On generalizations of separating and splitting families

Daniel Condon[*]

Georgia Institute of Technology
Atlanta, Georgia, U.S.A.

danielmcondon@gmail.com

Samuel Coskey

Boise State University
Boise, ID, U.S.A.

scoskey@nylogic.org

Luke Serafin

Carnegie Melon University
Pittsburgh, PA, U.S.A.

lserafin@alumni.cmu.edu

Cody Stockdale

Bucknell University
Lewisburg, PA, U.S.A.

stockdalecody@gmail.com

Submitted: Oct 31, 2015; Accepted: Aug 10, 2016; Published: Aug 19, 2016
Mathematics Subject Classifications: 05B30, 05D40, 51E30, 94B25

Abstract

Starting from the well-established notion of a separating family (or separating system) and the refinement known as a splitting family, we define and study generalizations called n-separating and n-splitting families, obtaining lower and upper bounds on their minimum sizes. For n-separating families our bounds are asymptotically tight within a linear factor, while for n-splitting families we provide partial results and open questions.

1 Introduction

If X is a finite set and $A, B \subseteq X$, we say that A is a *separator* of B if both $A \cap B \neq \emptyset$ and $A^c \cap B \neq \emptyset$, where A^c is the complement of A in X. If $\mathcal{F}$ is a family of subsets of X, we say $\mathcal{F}$ is a *separating family* if for all $B \subseteq X$ with more than one element, $\mathcal{F}$ contains a separator of B.

Separating families (also called separating systems) were first studied in [10] in connection with probabilistic questions about boolean algebras. Since then, such families have found applications in many areas, including combinatorial search, switching circuit theory, and coding theory. Numerous variations of separating families arise in the context of even further applications. Since small families are typically best-suited for applications, much of the theory revolves around finding bounds on the minimum size of the families. We refer the reader to [7] for an introduction to the many variants of separating families.

[*] All authors supported by NSF grant DMS-1359425.

One of the most extensively studied variants, and one which will be featured in this investigation, is the following: a family $\mathcal{F}$ of subsets of X is an (i, j)-*separating family* if for all $P, Q \subseteq X$ such that $|P| \leqslant i$, $|Q| \leqslant j$, and $P \cap Q = \emptyset$, there exists $A \in \mathcal{F}$ such that $P \subseteq A$ and $Q \subseteq A^c$, or vice versa. While this may seem like an entirely different use of the word "separating", it is standard in the literature and in fact a family of sets is separating in the sense of the first paragraph if and only if it is $(1, 1)$-separating. Applications of (i, j)-separating families arise in the theory of automata, see for instance [5].

A related notion is that of splitting families. If X is a finite set and $A, B \subseteq X$, we say A *splits* B (or A is a *splitter* of B) if $|A \cap B| = \lfloor |B|/2 \rfloor$ or $|A \cap B| = \lceil |B|/2 \rceil$. A family $\mathcal{F}$ of subsets of X is said to be a *splitting family* if for all $B \subseteq X$, $\mathcal{F}$ contains an element that splits B. In the definition of splitting, we allow rounding both up and down both for symmetry and convenience. Some authors have more strict rounding conventions, for example see [11].

Splitting families have a less illustrious history than separating families. They first appeared in Coppersmith's algorithm for computing the discrete logarithm in the low Hamming weight case (described in [12]). Coppersmith's algorithm only requires families that split sets of one fixed size; such families are studied in more detail in [8] and [2]. As far as we know, families that split all subsets of X have not been previously studied.

In this paper, we define and study generalizations of separating and splitting families, which we call n-separating and n-splitting families, respectively. A family $\mathcal{F}$ of subsets of X is called an *n-separating family* if for any collection $B_1, \ldots, B_n$ of subsets of X, if there exists a set A which is a separator for all of the B_i then $\mathcal{F}$ contains such a set A. Similarly, the family $\mathcal{F}$ will be called an *n-splitting family* if for any collection $B_1, \ldots, B_n$ of subsets of X, if the B_i share some single splitter then $\mathcal{F}$ contains such a splitter.

For each of these two concepts, we establish the relationship between the new notion and its familiar counterpart. Our greatest effort is devoted to finding bounds on the minimum size of n-separating and n-splitting families. We believe that both generalizations will find new applications related to the applications of separating and splitting families, and give a sample application of n-separating families to error-correcting codes in the next section.

Let us now briefly outline the organization and results of this article. In the next section we give a brief overview of separating families which includes notation, examples, and basic properties. We then investigate n-separating families, beginning with the question of which collections of sets actually do share a common separator. We establish the relationship between n-separating families and (i, j)-separating families for all i, j. Finally we establish the following lower and upper bounds on the minimum size of an n-separating family.

Theorem. *The minimum size of an n-separating family on a set of size k is $\Omega(2^n \log k)$ and $O(n 2^n \log k)$.*

In the third section, we investigate splitting and n-splitting families. Splitting families turn out to be more challenging to work with than separating families. Once again, the section begins by addressing the question of which collections can be split by a single set.

In this case, a complete characterization is given only when $n \leqslant 3$. We then establish the following lower and upper bounds on the minimum size of a 2-splitting family.

Theorem. *The minimum size of a* 2*-splitting family on a set of size* k *is* $\Omega(k)$ *and* $O(k^2)$.

We can also compute an analogous lower bound for the size of a 3-splitting family. However, analogous lower bounds in the cases where $n \geqslant 4$ have unfortunately not been established, nor have useful upper bounds in the cases where $n \geqslant 3$. Nevertheless, if the key results in lemma 22 and theorem 25 can be generalized to these higher cases, one would obtain the following.

Conjecture. For each n, the minimum size of an n-splitting family on a set of size k is $\Omega(k^{n/2})$ and $O(k^{n/2+1})$.

2 n-separating families

In this section, we introduce and develop the concept of an n-separating family. After recalling some basic properties of ordinary separating families, we give an explicit construction for 2-separating families. Connections with the existing notion of an (i, j)-separating family yield a lower bound on the minimum size of an n-separating family, and the probabilistic method provides an upper bound on the minimum size of an n-separating family. We also provide a potential application of n-separating families by giving a connection with error-correcting codes.

Before defining n-separating families in general, it is useful to consider the simplest case of a 1-separating family, which is the same as a separating family as defined in the introduction. The theory of separating families is well-understood, and so provides a good illustration of methods and visualizations which will be used in the more general theory of n-separating families.

Recall from the introduction that if $A, B \subseteq X$ then A is a *separator* of B if both $A \cap B \neq \emptyset$ and $A^c \cap B \neq \emptyset$. We will use the notation $\mathrm{Sep}(B) = \{A \subseteq X : A \text{ is a separator of } B\}$. A family $\mathcal{F} \subseteq \mathcal{P}P(X)$ is a *separating family* if for every $B \subseteq X$ such that $|B| \geqslant 2$, $\mathcal{F}$ contains an element of $\mathrm{Sep}(B)$.

Note that for a pair of sets $B_1 \subseteq B_2 \subseteq X$, we have $\mathrm{Sep}(B_1) \subseteq \mathrm{Sep}(B_2)$. It follows that $\mathcal{F}$ is a separating family on X if and only if $\mathcal{F}$ contains separators of all subsets of X with cardinality 2: the sets of larger cardinality inherit separators from their two element subsets.

Any family $\mathcal{F}$ of subsets of a finite set X can be visualized as a matrix by enumerating both the elements of X and the elements of $\mathcal{F}$; for $X = \{x_1, \ldots, x_k\}$ and $\mathcal{F} = \{A_1, \ldots, A_n\}$, the ij-entry of the matrix representation of $\mathcal{F}$ is 1 if $x_i \in A_j$, 0 otherwise. A matrix representation of the separating family $\mathcal{F} = \{\{1, 2, 3, 4\}, \{1, 2, 5, 6\}, \{1, 3, 5, 7\}\}$ on $X = \{1, 2, \ldots, 8\}$ is given in table 1.

This matrix representation can be used to quickly obtain the minimum size of a separating family over $[k]$. First note that a matrix represents a separating family if and only if it has distinct columns; the separator of a pair $\{x, y\} \subseteq [k]$ will correspond to a row where the columns for x and y differ. It can now easily be seen that the minimum size of a separating family over $[k]$ is at least $\lceil \log k \rceil$, for the matrix representation of such a separating family must contain at least $\lceil \log k \rceil$ rows in order to have distinct columns. Furthermore, this bound is attainable: take the separating family corresponding to the $k \times \lceil \log k \rceil$-matrix with the $\lceil \log k \rceil$-bit binary representations of 0 through $k - 1$ as columns. The case $k = 8$ is represented in table 1.

●	●	●	●				
●	●			●	●		
●		●		●		●	

Table 1: The matrix representation of the family $\mathcal{F} = \{\{1, 2, 3, 4\}, \{1, 2, 5, 6\}, \{1, 3, 5, 7\}\}$. The $\bullet$ symbol denotes a 1 and an empty square denotes a 0.

Since one can always permute the enumerations of X and of $\mathcal{F}$, a separating family will have multiple matrix representations. Furthermore, since $A \in \text{Sep}(B)$ if and only if $A^c \in \text{Sep}(B)$, it is also natural to say families are equivalent if they contain the same subsets of X up to complements. Thus let us say that $m \times k$ binary matrices are equivalent if they lie in the same orbit the group G generated by row permutations, column permutations, and row complements. The quotient of G by the subgroup of just the column permutations turns out to be a familiar group: the symmetry group of the m-dimensional hypercube (assuming $2^m \leqslant k$). Thus we may also think of separating families as subsets of a Hamming cube, with vertices corresponding to columns of the matrix representation. Row permutations of a matrix correspond to rotations of the Hamming cube, while row complementations correspond to reflections. Thus the number of pairwise inequivalent separating families over $[k]$ of size m is the same as the number of 2-colorings of an m-dimensional Hamming cube which are distinct up to cube symmetry, where exactly k of the vertices are colored red (the remainder being black). This quantity is computed using Pólya enumeration in [1] and [6]. The resulting formulae are very complicated, and it is natural to ask whether computing the number of separating families over $[k]$ of size m is an NP-hard problem, though we do not pursue this question.

We now turn from ordinary separating families to their generalization, n-separating families. The idea is to require that any collection of n sets share some common separator contained in the family. Using the notation $\text{Sep}(B_1, \dots, B_n) = \bigcap_{i=1}^{n} \text{Sep}(B_i)$, we want to study families $\mathcal{F}$ such that for all collections $B_1, \dots, B_n \subset X$ with each $|B_i| \geqslant 2$, $\mathcal{F}$ contains an element of $\text{Sep}(B_1, \dots, B_n)$. However, $\text{Sep}(B_1, \dots, B_n)$ could be empty, as is the case with $B_1 = \{1, 2\}$, $B_2 = \{2, 3\}$, and $B_3 = \{3, 1\}$. This motivates the following.

Definition 1. A collection $B_1, \dots, B_n$ of subsets of $[k]$ is *separable* if $\text{Sep}(B_1, \dots, B_n)$ is nonempty.

In the next result we give a graph-theoretic characterization of separability. For technical reasons let us establish the convention that a graph contains no isolated nodes, that is, a graph is completely determined by its edge set.

Proposition 2. *A collection $B_1, \dots, B_n$ of subsets of $[k]$ is separable if and only if there exist pairs $b_1 \subseteq B_1, \dots, b_n \subseteq B_n$ such that the graph with edge set $\{b_1, \dots, b_n\}$ is bipartite.*

Proof. ($\Leftarrow$) Suppose $b_1 \subseteq B_1, \dots, b_n \subseteq B_n$ are such that $G = \{b_1, \dots, b_n\}$ is the edge set of a bipartite graph. Fix a 2-coloring f of the vertices of G. Then the set $f^{-1}[0] \in \mathrm{Sep}(b_1, \dots, b_n) \subseteq \mathrm{Sep}(B_1, \dots, B_n)$.

($\Rightarrow$) Suppose $B_1, \dots, B_n$ are given and let $A \in \mathrm{Sep}(B_1, \dots, B_n)$. For each $i \in [n]$, let $\alpha_i \in A \cap B_i$ and $\beta_i \in A^c \cap B_i$, and define $b_i = \{\alpha_i, \beta_i\}$. Letting G be the graph with edge set $\{b_1, \dots, b_n\}$, we have that $b_i \subseteq B_i$ and the function $f \colon V(G) \to \{0, 1\}$ given by $f(\alpha_i) = 0$ and $f(\beta_i) = 1$ for $i \in [n]$ is a 2-coloring of G. □

Remark 3. The problem of recognizing whether a collection is separable is NP-complete. Indeed, note that a collection $B_1, \dots, B_n$ is separable if and only if, when the collection is viewed as a hypergraph, it is 2-colorable. The problem of recognizing hypergraph 2-colorability is known to be NP-complete; see [9].

We are now prepared to define n-separating families.

Definition 4. A family $\mathcal{F} \subseteq \mathcal{P}[k]$ is an *n-separating* family if, for every separable collection $B_1, \dots, B_n \subseteq [k]$, $\mathcal{F}$ contains an element of $\mathrm{Sep}(B_1, \dots, B_n)$.

As in the case of 1-separating families, a family $\mathcal{F}$ is an n-separating family if and only if for every separable collection of n pairs $b_1, \dots, b_n$, $\mathcal{F}$ contains an element of $\mathrm{Sep}(b_1, \dots, b_n)$

Natural examples of n-separating families which are not too large are not immediately apparent, but the following simple construction does allow us to give modest-sized examples of 2-separating families.

Theorem 5. *If $\mathcal{F}$ is a separating family on $[k]$, and $\mathcal{F}' = \{A \triangle B : A, B \in \mathcal{F}\}$, then $\mathcal{F} \cup \mathcal{F}'$ is a 2-separating family.*

Proof. Let b_1, b_2 be two pairs in $[k]$ and let A_1, $A_2 \in \mathcal{F}$ be separators of b_1, b_2, respectively. Often, either A_1 or A_2 is in $\mathrm{Sep}(b_1, b_2)$. Otherwise, A_1 contains precisely one element in b_1 and A_2 contains both or zero elements of b_1. Then $A_1 \triangle A_2$ contains precisely one element in b_1. By identical reasoning, $A_1 \triangle A_2$ contains precisely one element in b_2, and $A_1 \triangle A_2 \in \mathrm{Sep}(b_1, b_2)$. □

Example 6. The 2-separating family on $[k] = [8]$ obtained by applying the previous result to table 1 is $\mathcal{F} = \{\{1, 2, 3, 4\}, \{1, 2, 5, 6\}, \{1, 3, 5, 7\}, \{3, 4, 5, 6\}, \{2, 4, 5, 7\}, \{2, 3, 6, 7\}\}$. Its matrix representation is shown in table 2.

Remark 7. Theorem 5 provides a constructive upper bound of $O\left((\log k)^2\right)$ for the minimum size of a 2-separating family over $[k]$. The results of [13] can be used together with lemma 12 to improve upon this bound.

Before moving on to calculating lower and upper bounds for the minimum sizes of n-separating families, we briefly describe an application to error-correcting codes. A *d-code*

●	●	●	●				
●	●			●	●		
●		●		●		●	
		●	●	●	●		
	●		●	●		●	
	●	●			●	●	

Table 2: The matrix representation of the family in example 6.

is an $m \times k$ binary matrix whose columns have pairwise Hamming distances at least d. When used as an error-correcting code, the columns of a d-code can be used to detect up to $(d-1)$-many errors and correct up to $\lfloor (d-1)/2 \rfloor$-many errors in a message. These error-correcting codes were introduced by Hamming in [4].

Theorem 8. *If $\mathcal{F}$ is n-separating, then the matrix representation of $\mathcal{F}$ is a 2^{n-1}-code.*

Proof. The result follows from these two claims:

1. If $\mathcal{F}$ is an n-separating family on $[k]$ and $S \subseteq [k]$ with $|S| = n+1$, then for every subset $T \subseteq S$, either T or $S \smallsetminus T$ lies in $\mathcal{F} \restriction S = \{A \cap S : A \in \mathcal{F}\}$.

2. If $\mathcal{G}$ is a family of subsets of $[n+1]$ with the property that for every $T \subset [n+1]$ either T or T^c lies in $\mathcal{G}$, then the matrix representation of $\mathcal{G}$ is a 2^{n-1}-code.

For claim (1), enumerate the elements of T as $\{t_1, \dots, t_j\}$ and the elements of $S \smallsetminus T$ as $\{s_1, \dots, s_{n+1-j}\}$, and consider the matching given by the pairs $\{t_1, s_1\}, \dots, \{t_1, s_{n+1-j}\}$ together with $\{t_2, s_1\}, \dots, \{t_j, s_1\}$. Observe that these pairs determine a connected bipartite graph with parts T and $S \smallsetminus T$. Letting $A \in \mathcal{F}$ be a separator for this collection, we must have either $A \cap S = T$ or $A \cap S = S \smallsetminus T$, as desired.

For claim (2), note first that if M is the matrix representation of the full power set $\mathcal{P}[n+1]$, then the columns of M have pairwise Hamming distances exactly 2^n.

Next let M' be a matrix representation of $\mathcal{G}$ obtained by deleting half the rows of M. Specifically for each $A \subseteq [n+1]$ delete either the row corresponding to A or to A^c (it doesn't matter which one). To see that the columns of M' form a 2^{n-1}-code, note that for each pair of columns i, j, exactly half of the rows we deleted disagreed in coordinates i, j. Thus the Hamming distance between columns i, j of M' is exactly $2^n - \frac{1}{4}2^{n+1} = 2^{n-1}$, which completes the proof. □

We now proceed with the main task of finding lower and upper bounds on the minimum size of an n-separating family. We first calculate the upper bounds.

Theorem 9. *If $\mathcal{F}$ is an n-separating family of subsets of $[k]$, then $|\mathcal{F}| \leqslant \frac{2n \log k}{-\log(1-2^{-n})} + 1$. In particular, the minimum size of an n-separating family of subsets of $[k]$ is $O(2^n n \log k)$.*

The proof will make use of the probabilistic method, but first we need the following probability estimate.

Lemma 10. *Given a separable collection of n pairs, the probability p that a randomly chosen subset of $[k]$ is a separator of every pair in the collection is at least 2^{-n}.*

Proof. Let p_n be the minimum, over collections of pairs $b_1, \dots, b_n$, of the probability that a randomly chosen subset of $[k]$ is a separator of $b_1, \dots, b_n$. We shall show that $p_{n+1} \geqslant \frac{1}{2}p_n$, and the result follows by a simple induction. Let $b_1, \dots, b_{n+1}$ be a separable collection of pairs, and let p be the probability that a random set A is in $\mathrm{Sep}(b_1, \dots, b_{n+1})$. We show that $p \geqslant \frac{1}{2}p_n$ by considering several cases.

Case 1: b_{n+1} is disjoint from $b_1 \cup \cdots \cup b_n$. Then the event that $A \in \mathrm{Sep}(b_{n+1})$ is independent of the event that $A \in \mathrm{Sep}(b_1, \dots, b_n)$. Since the probability that $A \in \mathrm{Sep}(b_{n+1})$ is $\frac{1}{2}$, clearly $p \geqslant \frac{1}{2}p_n$.

Case 2: b_{n+1} shares exactly one element with $b_1 \cup \cdots \cup b_n$. Let x denote the shared element and y denote the other element of b_{n+1}. The event that A contains y is independent of the event that $A \in \mathrm{Sep}(b_1, \dots, b_n)$. Meanwhile if $x \in A$ then $A \in \mathrm{Sep}(b_{n+1})$ if and only if $y \notin A$, and if $x \notin A$ then $A \in \mathrm{Sep}(b_{n+1})$ if and only if $y \in A$. Together this again implies $p \geqslant \frac{1}{2}p_n$.

Case 3: b_{n+1} shares both its elements with $b_1 \cup \cdots \cup b_n$. Let G be the graph with edges $b_1, \dots, b_n$. If b_{n+1} is contained in a single connected component of G (Case 3a), then $\mathrm{Sep}(b_1, \dots, b_n) \subseteq \mathrm{Sep}(b_{n+1})$, giving $p \geqslant p_n$.

If b_{n+1} is divided between two components of G (Case 3b), then choose $i \leqslant n$ such that removing b_i doesn't disconnect any component of $G \cup \{b_{n+1}\}$. (Every graph either has a cycle or a leaf.) Adding b_i to the collection $(G \cup \{b_{n+1}\} \smallsetminus \{b_i\})$ yields one of the cases 1, 2, or 3a. Hence we again conclude that $p \geqslant \frac{1}{2}p_n$. □

We are now ready to establish the upper bound.

Proof of Theorem 9. Let N denote the number of separable collections of n pairs. Note $N \leqslant \binom{k}{2}^n$. By lemma 10, the probability that a random set A is a separator of each pair in a given collection is at least $p = 2^{-n}$.

Given any separable collection $\{b_1, \dots, b_n\}$ of n pairs, the probability that a randomly chosen set is *not* in $\mathrm{Sep}(b_1, \dots, b_n)$ is at most $1 - p$. For a family of sets $\mathcal{F}$, let $E_{\mathcal{F}}$ be the event that no set in $\mathcal{F}$ is in $\mathrm{Sep}(b_1, \dots, b_n)$. Observe that for a family $\mathcal{F}$ of size m, the probability of $E_{\mathcal{F}}$ is at most $(1-p)^m$.

It follows from the linearity of expectation that if $\mathcal{F}$ is a family of size m, then the expected number of separable collections of pairs $\{b_1, \dots, b_n\}$ such that $\mathcal{F}$ does not contain an element of $\mathrm{Sep}(b_1, \dots, b_n)$ is at most $N(1-p)^m$. For sufficiently large m, $N(1-p)^m < 1$, which indicates that there exists a family $\mathcal{F}$ of size m where the number of collections of separable pairs $\{b_1, \dots, b_n\}$ such that $\mathcal{F}$ does not contain an element of $\mathrm{Sep}(b_1, \dots, b_n)$ is 0. That is to say, $\mathcal{F}$ contains a separator of every such collection. It is easy to check that $N(1-p)^m < 1$ for $m > \frac{\log(N)}{-\log(1-p)} + 1$, giving an upper bound in the minimum size of an n-separating family. Using the upper bound on N and lower bound on p, there exists an

n-separating family $\mathcal{F}$ such that

$$\begin{aligned} |\mathcal{F}| &\leqslant \frac{\log(\binom{k}{2}^n)}{-\log(1-2^{-n})} + 1 \\ &\leqslant \frac{2n\log k}{-\log(1-2^{-n})} + 1. \end{aligned}$$

This implies the desired asymptotic bound. □

We now turn to the task of establishing a lower bound on the size of an n-separating family. Our lower bound will involve a comparison between the sizes of n-separating families and the well-studied (i, j)-separating families. We first recall this latter notion and explore the relationship between the two.

Definition 11. If $A, P, Q \subseteq [k]$, we say A *separates* P *from* Q if either $P \subseteq A$ and $Q \subseteq A^c$, or $Q \subseteq A$ and $P \subseteq A^c$. A family $\mathcal{F} \subseteq \mathcal{P}[k]$ is (i, j)*-separating* if for all $P, Q \subseteq [k]$ with $|P| \leqslant i$, $|Q| \leqslant j$ and $P \cap Q = \emptyset$, there exists an element of $\mathcal{F}$ that separates P from Q.

We hope the reader will excuse this overloading of the term "separating"—note that both 1-separating and $(1, 1)$-separating are equivalent to separating, so both notions generalize ordinary separating but in different directions. In the next several results, we establish the full set of implications between the notions of (i, j)-separating and n-separating families. Specifically we prove that all of the implications described in Figure 1 hold, and that no other implications hold. This situation helps confirm that the notion of n-separating is interesting in its own right.

We begin by establishing the implications that do hold.

Lemma 12. *1. If $\mathcal{F}$ is (i, j)-separating then for every $i' \leqslant i$ and $j' \leqslant j$, $\mathcal{F}$ is (i', j')-separating.*

2. If $\mathcal{F}$ is (n, n)-separating then $\mathcal{F}$ is n-separating.

3. If $\mathcal{F}$ is $(i + j - 1)$-separating, then $\mathcal{F}$ is (i, j)-separating.

Proof. (1) This is clear from the definition.

(2) Let $\mathcal{F}$ be (n, n)-separating and let $b_1, \ldots, b_n$ be a separable collection of pairs. Then $b_1, \ldots, b_n$ form the edges of a bipartite graph with parts P, P'. Since $|P|, |P'| \leqslant n$, we can find $A \in \mathcal{F}$ which separates P, P'. Clearly $A \in \mathrm{Sep}(b_1, \ldots, b_n)$ and so $\mathcal{F}$ is n-separating.

(3) Let $\mathcal{F}$ be $(i+j-1)$-separating and let P, Q be disjoint sets of sizes i, j respectively. We can build a connected bipartite graph G with parts P and Q using $i + j - 1$ edges $b_1, \ldots, b_{i+j-1}$. (For this, place an edge from a fixed $p_0 \in P$ to each element in Q, and an edge from a fixed $q_0 \in Q$ to each element in $P \setminus \{p_0\}$.) Let $A \in \mathcal{F} \cap \mathrm{Sep}(b_1, \ldots, b_{i+j-1})$. Notice that A separates P from Q, so $\mathcal{F}$ is (i, j)-separating. □

Theorem 13. *The only implications between (i, j)-separating and n-separating notions are those established in lemma 12.*

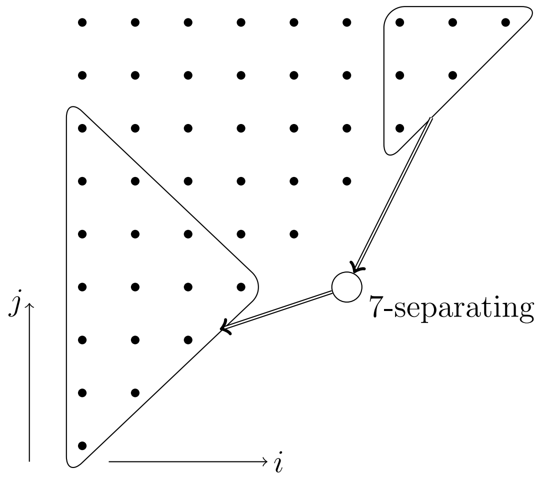


Figure 1: Diagram of separability notions. A filled dot in coordinate (i, j) represents the notion of (i, j)-separating. Implications between the filled dots go down and to the left. The outlined triangular regions are determined by satisfying the inequalities $i + j \leqslant 8$ and by $i, j \geqslant 7$.

Proof. The following is a series of counterexamples to the remaining implications.

- There exists an n-separating family which is not (n, n)-separating.

Fix disjoint sets $B, B' \subseteq [k]$ with $|B| = |B'| = n$, and let $\mathcal{F} = \bigcup_{i=1}^{n} \binom{[k]}{i} \smallsetminus \{B, B'\}$ (here $\binom{S}{i}$ denotes the subsets of S of cardinality i). Then the pair B, B' witnesses that $\mathcal{F}$ is not (n, n)-separating.

To see that $\mathcal{F}$ is n-separating, let $b_1, \ldots, b_n$ be a separable collection of pairs. Then these pairs make up the edges of a bipartite graph with two parts P, P'. If either $|P| < n$ or $|P'| < n$, then assuming without loss of generality that $|P| < n$, we have that $P \in \mathcal{F}$ and is also in $\mathrm{Sep}(b_1, \ldots, b_n)$.

On the other hand, if $|P| = |P'| = n$ then the pairs $b_1, \ldots, b_n$ must be pairwise disjoint. In this case, if for some $i_0 \leqslant n$, b_{i_0} contains an element x which is not in $B \cup B'$, choose any sequence $x_i \in b_i$ with $x_{i_0} = x$. Then the set $A = \{x_1, \ldots, x_n\}$ lies in $\mathcal{F}$ and $\mathrm{Sep}(b_1, \ldots, b_n)$. If some $b_{i_0} \subseteq B$ or B', then any selection of $x_i \in b_i$ will yield a set $A = \{x_1, \ldots, x_n\} \in \mathrm{Sep}(b_1, \ldots, b_n)$. Finally if every b_i meets both B and B', write $b_1 = \{x, y\}$, where $x \in B$ and $y \in B'$, and observe that $A = (B \smallsetminus \{x\}) \cup \{y\}$ lies in $\mathcal{F}$ and in $\mathrm{Sep}(b_1, \ldots, b_n)$.

- If $i+j \geqslant n+2$, then there exists an n-separating family which is not (i, j)-separating.

Fix disjoint sets $B, B' \subseteq [k]$ with $|A| = i$ and $|B| = j$, and let $\mathcal{F} = \mathcal{P}[k] \smallsetminus \{C : C \supseteq B \text{ or } C \supseteq B'\}$. Then $\mathcal{F}$ is not (i, j)-separating since no element of $\mathcal{F}$ contains either B or B'.

To see that $\mathcal{F}$ is n-separating, let $b_1, \ldots, b_n$ be a separable collection of pairs. Again these pairs make up the edges of a bipartite graph G with parts P, P'. If either P or P' lies in $\mathcal{F}$, then we are done. Otherwise, we can suppose that $P \supseteq B$ and $P' \supseteq B'$. Since $i + j \geqslant n + 2$, the set $B \cup B'$ cannot lie in a single connected component of G. Letting H be a component of G which meets $B \cup B'$, we can view G as a bipartite graph with parts $P \triangle H$ and $P' \triangle H$. Then at least one of these sets lies in $\mathcal{F}$, and is in $\mathrm{Sep}(b_1, \ldots, b_n)$.

- There exists an $(n-1, j)$-separating family which is not n-separating.

An example is $\mathcal{F} = \binom{[k]}{n-1}$.

- There exists an n-separating family which is not $(n+1)$-separating.

An example is $\mathcal{F} = \binom{[k]}{n}$.

- Let $i \leqslant j$. There exists an (i, j)-separating family which is not $(i+1, j)$-separating, and there exists an (i, j)-separating family which is not $(i, j+1)$-separating.

The family $\mathcal{F} = \binom{[k]}{i}$ is (i, j)-separating and not $(i+1, j)$-separating. The family $\mathcal{G} = \binom{[k]}{k-j}$ is (i, j)-separating and not $(i, j+1)$-separating.

- Let $i < i' \leqslant j' < j$. Then there exists an (i, j)-separating family which is not (i', j')-separating, and there exists an (i', j')-separating family which is not (i, j)-separating.

For the first statement, an example is given by $\mathcal{F} = \binom{[k]}{i}$.

For the second statement, fix $B \subseteq [k]$ with $|B| = i$ and set $\mathcal{G} = (\binom{[k]}{i'} \smallsetminus \{C : C \supseteq B\}) \cup \binom{[k] \smallsetminus B}{j'}$. Now let $P, Q \subseteq [k]$ be disjoint with $|P| \leqslant i'$, $|Q| \leqslant j'$. If $B \subseteq P$, then there is $A \in \mathcal{G}$ such that $Q \subseteq A$, and A separates P from Q. If $B \nsubseteq P$, then there is $A \in \mathcal{G}$ such that $P \subseteq A$, and again A separates P from Q. Thus $\mathcal{G}$ is (i', j')-separating.

On the other hand, fix $B' \subseteq [k]$ with $|B'| = j$ and $B \cap B' = \emptyset$. Since no set in $\mathcal{G}$ contains B, any set in $\mathcal{G}$ that would separate B from B' must contain B'. This is not possible since the sets in $\mathcal{G}$ have cardinality at most j' and $j' < j$. Thus $\mathcal{G}$ is not (i, j)-separating.

This concludes the proof of theorem 13. □

The implications established above can be used to convert the bounds on (i, j)-separating families given in [3] into bounds on n-separating families. The upper bound obtained in this way is not as tight as the upper bound already given in theorem 9. On the other hand, the lower bound obtained is the following.

Theorem 14. *The minimum size of an n-separating family has lower bound $\Omega(2^n \log k)$.*

Proof. By lemma 12, every n-separating family is an $(n/2, n/2)$-separating family. By theorem 3 of [3], the minimum size of an $(n/2, n/2)$-separating family has lower bound $\Omega(2^n \log k)$, as desired. □

3 n-splitting families and splittability

This section concerns splitting families and the new concept of n-splitting families. We begin with a simple upper bound on the minimum size of a splitting family. This result does not easily generalize to n-splitting families, so instead we use the probabilistic method to provide an upper bound on the minimum size of a 2-splitting family. To find lower bounds on the minimum size of a splitting family, we use the "volume method". With the help of a partial characterization of n-splittable collections, this method generalizes to 2- and 3-splitting families. We conclude with conjectures regarding bounds on the size of n-splitting families for arbitrary n.

Definition 15. A family $\mathcal{F} \subseteq \mathcal{P}[k]$ is a *splitting family* if, for all $B \subset [k]$, there exists $A \in \mathcal{F}$ such that $|A \cap B| = \lfloor |B|/2 \rfloor$ or $\lceil |B|/2 \rceil$.

When either of the latter two conditions holds, we say that A *splits* B or A is a *splitter* of B. We generalize this concept to that of an n-splitting family in an analogous manner to the generalization from separating to n-separating families.

Definition 16. A collection $\{B_1, \dots, B_n\} \subseteq \mathcal{P}[k]$ is *splittable* if there exists $A \subseteq [k]$ which splits each B_i.

Definition 17. A family $\mathcal{F}$ of subsets of $[k]$ is *n-splitting* if, for every splittable collection $\{B_1, \dots, B_n\} \subseteq \mathcal{P}[k]$, there exists $A \in \mathcal{F}$ which splits each B_i.

Note that ordinary splitting is equivalent to 1-splitting. We first address the case of ordinary splitting.

Theorem 18. *The minimum size of a splitting family of subsets of* $[k]$ *has upper bound* $\lceil k/2 \rceil$.

The construction below is attributed to Coppersmith in [12]; our statement and corresponding argument are slightly more general than the one found there.

Proof. For each i define $A_i = \{i, \dots, i + \lceil k/2 \rceil - 1\}$. Letting $\mathcal{F} = \{A_i : 1 \leqslant i \leqslant \lceil k/2 \rceil\}$, we claim that $\mathcal{F}$ is a splitting family. Let $B \subseteq [k]$ be given and define the function $f(i) = |A_i \cap B| - |A_i^c \cap B|$. We seek i such that $f(i) \in \{-1, 0, 1\}$, since this implies that A_i splits B.

To see there is such an i, we first claim that $f(i) - f(i+1) \in \{-2, 0, 2\}$. For this, note that $A_i \triangle A_{i+1} = \{i, i + \lceil k/2 \rceil\}$. If both or neither of these two points lie in B, then $f(i+1) = f(i)$; and if exactly one of these two points lies in B, then $f(i+1) = f(i) \pm 2$.

We can use similar reasoning to conclude that $f(1) + f(\lceil k/2 \rceil) \in \{-2, 0, 2\}$. Indeed, $A_1^c \triangle A_{\lceil k/2 \rceil}^c$ is $\{\lceil k/2 \rceil, k\}$ when k is even, and $\{\lceil k/2 \rceil\}$ when k is odd. Once again, if zero or two of these points lie in B, then $f(1) + f(\lceil k/2 \rceil) = 0$; and if exactly one of these points lies in B, then $f(1) + f(\lceil k/2 \rceil) = \pm 2$.

In sum, the sequence $f(1), \dots, f(\lceil k/2 \rceil)$ begins at $f(1)$, has step sizes at most 2, and ends at either $-f(1)$ or $-f(1) \pm 2$. It follows that there exists i such that $f(i) \in \{-1, 0, 1\}$, as desired. □

We conjecture the upper bound given above is sharp; however, we have only established a lower bound of $\Omega(\sqrt{k})$. In order to obtain this estimate, we shall use the volume method for computing lower bounds. This method was used together with more advanced techniques in [3] to obtain lower bounds for (i,j)-separating families. For fixed n, we say the *volume* of a set $A \subseteq [k]$ (as an n-splitter) is the number of distinct collections $\{B_1,\dots,B_n\} \subseteq \mathcal{P}[k]$ such that A is a splitter of each B_i in the collection.

Lemma 19. *Suppose there are at least N splittable subcollections of $\mathcal{P}[k]$ of size at least n and the maximum volume of any subset of $[k]$ (as an n-splitter) is v. If $\mathcal{F}$ is an n-splitting family on $[k]$, then*

$$|\mathcal{F}| \geqslant N/v.$$

The proof of this lemma is trivial: a family $\mathcal{F} \subseteq \mathcal{P}[k]$ splits at most $|\mathcal{F}|v$ collections of n distinct subsets of $[k]$, and there are at least N collections to be split. Therefore an n-splitting family must have size at least N/v. We now apply the volume method to find our lower bound on the size of an ordinary splitting family.

Theorem 20. *The minimum size of a splitting family of subsets of $[k]$ is $\Omega(\sqrt{k})$.*

Proof. Since the minimum size of a splitting family is monotone in k, we may suppose that k is even for the purpose of asymptotics. In this case, it is not difficult to see that the 1-splitters of maximum volume are of size $k/2$. To see this is true of splitting even-sized sets, one may simply compute that $\binom{k/2}{t/2}^2 \geqslant \binom{k/2+j}{t/2}\binom{k/2-j}{t/2}$. To establish it is also true of splitting odd-sized sets, note that if a set splits the maximum number of even-sized sets, then it splits the maximum number of sets.

Now, if $A \subseteq [k]$ has size $k/2$, then the volume v of A is given by

$$v = \sum_i \binom{k/2}{i}\binom{k/2}{i} + \sum_i \binom{k/2}{i}\binom{k/2}{i+1} + \sum_i \binom{k/2}{i+1}\binom{k/2}{i}.$$

Note that neither of the latter two terms are larger than the first since the inequality $2AB \leqslant A^2 + B^2$ implies

$$2\sum_i \binom{k/2}{i}\binom{k/2}{i+1} \leqslant \sum_i \binom{k/2}{i}^2 + \sum_i \binom{k/2}{i+1}^2 = 2\sum_i \binom{k/2}{i}^2.$$

It follows that the volume of A satisfies

$$v \leqslant 3\sum_i \binom{k/2}{i}^2 = 3\binom{k}{k/2}.$$

Applying the standard Stirling approximation that $\binom{k}{k/2} \sim 2^k/\sqrt{k}$, we conclude that v is $O(2^k/\sqrt{k})$. Meanwhile, the number of sets to be split is $N = 2^k$. By lemma 19, the volume lower bound is $\Omega(\sqrt{k})$, as desired. □

We remark that every collection consisting of just two sets B_1, B_2 is splittable: simply choose $D \subset B_1 \cap B_2$ of size $|D| = \lceil |B_1 \cap B_2|/2 \rceil$, choose $E \subset B_1 \smallsetminus B_2$ of size $|E| = \lfloor |B_1 \smallsetminus B_2|/2 \rfloor$, and choose $F \subset B_2 \smallsetminus B_1$ of size $|F| = \lfloor |B_2 \smallsetminus B_1|/2 \rfloor$. It is easy to see that $A = D \cup E \cup F$ is a splitter for both B_1 and B_2. This fact together with the method of theorem 20 gives the following lower bound on the size of 2-splitting families.

Theorem 21. *The minimum size of a* 2*-splitting family of subsets of* $[k]$ *is* $\Omega(k)$*.*

Proof. The calculation is similar to that of theorem 20. This time we consider the volume of $A \subseteq [k]$ as a 2-splitter. Once again, we assume that k is even and note that the 2-splitters of maximum volume are of size $k/2$. By a straightforward computation similar to that of theorem 20, the maximum volume v among pairs of subsets of $[k]$ satisfies

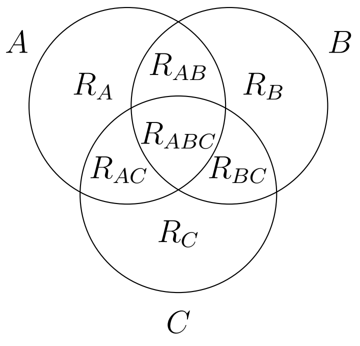
$$v \leqslant \left(3 \sum_i \binom{k/2}{i}\binom{k/2}{i}\right)^2 .$$

Since every subcollection of $\mathcal{P}[k]$ with at most two elements is splittable, and the number of such subcollections is $N = (2^k)^2$, we can use the same Stirling approximation as before to obtain a bound of $\mathcal{F} \geqslant (2^k)^2/v$ where $(2^k)^2/v$ is $\Omega\left((2^k)^2/(2^k/\sqrt{k})^2\right) = \Omega(k)$, as desired. □

The technique of splitting each sector of the Venn diagram of the B_i separately (described just above theorem 21) does not work in the case of n-splitting for general collections of three or more sets. For example, the collection $B_1 = \{1,2\}$, $B_2 = \{2,3\}$, $B_3 = \{3,1\}$ is not splittable. The next result loosely states that for collections of size three, this example is the only obstacle to splittability.

Lemma 22. *The collection* $\{A,B,C\} \subseteq \mathcal{P}[k]$ *is not splittable if and only if* $|A \cap B \cap C^c|$, $|A \cap B^c \cap C|$, *and* $|A^c \cap B \cap C|$ *are all odd, and there are no other elements in* $A \cup B \cup C$ *besides those elements in the three sets listed.*

Proof. We will make numerous references to the seven disjoint regions of the Venn diagram of A, B, C; and for convenience, they are labeled according to the figure shown below.



We first show that if R_{AB}, R_{BC}, and R_{AC} all have odd size, and $R_A = R_B = R_C = R_{ABC} = \emptyset$, then A, B, C is not splittable. Indeed, suppose towards a contradiction that S splits A, B, and C. Without loss of generality, suppose $|S \cap R_{AB}| > |R_{AB}|/2$. It follows that $|S \cap R_{BC}| < |R_{BC}|/2$ and that $|S \cap R_{AC}| > |R_{AC}|/2$. This implies that S does not split $R_{AB} \cup R_{AC}$— a contradiction because $R_{AB} \cup R_{AC} = A$ under our hypotheses.

Conversely, we show that if $\{A, B, C\}$ does not have this configuration, then the collection is splittable. First consider the case where $R_A = R_B = R_C = \emptyset$. If all four of the sectors R_{AB}, R_{BC}, R_{AC}, and R_{ABC} have even size, then we can build a splitter by simply taking the union of half of the elements in each sector. If just one of these four sectors has odd size, then follow the same procedure, rounding half of the number of elements we take in the union from the odd-sized sector either up or down. If two of these four sectors have odd size, then again follow the same procedure, rounding half of the number of elements from some odd-sized sector up and rounding the number of elements from the other odd-sized sector down. This leaves only the following three subcases shown in Figure 2.

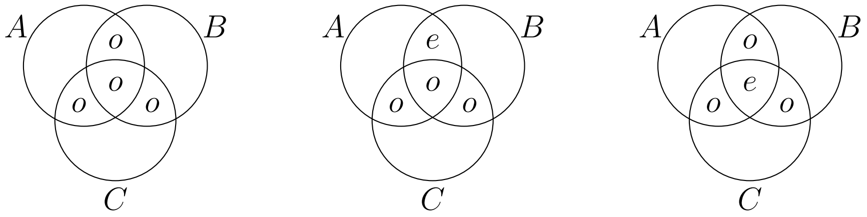


Figure 2: Subcases 1, 2, and 3 from left to right. The symbols e and o denote even and odd sized sectors.

In subcases 1 and 2, we round the number of elements taken from sector R_{ABC} up, and the number of elements from each of the other odd-sized sectors down. In subcase 3, we know that $|R_{ABC}| \geqslant 2$ (or else we are in the already-established unsplittable case). Thus in this case, we can build a splitter S with $|S \cap R_{ABC}| = |R_{ABC}|/2 + 1$, and each intersection of S with R_{AB}, R_{BC}, and R_{AC} having $\lfloor |R_{AB}|/2 \rfloor$, $\lfloor |R_{AB}|/2 \rfloor$, and $\lfloor |R_{AC}|/2 \rfloor$ elements respectively.

We next consider the case where at least one of R_A, R_B, or R_C is nonempty. If there exists a splitter S for the configuration $\{A \smallsetminus R_A, B \smallsetminus R_B, C \smallsetminus R_C\}$, then we can build a splitter S' of $\{A, B, C\}$ by letting $S' \supset S$ and suitably rounding the number of elements in intersection of S' with each of R_A, R_B, and R_C either up or down. Finally, if $\{A \smallsetminus R_A, B \smallsetminus R_B, C \smallsetminus R_C\}$ is not splittable, then by the above analysis it must be that R_{AB}, R_{BC}, R_{AC} all have odd size and $R_{ABC} = \emptyset$. Suppose for concreteness that $R_A \neq \emptyset$. Then we can build a splitter S such that $|S \cap R_{AB}| = \lfloor |R_{AB}|/2 \rfloor$, $|S \cap R_{BC}| = \lceil |R_{BC}|/2 \rceil$, $|S \cap R_{AC}| = \lfloor |R_{AC}|/2 \rfloor$, and $|S \cap R_A| = \lceil (|R_A| + 1)/2 \rceil$. This completes the proof. □

With this lemma, the lower bounds found for the sizes of 1-splitting and 2-splitting families in theorems 20 and 21 can be generalized to get lower bounds on the size of 3-splitting families.

Theorem 23. *The minimum size of a* 3*-splitting family of subsets of* $[k]$ *is* $\Omega(k^{3/2})$*.*

Proof. The reasoning of theorem 20 quickly shows that if v is the maximum volume of a 3-splitter then v satisfies

$$v \leqslant \left(3\sum \binom{k/2}{i}\binom{k/2}{i}\right)^3,$$

and hence (ignoring constants)

$$v \sim (2^k/\sqrt{k})^3.$$

However, since not every collection $\{B_1, B_2, B_3\} \subseteq \mathcal{P}[k]$ is splittable, N is more difficult to compute. It suffices to show that at least half of the collections $\{B_1, B_2, B_3\} \subseteq \mathcal{P}[k]$ are splittable. (The true fraction is significantly larger, but any constant will suffice).

Indeed, we can find an injection from the set of unsplittable collections to the set of splittable collections. Given an unsplittable collection $\{B_1, B_2, B_3\}$ we map it to $\{B_1 \cap B_2 \cap B_3^c, B_1 \cap B_2^c \cap B_3, B_1^c \cap B_2 \cap B_3^c\}$. The latter collection is disjoint and hence splittable. Moreover, by theorem 22, these are the only nonempty sectors of the Venn diagram of B_1, B_2, B_3. Thus this map is injective.

We have now shown that $N \geqslant (2^k)^3/2$, and this can be used as in the last two arguments to establish the desired volume lower bound. □

Conjecture. The minimum size of an n-splitting family of subsets of $[k]$ is $\Omega(3^{-n}k^{n/2})$.

We also conjecture that the problem of deciding whether an arbitrary collection of sets is splittable is NP-complete.

The remainder of this section is devoted to establishing an upper bound on the minimum size of a 2-splitting family. We conclude by stating a conjecture concerning an analogous upper bound on the minimum size of an n-splitting family.

Theorem 24. *The minimum size of a* 2*-splitting family of subsets of* $[k]$ *is* $O(k^2)$*.*

For this, we will need the following key result. First note that the number of splitters of a pair of sets S, T depends only on $|S|$, $|T|$, and $|S \cap T|$. If $|S|$ and $|T|$ are fixed in advance, then this number depends only on $|S \cap T|$.

Theorem 25. *If* $s + t \leqslant k$*, then the number of splitters of a pair of sets* S, T *with sizes* s, t *respectively is a nondecreasing function of* $b = |S \cap T|$*.*

Although the statement of theorem 25 feels intuitive, our proof is somewhat technical and is divided into several cases. We first consider the case where s, t are both even. For the proof of this case, fix sets S, T as in the statement of the theorem and elements $x \in S \setminus T$ and $y \in T \setminus S$; we may assume these elements exist since otherwise $S \cap T$ would already be as large as possible. We will also need to consider the pairs $S' = S \setminus \{x\}$, $T' = T \setminus \{y\}$, and $S'' = S$, $T'' = (T \setminus \{y\}) \cup \{x\}$. Refer to Figure 3 to visualize these three configurations.

Lemma 26. *Suppose* s *and* t *are both even.*

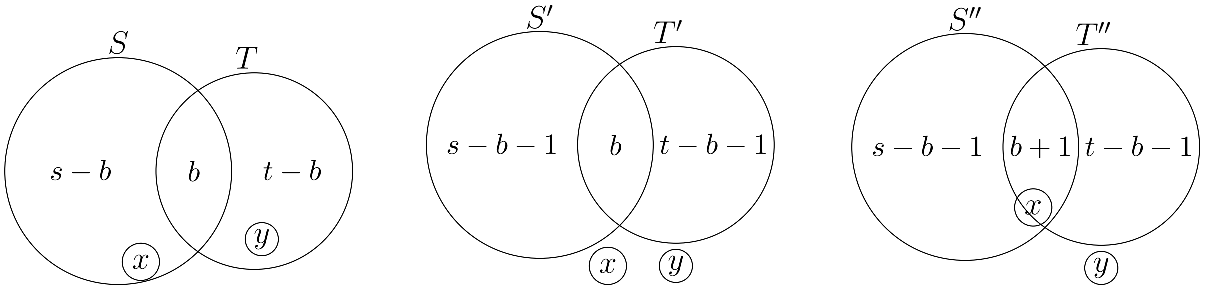


Figure 3: A diagram of the configurations used in the proof of theorem 25 when s, t are even. The elements of $\mathcal{A}$ split the configuration on the left, the elements of $\mathcal{B}$ split the configuration in the middle, and the elements of $\mathcal{C}$ split the configuration on the right.

- *If $\mathcal{A}$ is the family of all sets that simultaneously split S, T and $\mathcal{B}$ is the family of all sets that simultaneously split S', T', then $4|\mathcal{A}| = |\mathcal{B}|$.*
- *If $\mathcal{C}$ is the family of all sets that simultaneously split S'', T'', then $\frac{1}{4}|\mathcal{B}| \leqslant |\mathcal{C}|$.*

Proof. Since the elements of $[k] \setminus (T \cup S)$ have no effect on our claims, we may assume without loss of generality that $s+t-b=k$. Begin by noting that $\mathcal{A} \subseteq \mathcal{B}$. Indeed, suppose that $A \in \mathcal{A}$, so that A splits S and T. If $x \in A$ then $|A \cap S'| = \lfloor (t-1)/2 \rfloor$, and if $x \notin A$ then $|A \cap S'| = \lceil (t-1)/2 \rceil$. The corresponding statement for T' also holds, so A splits S' and T'.

In fact, for each element $A \in \mathcal{A}$, we can generate four distinct elements of $\mathcal{B}$. For this, given $A \in \mathcal{A}$ let A_1, A_2, A_3, A_4 be sets such that $A_i \cap (T' \cup S') = A \cap (T' \cup S')$ and $x \in A_1 \cap A_2 \cap A_3^c \cap A_4^c$ and $y \in A_1^c \cap A_2 \cap A_3 \cap A_4^c$. If $B \in \mathcal{A}$ and $B \neq A$, let B_i denote the four splitters generated in this manner from B. It is clear that $A_i \neq B_j$ for $i \neq j$. Moreover since s, t are even, A and B must disagree on $S' \cup T'$, and so $A_i \neq B_i$ as well. Lastly, it is not difficult to see that every element of $\mathcal{B}$ is of the form A_i for some $A \in \mathcal{A}$, which concludes the proof that $4|\mathcal{A}| = |\mathcal{B}|$.

For the second statement, it suffices to show that at least one fourth of the elements of $\mathcal{B}$ are in fact elements of $\mathcal{C}$. Let $B \in \mathcal{B}$, so that B splits both S' and T'. Observe that B is a splitter of both S'' and T'' if and only if either of the following conditions hold:

- $|B \cap S'| = s/2$, $|B \cap T'| = t/2$, and $x \notin B$; or
- $|B \cap S'| = s/2-1$, $|B \cap T'| = t/2-1$, and $x \in B$.

We claim that at least half of the elements of $\mathcal{B}$ satisfy either $|B \cap S'| = s/2$, $|B \cap T'| = t/2$ or else $|B \cap S'| = s/2-1$, $|B \cap T'| = t/2-1$. Once this claim is established, the proof will be complete because the conditions $x \notin B$ and $x \in B$ are independent of these and occur exactly half the time.

The number of elements of $\mathcal{B}$ that satisfy either $|B \cap S'| = s/2$, $|B \cap T'| = t/2$ or else $|B \cap S'| = s/2-1$, $|B \cap T'| = t/2-1$ is

$$\sum_{i=0}^{b} \binom{b}{i} \left[\binom{s-b-1}{s/2-i} \binom{t-b-1}{t/2-i} + \binom{s-b-1}{s/2-i-1} \binom{t-b-1}{t/2-i-1} \right].$$

On the other hand, the number of splitters that satisfy either $|B \cap S'| = s/2 - 1$, $|B \cap T'| = t/2$ or else $|B \cap S'| = s/2$, $|B \cap T'| = t/2 - 1$ is given by

$$\sum_{i=0}^{b} \binom{b}{i} \left[\binom{s-b-1}{s/2-i-1} \binom{t-b-1}{t/2-i} + \binom{s-b-1}{s/2-i} \binom{t-b-1}{t/2-i-1} \right].$$

We shall show that the first sum is greater than or equal to the second sum, and in fact that this is true term-by-term. Taking the i^{th} term of the first sum minus the i^{th} term in the second sum and factoring, this desired conclusion is equivalent to the following:

$$\left[\binom{s-b-1}{s/2-i} - \binom{s-b-1}{s/2-i-1} \right] \left[\binom{t-b-1}{t/2-i} - \binom{t-b-1}{t/2-i-1} \right] \geqslant 0$$

By the symmetric unimodal property of the binomial coefficients, both of the terms in the above product are negative for $i < b/2$ and both are nonnegative for $i \geqslant b/2$, so the inequality is always true. This completes the proof of the claim, and therefore the proof that at least one fourth of the elements of $\mathcal{B}$ also lie in $\mathcal{C}$. □

We now consider the case where s is odd and t is even. Once again we may let $x \in S \smallsetminus T$ and $y \in T \smallsetminus S$. We may also assume there exists $z \in [k] \smallsetminus (S \cup T)$; if there is not such an element, then we artificially add the element z to $[k]$. We shall need the sets $S' = S \cup \{z\}$ and $T' = T \cup \{x\} \smallsetminus \{y\}$.

Lemma 27. *Suppose s is odd and t is even.*

- *If $\mathcal{A}$ is the family of all sets that split S, T simultaneously and $\mathcal{A}'$ is the family of all sets that split S', T simultaneously, then $|\mathcal{A}| = 2|\mathcal{A}'|$.*
- *If $\mathcal{C}'$ is the family of all sets that split S', T' simultaneously, then $|\mathcal{A}'| \leqslant |\mathcal{C}'|$.*
- *If $\mathcal{C}$ is the family of all sets that split S, T' simultaneously, then $|\mathcal{C}| = 2|\mathcal{C}'|$.*

Proof. For the first statement, if $B \in \mathcal{B}$, then both $B \cup \{x\}$ and $B \smallsetminus \{x\}$ lie in $\mathcal{A}$. On the other hand if $A \in \mathcal{A}$ then exactly one of $A \cup \{x\}$ or $A \smallsetminus \{x\}$ lies in $\mathcal{B}$. This shows that there are exactly two elements of $\mathcal{A}$ for every element of $\mathcal{B}$, so $|\mathcal{A}| = 2|\mathcal{B}|$.

Now, the second statement is an instance of lemma 26, applied to the families $\mathcal{A}'$ and $\mathcal{C}'$.

The third statement is an instance of the first statement. □

We are now ready to complete the proof of the key result.

Proof of Theorem 25. Let $\mathcal{A}$ be the set of splitters of S, T where $|S| = s$, $|T| = t$, and $|S \cap T| = b$, and let $\mathcal{B}$ be the set of splitters of S', T' where $|S'| = s$, $|T'| = t$, and $|S' \cap T'| = b + 1$. We wish to show that $|\mathcal{A}| \leqslant |\mathcal{B}|$. The case where s, t are both even is handled by lemma 26, and the cases where just one of s, t is even is handled by lemma 27. In the remaining case (where s, t are both odd), we can use a method identical to the proof of lemma 27. More specifically, adjoin a new element z to T and then apply the statement of lemma 27. □

Finally, we can establish our upper bound for the minimum size of a 2-splitting family.

Proof of Theorem 24. If $T \subseteq [k]$ with $|T| = t$, then by a Stirling-type approximation computed in Section 2 of [12], the probability that T is split by a random subset of $[k]$ has a lower bound of $c/\sqrt{t}$ where c is a constant not depending on k.

Next, if $S \subseteq [k]$ with $|S| = s$ and $s + t \leqslant k$, then by theorem 25, the probability $p_{S,T}$ that a random set simultaneously splits S and T is minimized when $S \cap T = \emptyset$. In this case, the event that S is split and the event that T is split are independent, and so $p_{S,T} = p_s p_t$, where p_n is the probability that a randomly chosen subset of $[k]$ splits a given set of size n. Minimizing over the possible sizes s and t, we have that $p_{S,T}$ has lower bound c^2/k.

If it is not the case that $s + t \leqslant k$, then we instead regard S and T as subsets of $[2k]$ and again apply theorem 25. Since the estimate $c/\sqrt{t}$ didn't depend on k, we again obtain the desired lower bound of c^2/k.

Now invoke the probabilistic method calculation of the proof of theorem 9, this time using the values $p = c^2/k$ and $N = (2^k)^2$ (the number of ordered pairs of subsets of $[k]$). The calculation gives a 2-splitting family $\mathcal{F}$ of subsets of $[k]$ which satisfies

$$\begin{aligned} |\mathcal{F}| &< \frac{\log((2^k)^2)}{-\log(1-c^2/k)} + 1 \\ &\leqslant \frac{2k}{-\log(1-c^2/k)} + 1 \end{aligned}$$

This latter expression is $O(k^2)$, as desired. □

We close by conjecturing that the analog of theorem 25 holds for configurations of n sets as well. If this conjecture holds, one can easily obtain, for each n, an upper bound of $O(k^{n/2+1})$ on the minimum size of an n-splitting family on $[k]$.

Conjecture. If $B_1, \ldots, B_n$ is a collection of subsets of $[k]$, then the number of splitters of $B_1, \ldots, B_n$ is minimized when the collection is pairwise disjoint.

Acknowledgements

This article represents a part of the authors' work during the Boise State University 2014 math REU program. The authors were supported by NSF grant DMS-1359425, and the Boise State University mathematics department.

References

[1] William Y. C. Chen. Induced cycle structures of the hyperoctahedral group. *SIAM J. Discrete Math.*, 6(3):353–362, 1993.

[2] D. Deng, D. R. Stinson, P. C. Li, G. H. J. van Rees, and R. Wei. Constructions and bounds for (m, t)-splitting systems. *Discrete Math.*, 307(1):18–37, 2007.

[3] Michael L. Fredman and János Komlós. On the size of separating systems and families of perfect hash functions. *SIAM J. Algebraic Discrete Methods*, 5(1):61–68, 1984.

[4] R.W. Hamming. Error detecting and error correcting codes. *Bell System Technical Journal, The*, 29(2):147–160, April 1950.

[5] Michael A. Harrison. *Introduction to switching and automata theory*. McGraw-Hill Book Co., New York-Toronto-London, 1965.

[6] Michael A. Harrison and Robert G. High. On the cycle index of a product of permutation groups. *J. Combinatorial Theory*, 4:277–299, 1968.

[7] G. O. H. Katona. Combinatorial search problems. In *Survey of combinatorial theory (Proc. Internat. Sympos., Colorado State Univ., Fort Collins, Colo., 1970)*, pages 285–308. North-Holland, Amsterdam, 1973.

[8] Alan C. H. Ling, P. C. Li, and G. H. J. van Rees. Splitting systems and separating systems. *Discrete Math.*, 279(1-3):355–368, 2004. In honour of Zhu Lie.

[9] L. Lovász. Coverings and coloring of hypergraphs. In *Proceedings of the Fourth Southeastern Conference on Combinatorics, Graph Theory, and Computing (Florida Atlantic Univ., Boca Raton, Fla., 1973)*, pages 3–12. Utilitas Math., Winnipeg, Man., 1973.

[10] A. Rényi. On random generating elements of a finite Boolean algebra. *Acta Sci. Math. Szeged*, 22:75–81, 1961.

[11] Dongyoung Roh and Sang Geun Hahn. Constructions for uniform $(m, 3)$-splitting systems. *Math. Commun.*, 17(2):639–654, 2012.

[12] 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.

[13] D. R. Stinson, Tran van Trung, and R. Wei. Secure frameproof codes, key distribution patterns, group testing algorithms and related structures. *J. Statist. Plann. Inference*, 86(2):595–617, 2000. Special issue in honor of Professor Ralph Stanton.