COLLOQUIUM MATHEMATICUM

VOL. 74 1997 NO. 1

A REMARK ON VAPNIK–CHERVONIENKIS CLASSES

BY

AGATA SMOKTUNOWICZ (WARSZAWA)

We show that the family of all lines in the plane, which is a VC class of index 2, cannot be obtained in a finite number of steps starting with VC classes of index 1 and applying the operations of intersection and union. This confirms a common belief among specialists and solves a question asked by several authors.

Notations and definitions. For an abstract set I the family of all subsets of I is denoted by 2^I. The indicator function of a set A is denoted by χ_A; $\#A$ denotes the cardinality of A and A' stands for $I \setminus A$. If $\mathcal{F}, \mathcal{G} \subset 2^I$ then we put $\mathcal{F}' = \{A' : A \in \mathcal{F}\}$, $\mathcal{F} \wedge \mathcal{G} = \{A \cap B : A \in \mathcal{F},\ B \in \mathcal{G}\}$, $\mathcal{F} \vee \mathcal{G} = \{A \cup B : A \in \mathcal{F},\ B \in \mathcal{G}\}$. If $A \subset I$ and $\mathcal{F} \subset 2^I$ then $\mathcal{F} \wedge A = \mathcal{F} \wedge \{A\} = \{A \cap B : B \in \mathcal{F}\}$.

If $\mathcal{F} \wedge A = 2^A$ then we say that $\mathcal{F}$ *shatters* A. The *index* of $\mathcal{F}$ is defined by $\mathrm{vc}(\mathcal{F}) = \sup\{\#A : A \subset I,\ \mathcal{F} \text{ shatters } A\}$. We will say that $\mathcal{F}$ is a *VC class* if $\mathrm{vc}(\mathcal{F}) < \infty$.

A family $\mathcal{F} \subset 2^I$ is said to be a *chain* (on I) if $\mathcal{F}$ is linearly ordered by inclusion, i.e., for each $A, B \in \mathcal{F}$ either $A \subset B$ or $B \subset A$. If $\mathcal{F}_1, \ldots, \mathcal{F}_p$ are chains on I then $\mathcal{F}_1 \wedge \ldots \wedge \mathcal{F}_p$ is called a *p-chain*.

If J is a finite set and f, g are functions on J then $\langle f, g\rangle = \sum_{a \in J} f(a)g(a)$.

Introduction. The concept of VC class, with a different notation, was introduced by Vapnik and Chervonienkis for purposes of the theory of empirical distributions and it plays an important role there (cf. [1], also [5], Ch. XIV). Later on, it found applications in other branches of mathematics. It is a fundamental concept in the theory of learning and in the theory of additive processes. Knowledge of the structure of VC classes would have many strong consequences for these theories. However, this seems to be a rather hopeless task. Only in the case of VC classes of index 1 we know a complete description of their structure (cf. [1], Sect. 4.4). If $\mathcal{F}$ is a chain then $\mathrm{vc}(\mathcal{F}) = 1$; conversely, if $\emptyset, I \in \mathcal{F}$ and $\mathrm{vc}(\mathcal{F}) = 1$ then $\mathcal{F}$ is a chain.

1991 *Mathematics Subject Classification*: Primary 60C05.

If $\mathcal{F}, \mathcal{G}$ are VC classes then $\mathcal{F}'$, $\mathcal{F} \wedge \mathcal{G}$, $\mathcal{F} \vee \mathcal{G}$ are also VC classes. Therefore a natural question is whether each VC class can be obtained by the operations $'$, $\wedge$, $\vee$ starting from families with lower index. If so, then we would be able to obtain each VC class applying the operations $\wedge$, $\vee$ and starting with chains, because as will be shown in the next section, for each VC class $\mathcal{F}$ of index 1 there are chains $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \mathcal{G}_4$ such that $\mathcal{F} \subset (\mathcal{G}_1 \wedge \mathcal{G}_2) \vee \mathcal{G}_3 \vee \mathcal{G}_4$.

The following question was asked by S. Kwapień [4]:

Is it true that for each VC class $\mathcal{F}$ there exist $k \in \mathbb{N}$ and k-chains $\mathcal{G}_1, \ldots, \mathcal{G}_k$ such that $\mathcal{F} \subset \mathcal{G}_1 \vee \ldots \vee \mathcal{G}_k$?

A more general question was asked by J. Hoffmann-Jørgensen, K.-L. Su, and R. L. Taylor [3] (cf. Remark (2) after Theorem 2.5):

Is it true that for each VC class $\mathcal{F}$ of subsets of I there are numbers $k \in \mathbb{N}$, $r > 0$ and a k-chain $\mathcal{G}$ such that for each finite subset $J \subset I$ and each $A \in \mathcal{F}$ we can find real numbers λ_i and $B_i \in \mathcal{G}$, $i = 1, \ldots, m$, with $\sum_{i=1}^m |\lambda_i| \le r$ such that

$$\chi_{A \cap J} = \sum_{i=1}^{m} \lambda_i \chi_{B_i \cap J} \ ? \tag{1}$$

To see that it is really more general assume that $\mathcal{F} \subset \mathcal{G}_1 \vee \ldots \vee \mathcal{G}_p$ where $\mathcal{G}_i$, $i = 1, \ldots, p$, are p-chains. If $A = B_1 \cup \ldots \cup B_p$ where $A \in \mathcal{F}$, $B_i \in \mathcal{G}_i$, $i = 1, \ldots, p$, then

$$\chi_A = \sum_{i=1}^{p} (-1)^{i+1} \sum_{l_1 > \ldots > l_i} \chi_{B_{l_1} \cap \ldots \cap B_{l_i}}$$

and $B_{l_1} \cap \ldots \cap B_{l_i} \in \mathcal{G}_1 \wedge \ldots \wedge \mathcal{G}_p$ if $I \in \mathcal{G}_i$, $i = 1, \ldots, p$, which can be assumed without loss of generality; thus χ_A has the representation (1) with $r = 2^p$, $k = p^2$ and $\mathcal{G} = \mathcal{G}_1 \wedge \ldots \wedge \mathcal{G}_p$, which is a p^2-chain.

The aim of the paper is to show that the above questions have negative answers.

Deriving VC classes from chains. Let $\mathcal{L}$ denote the family of all lines in the plane $\mathbb{R}^2$. It is a VC class of index 2. We will show that it is a counterexample to the questions from the introduction. Namely, we will prove the following

THEOREM. *Let $\mathcal{G}$ be a k-chain of subsets of $\mathbb{R}^2$ and $r > 0$. If p and n are integers such that $p > rk$, $n > (4p^4k)^{2k}$ and $J = \{(i,j) \in \mathbb{R}^2 : i = 1, \ldots, n,\ j = 1, \ldots, p\}$ then there exists a line $L \in \mathcal{L}$ such that the indicator function $\chi_{L \cap J}$ cannot be written in the form $\chi_{L \cap J} = \sum_{i=1}^m \lambda_i \chi_{B_i \cap J}$ where $\sum_{i=1}^m |\lambda_i| \le r$ and $B_i \in \mathcal{G}$.*

Proof. The well known Erdős–Szekeres Theorem [2] states that each sequence of numbers of length N contains a monotone subsequence of length at least $\sqrt{N}$. Since each maximal chain on a set T may be identified with a linear ordering of T this theorem can be interpreted as follows: given two maximal chains $\mathcal{F}_0$, $\mathcal{F}_1$ on T there exists a subset $T_0 \subset T$ with $\#T_0 \geq \sqrt{\#T}$ such that either $\mathcal{F}_1 \wedge T_0 = \mathcal{F}_0 \wedge T_0$ or $\mathcal{F}_1 \wedge T_0 = \mathcal{F}_0' \wedge T_0$. Hence by an easy induction we can prove that if $\mathcal{F}_0, \mathcal{F}_1, \ldots, \mathcal{F}_k$ are chains on T and $\mathcal{F}_0$ is maximal then there exists $T_0 \subset T$ with $\#T_0 \geq (\#T)^{1/2^k}$ such that for each $i = 1, \ldots, k$ we have either $\mathcal{F}_i \wedge T_0 \subset \mathcal{F}_0 \wedge T_0$ or $\mathcal{F}_i \wedge T_0 \subset \mathcal{F}_0' \wedge T_0$. Let $\mathcal{G} = \mathcal{F}_1 \wedge \ldots \wedge \mathcal{F}_k$ where $\mathcal{F}_1, \ldots, \mathcal{F}_k$ are chains on $\mathbb{R}^2$ defining $\mathcal{G}$. Applying the above we deduce that for each $q \geq 2$, $q \in \mathbb{N}$ we can split J into disjont sets: $J = J_0 \cup J_1 \cup \ldots \cup J_l$ such that $\#J_0 < q^{2^k}$, $\#J_j = q$ for $j = 1, \ldots, l$ and either $\mathcal{F}_i \wedge J_j \subset \mathcal{F}_0 \wedge J_j$ or $\mathcal{F}_i \wedge J_j \subset \mathcal{F}_0' \wedge J_j$ where $\mathcal{F}_0$ is a fixed maximal chain on J.

Let $\mathcal{K} = \{L \cap J : L \in \mathcal{L},\ \#L \cap J = p\}$. If $1 \leq i, j \leq n$ are integers such that $\frac{1-i}{p-1} \leq j - i \leq \frac{n-i}{p-1}$ then the line L which contains the points $(i, 1)$ and $(j, 2)$ satisfies $L \cap J \in \mathcal{K}$. Therefore

$$\#\mathcal{K} \geq n\left[\frac{n-1}{p-1}\right] \geq \frac{n^2}{p}. \tag{2}$$

We will show that not for all $K \in \mathcal{K}$,

$$\chi_K = \sum_{i=1}^{m} \lambda_i \chi_{B_i \cap J}, \tag{3}$$

where $\sum_{i=1}^m |\lambda_i| \leq r$, $B_i \in \mathcal{G}$, $i = 1, \ldots, m$.

Assume the contrary; we will show that it leads to a contradiction with (2).

Let $\mathcal{K}_0 = \{K \in \mathcal{K} : K \cap J_0 \neq \emptyset\}$. Then $\#\mathcal{K}_0$ is less than the number of pairs of elements in J such that the first one is in J_0.

Hence $\#\mathcal{K}_0 \leq \#J_0 \#J \leq q^{2^k} np$. For $K \in \mathcal{K}$ define a function g_K on J by $g_K = \sum_{a \in K}(\chi_{\{a\}} - \chi_{\{a^*\}})$ where a and a^* are in the same J_i, a^* is the immediate successor of a in the linear order on J_i defined by $\mathcal{F}_0$ if a is not the last element of J_i, and a^* is the immediate predecessor of a otherwise.

Let $\mathcal{K}_- = \{K \in \mathcal{K} \setminus \mathcal{K}_0 : \langle \chi_K, g_K \rangle = \sum_{a \in J} \chi_K(a) g_K(a) < p\}$.

If $K \in \mathcal{K}_-$ then $a^* \in K$ for some $a \in K$ and therefore $\#\mathcal{K}_-$ is less than or equal to the number of pairs $\{a, a^*\}$ such that $a \in J_i$ for some $i = 1, \ldots, l$. Thus $\#\mathcal{K}_- \leq ql \leq np$. Finally, let $K \in \mathcal{K} \setminus (\mathcal{K}_0 \cup \mathcal{K}_-)$. Then assuming a representation as in (3) we obtain $p \leq \langle \chi_K, g_K \rangle = \sum_{i=1}^m \lambda_i \langle \chi_{B_i \cap J}, g_K \rangle$. Since $\sum_{i=1}^m |\lambda_i| \leq r$ we have $|\langle \chi_{B_i \cap J}, g_K \rangle| \geq p/r > k$ for some $1 \leq i \leq m$. Let $B_i = A_1 \cap \ldots \cap A_k$ where $A_j \in \mathcal{F}_j$, $j = 1, \ldots, k$. If $\langle \chi_{B_i \cap J}, g_K \rangle = \sum_{a \in K}(\chi_{B_i \cap J}(a) - \chi_{B_i \cap J}(a^*)) > k$ then there exists $1 \leq j \leq k$ such that for at least two elements $a \in K$ we have $a \in A_j$ and $a^* \notin A_j$. Similarly, if $\langle \chi_{B_i \cap J}, g_K \rangle < -k$ then there exists $1 \leq j \leq k$ such that for at least two elements $a \in K$ we have $a \notin A_j$ and $a^* \in A_j$. Therefore $\#\mathcal{K} \setminus (\mathcal{K}_0 \cup \mathcal{K}_-)$ is at most the number of pairs $\{a,b\} \subseteq J \setminus J_0$ such that there exist $1 \le j \le k$ and $A \in \mathcal{F}_j$ such that either $a, b \in A$ and $a^*, b^* \notin A$, or $a, b \notin A$ and $a^*, b^* \in A$.

If $1 \le t, s \le l$ and $1 \le j \le k$ are fixed integers then the number of pairs $\{a,b\}$ such that $a \in J_t$, $b \in J_s$ and either $a, b \in A$, $a^*, b^* \notin A$ or $a, b \notin A$, $a^*, b^* \in A$ for some $A \in \mathcal{F}_j$ does not exceed $2q$. This is so because if A is fixed then there is at most one such pair and since $\mathcal{F}_j$ is a chain, $\#\mathcal{F}_j \wedge (J_s \cup J_t) \le \#J_s + \#J_t \le 2q$. Thus $\#\mathcal{K} \setminus (\mathcal{K}_0 \cup \mathcal{K}_{\_}) \le l^2 k 2q$. Since $l \le np/q$ we finally obtain

$$\#\mathcal{K} \le q^{2^k} np + np + 2k(np)^2/q$$

and this contradicts (2) if we choose q such that

$$16kp^3 < q < \left(\frac{1}{4}\frac{n}{p^2}\right)^{1/(2k)}. \ \blacksquare$$

To prove the statement opening the paper we have to prove the claim about VC classes of index 1 from the introduction which is contained in the following

Proposition. *Let $\mathcal{F}$ be a VC class of index* 1 *of subsets of I. If $\emptyset \in \mathcal{F}$ then there exists a* 2*-chain $\mathcal{G}$ on I such that $\mathcal{F} \subset \mathcal{G}$. In general, there are* 2*-chains $\mathcal{G}_1, \mathcal{G}_2$ on I such that $\mathcal{F} \subset \mathcal{G}_1 \vee \mathcal{G}_2'$.*

Proof. If $\mathcal{F}$ is a VC class of index 1 and $A \in \mathcal{F}$ then $\mathcal{F} \wedge A'$ and $\mathcal{F}' \wedge A$ are VC classes of index 1 which contain $\emptyset$; moreover, $\mathcal{F} \subset (\mathcal{F} \wedge A') \vee ((\mathcal{F}' \wedge A)' \wedge A)$. Thus the second statement is an easy consequence of the first one.

In the case when I is a finite set the first statement can be proved by induction on the number of elements in $\mathcal{F}$ as follows. If $\#\mathcal{F} = 2$ there is nothing to prove. If $\#\mathcal{F} > 2$ choose $A, B \in \mathcal{F}$ such that $B \setminus A \ne \emptyset$ and $A \cap B$ is maximal in $\{F \cap G : F, G \in \mathcal{F},\ F \ne G\}$, i.e. $A \cap B$ is strictly contained in no other member of that family. Using the fact that $\mathrm{vc}(\mathcal{F}) = 1$ and that $\emptyset \in \mathcal{F}$ we prove easily that $D = (A \setminus B) \cup (B \setminus A)$ is disjoint from each $C \in \mathcal{F}$, $C \ne A, B$. By the induction assumption there exist chains $\mathcal{P}_1, \mathcal{P}_2$ such that $\mathcal{F} \setminus \{A\} \subset \mathcal{P}_1 \wedge \mathcal{P}_2$. Let $B = P_1 \cap P_2$ where $P_i \in \mathcal{P}_i$, $i = 1, 2$. Define two new chains on I:

$$\begin{aligned}
\mathcal{Q}_1 &= \{P \setminus D : P \in \mathcal{P}_1, P \subset P_1\} \cup \{P_1 \setminus (A \setminus B)\} \\
&\quad \cup \{P \cup D : P \in \mathcal{P}_1, P_1 \subset P\}, \\
\mathcal{Q}_2 &= \{P \setminus D : P \in \mathcal{P}_2, P \subset P_2\} \cup \{(P_2 \cup (A \setminus B)) \setminus (B \setminus A)\} \\
&\quad \cup \{P \cup D : P \in \mathcal{P}_2,\ P_2 \subset P\}.
\end{aligned}$$

For $R_i \in \mathcal{P}_i$, $i = 1, 2$, put

$$R_i^{\sim} = \begin{cases} R_i \setminus D & \text{if } R_i \subset P_i, \\ R_i^{\sim} = R_i \cup D & \text{if } P_i \subset R_i \text{ and } P_i \ne R_i. \end{cases}$$

Since $A=(P_1\cup D)\cap((P_2\cup(A\setminus B))\setminus(B\setminus A))\in\mathcal{Q}_1\wedge\mathcal{Q}_2$, $B=(P_1\setminus(A\setminus B))\cap(P_2\cup D)\in\mathcal{Q}_1\wedge\mathcal{Q}_2$ and since for each $C\in\mathcal{F}$, $C\neq A,B$ with $C=R_1\cap R_2$, where $R_i\in\mathcal{P}_i, i=1,2$, we have $C=R_1^\sim\cap R_2^\sim\in\mathcal{Q}_1\wedge\mathcal{Q}_2$; this is because by the maximality of $A\cap B$ at least one of the inclusions $R_i\subset P_i, i=1,2$, holds and $C\cap D=\emptyset$. Thus $\mathcal{F}\subset\mathcal{Q}_1\wedge\mathcal{Q}_2$ and the induction is completed.

To prove the case of infinite I it is enough to prove that if $\mathcal{F}$ is a family on I such that for each finite $J\subset I$ there are chains $\mathcal{P}_J^1,\mathcal{P}_J^2$ on J with $\mathcal{F}\wedge J\subset\mathcal{P}_J^1\wedge\mathcal{P}_J^2$ then $\mathcal{F}$ is contained in a 2-chain on I. The proof follows easily by the method of ultrafilters. Let $\mathcal{H}=\{J\subset I:\#J<\infty\}$ and let $\mathbf{h}$ be an ultrafilter on $\mathcal{H}$, i.e., $\mathbf{h}$ is any family of subsets of $\mathcal{H}$ which satisfies:

0. $\emptyset\notin\mathbf{h}$,
1. if $J\in\mathcal{H}$ then $\{K\in\mathcal{H}:J\subset K\}\in\mathbf{h}$,
2. if $\mathcal{G}_1,\mathcal{G}_2\in\mathbf{h}$ then $\mathcal{G}_1\cap\mathcal{G}_2\in\mathbf{h}$,
3. if $\mathcal{G}\subset\mathcal{H}$ then either $\mathcal{G}\in\mathbf{h}$ or $\mathcal{H}\setminus\mathcal{G}\in\mathbf{h}$.

Given any family $(A_J)_{J\in\mathcal{H}}$ of sets we define $\operatorname{Lim}_{\mathbf{h}} A_J=\{(a_J)_{J\in\mathcal{H}}: a_J\in A_J$ for each $J\in\mathcal{H}\}$ where we identify two elements $(a_J)_{J\in\mathcal{H}}$, $(b_J)_{J\in\mathcal{H}}$ whenever $\{J\in\mathcal{H}:a_J=b_J\}\in\mathbf{h}$.

If $\mathcal{A}_J$ is a class of subsets of A_J for each $J\in\mathcal{H}$ then we can identify $\operatorname{Lim}_{\mathbf{h}}\mathcal{A}_J$ with a class of subsets of $\operatorname{Lim}_{\mathbf{h}} A_J$; the identification is given by the relation

$$(a_J)_{J\in\mathcal{H}}\in(A_J)_{J\in\mathcal{H}}\equiv\{J\in\mathcal{H}:a_J\in A_J\}\in\mathbf{h}.$$

Let $\bar I=\operatorname{Lim}_{\mathbf{h}} J$, $\overline{\mathcal{F}}=\operatorname{Lim}_{\mathbf{h}}\mathcal{F}\wedge J$ and for $i=1,2$ let $\overline{\mathcal{P}}_i=\operatorname{Lim}_{\mathbf{h}}\mathcal{P}_J^i$. It is very easy to check that $\overline{\mathcal{P}}_i$, $i=1,2$, are chains on $\bar I$ and $\overline{\mathcal{F}}\subset\overline{\mathcal{P}}_1\wedge\overline{\mathcal{P}}_2$. Moreover, I can be identified with a subset of $\bar I$ by the relation

$$i=(i_J)_{J\in\mathcal{H}}\equiv\{J\in\mathcal{H}:i=i_J\}\in\mathbf{h}.$$

We check easily that with this identification, $\mathcal{F}\subset\overline{\mathcal{F}}\wedge I$. ∎

Corollary. *The class $\mathcal{L}$ of index 2 cannot be obtained by applying the operations $\wedge,\vee$ a finite number times to VC classes of index 1.*

Remark. We do not know if there is a VC class of index 3 which cannot be obtained from VC classes of index 2 by applying the operations $\wedge,\vee$ a finite number times. It seems that the family of all planes in $\mathbb{R}^3$ is a good candidate.

Acknowledgments. The author wishes to express her gratitude to Professors Richard M. Dudley and Stanisław Kwapień for their helpful comments and suggestions which have improved the original manuscript and to Prof. Richard M. Dudley for making available to her two sections of his forthcoming monograph.

REFERENCES

[1] R. M. Dudley, *Uniform Central Limit Theorems*, Cambridge University Press, to appear.
[2] P. Erdős and G. Szekeres, *A combinatorial problem in geometry*, Compositio Math. 2 (1939), 463–470.
[3] J. Hoffmann-Jørgensen, K.-L. Su and R. L. Taylor, *The law of large numbers and the Ito–Nisio theorem for vector valued random fields*, J. Theoret. Probab. 10 (1997), 145–183.
[4] S. Kwapień, *On maximal inequalities for sums of independent random variables*, in: XIII Jubileuszowy Zjazd Matematyków Polskich, Referaty, Wydawnictwa PTM, 1994 (in Polish).
[5] M. Ledoux and M. Talagrand, *Probability in Banach Spaces*, Springer, 1991.

Institute of Mathematics
University of Warsaw
Banacha 2
02-097 Warszawa, Poland

Received 5 November 1996;
revised 2 December 1996