Chapter 2
A Paper that Created Three New Fields: Teoriya Veroyatnosteĭ i Ee Primeneniya 16(2), 1971, pp. 264–279

R.M. Dudley

Abstract This is an introduction, with remarks, to the paper by Vapnik and Chervonenkis in which they gave proofs for the strikingly innovative statements they had announced in Doklady Akademii Nauk SSSR three years earlier.

I apologize in advance for using some notations and terminology which differ from those of Vapnik and Chervonenkis (1971).

Let F be an infinite set and $\mathcal{C}$ a collection of subsets of F. For each finite set $A \subset F$ let $\Delta^{\mathcal{C}}(A)$ be the number of subsets of A of the form $B \cap A$ for $B \in \mathcal{C}$. For each n let $m^{\mathcal{C}}(n)$ be the maximum of $\Delta^{\mathcal{C}}(A)$ over all $A \subset F$ with cardinality $|A| = n$. $\mathcal{C}$ is said to *shatter* A if $\Delta^{\mathcal{C}}(A) = 2^{|A|}$. Let $S(\mathcal{C})$ be the largest cardinality of a shattered set, if it is finite, otherwise $S(\mathcal{C}) = +\infty$. The class $\mathcal{C}$ is called a *Vapnik–Chervonenkis* or *VC* class if and only if $S(\mathcal{C}) < +\infty$. Then let $V(\mathcal{C}) = S(\mathcal{C}) + 1$, the smallest m so that no set A with $|A| = m$ is shattered.

By defining these notions for a general set F, Vapnik and Chervonenkis founded three fields or branches:

1. A subfield of combinatorics, studying VC classes in themselves. This subfield has only, to my knowledge, a small literature relative to the other two fields to be mentioned, containing for example the papers by Stengle and Yukich [4] and Laskowski [3].
2. A subfield of probability, called *empirical processes*, in which the empirical measure $P_n = \frac{1}{n}\sum_{j=1}^{n} \delta_{X_j}$ and process $\sqrt{n}(P_n - P)$ are considered for i.i.d. random elements X_j taking values in general spaces, and limit theorems are considered, first laws of large numbers $\sup_{C\in\mathcal{C}} |(P_n - P)(A)| \to 0$ in probability (weak law of large numbers) or almost surely (strong law), where $\mathcal{C}$ is a VC class of sets and suitable measurability conditions hold, later extended to other classes of sets and functions; and central limit theorems in which $\sqrt{n}(P_n - P)$ is shown to converge in distribution with respect to uniform convergence over $\mathcal{C}$, or later over suitable

R.M. Dudley (✉)
Department of Mathematics, Massachusetts Institute of Technology,
Cambridge, MA, USA
e-mail: rmd@math.mit.edu

V. Vovk et al. (eds.), *Measures of Complexity*,
DOI 10.1007/978-3-319-21852-6_2

classes $\mathcal{F}$ of functions, to a Gaussian process G_P, e.g., van der Vaart and Wellner [5] and Dudley [2].

3. Although the field of Pattern Recognition had existed earlier, the introduction of Vapnik–Chervonenkis methods, as in their 1974 book [6], gave a new impetus and direction, providing a major new branch of that field. The contributions of Vapnik and Chervonenkis to what is now called Machine Learning were not limited to those flowing from the definition of VC classes of sets, although I am not qualified to describe the contributions in detail.

My comments here will only be about probability limit theorems. The 1971 paper, for uniform convergence over VC classes of sets, states a weak law of large numbers (Corollary of Theorem 2), and later a strong law (Theorem 3, in light of Theorem 1). My first publication using VC classes, in 1978 [1], gave uniform central limit theorems over classes of sets, in particular, VC classes satisfying a measurability condition. Although the main results were correct, there were numerous errors in the details, some in the published (1979) Corrections, and others repaired later, for example in [2].

Having acknowledged these errors, I take the liberty of pointing one out in the 1971 Vapnik and Chervonenkis paper: their measurability condition on $\mathcal{C}$ is that $\sup_{A\in\mathcal{C}} |(P_n - P)(A)|$ is measurable (next to last paragraph of the Introduction). That this is insufficient is shown in the introduction to Chap. 5 of [2], in the first edition as well. What is actually needed in the proof is that if P_n' and P_n'' are two independent versions of P_n, then $\sup_{A\in\mathcal{C}} |(P_n' - P_n'')(A)|$ is measurable (see the next display after (11) in the 1971 paper).

References

1. Dudley, R.M.: Central limit theorems for empirical measures. Ann. Probab. **6**, 899–929 (1978). Corrections, ibid. **7**, 909–911 (1979)
2. Dudley, R.M.: Uniform Central Limit Theorems, 2nd edn. Cambridge University Press, Cambridge (2014)
3. Laskowski, M.C.: Vapnik-Chervonenkis classes of definable sets. J. Lond. Math. Soc. Ser. 2 **45**, 377–384 (1992)
4. Stengle, G., Yukich, J.E.: Some new Vapnik–Chervonenkis classes. Ann. Stat. **17**, 1441–1446 (1989)
5. van der Vaart, A.W., Wellner, J.A.: Weak Convergence and Empirical Processes. Springer, New York (1996)
6. Vapnik, V.N., Chervonenkis, A.Y.: Теория распознавания образов: Статистические проблемы обучения (Theory of Pattern Recognition: Statistical Problems of Learning: in Russian). Nauka, Moscow (1974). German translation: Theorie der Zeichenerkennung, transl. K.G. Stöckel and B. Schneider, ed. S. Unger and B. Fritzsch, Akademie Verlag, Berlin (1979)

Springer

http://www.springer.com/978-3-319-21851-9

Measures of Complexity
Festschrift for Alexey Chervonenkis
Vovk, V.; Papadopoulos, H.; Gammerman, A. (Eds.)
2015, XXXI, 399 p. 47 illus., 30 illus. in color., Hardcover
ISBN: 978-3-319-21851-9