What is the minimal cardinal of a family which shatters all $d$-subsets of a finite set?

N. Chevallier and A. Fruchard

September 8, 2015

In this note, $d \leq n$ are positive integers. Let $S$ be a finite set of cardinal $|S| = n$ and let $2^S$ denote its power set, i.e. the set of its subsets. A $d$-subset of $S$ is a subset of $S$ of cardinal $d$. Let $\mathcal{F} \subseteq 2^S$ and $A \subseteq S$. The trace of $\mathcal{F}$ on $A$ is the family $\mathcal{F}_A = \{ E \cap A ; E \in \mathcal{F} \}$. One says that $\mathcal{F}$ shatters $A$ if $\mathcal{F}_A = 2^A$. The VC-dimension of $\mathcal{F}$ is the maximal cardinal of a subset of $S$ that is shattered by $\mathcal{F}$ [7]. The following is well-known [7, 4, 5]:

**Theorem 1.** (Vapnik-Chervonenkis, Sauer, Shelah)
If $\text{VC-dim}(\mathcal{F}) \leq d$ (i.e. if $\mathcal{F}$ shatters no $(d + 1)$-subset of $S$) then $|\mathcal{F}| \leq c(d, n)$, where

$$c(d, n) = \binom{n}{0} + \cdots + \binom{n}{d}.$$  

Moreover this bound is tight: It is achieved e.g. for $\mathcal{F} = \binom{S}{\leq d}$, the family of all $k$-subsets of $S$, $0 \leq k \leq d$.

A first natural question is:

**Question 1.** Assume a family $\mathcal{F} \subseteq 2^S$ is maximal for the inclusion among all families of VC-dimension at most $d$. Does $\mathcal{F}$ always have the maximal possible cardinal $c(d, n)$?

Let us define the index of $\mathcal{F}$ as follows:

$$\text{Ind} \mathcal{F} = \max\{ d \in \{0, ..., n\} ; \mathcal{F} \text{ shatters all } d\text{-subsets of } S \}. $$

Let $C(d, n) = \min\{ |\mathcal{F}| ; \text{Ind} \mathcal{F} = d \}$. For instance, we have $C(1, n) = 2$, with the (only possible) choice $\mathcal{F} = \{ \emptyset, S \}$. Of course we have $2^d \leq C(d, n) \leq 2^n$. The question is:

**Question 2.** Give the exact value of $C(d, n)$ for $2 \leq d \leq n$. If this is not possible, give lower and upper bounds as accurate as possible.

A well-known duality yields another formulation of Question 2. Let $\varphi : S \to 2^F$, $a \mapsto \{ E \in \mathcal{F} ; a \in E \}$ and set $S = \varphi(S)$. In this manner, we have for all $a \in S$ and all $E \in \mathcal{F}$:

$$a \in E \Leftrightarrow E \in \varphi(a). \quad (1)$$

One can check that $\mathcal{F}$ shatters $A \subseteq S$ if and only if, for every partition $(B, C)$ of $A$ (i.e. $A = B \cup C$ and $B \cap C = \emptyset$) the intersection $\left( \bigcap_{b \in B} \varphi(b) \right) \cap \left( \bigcap_{c \in C} \overline{\varphi(c)} \right)$ is nonempty, where the notation $\overline{Y}$ stands for $\mathcal{F} \setminus Y$.

If $\text{Ind} \mathcal{F} \geq 2$, then $\varphi$ is a one-to-one correspondence from $S$ to $S$, hence we have $\log n \leq C(d, n)$ for all $2 \leq d \leq n$, where $\log$ denotes the logarithm in base 2.
The case $d = 2$. Using for instance the binary expansion, it is easy to show that the order of magnitude of $C(2, n)$ is actually $\log n$. The next statement refines this.

Proposition 2. If $n = \frac{1}{2} \binom{2l}{l} = \binom{2l-1}{l-1}$, then $C(2, n) = 2l$.

Proof. (Recall the notation $\overline{A} = F \setminus A$.) We first prove by contradiction that $C(2, n) > 2l - 1$. Actually, if a family $F$ of subsets of $S$ shatters all 2-subsets of $S$, then the image $S \subseteq 2^F$ of $S$ by $\varphi$ must satisfy

\[ \forall A \neq B \in S, A \cap B, A \cap \overline{B}, \overline{A} \cap B, \text{ and } \overline{A} \cap \overline{B} \text{ are nonempty.} \]  

(2)

In particular $S$ is a Sperner family of $F$ (i.e. an antichain for the partial order of inclusion; one finds several other expressions in the literature: ‘Sperner system’, ‘independent system’, ‘clutter’, ‘completely separating system’, etc.). For a survey on Sperner families and several generalizations, we refer e.g. to [1] and the references therein.

Assume now that $|F| = 2l - 1$; it is known [6] [2] [3] that all Sperner families of $F$ have a cardinal at most $\binom{2l-1}{l-1}$, and that there are only two Sperner families of maximal cardinal: the families $\binom{F}{l}$ and $\binom{F}{l'}$, i.e. of $(l-1)$-subsets, resp. $l$-subsets of $F$. However, none of these families satisfies both $A \cap B$ and $\overline{A} \cap \overline{B}$ nonempty in (2). As a consequence, we must have $|F| \geq 2l$.

Conversely, let $S = \{a_1, \ldots, a_n\}$, consider $\binom{\{1, \ldots, 2l\}}{l}$, the set of $l$-subsets of $\{1, \ldots, 2l\}$, and choose one element in each pair of complementary $l$-subsets. We then obtain a family \{A_1, \ldots, A_n\} which satisfies (2). Now we set $F = \{E_1, \ldots, E_{2l}\}$, with $E_i = \{a_j \mid i \in A_j\}$. The characterization [1] shows that $F$ shatters every 2-subset of $S$.

The proof of the following statement is straightforward.

Corollary 3. If $\binom{2l-1}{l-1} < n \leq \binom{2l-1}{l}$, then $2l \leq C(2, n) \leq 2l + 2$.

The upper bound can be slightly improved: One can prove that, if $\binom{2l-1}{l-1} < n \leq \binom{2l}{l-1}$, then $2l \leq C(2, n) \leq 2l + 1$.

Question 3. It seems that we have $C(2, n) = k$ if and only if $\binom{k-2}{(\lfloor k/2 \rfloor - 1)} \leq n \leq \binom{k-1}{(\lfloor k/2 \rfloor - 1)}$, where $\lfloor x \rfloor$ denotes the integer part of $x$. Is it true? Is it already known?

The first values are $C(2, 2) = C(2, 3) = 4$, $C(2, 4) = 5$, $C(2, 5) = \cdots = C(2, 10) = 6$. Computer seems to be useless, at least for a naive treatment. Already in order to obtain $C(2, 11) = 7$, we would have to verify that $C(2, 11) > 6$, i.e. to find, for each of the $\binom{21}{6} \approx 10^{17}$ families $F$ in $2^S$ some 2-subset that is not shattered by the family. (Alternatively, in the dual statement, we have to check “only” $\binom{21}{6} \approx 7.10^{11}$ families $S$ in $2^F$.)

The case $d \geq 3$. From now, we assume $n \geq 4$.

Proposition 4. For all $3 \leq d < n$, we have $C(d, n) \leq \frac{2d}{d!} (3 \log n)^d$.

The constant 3 can be improved. The proof below shows that, for all $a > 1$ and all $n$ large enough, $C(d, n) \leq \frac{2d}{d!} (a \log n)^d$.

Proof. Let $F_0 \subseteq 2^S$ be a minimal separating system of $S$, i.e. such that, for all $a, b \in S$ there exists $E_a^0 \in F_0$ which satisfies $b \notin E_a^0 \supseteq a$. Since this amounts to choosing $F_0$ minimal such that $S = \varphi(S)$ is a Sperner family for $F_0$, we know that $|F_0| = N$ if and only if $\binom{N-1}{\lfloor (N-1)/2 \rfloor} < n \leq \binom{N}{\lfloor N/2 \rfloor}$, hence $N := |F_0| \leq 2 + \log n + \frac{1}{2} \log \log n \leq 3 \log n$ since $n \geq 4$. We assume
Given two disjoint subsets $B$ and $C$ of $S$ such that $|B \cup C| = d$, the set $E_B^C = \bigcap_{c \in C} \left( \bigcup_{b \in B} E_b^c \right)$ contains $B$ and does not meet $C$. Let $\mathcal{F}$ be the collection of all such sets $E_B^C$; then $\mathcal{F}$ shatters all subsets of $S$ of cardinal at most $d$.

To estimate $|\mathcal{F}|$, we consider $\mathcal{F}_k$ the collection of all such sets $E_B^C$, with $|B| = k$ (and thus $|C| = d - k$). We have $|\mathcal{F}_k| = \binom{N}{k} \binom{N - k}{d - k}$ (with $N = |\mathcal{F}_0|$). Then we choose $\mathcal{F} = \bigcup_{k=0}^d \mathcal{F}_k$. We obtain $|\mathcal{F}| \leq \sum_{k=0}^d \binom{N}{k} \binom{N - k}{d - k} = \binom{N}{d}^2 \leq \frac{2^d}{d!} N^d \leq \frac{2^d}{d!} (3 \log n)^d$. 

**Question 4.** Is $(\log n)^{\lfloor d/2 \rfloor (\lfloor (d+1)/2 \rfloor)}$ the right order of magnitude for $C(d, n)$?

By constructing auxiliary Sperner families from $S$, it is possible to give a better lower bound for $C(d, n)$ than only $C(d, n) \geq C(2, n)$. For instance, in the case $d = 3$, for all distinct $A, B, C \in S$, we must have $A \cap B \not\subseteq C$. One can check that this implies that the family $\{A \cap B : A, B \in S\}$ is a Sperner family, therefore we obtain $\binom{n}{2} \leq \binom{C(3, n)}{\lfloor C(3, n)/2 \rfloor}$. Unfortunately, this does not modify the order of magnitude. Already in this case $d = 3$, we do not know whether $C(3, n)$ is of order $\log n$, $(\log n)^2$, or an intermediate order of magnitude. Another formulation is:

**Question 5.** Prove or disprove: There exists $C > 0$ such that, for all $k \in \mathbb{N}$, if $\mathcal{F}$ is a finite set of cardinal $k$ and $\mathcal{S} \subseteq 2^\mathcal{F}$ satisfies $\forall A, B, C \in \mathcal{S}$, $A \cap B \not\subseteq C$, then $|\mathcal{S}| \leq C 2^{C \sqrt{k}}$.

**References**

[1] P. Borg, Intersecting families of sets and permutations: a survey. *Int. J. Math. Game Theory Algebra* 21 (2012) 543–559.

[2] G. Katona, On a conjecture of Erdős and a stronger form of Sperner’s theorem. *Studia Sci. Math. Hungar.* 1 (1966) 59–63.

[3] D. Lubell, A short proof of Sperner’s theorem, *J. Combin Theory* 1 (1966) 299.

[4] N. Sauer, On the density of families of sets, *J. Combin. Theory* 25 (1972) 80–83.

[5] S. Shelah, A combinatorial problem, stability and order for models and theories in infinite languages, *Pacific J. Math.* 41 (1972) 247–261.

[6] E. Sperner, Ein Satz über Untermengen einer endlichen Menge, *Math. Zeitschrift* 27 (1928) 544–548.

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