Dependence and the Vapnik-Chervonenkis Property

Shelah's notion of dependence is equivalent to a combinatorial notion, the *Vapnik-Chervonenkis property*, introduced by Vapnik and Chervonenkis at the same time as Shelah discovered dependence. This equivalence can be stated as follows. Let M be a model of the complete L-theory T, let the relation $\Phi \subseteq M^{m+n}$ be definable, and for $x \in M^m$, put

$$\Phi(x) \;:=\; \{y \in M^n : \; (x,y) \in \Phi\}.$$

We say that Φ is dependent (in M) if $\phi(x;y)$ is dependent in the $L(M)$-theory of M, where $\phi(x;y)$ is an $L(M)$-formula defining Φ.

Theorem 0.1. *The following are equivalent:*

(1) Φ *is dependent;*
(2) *there is a* $d \in \mathbb{N}$ *such that every* d*-element set* $D \subseteq M^n$ *has a subset that is not of the form* $\Phi(x) \cap D$ *with* $x \in M^m$*;*
(3) *there is a* $d \in \mathbb{N}$ *such that for all sufficiently large* k *each* k*-element set* $F \subseteq M^n$ *has at most* k^d *subsets of the form* $\Phi(x) \cap F$ *with* $x \in M^m$*.*

The remarkable thing about (3) is that the total number of subsets of a k-element set is 2^k, which grows exponentially with k, and thus much faster than the polynomial function k^d. So (3) says that for large k only very few subsets of any k-element set F are of the form $\Phi(x) \cap F$. We are going to indicate why this equivalence holds. Next, we describe how this relates to probability theory (more precisely, statistical learning theory), which is the back ground for the work of Vapnik and Chervonenkis.

A combinatorial dichotomy

Let X be an infinite set, and let $X^{(k)}$ be the collection of k-element subsets of X. Let $\mathcal{C}$ be a collection of subsets of X. Given finite $F \subseteq X$, put

$$\mathcal{C} \cap F \;:=\; \{C \cap F : \; C \in \mathcal{C}\},$$

the set of intersections of sets in $\mathcal{C}$ with F. So $|\mathcal{C} \cap F| \le 2^{|F|}$. We define the function $f_{\mathcal{C}} : \mathbb{N} \to \mathbb{N}$ by

$$f_{\mathcal{C}}(k) := \max\{|\mathcal{C} \cap F| : \; F \in X^{(k)}\},$$

so $0 \le f_{\mathcal{C}}(k) \le 2^k$ for all k. We have the following surprising dichotomy.

Theorem 0.2. *Either* $f_{\mathcal{C}}(k) = 2^k$ *for all* k*, or there is* $d \in \mathbb{N}$ *such that* $f_{\mathcal{C}}(k) \le k^d$ *for all sufficiently large* k*.*

Theorem 0.2 is an easy consequence of the lemma below. Put

$$p_d(k) := \sum_{i<d} \binom{k}{i}$$

where by convention $\binom{k}{i} := 0$ if $k < i$. So $p_d(k)$ is the number of subsets of size $< d$ of a k-element set, and is a polynomial function of k of degree

$d-1$. For example, if $\mathcal{C}$ is the collection of subsets of X of size $< d$, then $f_{\mathcal{C}}(k) = p_d(k)$ for all k, in particular, $f_{\mathcal{C}}(k) = 2^k$ for $k \le d$.

Lemma 0.3. *Let $|F| = k$ and let $\mathcal{F}$ be a collection of subsets of F such that*

$$|\mathcal{F}| > p_d(k), \qquad d \le k.$$

Then F has a subset D such that $|D| = d$ and $\mathcal{F} \cap D = \{$all subsets of $D\}$.

The bound $p_d(k)$ here is sharp: if $|F| = k$ and $d \le k$, then the collection $\mathcal{F}$ of subsets of F of size $< d$ has $p_d(k)$ elements, and violates the conclusion of the lemma. The proof of the lemma goes by induction on k and is an attractive exercise. Hint: the result holds trivially for $d = 0$ and $d = k$. Let $0 < d < k$, pick a point $x \in F$, put $E' := E \setminus \{x\}$ for $E \in \mathcal{F}$, and consider the map $E \mapsto E' \ : \ \mathcal{F} \to \mathcal{F}' := \{E' : \ E \in \mathcal{F}\}$. We also leave it as an exercise to derive Theorem 0.2 from Lemma 0.3. Note also that Theorem 0.2 yields the equivalence of (2) and (3) in Theorem 0.1.

Vapnik-Chervonenkis Classes and Dependence

With X and $\mathcal{C}$ as above we say that $\mathcal{C}$ is a *Vapnik-Chervonenkis class* (or VC-class) if $f_{\mathcal{C}}(d) < 2^d$ for some d, and then we define its VC-index $V(\mathcal{C})$ to be the least such d. The derivation of Theorem 0.2 from Lemma 0.3 shows that if $\mathcal{C}$ is a VC-class, then $f_{\mathcal{C}}$ has polynomial growth of degree $< V(\mathcal{C})$.

Besides X we fix a second infinite set Y and consider a collection $\mathcal{G}$ of subsets of Y. We say that $\mathcal{G}$ is *independent* if for each k there are sets $S_1, \dots, S_k \in \mathcal{G}$ such that

$$S_1^{e(1)} \cap \cdots \cap S_k^{(k)} \neq \emptyset$$

for all 2^k assignments $e : \{1, \dots, k\} \to \{-1, 1\}$. (Here, for a set $S \subseteq Y$ we put $S^1 := S$ and $S^{-1} = Y \setminus S$.) If $\mathcal{G}$ is not independent, we call $\mathcal{G}$ *dependent* and let $D(\mathcal{G})$, the *dependency index of* $\mathcal{G}$, be the least k such that for all $S_1, \dots, S_k \in \mathcal{G}$ some intersection as displayed above is empty.

To discuss the connection between the VC-property and dependence we consider a relation $\Phi \subseteq X \times Y$, with reverse

$$\breve{\Phi} = \{(y, x) : \ (x, y) \in \Phi\} \subseteq Y \times X.$$

For $x \in X$ and $y \in Y$, put

$$\Phi(x) := \{y \in Y : \ (x, y) \in \Phi\}, \qquad \breve{\Phi}(y) := \{x \in X : \ (x, y) \in \Phi\}.$$

Then we have a collection $\Phi(X) := \{\Phi(x) : \ x \in X\}$ of subsets of Y and a collection $\breve{\Phi}(Y) := \{\breve{\Phi}(y) : \ y \in Y\}$ of subsets of X.
We leave it as an exercise to check the following basic equivalence:

$$\Phi(X) \text{ is dependent } \iff \breve{\Phi}(Y) \text{ is a VC-class},$$

and to show that if $\Phi(X)$ is dependent, then $D\big(\Phi(X)\big) = V\big(\breve{\Phi}(Y)\big)$. We also have symmetry of dependence, and thus of the VC-property:

$$\Phi(X) \text{ is dependent } \iff \breve{\Phi}(Y) \text{ is dependent},$$

and if $\Phi(X)$ is dependent, then $D\big(\breve{\Phi}(Y)\big) \le 2^{D\left(\Phi(X)\right)}$. So all four conditions

$$\text{“}\Phi(X) \text{ is a VC-class”}, \quad \text{“}\breve{\Phi}(Y) \text{ is a VC-class”},$$
$$\text{“}\Phi(X) \text{ is dependent”}, \quad \text{“}\breve{\Phi}(Y) \text{ is dependent”}$$

are equivalent. When these conditions are satisfied we say that the relation Φ is *dependent*. Please check that for a definable relation $\Phi \subseteq M^{m+n}$ and taking $X = M^m$ and $Y = M^n$, this agrees with dependence of Φ in the model M. This yields the equivalence of (1) with (2) and (3) in Theorem 0.1.