Minimality of Uniform Splitting Families

Liljana Babinkostova[1], Gabriel Currier[2], Colin Okasaki[3], Marion Scheepers[1]

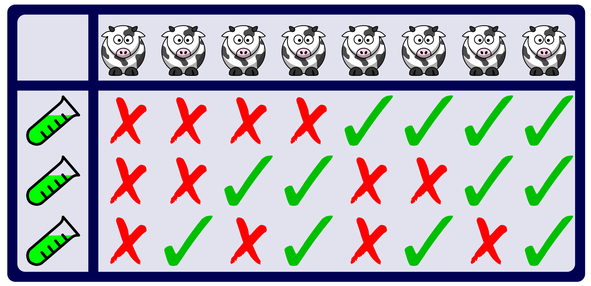
[1]Boise State University, [2]Pomona College, [3]Harvey Mudd College

A Bovine Blood-Test

Suppose you are the caretaker for a herd of 8 cows. You have noticed that exactly one cow has fallen ill, but you need to order blood tests to determine which it is. If you can mix several cows' blood into one sample, so that the test determines whether *any* of them are sick, how many tests do you need?

Numbering the cows 1–8, we claim the following scheme is optimal:

$$\mathcal{T} = \{\{5,6,7,8\},\{3,4,7,8\},\{2,4,6,8\}\}.$$



The Basics of Splitting

Uniform splitting families were originally conceived for use in cryptographic applications, where they provide an efficient attack on the low Hamming weight discrete log problem [1]. Our research focuses on minimality of these families — an interesting property in its own right. We are primarily interested in large minimal splitting families: what is the maximum size they can achieve, and how are such large families structured?

Splitting Families

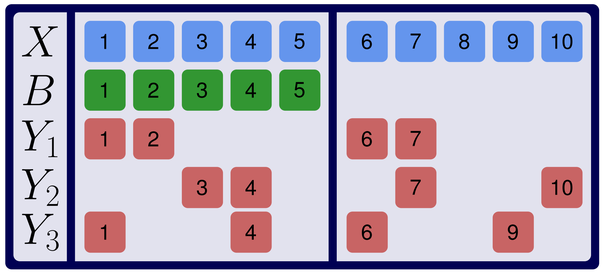
Let X be a set of size m. Let a subset of X be called a **block.** Then a set $\mathcal{F}$ of blocks is called an $(\mathbf{m},\mathbf{t})$-**splitting family** if for all $Y \subset X$ with $|Y| = t$, there exists a block $B \in \mathcal{F}$ so that $|B \cap Y| = \lfloor t/2 \rfloor$. The family $\mathcal{F}$ will be called **uniform** if all of its blocks have size $\lfloor m/2 \rfloor$.

Minimality

Let $\mathcal{F}$ be an (m,t)-uniform splitting family. We call $\mathcal{F}$ **minimal** if there does not exist a block $B \in \mathcal{F}$ so that $\mathcal{F} \backslash B$ is an (m,t)-uniform splitting family.

What Does Splitting Look Like?

Let $m = 10$ and $t = 4$. Then the following t-sets Y_1, Y_2, and Y_3 are all split by the block B, because they all have exactly half of their elements in B.



Splitting Families for $t = 2$

Let $N_{\min}(m,t)$ and $N_{\max}(m,t)$ denote the minimum and maximum sizes of minimal splitting families. For $t = 2$ we have exact values for both $N_{\min}$ and $N_{\max}$.

Bounds for $t = 2$

Let $m \geq 6$ be an integer. Then,

$$N_{\min}(m,2) = \lceil \log_2(m) \rceil$$

and

$$N_{\max}(m,2) = m - 3.$$

This $N_{\min}$ construction is how we are able to achieve the result for our cow problem to the left.

Splitting Families for Larger t

Although we have not been able to achieve the same strict bound for larger t, we do have some results limiting the possible values of $N_{\min}$ and $N_{\max}$.

Bound on N_{max}

Let m and t be even positive integers. Then,

$$N_{\max}(m,t) \leq \frac{2}{3}\binom{m}{t}.$$

However, the constructions that we've created have been considerably smaller than this bound. The following theorem gives the largest such construction:

Construction for $(m, m/2)$-families

Let m be an even positive integer. Then, there exists an $(m, m/2)$ splitting family of size $\binom{m/2}{m/4}$.

This theorem is proved by a construction which is most easily represented using incidence matrix notation. Given an (m,t)-splitting family $\mathcal{F}$, its incidence matrix is a $|\mathcal{F}| \times m$ matrix, where each row represents a block in $\mathcal{F}$ and each column represents an element of X. At position (i,j) we will place a 1 if block i contains element j and a 0 if it does not. For example if $m = 8$ we could represent the following set using vector notation:

$$\{1,3,5,7\} \iff \left[1\,0\,1\,0\,1\,0\,1\,0\right].$$

When expressed in incidence matrix notation, the construction for the above theorem is fairly simple and results in matrices similar to the following:

$$\left[\begin{array}{cc|cc|cc|cc} 1&1&1&1&0&0&0&0\\ 1&1&1&0&1&0&0&0\\ 1&1&1&0&0&1&0&0\\ 1&1&0&1&1&0&0&0\\ 1&1&0&1&0&1&0&0\\ 1&1&0&0&1&1&0&0 \end{array}\right]$$

It's unclear exactly how large these families can become, but it seems as though for $t = m/2$, this construction may in fact be as large as is possible.

Bounds for $N_{\min}$

Based on techniques used in [2], we were able to find a logarithmic bound for $N_{\min}(m,t)$. Although there are many examples of minimum-sized splitting families showing that this bound is not exact, no better bounds are known for arbitrary m and t.

Generalized Log Bound

Let m and t be even positive integers. Then,

$$N_{\min}(m,t) \leq \frac{t \log_2 m}{-\log_2\left(1 - \binom{t}{t/2}\left(\frac{1}{2}\right)^t\right)}.$$

An analogous result holds for m and t odd.

However, this bound is still far from the actual value of $N_{\min}$. The following are bounds on $N_{\min}$ that have been produced experimentally:

m	t	$N_{\min}$	m	t	$N_{\min}$
10	3	5	10	4	5
12	3	6	12	4	6
14	3	6	14	4	7
16	3	6	16	4	6
18	3	≤ 7	18	4	≤ 9
20	3	≤ 7	20	4	≤ 8
22	3	≤ 8	22	4	≤ 11
24	3	≤ 8	24	4	≤ 9

Future Work

- Develop further constructions for large minimal splitting families.
- Improve upper bounds on $N_{\max}$.
- Extend results to more general versions of splitting families.
- Prove the following theorem for the creation of large splitting families:

Construction for (m,t)-families

Let t be an even integer. Then for any sufficiently large m there exists a uniform (m,t) splitting family of size

$$\binom{\frac{m-t}{2}}{t/2}.$$

References

[1] D. R. Stinson. Some baby-step giant-step algorithms for the low Hamming weight discrete logarithm problem. *Math. Comp.*, 71(237):379-391 (electronic), 2002.

[2] A. C.H. Ling, P. C. Li, G. H. J. van Rees. Splitting systems and separating systems. *Discrete Math* 279 (2004) 355-368,

Acknowledgments

Funding for this research was provided by the National Science Foundation under Award No. DMS 1359425 and by Boise State University.