sqrt{\log\log u^{-1}}}{\omega(u)} \leq 1$$

for all $u \in (0,1)$. The definition of r_n also implies that $r_n \geq \sqrt{(\log\log r_n^{-1})/n}$, and this immediately gives

$$\frac{\log\log r_n^{-1}}{\sqrt{n}\omega(r_n)} \leq 1.$$

Now, the result follows from Theorem 4. □

We conclude with an analogue for functions of part of Theorem 5.1 in [2].

Theorem 10. *Let $\mathcal{F}$ be a measurable VC class of functions taking values on* $[0,1]$. *Let $r_n \to 0$ and $nr_n^2 \to \infty$, and let $q_n \downarrow 1$ be such that*

$$\log\frac{1}{q_n - 1} = o(nr_n^2).$$

Assume $g_{q_n}(r)$ is nonincreasing. Then, if

$$\frac{\log g_{q_n}(r_n)}{nr_n^2} \to 0,$$

we have

$$\sup_{f\in\mathcal{F}, Pf>r_n^2}\left|\frac{P_nf}{Pf} - 1\right| \to 0 \ \text{ in Pr.}$$

Proof. We check the conditions of Theorem 6 with $s_n = 1$. We will apply the corollary to Theorem 7 with $\sigma = \sup_f\sqrt{Pf}$, and envelope the square root of the natural envelope, which we can since the functions in $\mathcal{F}$ take values on $[0,1]$. Obviously $nr_n^2 > A^2$ from some $n = n(A)$ on. Then, for these values of n, by the corollary to Theorem 7,

$$\begin{aligned} E_{n,q_n}(r_n, 1] &\leq \sup_{r>r_n}\frac{1}{r^2}\mathbb{E}\,\|P_n - P\|_{\mathcal{F}(r/q_n, r)} \\ &\leq K\sup_{r>r_n}\left(\sqrt{\frac{\log Ag_{q_n}(r)}{nr^2}} + \frac{\log Ag_{q_n}(r)}{nr^2}\right) \end{aligned}$$

for some constant K that depends only on A and v. But the monotonicity of g_{q_n} implies that this sup is attained at $r = r_n$, which implies that $E_{n,q_n}(r_n, 1] \to 0$. Now, Theorem 6 implies the result. □

When specialized to VC classes of sets, the last three theorems completely recover the results of [2] mentioned in this section, up to constants. In particular then, one gets the classical results for the empirical distribution function and the empirical measure of intervals when P is uniform on $[0,1]^d$. For example, if $\mathcal{F}_1 = \{I_{[0,a]} : 0 \leq a \leq 1, \prod_{i=1}^d a_i \leq 1/2\}$, (here $a = (a_1, \ldots, a_d)$ and $[0,a] = \{(x_1, \ldots, x_d) : 0 \leq x_i \leq a_i, i = 1, \ldots, d\}$) then we take $\sigma_P^2[0,a] = P[0,a]$, and we find that $F_{q,r}(x) = 1\{x \in [0,1]^d : \prod_{j=1}^d x_j \le r^2\}$ so, with $X = (X_1, \ldots, X_d)$ and $X_i \sim \text{Uniform}[0,1]$,

$$\begin{aligned}
\|F_{q,r}\|^2_{L_2(P)} &= P(X_1 \cdots X_d \le r^2) \\
&= P(-\log X_1 - \cdots - \log X_d > -\log(r^2)) \\
&= P(\text{Gamma}(d,1) > -\log(r^2)) \\
&= P(\text{Poisson}(-\log(r^2)) < d) = \sum_{j=0}^{d-1} r^2 \frac{(-\log(r^2))^j}{j!} \\
&\sim r^2 \frac{(2\log(1/r))^{d-1}}{(d-1)!} \qquad \text{as} \quad r \downarrow 0\,.
\end{aligned}$$

So, we have $g_q(r) \simeq (2\log r^{-1})^{(d-1)/2}/\sqrt{(d-1)!}$, $r_n \simeq \sqrt{(\log\log n)/n}$ in Theorems 8 and 9, $b_n \simeq \sqrt{\log\log n}$ and $\omega(r) \simeq r\sqrt{\log\log r^{-1}}$. For Theorem 10, we can take any r_n such that $nr_n^2 \to \infty$ for $d = 1$ (thus recovering a result of [35]), and such that $nr_n^2/\log\log n \to \infty$ for $d > 1$. Likewise, if $\mathcal{F}_2 = \{I_{[a,b]} : 0 \le a_i \le b_i \le 1, \prod_{i=1}^d (b_i - a_i) \le 1/2\}$, then $\|F_{q,r}\|_{L_2(P)} = 1$, $g_q(r) \simeq 1/r$, $r_n \simeq \sqrt{(\log n)/n}$ in Theorems 8 and 9, $b_n \simeq \sqrt{\log n}$ and $\omega(r) \simeq r\sqrt{\log r^{-1}}$; and for Theorem 10, we can take any r_n such that $nr_n^2/\log n \to \infty$.

The results in this section and the previous examples illustrate one of the main points of this article, namely, that very general theorems, that apply to classes of functions that may not even be VC and which have very simple proofs, are sharp (at least up to constants) when specialized to VC classes of sets and functions and, in particular, to the classical settings of distribution functions and the empirical measure of intervals.

7. An Oracle Inequality for Regression via Ratio Bounds

Here we give an application of the ratio bounds in Section 6 to a statistical problem in the setting of nonparametric regression. The type of inequality we prove in this section provides an 'in probability' type of 'oracle inequality' for a simple version of this type of problem. For a nice introduction to oracle inequalities more generally, see [16]. The flavor of our result here is somewhat akin to the results of [18]. For an example of some L_2-type oracle inequalities see e.g. [19]. Massart in [23] develops a very general framework for oracle inequalities in many statistical problems including regression.

Consider the following regression model:

$$Y_i = f_0(X_i) + \xi_i$$

where the variables X_i are i.i.d. with law P, f_0 is a bounded measurable function and the variables ξ_i's are i.i.d. $N(0,1)$ (other distributions are possible), independent from the variables X_j. For a class of functions $\mathcal{F}$ define

$$\widehat{f}_n \equiv \mathrm{argmin}_{f\in\mathcal{F}} n^{-1}\sum_{j=1}^{n}(Y_j - f(X_j))^2 ,$$

and

$$\overline{f} \equiv \mathrm{argmin}_{f\in\mathcal{F}} P(f-f_0)^2 ,$$

(assuming, for simplicity, the existence of the argmins). Since the only norms occurring in this section are $L_2(P)$ norms, we set, from here on, $\|\cdot\| = \|\cdot\|_{L_2(P)}$.

Theorem 11. *Suppose that $\mathcal{F}$ is a measurable class of functions taking values in $[0,1]$, and with $L_2(P_n)$ metric entropies bounded a.s. by $A\tau^{-\alpha}$ with $0<\alpha<2$, as in Corollary 1, and let $C = C(2^\alpha A, \alpha, 3/2)$ be as defined in this corollary. Then there exist constants $C_i = C_i(A,\alpha)$, $i = 1,2,3$, depending only on A and α, such that for all $n\in\mathbb{N}$ and $\varepsilon\in(0,1/3]$ satisfying $\log\log_{3/2}(n\varepsilon^2/C^2)^{1/(2+\alpha)} \le (n\varepsilon^2/C^2)^{2\alpha/(2+\alpha)}$, the bound*

$$\|\widehat{f}_n - f_0\|^2 \le \frac{1+2\varepsilon}{1-2\varepsilon}\|\overline{f}-f_0\|^2 + \frac{C_1(A,\alpha)}{\varepsilon(n\varepsilon^2)^{2/(2+\alpha)}}$$

holds with probability at least $1-\tau_n$, where

$$\tau_n = C_2\exp\left\{-C_3(n\varepsilon^2)^{\alpha/(2+\alpha)}\right\}.$$

Proof. Let $C = C(2^\alpha A,\alpha,3/2)$ and $D = D(3/2) = 27/5$ be constants as defined in Corollary 1 and, given $0<\varepsilon\le 1/3$, set $\delta_n := (n\varepsilon^2/C^2)^{-2/(2+\alpha)}$. On the event

$$L_n := \left\{\|\widehat{f}_n - \overline{f}\|^2 < \delta_n\right\}$$

we have

$$\|\widehat{f}_n - f_0\| \le \sqrt{\delta_n} + \|\overline{f}-f_0\| .$$

This yields, using $ab\le(a^2+b^2)/2$,

$$\begin{aligned}\|\widehat{f}_n - f_0\|^2 &\le \|\overline{f}-f_0\|^2 + 2\sqrt{\delta_n}\|\overline{f}-f_0\| + \delta_n\\ &= \|\overline{f}-f_0\|^2 + 2\sqrt{\frac{\delta_n}{\varepsilon}}\sqrt{\varepsilon}\|\overline{f}-f_0\| + \delta_n\\ &\le (1+\varepsilon)\|\overline{f}-f_0\|^2 + \left(1+\frac{1}{\varepsilon}\right)\delta_n . \qquad (7.1)\end{aligned}$$

So, if $\mathbb{P}(L_n)\ge 1-\tau_n$, the theorem is proved. Otherwise, we must look at $\|\widehat{f}_n - f_0\|$ on L_n^c. To this end, we first note that, letting P and P_n denote respectively the law of (X,ξ) and the empirical measure of the variables (X_i,ξ_i), $i=1,\ldots,n$, when both types of variables occur (and only the law of X and the empirical measure of the X_i when ξ or ξ_i are not present),

$$\begin{aligned}\|\widehat{f}_n - f_0\|^2 &= \|\overline{f} - f_0\|^2 + P\left[(\widehat{f}_n - Y)^2 - (\overline{f} - Y)^2\right] \\ &\le \|\overline{f} - f_0\|^2 + (P - P_n)\left[(\widehat{f}_n - Y)^2 - (\overline{f} - Y)^2\right],\end{aligned}$$

where the identity follows from the orthogonality with respect to P between ξ and any square integrable function of X, and the inequality holds because $\widehat{f}_n$ minimizes the empirical squared error. The idea is to apply the ratio bound in Corollary 1 to the last term, but we cannot apply it directly (since the class of functions involved is not bounded because of the Gaussian noise) and some modifications are needed. Note that

$$(\widehat{f}_n - Y)^2 - (\overline{f} - Y)^2 = (\widehat{f}_n - f_0)^2 - (\overline{f} - f_0)^2 - 2(\widehat{f}_n - \overline{f})\xi,$$

and therefore, that

$$\begin{aligned}(P - P_n)\left[(\widehat{f}_n - Y)^2 - (\overline{f} - Y)^2\right] &\le (P - P_n)\left[(\widehat{f}_n - f_0)^2 \vee \delta_n - (\overline{f} - f_0)^2 \vee \delta_n\right] \\ &\quad + 2P_n\left[(\widehat{f}_n(x) - \overline{f}(x))y\right] + 2\delta_n.\end{aligned}$$

If we apply Corollary 1 for the class

$$\mathcal{G} \equiv \{(f - f_0)^2 \vee \delta_n : \ f \in \mathcal{F}\}$$

to the first term and with $q = 3/2$ (note that the $L_2(Q)$ entropies of $\mathcal{G}$ are dominated by $\overline{A}\tau^{-\alpha} = 2^\alpha A\tau^{-\alpha}$), we obtain that the two bounds

$$(P - P_n)\left((\widehat{f}_n - f_0)^2 \vee \delta_n\right) \le \varepsilon\|\widehat{f}_n - f_0\|^2 + \varepsilon\delta_n$$

and

$$(P_n - P)\left((\overline{f}_n - f_0)^2 \vee \delta_n\right) \le \varepsilon\|\overline{f}_n - f_0\|^2 + \varepsilon\delta_n$$

hold together with probability at least

$$1 - \frac{C^2 D}{n\varepsilon^2\delta_n}\exp\left\{-\frac{n\varepsilon^2\delta_n}{3C^2}\right\}.$$

We thus conclude that the inequality

$$(1 - \varepsilon)\|\widehat{f}_n - f_0\|^2 \le (1 + \varepsilon)\|\overline{f} - f_0\|^2 + \frac{2}{n}\sum_{i=1}^n \xi_i(\widehat{f}_n - \overline{f})(X_i) + 2(1 + \varepsilon)\delta_n \tag{7.2}$$

holds with probability at least

$$1 - \frac{C^2 D}{n\varepsilon^2\delta_n}\exp\left\{-\frac{n\varepsilon^2\delta_n}{3C^2}\right\}.$$

Now we need to bound the second term on L_n^c. First we consider

$$A_n := \frac{2}{n}\sum_{j=1}^n \xi_j(\widehat{f}_n(X_j) - \overline{f}(X_j))I\left(P_n(\widehat{f}_n - \overline{f})^2 \le \delta_n\right).$$

We have

$$|A_n| \le \frac{2}{\sqrt{n}} \sup_{f,g\in\mathcal{F}, P_n(f-g)^2\le\delta_n} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i (f(X_i) - g(X_i)) \right|,$$

and by the Gaussian entropy bound, if we let $\mathbb{E}_\xi$ denote conditional expectation given the variables X_i,

$$\begin{aligned}
\mathbb{E}_\xi |A_n| &\le \frac{2K}{\sqrt{n}} \int_0^{\sqrt{\delta_n}} \sqrt{\log N(\mathcal{F}-\mathcal{F}, L_2(P_n), \tau)}\, d\tau \\
&\le \frac{2K}{\sqrt{n}} \int_0^{\sqrt{\delta_n}} \sqrt{\log N^2(\mathcal{F}, L_2(P_n), \tau/2)}\, d\tau \\
&\le \frac{\kappa'(A,\alpha)}{\sqrt{n}} \int_0^{\sqrt{\delta_n}} \sqrt{\tau^{-\alpha}}\, d\tau \\
&\le \frac{\kappa(A,\alpha)\delta_n^{(1-\alpha/2)/2}}{\sqrt{n}},
\end{aligned}$$

where K is a universal constant and κ and κ' are constants that depend only on A and α, and we can assume $\kappa(A,\alpha) \ge 1$. Then, the Borell-Sudakov-Tsirel'son inequality (cf. [22], page 57, comments following (3.2) or [33], Proposition A.2.1 and its proof) gives

$$\mathbb{P}_\xi \left\{ |A_n| \ge \frac{\kappa(A,\alpha)\sqrt{\delta_n^{1-\alpha/2}}}{\sqrt{n}} + \sqrt{\frac{8\delta_n t}{n}} \right\} \le e^{-t},$$

so that, taking $t = \kappa^2(A,\alpha)\delta_n^{-\alpha/2}$, we obtain that

$$|A_n| \le 4\kappa(A,\alpha) \sqrt{\frac{\delta_n^{1-\alpha/2}}{n}} \tag{7.3}$$

with probability at least

$$1 - \exp\left\{-\kappa^2(A,\alpha)\delta_n^{-\alpha/2}\right\} = 1 - \exp\left\{-\kappa^2(A,\alpha)(n\varepsilon^2/C^2)^{\alpha/(2+\alpha)}\right\}.$$

Let now

$$B_n := \frac{2}{n} \sum_{j=1}^n \xi_j (\widehat{f}_n(X_j) - \overline{f}(X_j)) I\left(P_n(\widehat{f}_n - \overline{f})^2 > \delta_n\right),$$

which we decompose as

$$B_n = \left\{P_n(\widehat{f}_n - \overline{f})^2\right\}^{1/2} \frac{\frac{2}{n}\sum_{i=1}^n \xi_i(\widehat{f}_n - \overline{f})(X_i)}{\left\{P_n(\widehat{f}_n - \overline{f})^2\right\}^{1/2}} I\left(P_n(\widehat{f}_n - \overline{f})^2 > \delta_n\right).$$

By using the ratio bound again (Corollary 1), it follows that the intersection of L_n^c with the event that

$$\left\{P_n(\widehat{f}_n - \overline{f})^2\right\}^{1/2} \le (1+\varepsilon)^{1/2}\|\widehat{f}_n - \overline{f}\|$$

has probability at least

$$\mathbb{P}(L_n^c) - (C^2 D/(n\varepsilon^2\delta_n))\exp\left\{-n\varepsilon^2\delta_n/(3C^2)\right\}.$$

So, the intersection of L_n^c with the event that

$$B_n \le \frac{2(1+\varepsilon)^{1/2}\|\widehat{f}_n - \overline{f}\|}{\sqrt{n}} \cdot \sup_{\substack{f,g\in\mathcal{F}\\ P_n(f-g)^2>\delta_n}} \frac{\sum_{i=1}^n \xi_i(f-g)(X_i)/\sqrt{n}}{\{P_n(f-g)^2\}^{1/2}}$$

also has at least this probability. To estimate

$$C_n := \sup_{\substack{f,g\in\mathcal{F}\\ P_n(f-g)^2>\delta_n}} \frac{\sum_{i=1}^n \xi_i(f-g)(X_i)/\sqrt{n}}{\{P_n(f-g)^2\}^{1/2}}$$

we proceed as in Lemmas 1 and 2, and we isolate the computation:

Lemma 4. *Let* $\xi := (\xi_1,\ldots,\xi_n)$ *be* $N(0,I)$ *and let*

$$\mathcal{A}_n := \left\{\left(\frac{(f-g)(X_i)/\sqrt{n}}{\{P_n(f-g)^2\}^{1/2}} : i=1,\ldots,n\right) : f,g\in\mathcal{F}, P_n(f-g)^2 > \delta_n\right\}.$$

Then, we have that, for the constant $\kappa(A,\alpha)$ *in the bound (7.3) for* A_n *and for any* $q>1$,

$$\sup_{a\in\mathcal{A}_n} |\langle a,\xi\rangle| \le 2\kappa(A,\alpha)q\delta_n^{-\alpha/4}$$

with probability at least $1 - 2\exp\left\{-\kappa^2(A,\alpha)q^2\delta_n^{-\alpha/2} + 2\log\log_q \delta_n^{-1/2}\right\}$.

Proof of the Lemma. Set $r_n = \sqrt{\delta_n}$, let $q>1$ be such that $r_n q^{l_n} = 1$, l_n a positive integer, and, for $r \ge r_n$, let

$$\mathcal{A}_n(r) := \left\{a \in \mathcal{A}_n : \sqrt{\mathbb{E}_\xi\langle a,\xi\rangle^2} \in (r/q, r]\right\}.$$

Then, by the Gaussian entropy bound we have, as above,

$$\begin{aligned}\mathbb{E}_\xi \|\langle a,\xi\rangle\|_{\mathcal{A}_n(r)} &\le \frac{q}{r}\mathbb{E}_\xi\left(\left|\sum_{i=1}^n \xi_i(f-g)(X_i)/\sqrt{n}\right| I\left(P_n(f-g)^2 \le r^2\right)\right)\\ &\le \kappa(A,\alpha)qr^{-\alpha/2}/2,\end{aligned}$$

Hence, again by Borell-Sudakov-Tsirel'son,

$$\mathbb{P}_\xi\left\{\|\langle a,\xi\rangle\|_{\mathcal{A}_n(r)} > \kappa(A,\alpha)qr^{-\alpha/2}/2 + \sqrt{2t}\right\} \le e^{-t}.$$

So, if

$$E_j^+(t) = \left\{ \|\langle a, \xi\rangle\|_{\mathcal{A}_n(r_n q^j)} \le \kappa(A,\alpha) q (r_n q^j)^{-\alpha/2}/2 + \sqrt{2(t + 2\log j)} \right\},$$

we have

$$\mathbb{P}\left(\bigcap_{j=1}^{\infty} E_j^+(t) \right) \ge 1 - 2e^{-t},$$

and, on the event $\bigcap_{j=1}^{\infty} E_j^+(t)$,

$$\forall j, \forall a \in \mathcal{A}_n(r_n q^j), \quad |\langle a, \xi\rangle| \le C(A,\alpha) q r_n^{-\alpha/2}/2 + \sqrt{2\left(t + 2\log\log_q \frac{1}{r_n}\right)}$$

(see the proof of Lemma 1). If we now take

$$t = \kappa^2(A,\alpha) q^2 r_n^{-\alpha} - 2\log\log_q r_n^{-1},$$

the lemma follows (we are making the tacit assumption that this quantity is positive; if it is not positive, the lemma is true but meaningless). $\square$

Thus, we conclude that, at least with the large probability prescribed by the lemma, we have (with $q = 3/2$)

$$C_n \le 3\kappa(A,\alpha)\delta_n^{-\alpha/4}.$$

Hence, the probability of the intersection of L_n^c with the event

$$B_n \le \frac{6\kappa(A,\alpha)(1+\varepsilon)^{1/2}}{n^{1/2}\delta_n^{\alpha/4}} \|\widehat{f}_n - \overline{f}\| \tag{7.4}$$

has probability at least

$$\begin{aligned} \mathbb{P}(L_n^c) \ &- \ (D/(n\varepsilon^2/C^2)^{\alpha/(2+\alpha)}) \exp\left\{ -(n\varepsilon^2/C^2)^{\alpha/(2+\alpha)}/3 \right\} \\ &-2\exp\left\{ -\frac{9}{4}\kappa^2(A,\alpha)(n\varepsilon^2/C^2)^{\alpha/(2+\alpha)} + 2\log\log_{3/2}(n\varepsilon^2/C^2)^{1/(2+\alpha)} \right\}, \end{aligned}$$

where we have replaced δ_n by its value $(n\varepsilon^2/C^2)^{-2/(2+\alpha)}$.

Collecting the bounds (7.2)-(7.4) together with their probabilities, and using $\|\widehat{f}_n - \overline{f}\| \le \|\widehat{f}_n - f_0\| + \|\overline{f} - f_0\|$, we obtain that the intersection of L_n^c with the event that

$$\begin{aligned} &(1-\varepsilon)\|\widehat{f}_n - f_0\|^2 \\ &\quad \le \ (1+\varepsilon)\|\overline{f} - f_0\|^2 + \frac{2(1+\varepsilon)}{(n\varepsilon^2/C^2)^{2/(2+\alpha)}} + 4\kappa(A,\alpha)\frac{\varepsilon/C}{(n\varepsilon^2/C^2)^{2/(2+\alpha)}} \\ &\qquad + \ 6\kappa(A,\alpha)\frac{\varepsilon(1+\varepsilon)^{1/2}/C}{(n\varepsilon^2/C^2)^{1/(2+\alpha)}} \left(\|\widehat{f}_n - f_0\| + \|\overline{f} - f_0\| \right) \end{aligned}$$

has probability at least

$$\mathbb{P}(L_n^c) - 2(D/(n\varepsilon^2/C^2)^{\alpha/(2+\alpha)}) \exp\left\{-(n\varepsilon^2/C^2)^{\alpha/(2+\alpha)}/3\right\}$$
$$-3\exp\left\{-\kappa^2(A,\alpha)(n\varepsilon^2/C^2)^{\alpha/(2+\alpha)} + 2\log\log_{3/2}(n\varepsilon^2/C^2)^{1/(2+\alpha)}\right\}.$$

Since D is a constant and C depends only on A and α and can be taken to be at least 1, it is clear that, under the assumption in the theorem about n and ε, we can find $C_2(A,\alpha)$ and $C_3(A,\alpha)$ such that the above probability is at least $\mathbb{P}(L_n^c) - \tau_n$, with τ_n as in the statement of the theorem. Combining with the bound (7.10), that holds on L_n, we obtain that

$$\begin{aligned}
&(1-\varepsilon)\|\widehat{f}_n - f_0\|^2 \\
&\quad\le (1+\varepsilon)\|\overline{f} - f_0\|^2 + \frac{1+1/\varepsilon}{(n\varepsilon^2/C^2)^{2/(2+\alpha)}} + 4\kappa(A,\alpha)\frac{\varepsilon/C}{(n\varepsilon^2/C^2)^{2/(2+\alpha)}} \\
&\qquad + 6\kappa(A,\alpha)\frac{\varepsilon(1+\varepsilon)^{1/2}/C}{(n\varepsilon^2/C^2)^{1/(2+\alpha)}}\left(\|\widehat{f}_n - f_0\| + \|\overline{f} - f_0\|\right)
\end{aligned}$$

holds with probability at least $1-\tau_n$. Using $ab \le (a^2+b^2)/2$, $0 < \varepsilon \le 1/3$, and collecting terms, the above inequality implies the following one:

$$\begin{aligned}
(1-2\varepsilon)\|\widehat{f}_n - f_0\|^2 \quad &\le \quad (1+2\varepsilon)\|\overline{f} - f_0\|^2 \\
&\quad + \left(\varepsilon\lambda(A,\alpha) + \frac{2C^{4/(2+\alpha)}}{\varepsilon}\right)\frac{1}{(n\varepsilon)^{2/(2+\alpha)}},
\end{aligned}$$

where

$$\lambda(A,\alpha) := 4\kappa(A,\alpha)C^{(2-\alpha)/(2+\alpha)} + 48\kappa^2(A,\alpha)C^{-2\alpha/(2+\alpha)}.$$

Dividing both sides by $1-2\varepsilon \ge 1/3$, the bound in the theorem follows e.g. for $C_1(A,\alpha) = \lambda(A,\alpha)/9 + 2C^{4/(2+\alpha)}/3$. $\square$

The rate prescribed in Theorem 11 obviously depends on the complexity of the class, in particular, a better rate obtains for VC type classes. Using Corollary 2 instead of Corollary 1 in the above proof, and taking

$$\delta_n = \frac{\log\frac{(2/3)^4 A\sqrt{n}\varepsilon}{C\sqrt{v}}}{(2/3)^8 n\varepsilon^2/(vC^2)}$$

gives the following:

Theorem 12. *Let $\mathcal{F}$ be a measurable class of functions taking values in $[0,1]$ satisfying the entropy condition of Corollary 2 with $v \ge 1$ and $A \ge 2\sqrt{ev}$. Let n and $C^2(3/2)^8/n < \varepsilon^2 \le 1/9$ be such that $\delta_n \le 1/(2A^2)$, and assume $\log\delta_n^{-1} >$*

$\log\log_{3/2}\delta_n^{-1/2}$. *Then, there exist constants* $C_i = C_i(A, v) > 1$, $i = 1, 2, 3$, *depending only on* A *and* v, *such that*

$$\|\widehat{f}_n - f_0\|^2 \le \frac{1+2\varepsilon}{1-2\varepsilon}\|\overline{f} - f_0\|^2 + C_1\frac{\log(n\varepsilon^2)}{n\varepsilon^3},$$

with probability at least $1 - C_2(n\varepsilon^2)^{-C_3}$.

Acknowledgement. We are grateful to Olivier Bousquet for several interesting exchanges on different aspects of this article and for making his manuscript [9] available to us.

References

[1] Alexander, K.S., *Rates of growth for weighted empirical processes.* In: Proc. of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Volume II, L. Le Cam and R. Olshen (Eds.) (1985), Wadsworth, Inc., 475–493.

[2] Alexander, K.S., *Rates of Growth and Sample Moduli for Weighted Empirical Processes Indexed by Sets.* Probability Theory and Related Fields **75** (1987), 379–423.

[3] Alexander, K.S., *The Central Limit Theorem for Weighted Empirical Processes Indexed by Sets.* J. Multivariate Analysis **22** (1987), 313–339.

[4] Bartlett, P. and Lugosi, G., *An inequality for uniform deviations of sample averages from their means.* Statistics and Probability Letters **44** (1999), 55–62.

[5] Bartlett, P., Bousquet, O. and Mendelson, S. *Localized Rademacher Complexities.* In: Computational Learning Theory, Lecture Notes in Artificial Intelligence (2002), Springer, 44–58.

[6] Bercu, B., Gassiat, E., and Rio, E. *Concentration inequalities, large and moderate deviations for self-normalized empirical processes.* Annals of Probability **30**, 1576-1604.

[7] Bousquet, O. *Concentration Inequalities and Empirical Processes Theory Applied to the Analysis of Learning Algorithms*, Ph.D. Thesis, Ecole Polytechnique, Paris (2002).

[8] Bousquet, O. *A Bennett concentration inequality and its applications to empirical processes.* C.R. Acad. Sci. Paris, Ser. I 334 (2002), 495–500.

[9] Bousquet, O. Concentration Inequalities for Sub-Additive Functions Using the Entropy Method (2003), Preprint.

[10] Bousquet, O., Koltchinskii, V. and Panchenko, D. *Some Local Measures of Complexity of Convex Hulls and Generalization Bounds.* In: Computational Learning Theory, Lecture Notes in Artificial Intelligence (2002), Springer, 59–73.

[11] Einmahl, U. and Mason, D., *An empirical process approach to the uniform consistency of kernel type function estimators.* J. Theor. Probab. **13** (2000), 1-37.

[12] Giné, E. and Guillou, A., *On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals.* Ann. I. H. Poincaré **4** (2001), 503-522.

[13] Giné, E. and Zinn, J., *Some limit theorems for empirical processes.* Ann. Probab. **12** (1984), 929-989.

[14] Giné, E. and Zinn, J., *Lectures on the central limit theorem for empirical processes.* Probability and Banach Spaces, Zaragoza, Spain. Lecture Notes in Math., **1221** (1986), 50-113.

[15] Haussler, D., *Decision theoretic generalizations of the PAC model for neural nets and other learning applications.* Information and Computation **100** (1992), 78–150.

[16] Johnstone, I. M. *Oracle inequalities an nonparametric function estimation.* Proceedings of the International Congress of Mathematicians. Vol. III, Berlin (1998). Doc. Math. **1998**, Extra Vol. III, 267–278.

[17] Klein, T. *Une inégalité de concentration á gauche pour les processus empiriques.* C. R. Acad. Sci. Paris, Ser. I, (2002), 334:500–505.

[18] Kneip, A. *Ordered linear smoothers.* Ann. Statist. **22**, 835–866.

[19] Kohler, M. *Inequalities for uniform deviations of averages from expectations with applications to nonparametric regression*. J. of Statistical Planning and Inference **89** (2000), 1–23.

[20] Koltchinskii, V. and Panchenko, D. *Empirical Margin Distributions and Bounding the Generalization Error of Combined Classifiers*. Ann. Statist. **30** (2002), 1–50.

[21] Koltchinskii, V. *Bounds on Margin Distributions in Learning Problems.* Ann. Inst. H. Poincaré, (2003) to appear.

[22] Ledoux, M. and Talagrand, M., *Probability in Banach spaces.* Springer, New York, 1991.

[23] Massart, P. *Some applications of concentration inequalities in statistics.* Annales de la Faculté des Sciences de Tolouse IX (2000), 245–303.

[24] Mendelson, S. *Rademacher averages and phase transitions in Glivenko–Cantelli classes* IEEE Transactions on Information Theory (2002), to appear.

[25] Panchenko, D. *Concentration inequalities in product spaces and applications to statistical learning theory*, Ph.D. Thesis, University of New Mexico, Albuquerque (2002).

[26] Panchenko, D. *Some extensions of an inequality of Vapnik and Chervonenkis.* Electronic Communic. in Probab. **7** (2002).

[27] Panchenko, D. *Symmetrization Approach to Concentration Inequalities for Empirical Processes*. Ann. Probab., (2003) to appear.

[28] Pollard, D., *Uniform ratio limit theorems for empirical processes* Scandinavian J. Statistics **22** (1995), 271–278.

[29] Rio, E., *Une inégalité de Bennett pour les maxima de processus empiriques* Colloque en l'honneur de J. Bretagnolle, D. Dacunha-Castelle et I. Ibragimov (2001).

[30] Talagrand, M., *Sharper bounds for Gaussian and empirical processes.* Ann. Probab. **22** (1994), 28-76.

[31] Talagrand, M., *New concentration inequalities in product spaces.* Invent. Math. **126** (1996), 505-563.

[32] van de Geer, S.A. *Applications of Empirical Processes Theory.* Cambridge University Press, Cambridge (2000).

[33] van der Vaart, A. W. and Wellner, J. A. *Weak Convergence and Empirical Processes.* Springer Verlag, New York

[34] Vapnik, V.N. *Statistical Learning Theory.* John Wiley & Sons, New York (1998).

[35] Wellner, J.A. *Limit theorems for the ratio of the empirical distribution function to the true distribution function.* Z. Wahrscheinlichkeitstheor. Verw. Geb. **45** (1978), 108–123.

Departments of Mathematics and Statistics, University of Connecticut, Storrs, CT 06269, USA
E-mail address: `gine@uconnvm.uconn.edu`

Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131-1141, USA
E-mail address: `vlad@math.unm.edu`

University of Washington, Department of Statistics, Box 354322, Seattle, Washington 98195-4322, USA
E-mail address: `jaw@stat.washington.edu`