Three proofs of Sauer-Shelah Lemma

Let $\mathcal{H}$ be a hypothesis class, i.e. a class of functions from $\Omega \to \{0,1\}$. Each hypothesis can be thought of as a subset of Ω. For any finite $S \subseteq \Omega$, let $\Pi_{\mathcal{H}}(S) = \{h \cap S : h \in \mathcal{H}\}$. We call $\Pi_{\mathcal{H}}(S)$ the *projection* of $\mathcal{H}$ on S. Equivalently, suppose $S = \{x_1, \dots, x_m\}$, let

$$\Pi_{\mathcal{H}}(S) = \{[h(x_1), \dots, h(x_m)] \mid h \in \mathcal{H}\}$$

and call $\Pi_{\mathcal{H}}(S)$ the set of all *dichotomies* (or *behaviors*) on S *realized by* (or *induced by*) $\mathcal{H}$. A set S is *shattered* by $\mathcal{H}$ if $|\Pi_{\mathcal{H}}(S)| = 2^{|S|}$. Note that, if S is shattered then every subset of S is shattered.

Definition 0.1 (VC-dimension)**.** The *VC-dimension* of $\mathcal{H}$ is defined to be

$$\text{VCD}(\mathcal{H}) = \max\{|S| : S \text{ shattered by } \mathcal{H}\}.$$

The following lemma was first proved by Vapnik-Chervonenkis [5], and rediscovered many times (Sauer [3], Shelah [4]), among others. It is often called the Sauer lemma or Sauer-Shelah lemma in the literature. (Sauer said that Paul Erdös posed the problem.)

Lemma 0.2 (Sauer lemma)**.** *Suppose* $\text{VCD}(\mathcal{H}) = d < \infty$. *Define*

$$\Pi_{\mathcal{H}}(m) = \max\{|\Pi_{\mathcal{H}}(S)| : S \subseteq \Omega, |S| = m\}$$

(i.e., $\Pi_{\mathcal{H}}(m)$ *is the maximum size of a projection of* $\mathcal{H}$ *on an* m*-subset of* Ω*.) Then,*

$$\Pi_{\mathcal{H}}(m) \le \Phi_d(m) := \sum_{i=0}^{d} \binom{m}{d} \le \left(\frac{em}{d}\right)^d = O(m^d)$$

(Note that, if $\text{VCD}(\mathcal{H}) = \infty$, then $\Pi_{\mathcal{H}}(m) = 2^m, \forall m$)

Proof #1: The inductive proof (not nice!) We induct on $m + d$. For $h \in \mathcal{H}$, define $h_S = h \cap S$. The $m = 0$ and $d = 0$ cases are trivial. Now consider $m > 0, d > 0$. Fix an arbitrary element $s \in S$. Define

$$\mathcal{H}' = \{h_S \in \Pi_{\mathcal{H}}(S) \mid s \notin h_S,\ h_S \cup \{s\} \in \Pi_{\mathcal{H}}(S)\}$$

Then,

$$|\Pi_{\mathcal{H}}(S)| = |\Pi_{\mathcal{H}}(S - \{s\})| + |\mathcal{H}'| = |\Pi_{\mathcal{H}}(S - \{s\})| + |\Pi_{\mathcal{H}'}(S)|$$

Since $\text{VCD}(\mathcal{H}') \le d - 1$, by induction we obtain

$$|\Pi_{\mathcal{H}}(S)| \le \Phi_d(m-1) + \Phi_{d-1}(m) = \Phi_d(m).$$

□

The shifting technique is a very powerful proof technique in extremal set theory. See [1,2], for example. Recently the technique has found applications in the harmonic analysis of Boolean functions. It's good to get a glimpse of the technique.

Proof #2: a proof by shifting. Let $\mathcal{F} = \Pi_{\mathcal{H}}(S)$, then $\mathcal{F}$ is a family of subsets of $[m]$. Without loss of generality, we assume $m > d$, because if $m \leq d$ then $\Phi_d(m) = 2^m$ and the inequality is trivial..

We will use "shifting" to construct a family $\mathcal{G}$ of subsets of $[m]$ satisfying the following three conditions:

1. $|\mathcal{G}| = |\mathcal{F}|$
2. If $A \subset S$ is shattered by $\mathcal{G}$ then A is shattered by $\mathcal{F}$
3. If $A \in \mathcal{G}$, then every subset of A is in $\mathcal{G}$. (The technical term of this is that G is an *order ideal* of the Boolean algebra lattice. Another term is *"closed under containment."*)

So, instead of upperbounding $|\mathcal{F}|$ we can just upperbound $\mathcal{G}$. Every member of $\mathcal{G}$ is shattered by $\mathcal{G}$ and thus every member of $\mathcal{G}$ is shattered by $\mathcal{F}$. Thus, every member of $\mathcal{G}$ has size at most d, implying $|\mathcal{G}| \leq \Phi_d(m)$ as desired.

We next describe the *shifting* operation which achieves 1, 2, 3 by an algorithm.

```
1: for i = 1 to m do
2:   for F ∈ F do
3:     if F − {i} ∉ F then
4:       Replace F by F − {i}
5:     end if
6:   end for
7: end for
8: Repeat steps 1–7 until no further changes is possible.
```

The algorithm terminates because some set gets smaller at each step. Properties 1 and 3 are easy to verify.

We verify 2. Let A be shattered by $\mathcal{F}$ *after* executing lines 2–6 at any point in the execution. We will show that A must have been shattered by $\mathcal{F}$ *before* the execution. Let i be the element examined in that iteration. To avoid confusion, let $\mathcal{F}'$ be the set family after the iteration. We can assume $i \in A$, otherwise the iteration does not affect the "shatteredness" of A.

Let R be an arbitrary subset of A. We know there's $F' \in \mathcal{F}'$ such that $F' \cap A = R$. If $i \in R$, then $F' \in \mathcal{F}$. Suppose $i \notin R$. There is $T \in \mathcal{F}'$ such that $T \cap A = R \cup \{i\}$. This means $T - \{i\} \in F$, or else T would have been replaced in step 4. But, $T - \{i\} \cap A = R$ as desired. $\square$

I found the next proof from Tim Gowers' sample Wiki-trick entry[1]. The proof is by Peter Frankl and Janos Pach.

Proof #3: dimensionality argument. Let $\mathcal{F} = \Pi_{\mathcal{H}}(S)$, then $\mathcal{F}$ is a family of subsets of $[m]$. Without loss of generality, we assume $m > d$, Let $\binom{[m]}{\leq d}$ denote all subsets of $[m]$ of size at most d. There are $\Phi_d(m)$ such sets. For each $F \in \mathcal{F}$, associate a function $g_F : \binom{[m]}{\leq d} \to \mathbb{R}$ defined as follows. For each $X \in \binom{[m]}{\leq d}$, $g_F(X) = 1$ if $X \subseteq F$, and $g_F(X) = 0$ otherwise. The functions g_F can naturally be viewed as vectors in the space $\mathbb{R}^{\Phi_d(m)}$. We prove that these vectors are linearly independent, which implies $|\mathcal{F}| \leq \Phi_d(m)$.

[1]http://gowers.wordpress.com/2008/07/31/dimension-arguments-in-combinatorics/

Suppose to the contrary that there are coefficients α_F, not all zero, such that $\sum_{F\in\mathcal{F}} \alpha_F g_F = 0$, i.e. the g_F are not linearly independent. We derive the contradiction that there is a subset $Y \subseteq [m]$, $|Y| \geq d+1$ which is shattered by $\mathcal{F}$. For convenience, for any set Z we define

$$\sigma(Z) = \sum_{\substack{F\in\mathcal{F} \\ Z\subseteq F}} \alpha_F.$$

First, for any $X \in \binom{[m]}{\leq d}$, we have

$$0 = \sum_{F\in\mathcal{F}} \alpha_F g_F(X) = \sum_{\substack{F\in\mathcal{F} \\ X\subseteq F}} \alpha_F = \sigma(X).$$

Hence, $\sigma(X) = 0$ for every $|X| \leq d$. Let $Y \subseteq [m]$ be a minimum-sized subset of $[m]$ such that $\sigma(Y) \neq 0$. Then, certainly $|Y| \geq d+1$. (If F is a maximum-sized member of $\mathcal{F}$ for which $\alpha_F \neq 0$, then $\sigma(F) \neq 0$; thus, Y is well-defined.) We prove that Y is shattered by $\mathcal{F}$.

Consider any subset $Z \subseteq Y$. To show that there is some $F \in \mathcal{F}$ for which $F \cap Y = Z$, we prove that

$$\sum_{\substack{F\in\mathcal{F} \\ Z=F\cap Y}} \alpha_F \neq 0.$$

The following is a well-known identity in distributive lattice theory, which is basically just an inclusion-exclusion formula:

$$\sum_{\substack{F\in\mathcal{F} \\ Z=F\cap Y}} \alpha_F = \sum_{Z\subseteq W\subseteq Y} (-1)^{|W-Z|}\sigma(W).$$

Now, since $\sigma(W) = 0$ for all $Z \subseteq W \subset Y$, we conclude that

$$\sum_{\substack{F\in\mathcal{F} \\ Z=F\cap Y}} \alpha_F = (-1)^{|Y-Z|}\sigma(Y) \neq 0.$$

□

References

[1] P. FRANKL, *The shifting technique in extremal set theory*, in Surveys in combinatorics 1987 (New Cross, 1987), vol. 123 of London Math. Soc. Lecture Note Ser., Cambridge Univ. Press, Cambridge, 1987, pp. 81–110.

[2] P. FRANKL, *Shadows and shifting*, Graphs Combin., 7 (1991), pp. 23–29.

[3] N. SAUER, *On the density of families of sets*, J. Combinatorial Theory Ser. A, 13 (1972), pp. 145–147.

[4] S. SHELAH, *A combinatorial problem; stability and order for models and theories in infinitary languages*, Pacific J. Math., 41 (1972), pp. 247–261.

[5] V. N. VAPNIK AND A. Y. CHERVONENKIS, *Theory of uniform convergence of frequencies of events to their probabilities and problems of search for an optimal solution from empirical data*, Avtomat. i Telemeh., (1971), pp. 42–53.