Random processes via the combinatorial dimension: introductory notes

M. Rudelson*   R. Vershynin†

Abstract

This is an informal discussion on one of the basic problems in the theory of empirical processes, addressed in our preprint *"Combinatorics of random processes and sections of convex bodies"*, which is available at ArXiV and from our web pages.

Arguably, the central problem of the theory of empirical processes is the following:

> *Describe the classes of functions on which the classical limit theorems of probability hold uniformly.*

Precisely, for a given probability space (Ω, μ), one looks at the sequence of independent samples (X_i) with values in Ω and distributed according to the law μ. Then the problem is to describe the classes F of functions $f : \Omega \to \mathbb{R}$ for which the sequence of real valued random variables $f(X_i)$ satisfies the classical limit theorems of probability uniformly over $f \in F$. The two classical limit theorems we have in mind are the law of large numbers, for which such classes are called *Glivenko-Cantelli*, and the central limit theorem, which gives rise to *Donsker* classes.

In practice one does not know the law μ according to which the samples X_i are drawn. Thus a particularly vital question is to know what classes are Glivenko-Cantelli or Donsker for every μ. Such classes are called *universal* or even *uniform* if the convergence in the corresponding limit theorems is uniform over all μ. We refer the reader to Chapter 14.3 of the book of Ledoux and Talagrand [LT] for a brief introduction to the theory of empirical processes and to the book of Dudley [Du 99] for a comprehensive account.

*Department of Mathematics, University of Missouri, Columbia, MO 65211, USA; e-mail: rudelson@math.missouri.edu

†Deptartment of Mathematics, University of California, Davis, CA 95616, USA; e-mail: vershynin@math.ucdavis.edu

Vapnik and Chervonenkis were first to realize that the problem above is intimately connected with the combinatorics of the class F. In their pioneering works [VC 68, VC 71, VC 81], they quantified the size of a class F by what now is called the *Vapnik-Chervonenkis dimension.* This dimension makes sense only for Boolean classes, i.e. for classes of $\{0,1\}$-valued functions. A natural extension of this notion to general classes is possible (although not unique); the resulting concept is called by Talagrand "the quantity of fundamental importance" [T 02]. This quantity measures how much F oscillates.

Precisely, for a given $t \geq 0$, a subset σ of Ω is called t-shattered by a class F if there exists a level function h on σ such that, given any partition $\sigma = \sigma_- \cup \sigma_+$, one can find a function $f \in F$ with $f(x) \leq h(x)$ if $x \in \sigma_-$ and $f(x) \geq h(x) + t$ if $x \in \sigma_+$. The *combinatorial dimension* of F, denoted by $v(F,t)$, is the maximal cardinality of a set t-shattered by F. Simply speaking, $v(F,t)$ is the maximal size of a set on which F oscillates in all possible $\pm t/2$ ways around some level h.

For Boolean classes, this is clearly the classical Vapnik-Chervonenkis dimension in the whole nontrivial range $0 < t < 1$. For general classes, the combinatorial dimension appears implicitely in the work of Talagrand [T 92] and explicitely in the paper of Alon et al [ABCH].

Intuitively, the smaller the combinatorial dimension of a class F, the less F oscillates, hence the better chances are for F to be a Donsker class. Recalling the known methods such as Dudley's integral inequality and Sudakov minoration, it is not difficult to locate the obstacles for the class to be universal Donsker and, in fact, to guess the optimal description of such classes:

Conjecture 1 *For every uniformly bounded class F,*

$$\int_0^\infty \sqrt{v(F,t)}\, dt < \infty \;\Rightarrow\; F \text{ is universal Donsker} \;\Rightarrow\; v(F,t) = O(t^{-2}).$$

The history of the work related to this conjecture goes back to 70's. The right hand side, which makes sense for $t \to 0$, is known [Du 99]. The left hand side is difficult.

In 1978, R. Dudley proved Conjecture 1 for classes of $\{0,1\}$-valued functions, where it clearly reads as follows: F is a uniform Donsker class if and only if its combinatorial dimension $v(F,1)$ is finite (see [Du 99], [LT] 14.3). As Talagrand mentiones in [LT] 14.3, this remarkable characterization is one of the main results on the combinatorics of empirical processes on $\{0,1\}$ classes.

In his 1992 Inventiones paper, Talagrand proved Conjecture 1 for convex classes, however up to an additional factor of $\log^M(1/t)$ in the integrand [T 92]. In 2002, Talagrand rewrote his proof in [T 02], now valid for arbitrary classes, and asked about the optimal value of the absolute constant exponent M.

In their recent Inventiones paper [MV 03], Mendelson and Vershynin reduced the exponent of the logarithm in Talagrand's inequality to $M = 1/2$. This was achieved by a new argument, via construction of an unbalanced separated tree in F. However, these methods alone could not possibly remove the factor of $\sqrt{\log(1/t)}$.

The logarithmic factor is removed completely in the forthcoming preprint [RV]. Thus the optimal exponent is $M = 0$ and Conjecture 1 is true. Proving this requires, in addition to the combinatorial ingredients of [MV 03], a new iteration argument and a general probabilistic reduction scheme. Along the proof of Conjecture 1, its quantitative form would be especially helpful in practical applications. It can be written in words as follows:

> *In Dudley's entropy integral inequality, the entropy can be replaced by the combinatorial dimension.*

The argument is based on the comparison of the combinatorial dimension and the uniform entropy $D(F,t)$ (also known as Koltchinskii-Pollard entropy), which is the supremum of the metric entropies of F with respect to the metric of $L_2(\mu)$ over all probabilities μ. Even though the combinatorial dimension and the uniform entropy are not equivalent in general, the situation changes dramatically when we look at regular classes, those for which, say, $v(F, 2t) \le \frac{1}{2} v(F,t)$ for all $t > 0$. Talagrand proved the required equivalence for these classes, however with an additional factor of $\log^M(1/t)$, where M is an absolute constant [T 02]. In the Inventiones paper [MV 03] of Mendelson and Vershynin, the exponent was reduced to $M = 1$ even *without* the regularity asumption on the class. In general, this exponent is optimal.

In [RV] we prove that *with* the regularity assumption, the logarithmic factor can be removed completely; thus the optimal exponent in Talagrand's result is $M = 0$, hence

> *the uniform entropy and the combinatorial dimension are equivalent for all regular classes.*

The argument is similar to that for Conjecture 1: the iteration scheme works in both cases, for the integral and for the regular quantity.

References

[ABCH] N. Alon, S. Ben-David, N. Cesa-Bianchi, D. Hausser, *Scale sensitive dimensions, uniform convergence and learnability*, Journal of the ACM 44 (1997), 615–631

[Du 99] R.M. Dudley, *Uniform central limit theorems*, Cambridge University Press, 1999.

[LT] M. Ledoux and M. Talagrand, *Probability in Banach spaces*, Springer, 1991

[MV 03] S. Mendelson, R. Vershynin, *Entropy, dimension and the Elton-Pajor Theorem*, Inventiones Mathematicae 152 (2003), 37–55

[RV] M. Rudelson, R. Vershynin, *Combinatorics of random processes and sections of convex bodies*, preprint available at ArXiV `http://front.math.ucdavis.edu`, Banach Space Bulletin `http://www.math.okstate.edu/~alspach/banach` and our webpages, `http://www.math.ucdavis.edu/~vershynin` and `http://www.math.missouri.edu/~rudelson`

[T 92] M. Talagrand, *Type, infratype, and Elton-Pajor Theorem*, Inventiones Math. 107 (1992), 41–59

[T 02] M. Talagrand, *Vapnik-Chervonenkis type conditions and uniform Donsker classes of functions*, Ann. Probab. 31 (2003) 1565–1582

[VC 68] V. Vapnik, A. Chervonenkis, *Uniform convergence of frequencies of occurence of events to their probabilities*, Dokl. Akad. Nauk SSSR 181 (1968), 781–783 (Russian), Sov. Math. Doklady 9 (1968), 915–918 (English translation)

[VC 71] V. Vapnik, A. Chervonenkis, *On the uniform convergence of relative frequencies of events to their probabilities*, Theory Probab. Appl. 16 (1971), 264–280

[VC 81] V. Vapnik, A. Chervonenkis, *Necessary and sufficient conditions for the uniform convergence of empirical means to their expectations*, Theory Probab. Appl. 3 (1981), 532–553