On splitting and splittable families

SAMUEL COSKEY

Boise State University
Boise, ID, U.S.A.
scosky@gmail.com

BRYCE FREDERICKSON

Emory University
Atlanta, GA, U.S.A.
bfrede4@emory.edu

SAMUEL MATHERS

Chicago, IL, U.S.A.
smathers304@gmail.com

HAO-TONG YAN

New York, NY, U.S.A.
haotongyan23@gmail.com

Abstract

A set A is said to split a finite set B if exactly half the elements of B (up to rounding) are contained in A. We study the dual notions: (1) splitting family, which is a collection of sets such that any subset of $\{1, \ldots, k\}$ is split by a set in the family, and (2) splittable family, which is a collection of sets such that there is a single set A that splits each set in the family. We study the minimum size of a splitting family on $\{1, \ldots, k\}$, as well as the structure of splitting families of minimum size. We use a mixture of computational and theoretical techniques. We additionally study the related notions of ≤ 4-splitting families and 4-splitting families, and we provide lower bounds on the minimum size of such families. Next we investigate splittable families that are just on the edge of unsplittability in several senses. First, we study splittable families that have the fewest number of splitters. We give a complete characterization in the case of two sets, and computational results in the case of three sets. Second, we define a splitting game, and study splittable families for which a splitter cannot be found under adversarial conditions.

1   Introduction

This article concerns the dual notions of splitting families and splittable families, which arise naturally in several areas of combinatorics. Both types of families involve the following key notion:

Definition 1.1. Let A, B be subsets of $[k] = \{1, \ldots, k\}$. We say that A *splits* B, or A is a *splitter* of B, if $|A \cap B| = |B|/2$ for $|B|$ even, and $|A \cap B| = (|B| \pm 1)/2$ for $|B|$ odd.

A collection $\mathcal{A}$ of subsets of $[k]$ is said to be a *splitting family* on k if for any $B \subset [k]$ there exists an $A \in \mathcal{A}$ such that A splits B. Splitting families are also called splitting systems, and along with separating systems such families have been used to aid in combinatorial search algorithms [14], see also [7, 12]. In sections 2–3 below, we will study several questions surrounding splitting families and variations of splitting.

A collection $\mathcal{B}$ of subsets of $[k]$ is said to be a *splittable family* if there exists $A \subset [k]$ such that for each $B \in \mathcal{B}$ we have A splits B (we sometimes abbreviate this: A splits $\mathcal{B}$). Splittable families correspond to the families with combinatorial discrepancy ≤ 1, the optimal value. We refer to [5] as a starting point for discrepancy theory and its numerous applications. In sections 4–5, we will study several questions surrounding splittable families and variations of splittability.

Beginning with splitting families, our first investigation focuses on the question of the least size of a splitting family on k. Splitting families were used in [14] as an aid in an algorithm for solving the low Hamming weight discrete logarithm problem. In that article, the author presents a construction, attributed varyingly to Coppersmith and Galvin, of a splitting family on k with $\lceil k/2 \rceil$ sets. Specifically he shows that the family consisting of $A_i = \{i, i+1 \ldots, i + \lceil k/2 \rceil - 1\}$ for $i = 1, \ldots, \lceil k/2 \rceil$ is a splitting family on k. (We will refer to this as a *standard* splitting family.)

The article [14] left it unclear whether there exists a splitting family of smaller size than the standard splitting family. It is clear that a smaller splitting family would improve their algorithms, but it is not straightforward to observe whether such an object exists. In Section 2 we affirm that the standard splitting family is optimal in the sense that any splitting family on k has size at least $\lceil k/2 \rceil$. This result follows from the results of [1] and was pointed out to us by a colleague [16].

It is also natural to ask whether the standard splitting family is the *unique* splitting family of optimal size. We provide computational results essentially confirming this conjecture for $k \leq 16$ (with just one notable exception). We further show that under additional assumptions, the conjecture holds for all k.

In Section 3 we study two broad generalizations of splitting families which arise naturally in the above investigations. A collection $\mathcal{A}$ is said to be a ≤ 4-*splitting family* on k if for any $B \subset [k]$ such that $|B| \leq 4$ there exists an $A \in \mathcal{A}$ such that A splits B. The notion of 4-splitting family is defined analogously. We will provide lower bounds on the minimum size of a ≤ 4-splitting family and of a 4-splitting family. The latter bound corrects a minor error in [7].

Our second major area of investigation concerns the boundary between splittable and unsplittable families. It was shown in [4] (and elaborated in [3]) that the general problem of deciding whether a given family is splittable is NP-complete. It is then natural to ask which families $\mathcal{B}$ are "just splittable," that is, splittable but with the fewest number of distinct sets A that split $\mathcal{B}$. The answer to the question of which n-set families $\mathcal{B}$ are "just splittable" can be used to find bounds on the least size of an n-splitting family on $[k]$, see [6].

In Section 4 we give results on n-set families on $[k]$ with the fewest number of splitters for $n \leq 3$. When $n = 1$ and $\mathcal{B} = \{B\}$, the minimum number of splitters occurs when $|B| = k$ or $k - 1$, and the number of splitters is asymptotic to $2^k/\sqrt{k}$. When $n = 2$ and $\mathcal{B} = \{B_1, B_2\}$, it was shown in [6] that if $|B_i| \leq k/2$ then the minimum number of splitters occurs when B_1 and B_2 are disjoint. In this article we remove the cardinality constraints $|B_i| \leq k/2$ and find that the minimum number of splitters occurs when $|B_i| \approx 2k/3$ and $|B_1 \cap B_2| \approx k/3$. In both the disjoint and general cases, the number of splitters is asymptotic to $2^k/k$. We also calculate an analytical formula for approximating the number of splitters of an arbitrary two-set family $\mathcal{B}$.

When $n = 3$ the situation is somewhat more complex and we provide partial answers and computational results. Based solely on the short pattern $2^k/k^{1/2}$, $2^k/k^1$, the authors of [6] conjectured that for $n = 3$ the minimum number of splitters is asymptotic to $2^k/k^{3/2}$. Our computational results support this conjecture. However the explanation for the formula $2^k/k^{3/2}$ does not seem to be the same as the explanation for the cases $n = 1, 2$. Indeed, when $n \geq 3$ there exist unsplittable families, unlike when $n < 3$, and the "just splittable" families are very similar to the unsplittable ones in appearance.

In Section 5, we define and study a strategic variant of the notion of splittability. We define the *splitting game*, in which players Split and Skew are given a family $\mathcal{B}$ and collaborate to construct a set A. Split wins if A splits $\mathcal{B}$ and Skew wins otherwise. We establish several general lemmas that allow one to simplify the analysis of a given instance of the game. For families $\mathcal{B}$ of three or fewer sets, we give a complete characterization of when each player has a winning strategy. We also provide complete solutions to several special case studies, such as the tic-tac-toe style game which arises when the splitting game is played on a grid.

2   Splitting families

Recall from the introduction that a splitting family $\mathcal{A}$ on k is a collection of sets for which every subset of $[k]$ is split by some set in $\mathcal{A}$. In this section we investigate several questions surrounding the minimum size of a splitting family on k, as well as the structure of families of minimum size.

Recall from the introduction that for even k, the *standard splitting family* on k consists of $A_i = \{i, i + 1 \ldots, i + k/2 - 1\}$ for $i = 1, \ldots, k/2$. For any even k, the standard splitting family on k is in fact a splitting family, see for instance [8] and

[14]. As a consequence, for any k the minimum size of a splitting family on k has upper bound $\lceil k/2 \rceil$. The next result states that this bound is tight. We are grateful to Calvin Yost–Wolff for discovering the proof and allowing us to include it here.

Theorem 2.1 ([16]). *Any splitting family on k has size at least $k/2$. As a consequence, the minimum size of a splitting family on k is exactly $\lceil k/2 \rceil$.*

Proof. Suppose that $\mathcal{A}$ is a splitting family on k. For each $A \in \mathcal{A}$ let $\boldsymbol{v}_A$ be the ± 1 characteristic vector of A, that is, $\boldsymbol{v}_A(i) = 1$ for $i \in A$ and $\boldsymbol{v}_A(i) = -1$ for $i \notin A$. Define a polynomial $p(\boldsymbol{x})$ in k variables by

$$p(\boldsymbol{x}) = \prod_{A \in \mathcal{A}} \boldsymbol{v}_A \cdot \boldsymbol{x}.$$

Then p vanishes on every $0, 1$-vector with an even number of 1's. Next define

$$q(\boldsymbol{x}) = p((x_1 + 1)/2, \ldots, (x_k + 1)/2).$$

and note that q vanishes on every ± 1-vector with an even number of 1's.

Finally let $\bar{q}$ be the polynomial obtained from q by repeatedly replacing occurrences of x_i^2 with 1 in each monomial of q. Then $\bar{q}$ agrees with q on every ± 1-vector, and hence $\bar{q}$ vanishes on every ± 1-vector with an even number of 1's.

Now $\bar{q}$ is multilinear in the sense that it is affine in each coordinate. Moreover $\bar{q} \neq 0$: indeed, p does not vanish on any $0, 1$-vector with an odd number of 1's, so q does not vanish on any ± 1-vector with an odd number of 1's, so $\bar{q}$ does not vanish on any ± 1-vector with an odd number of 1's. It follows from [1, Lemma 2.1] that $\deg(\bar{q}) \geq k/2$. We can now conclude that $|\mathcal{A}| = \deg(p) = \deg(q) \geq \deg(\bar{q}) \geq k/2$. $\square$

Given this result, it is natural to wonder whether the standard splitting family on k is the unique splitting family of size $\lceil k/2 \rceil$. In the rest of this section, we present partial answers to this question.

We begin with computational results. In the following we say that a splitting family $\mathcal{A}$ on k is *uniform* if every set $A \in \mathcal{A}$ has size $|A| = \lfloor k/2 \rfloor$ or $\lceil k/2 \rceil$. Observe that the standard splitting family on k is uniform. We also say that splitting families $\mathcal{A}, \mathcal{A}'$ are *equivalent* if $\mathcal{A}'$ can be obtained from $\mathcal{A}$ by complementing sets in $\mathcal{A}$ and permuting the elements of $[k]$.

Proposition 2.2.   ○ *For all $k \leq 16$, every minimal splitting family is uniform.*

     ○ *Among even $k \leq 16$, there is just one minimal splitting family which is not equivalent to a standard splitting family, namely $k = 8$ and*

$$\mathcal{F} = \{\{1, 2, 3, 4\}, \{1, 2, 5, 6\}, \{3, 4, 5, 6\}, \{1, 3, 5, 7\}\}.$$

The method of search centered on the following notion. We say that splitting family $\mathcal{A} = \{A_1, \ldots, A_n\}$ on k is *extendable* if it is the restriction to k of some splitting family on $k + 1$. Our algorithm performed an exhaustive search of non-extendable

splitting families with a fixed number of sets. We do not see any immediate generalizations to the exceptional splitting family with $k = 8$. The code and its output may be found in [11].

In our next results we identify the structure of minimal splitting families under a strong additional hypothesis. To begin, we recall that if $\mathcal{A} = \{A_1, \ldots, A_n\}$ is a family of subsets of $[k]$, its *incidence matrix* is the $n \times k$ matrix A with a 1 in the (i, j) entry if $j \in A_i$, and a 0 otherwise. When every pair from $[k]$ is split by some set in $\mathcal{A}$, the columns of A are distinct elements of the Hamming cube $\{0, 1\}^n$. In this case, the *Hamming representation* of $\mathcal{A}$ is the graph on the subset $S \subset \{0, 1\}^n$ consisting of the columns of A, with adjacency defined by $s \sim t$ if and only if s, t differ in exactly one entry.

If $\bar{x}$ is a subset of the indices $1, \ldots, n$ we let $\delta_{\bar{x}}$ denote the element of $\{0, 1\}^n$ with 1's in the entries indicated by $\bar{x}$ and 0's everywhere else. We say that a set B_δ of such vectors is *split* if some index appears in exactly half (up to rounding) of the elements of B_δ with this representation. Thus, given a family $\mathcal{A}$ with Hamming representation S on $\{\delta_{\bar{x}_1}, \ldots, \delta_{\bar{x}_k}\}$, we can identify a subset $B \subset [k]$ with the subset $B_\delta = \{\delta_{\bar{x}_j} \mid j \in B\}$ of S, which is split if and only if some set in $\mathcal{A}$ splits B. In our graphical representations we stratify the elements $\delta_{\bar{x}}$ by their *weight*, $|\bar{x}|$. See Figure 1 for two examples of Hamming representations.
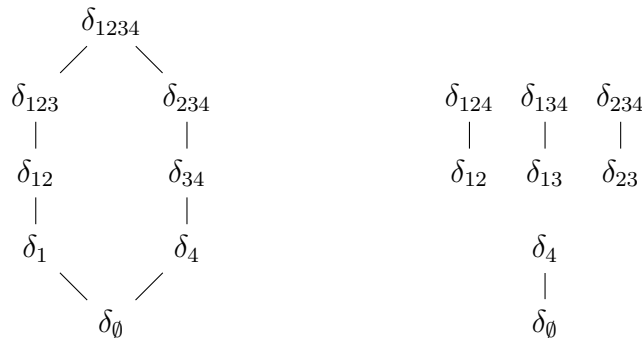

Figure 1: Left: The Hamming representation of the standard splitting family on $k = 8$. Right: The Hamming representation of the nonstandard minimal splitting family on $k = 8$.

In the following result, we say that a family $\mathcal{A}$ of subsets of $[k]$ is a *$\leq t$-splitting family* if every subset of $[k]$ of size at most t is split by some set in $\mathcal{A}$. We say that a ≤ 2-splitting family $\mathcal{A}$ is *connected* if its Hamming representation is a connected graph.

Theorem 2.3. *Let $\mathcal{A}$ be a connected ≤ 4-splitting family on k of minimum size. Then for k even, $\mathcal{A}$ is equivalent to the standard splitting family on k, and for k odd, $\mathcal{A}$ is equivalent to the standard splitting family on $k+1$ with the point $k+1$ removed. In particular, $|\mathcal{A}| = \lceil k/2 \rceil$.*

The proof consists of a series of lemmas. In the following, when $\bar{x}$ and $\bar{y}$ are sets of indices, we let $\bar{x}\bar{y}$ denote the union of the two sets.

Lemma 2.4 (Y lemma)**.** *Let S be the Hamming representation of a ≤ 4-splitting family $\mathcal{A}$. Then every element of S has degree at most 2 in S. In fact, S cannot contain an arrangement of the following four "Y" types, where $\bar{x}, \bar{y}, \bar{w}, \bar{u}, \bar{v}$ are pairwise disjoint, and $\bar{u}, \bar{v}, \bar{w}$ are nonempty:*

(a) $\{\delta_{\bar{x}}, \delta_{\bar{x}\bar{y}\bar{u}}, \delta_{\bar{x}\bar{y}\bar{v}}, \delta_{\bar{x}\bar{y}\bar{w}}\}$;

(b) $\{\delta_{\bar{x}}, \delta_{\bar{x}\bar{u}}, \delta_{\bar{x}\bar{u}\bar{v}}, \delta_{\bar{x}\bar{u}\bar{w}}\}$;

(c) $\{\delta_{\bar{x}\bar{u}}, \delta_{\bar{x}\bar{v}}, \delta_{\bar{x}\bar{u}\bar{v}}, \delta_{\bar{x}\bar{u}\bar{v}\bar{w}}\}$;

(d) $\{\delta_{\bar{x}\bar{u}\bar{v}}, \delta_{\bar{x}\bar{u}\bar{w}}, \delta_{\bar{x}\bar{v}\bar{w}}, \delta_{\bar{x}\bar{u}\bar{v}\bar{w}}\}$.
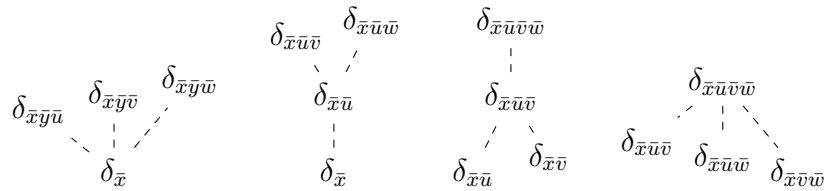
These four arrangements are shown in Figure 2.

Figure 2: Forbidden "Y" arrangements for the Hamming representation of a ≤ 4-splitting family; (a)–(d) appear left–right.

Proof. To prove the arrangements of types (a)–(d) are impossible, it is sufficient to observe that in each type (a)–(d), no index appears exactly twice among the subscripts, as this means that the arrangement is not split.

Now this implies every element of S has degree at most 2. To see this, suppose towards a contradiction that some element $s \in S$ is adjacent in the Hamming hypercube to three other elements $s_1, s_2, s_3 \in S$. We claim that the 4-element set $\{s, s_1, s_2, s_3\}$ is of one of the types (a)–(d). Indeed, each of s_1, s_2, s_3 has weight one higher or one lower than s. If all three are higher the set is of type (a), if two of the three are higher the set is of type (b), if two of the three are lower the set is of type (c), and if all three are lower the set is of type (d). In all cases we obtain a contradiction to the previous paragraph. $\square$
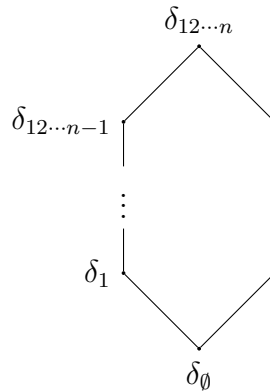
We remark that avoiding the types (a)–(d) does not guarantee a ≤ 4-splitting family. For example $\{\delta_1, \delta_2, \delta_3, \delta_4\}$ is not split and not of any of the types (a)–(d).

Lemma 2.5. *Let S be the Hamming representation of a ≤ 4-splitting family $\mathcal{A}$ with $\emptyset \notin \mathcal{A}$. If S is a simple cycle, then $\mathcal{A}$ is equivalent to the standard splitting family on $k = |S|$.*

Proof. By complementing sets of $\mathcal{A}$, we can assume $\delta_\emptyset \in S$. We first argue that the stratified graph of S has the shape of a standard arrangement, that is, S has a single element of maximum weight, and two elements at each intermediate weight

(see Figure 1 left). Indeed otherwise S must contain elements of the form $\delta_{\bar{x}a}, \delta_{\bar{x}}, \delta_{\bar{x}b}$ (informally, the cycle must have a local minimum at a location other than $\delta_\emptyset$). But then the set $\{\delta_\emptyset, \delta_{\bar{x}}, \delta_{\bar{x}a}, \delta_{\bar{x}b}\}$ is a Y of type (b), a contradiction.

Next we argue that the δ-labeling of the elements of S is equivalent to a standard arrangement up to renaming the sets. To begin we know the bottom element (least weight) is $\delta_\emptyset$ and we can say that the top element (greatest weight) is $\delta_{12\cdots n}$ where n is the total number of sets. Furthermore we can assume without loss of generality that the labels up the left-hand side are $\delta_1, \delta_{12}, \ldots, \delta_{12\cdots n-1}$.

$$
\begin{array}{c}
\delta_{12\cdots n} \\[6pt]
\diagup \quad \diagdown \\[4pt]
\delta_{12\cdots n-1} \mid \qquad \mid \\[4pt]
\vdots \qquad \vdots \\[4pt]
\delta_1 \mid \qquad \mid \\[4pt]
\diagdown \quad \diagup \\[6pt]
\delta_\emptyset
\end{array}
$$

To conclude, we need only show that the labels down the right-hand side are exactly $\delta_{2\cdots n}, \delta_{3\cdots n}, \ldots, \delta_n$. To see this, assume as an inductive hypothesis that the first $i-1$ vertices down the right-hand side are $\delta_{2\cdots n}, \ldots, \delta_{i\cdots n}$, for some $i \geq 1$.

We claim that the remaining vertices down the right-hand side must all omit i. Otherwise we would be able to find two distinct such vertices of the form $\{\delta_{\bar{x}}, \delta_{\bar{x}i}\}$. But then the set $\{\delta_{1\cdots i}, \delta_{\bar{x}}, \delta_{\bar{x}i}, \delta_{i\cdots n}\}$ is not split. This proves the claim.

Now we can conclude that the vertex immediately below $\delta_{i\cdots n}$ must be $\delta_{(i+1)\cdots n}$. This completes the inductive step, and the proof. $\qquad\square$

In the next result, if S is a Hamming representation of an n-set ≤ 4-splitting family $\mathcal{A}$, we say that S is *maximal* if whenever an element of $\{0,1\}^n$ is added to S the result is no longer a Hamming representation of a ≤ 4-splitting family.

Lemma 2.6. *Let S be the Hamming representation of a ≤ 4-splitting family $\mathcal{A}$ with $\emptyset \notin \mathcal{A}$. If S is a simple cycle, then S is maximal.*

Proof. By the previous lemma, it is sufficient to show that the standard arrangement $S = \{\delta_\emptyset, \delta_1, \ldots, \delta_{1\cdots n}, \ldots, \delta_n\}$ is maximal. Now suppose towards a contradiction that there exists $\delta_{\bar{x}} \notin S$ such that $S \cup \{\delta_{\bar{x}}\}$ is a Hamming representation of a ≤ 4-splitting family.

We first claim that either $1 \in \bar{x}$ or $n \in \bar{x}$. Indeed otherwise we would have that the set $\{\delta_\emptyset, \delta_1, \delta_n, \delta_{\bar{x}}\}$ is a Y of type (a). Thus we can suppose without loss of generality that $1 \in \bar{x}$.

Now let $i \in 2, \ldots, n$ be the least index such that $i \notin \bar{x}$. Then the set $\{\delta_{1\cdots i-2}, \delta_{1\cdots i-1}, \delta_{1\cdots i}, \delta_{\bar{x}}\}$ is a Y of type (b), a contradiction which completes the proof. $\qquad\square$

The lemmas so far imply that if S is the Hamming representation of a ≤ 4-splitting family, then S is either a single simple cycle or else a disjoint union of paths. Next we investigate constraints on the paths.

Lemma 2.7. *Let S be the Hamming representation of a ≤ 4-splitting family $\mathcal{A}$ with $\emptyset \notin \mathcal{A}$. If S is a path, then $\mathcal{A}$ is equivalent to the restriction to $k = |S|$ of the standard splitting family on $k' = 2|\mathcal{A}|$.*

Proof. By complementing some of the sets of $\mathcal{A}$, we can suppose without loss of generality that one end of the path S is $\delta_\emptyset$. We will show that S is a subset of the Hamming representation of a standard splitting family. By permuting the indices of the sets, we can suppose that the longest initial segment of increasing weights is $\delta_\emptyset, \delta_1, \ldots, \delta_{1\cdots n}$ for some n. If there are no more elements of S, then we are done.

Otherwise we can suppose that S turns after $\delta_{1\cdots n}$. Then this must be the last turn, since an additional turn would form a vee shape, which together with $\delta_\emptyset$ would form a Y of type (b). Now the inductive argument from Lemma 2.5 implies that the remaining points of S after $\delta_{1\cdots n}$ must be $\delta_{2\cdots n}, \ldots, \delta_{i\cdots n}$. It follows that $n = |\mathcal{A}|$, and that S is the restriction of the standard splitting family on n. $\square$

We are now ready to conclude the proof of the theorem.

Proof of Theorem 2.3. The lemmas together imply that $\mathcal{A}$ is either equivalent to the standard splitting family on k or the restriction to k of the standard splitting family on $k' = 2|\mathcal{A}|$. Since $\mathcal{A}$ is of minimum size, $|\mathcal{A}| \leq \lceil k/2 \rceil$, which implies that $k' = k$ or $k' = k + 1$ (if k is odd). $\square$

3   Splitting subsets of size 2 and 4

Recall from the previous section that $\mathcal{A}$ is a $\leq t$-splitting family on k if $\mathcal{A}$ splits all subsets of $[k]$ of size at most t. We will also say that $\mathcal{A}$ is a *t-splitting family* on k if $\mathcal{A}$ splits all subsets of $[k]$ of size exactly t. In [14], t-splitting families are used in an algorithm to solve the weight t discrete log problem. In this section we study the least size of a 4-splitting family and the least size of a ≤ 4-splitting family on k.

Beginning with 4-splitting families, the article [13] established a lower bound on the least size of a *uniform* 4-splitting family (that is, a 4-splitting family in which every set has size $\lfloor k/2 \rfloor$). Later, [7] stated that this lower bound holds for arbitrary 4-splitting families, citing the aforementioned article as proof. In the following result we repair this minor error and prove that the claimed lower bound does indeed hold for arbitrary 4-splitting families.

Theorem 3.1. *Let $k \geq 6$ and suppose $\mathcal{A}$ is a 4-splitting family on k. Then $|\mathcal{A}| \geq \log_2 k$.*

In the proof and later in the paper, we will use the following terminology. Let $\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$ be a family of subsets of $[k]$. For any $I \subset \{1, 2, \ldots, n\}$, the

Venn region corresponding to I is the set

$$\bigcap\{A_i \mid i \in I\} \cap \bigcap\{A_i^c \mid i \notin I\} \,.$$

The *multiplicity* of the Venn region corresponding to I, or any of its elements, is $|I|$. We will say two Venn regions corresponding to I, I' are *adjacent* if $|I \triangle I'| = 1$.

Proof. Assume towards a contradiction that $n < \log_2 k$ so that $k \geq 2^n + 1$.

We consider the subfamily $\mathcal{A}_0 = \{A_1, A_2, \ldots A_{n-1}\}$. Then $\mathcal{A}_0$ has 2^{n-1} Venn regions, and since $k > 2^n = 2 \cdot 2^{n-1}$ we conclude that some Venn region R of $\mathcal{A}_0$ contains at least three points x, y, z.

We can replace A_n with its complement if necessary, so we can assume without loss of generality that A_n contains at least two of x, y, z, say x and y. If $w \in [k]$ is any other point, consider the set $\{x, y, z, w\}$. $\mathcal{A}_0$ does not split it since $x, y, z \in R$. Therefore, since $\mathcal{A}$ is a 4-splitting family, A_n must split $\{x, y, z, w\}$, implying $z, w \notin A_n$. Since w was chosen arbitrarily, it follows that $A_n = \{x, y\}$. The $k - 3 \geq 2^n - 2$ points other than x, y, z are distributed among the 2^{n-1} Venn regions of $\mathcal{A}_0$; we consider several cases on how these points are distributed.

<u>Case 1:</u> There is a point in R besides x, y, z.

Let w be such a point. We know $w \notin A_n$. Since $k \geq 6$, we can find a fifth point w', and again $w' \notin A_n$. Then $\{x, z, w, w'\}$ is not split by A_n since A_n only contains x. Further, $\{x, z, w, w'\}$ is not split by any set from $\mathcal{A}_0$ since x, z, w are in the same Venn region of $\mathcal{A}_0$. Thus, $\{x, z, w, w'\}$ is not split by any set from $\mathcal{A}$, a contradiction.

<u>Case 2:</u> There are no points in R besides x, y, z and some Venn region of $\mathcal{A}_0$ other than R has at least three points.

Let R' be such a region and let $w, w', w'' \in R'$. Then $\{z, w, w', w''\}$ is not split by any set from $\mathcal{A}_0$. Moreover A_n contains none of these four points, so $\{z, w, w', w''\}$ is not split by any set from $\mathcal{A}$, a contradiction.

<u>Case 3:</u> There are no points in R besides x, y, z and no Venn region of $\mathcal{A}_0$ other than R has at least three points.

The $2^n - 2$ points other than x, y, z are distributed among the $2^{n-1} - 1$ Venn regions of $\mathcal{A}_0$ other than R. In this case, since $2^n - 2 = 2 \cdot (2^{n-1} - 1)$, every such region must have exactly two points. Choose distinct Venn regions R' and R'' which are adjacent to R. (This is possible because the assumption $k \geq 6$ implies $|\mathcal{A}| \geq 3$, so $|\mathcal{A}_0| \geq 2$.) Letting $w' \in R'$ and $w'' \in R''$, we have that $\{x, z, w', w''\}$ is not split by any set from $\mathcal{A}_0$. Moreover $\{x, z, w', w''\}$ is not split by $A_n = \{x, y\}$. Once again $\{x, z, w', w''\}$ is not split by any set from $\mathcal{A}$, a contradiction.

This completes the proof. $\qquad\square$

We now turn to ≤ 4-splitting families.

Theorem 3.2. *Let $k \geq 5$ and suppose $\mathcal{A}$ is a ≤ 4-splitting family on k. Then $|\mathcal{A}| \geq \log_2 k + 3 - \log_2 5$.*

Proof. Suppose $\mathcal{A} = \{A_1, \ldots, A_n\}$ and this time let $\mathcal{A}_0 = \{A_1, \ldots, A_{n-2}\}$ and $\mathcal{A}_1 = \{A_{n-1}, A_n\}$.

We first claim that no Venn region of $\mathcal{A}_0$ contains four points. Indeed if there were such x, y, z, w, then since $\mathcal{A}$ is 2-splitting we would have that each of x, y, z, w lie in distinct Venn regions of $\mathcal{A}_1$. Since we assumed $k \geq 5$, we can let u be any other point. Without loss of generality, u is in the same Venn region of $\mathcal{A}_1$ as x. But then the 4-element set $\{x, y, z, u\}$ is not split, a contradiction.

We next claim there cannot exist Venn regions R_1, R_2, R_3 of $\mathcal{A}_0$, each with three points, such that $R_1 \sim R_2 \sim R_3$. Indeed, otherwise we would be able to find x_i in each R_i such that all three x_i are in the same Venn region of $\mathcal{A}_1$. (This holds because three 3-element subsets of a 4-element set must have a point in common.) Suppose without loss of generality that $x_1, x_2, x_3 \in A_{n-1} \cap A_n$. Let y_2 be another element of R_2 which is in a Venn region of $\mathcal{A}_1$ adjacent to that of x_2. Then $\{x_1, x_2, y_2, x_3\}$ is not split by any set of $\mathcal{A}$. Indeed A_{n-1} and A_n contain three of these points, and $A_1, \ldots, A_{n-2}$ each agree on x_2, y_2 and at least one of x_1, x_3.

Now let S be the set of Venn regions of $\mathcal{A}_0$ with three points. We claim that $|S| \leq 2^{n-3}$. For this, observe that the adjacency graph of the 2^{n-2} Venn regions of $\mathcal{A}_0$ is isomorphic to a hypercube and thus may be decomposed into 2^{n-4} disjoint squares. If $|S| > 2^{n-3} = 2 \cdot 2^{n-4}$ then one of these squares must contain three elements of S. It follows that we can find three elements $R_1, R_2, R_3 \in S$ such that $R_1 \sim R_2 \sim R_3$. This contradicts the previous claim.

We now have that the number of elements k satisfies:

$$k \leq 3|S| + 2|S^c| = 2(|S| + |S^c|) + |S| \leq 2 \cdot 2^{n-2} + 2^{n-3} = 5 \cdot 2^{n-3}.$$

This implies the desired bound. $\qquad\square$

It is remarkable that the lower bounds in the previous two theorems are so close to one another. The distance between them is more visible if one considers the inverse functions, that is, the corresponding upper bounds on the k for which there exists a 4- or ≤ 4-splitting family on k of a given size. Of course, these bounds are unlikely to be tight.

4   Splittable families

Recall from the introduction that a family $\mathcal{B} = \{B_1, \ldots, B_n\}$ of subsets of $[k]$ is called *splittable* if there exists a single set $A \subset [k]$ such that A splits B_i for all $i \leq n$. While some splittable families have many distinct splitters, others are "just barely splittable" in the sense that they have very few splitters. In this section we investigate the question: which n-set splittable family on k has the fewest number of splitters, and what is this number? Solutions to this question can be used to find bounds on the least size of an "n-splitting family" (see [6] for details). Our solution is somewhat involved even when $n = 1, 2$, and we provide computational results when $n = 3$.

We first consider the case when $n = 1$, that is, $\mathcal{B} = \{B\}$ and $B \subset [k]$. If $|B|$ is even then the number of splitters of $\mathcal{B}$ is

$$2^{k-|B|} \cdot \binom{|B|}{|B|/2},$$

and if $|B|$ is odd then the number of splitters of $\mathcal{B}$ is

$$2^{k-|B|} \cdot \left(\binom{|B|}{|B|/2 + 1} + \binom{|B|}{|B|/2 - 1} \right) = 2^{k-|B|+1} \binom{|B|}{(|B|-1)/2}.$$

The following result is unsurprising, and we record the proof.

Proposition 4.1. *The number of splitters of $\mathcal{B} = \{B\}$ on $[k]$ is minimized when $|B| = k$ if k is even, and when $|B| = k - 1$ when k is odd. The minimum number of splitters is asymptotic to $2^k/\sqrt{k}$.*

Proof. We first claim that if $|B| = k - i$ is odd (and nonempty), then removing one point from B decreases the number of splitters. Indeed the number of splitters of B is

$$2^{i+1} \binom{k-i}{(k-i-1)/2},$$

and the number of splitters of B with a point removed is

$$2^{i+1} \binom{k-i-1}{(k-i-1)/2}.$$

It is not difficult to calculate that the latter expression is smaller than the first whenever $k - i \geq 3$. (Note that when $k - i = 1$, the two expressions are equal.)

In particular the number of splitters is minimized when $|B|$ is even. We next claim that if $|B| = k - i$ is even (and nonempty), then removing two points from B increases the number of splitters, that is:

$$2^i \binom{k-i}{(k-i)/2} < 2^{i+2} \binom{k-i-2}{(k-i-2)/2}.$$

This comes from applying Pascal's identity:

$$
\begin{aligned}
2^i \binom{k-i}{\frac{k-i}{2}} &= 2^i \left(\binom{k-i-2}{\frac{k-i-2}{2}-1} + 2\binom{k-i-2}{\frac{k-i-2}{2}} + \binom{k-i-2}{\frac{k-i-2}{2}+1} \right) \\
&< 2^{i+2} \binom{k-i-2}{\frac{k-i-2}{2}}.
\end{aligned}
$$

The desired result follows.

For the second statement, we now see that when k is even the minimum number of splitters of a 1-set family on k is exactly $\binom{k}{k/2}$. It is a standard application of Stirling's approximations to conclude that this is asymptotic to $2^k/\sqrt{k}$. $\qquad \square$

We now turn to the case when $n = 2$. We begin by fixing some notation. Given a family $\mathcal{B} = \{B_1, B_2\}$, we say the *arrangement* of $\mathcal{B}$ is the quadruple (a_1, b, a_2, d), where $a_1 = |B_1 \backslash B_2|$, $b = |B_1 \cap B_2|$, $a_2 = |B_2 \backslash B_1|$, and $d = |B_1^c \cap B_2^c|$ (see Figure 3). The number of splitters of $\mathcal{B}$ is determined by its arrangement, and we will often use the family and the arrangement interchangeably. We let $\mathrm{split}(\mathcal{B})$ and $\mathrm{split}(a_1, b, a_2, d)$ both denote the number of splitters of $\mathcal{B}$.
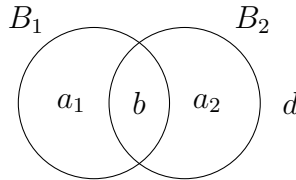

Figure 3: A family $\mathcal{B} = \{B_1, B_2\}$ with arrangement (a_1, b, a_2, d).

We now proceed to calculate a formula for $\mathrm{split}(a_1, b, a_2, d)$. Initially suppose that $|B_1| = a_1 + b$ and $|B_2| = a_2 + b$ are both even. Then if A contains i elements from $B_1 \cap B_2$, then A is a splitter if and only if A contains exactly $\frac{a_1+b}{2} - i$ elements from $B_1 \backslash B_2$ and $\frac{a_2+b}{2} - i$ elements from $B_2 \backslash B_1$. Thus the total number of splitters of $\mathcal{B}$ is given by:

$$\mathrm{split}(a_1, b, a_2, d) = 2^d \sum_{i=0}^{b} \binom{a_1}{\frac{a_1+b}{2} - i} \binom{b}{i} \binom{a_2}{\frac{a_2+b}{2} - i}.$$

(Here if $z \in \mathbb{N}$ and $x \neq 0, \ldots, z$ we define $\binom{z}{x} = 0$.)

Note that if $|B_1| = a_1 + b$ is odd and $|B_2| = a_2 + b$ is even, then

$$\mathrm{split}(a_1, b, a_2, d) = 2^d \sum_{\epsilon=0}^{1} \sum_{i=0}^{b} \binom{a_1}{\lfloor \frac{a_1+b}{2} \rfloor + \epsilon - i} \binom{b}{i} \binom{a_2}{\frac{a_2+b}{2} - i}$$

$$= 2^d \sum_{i=0}^{b} \binom{a_1 + 1}{\frac{(a_1+1)+b}{2} - i} \binom{b}{i} \binom{a_2}{\frac{a_2+b}{2} - i}$$

$$= \mathrm{split}(a_1 + 1, b, a_2, d).$$

Similarly if both $|B_1| = a_1 + b$ and $|B_2| = a_2 + b$ are odd, we have

$$\mathrm{split}(a_1, b, a_2, d) = \mathrm{split}(a_1 + 1, b, a_2 + 1, d).$$

We are now ready to give the characterization of 2-set families with the minimum number of splitters.

Theorem 4.2. *Let $k = a_1 + b + a_2 + d$ be fixed. Then $\mathrm{split}(a_1, b, a_2, d)$ is minimized by the following arrangements:*

- *if $k \equiv 0 \pmod 3$, minimum at $(m, m, m, 0)$, where $m = \frac{k}{3}$;*

- *if $k \equiv 1 \pmod 3$, minimum at $(m - 2, m, m, 0)$, where $m = \frac{k+2}{3}$, and;*

   ○ *if $k \equiv 2$ (mod 3), minimum at $(m, m, m + 2, 0)$, where $m = \frac{k-2}{3}$.*

Moreover, the minimum number is asymptotic to $2^k/k$.

The proof consists of a series of "point-moving" lemmas. Each one shows that given an arbitrary arrangement, we can find a way to nudge it towards the proposed minimum arrangement without increasing the number of splitters.

Lemma 4.3. *If $a_1 + b$ and $a_2 + b$ are even and $d \geq 0$, then for any permutation (a_1', b', a_2') of (a_1, b, a_2), we have $\mathrm{split}(a_1, b, a_2, d) = \mathrm{split}(a_1', b', a_2', d)$.*

Proof. It suffices to consider the case when $d = 0$, since otherwise both sides of the inequality simply have an extra factor of 2^d. Further it suffices to show that $\mathrm{split}(a_1, b, a_2, 0) = \mathrm{split}(b, a_1, a_2, 0)$, since by symmetry $\mathrm{split}(a_1, b, a_2, 0) = \mathrm{split}(a_2, b, a_1, 0)$, and all other permutations are generated by these two.

To show that $\mathrm{split}(a_1, b, a_2, 0) = \mathrm{split}(b, a_1, a_2, 0)$, fix a family $\mathcal{B} = \{B_1, B_2\}$ with arrangement $(a_1, b, a_2, 0)$ and let $\mathcal{B}' = \{B_1, B_2'\}$, where $B_2' = B_1 \triangle B_2$, so that $\mathcal{B}'$ has arrangement $(b, a_1, a_2, 0)$. Next define a mapping of sets $A \mapsto A'$ by $A' = (A \cap B_1) \cup ((B_2 \backslash B_1) \backslash A)$. We claim that $A \mapsto A'$ is an injection from the splitters of $\mathcal{B}$ to the splitters of $\mathcal{B}'$.

We first argue that if A splits $\mathcal{B}$ then A' splits $\mathcal{B}'$. Since A splits B_1 and $A' \cap B_1 = A \cap B_1$, A' splits B_1 as well. By the assumption that $a \equiv b \equiv c$ (mod 2), all of B_1, B_2, B_2' are even-sized. Now if A contains i elements from $B_1 \cap B_2$, then A contains $\frac{a+b}{2} - i$ elements from $B_1 \backslash B_2$ and $\frac{b+c}{2} - i$ elements from $B_2 \backslash B_1$. Thus A' contains i elements from $B_1 \backslash B_2'$ and $\frac{a+b}{2} - i$ elements from $B_1 \cap B_2'$ and $c - (\frac{b+c}{2} - i) = \frac{c-b}{2} + i$ elements from $B_2' \backslash B_1$. Thus A' contains exactly $\frac{a+c}{2}$ elements from B_2', so A' splits $\mathcal{B}'$.

Finally note that since $d = 0$, we can recover A from A' by the formula $A = (A' \cap B_1) \cup ((B_2 \backslash B_1) \backslash A')$. This shows that $A \mapsto A'$ is injective, so $\mathrm{split}(\mathcal{B}) \leq \mathrm{split}(\mathcal{B}')$. Moreover the symmetry of the inverse mapping shows that $\mathrm{split}(\mathcal{B}') \leq \mathrm{split}(\mathcal{B})$, concluding the proof. □

Recall that the definition of splitting allows rounding up or down in the case of odd-sized sets. Thus if a family has an odd-sized set, one would expect it to have more splitters than a similar family with even-sized sets only. The following result confirms this intuition.

Lemma 4.4. *Suppose that $a_1 + b$ and $a_2 + b$ are even, and let $\epsilon_1, \epsilon_2 \in \{-1, 0, 1\}$. Then*
$$\mathrm{split}(a_1, b, a_2, 0) \leq \mathrm{split}(a_1 + \epsilon_1, b - \epsilon_1 - \epsilon_2, a_2 + \epsilon_2, 0).$$

Proof. Let $\mathcal{B} = \{B_1, B_2\}$ be a family with arrangement $(a_1, b, a_2, 0)$, and let $\mathcal{B}' = \{B_1', B_2'\}$ be the family with arrangement $(a_1 + \epsilon_1, b - \epsilon_1 - \epsilon_2, a_2 + \epsilon_2, 0)$ that is derived from $\mathcal{B}$ according to the ϵ_i in the natural way. (For example, if $\epsilon_1 = 1$ and $\epsilon_2 = 0$, we can construct B' from B by selecting some point $x \in B_1 \cap B_2$ and placing it in $B_1' \backslash B_2'$.) It is easy to see that if A splits $\mathcal{B}$, then A splits $\mathcal{B}'$ too, completing the proof. □

The next result allows one to move points from outside the sets B_1, B_2 into the sets without increasing the number of splitters.

Lemma 4.5. *For fixed k, for any arrangement (a_1, b, a_2, d) on k, there exists an arrangement $(a_1', b', a_2', 0)$ on k such that*

$$\mathrm{split}(a_1', b', a_2', 0) \leq \mathrm{split}(a_1, b, a_2, d).$$

Proof. By Lemma 4.4, we need only consider the case when $a_1 + b$ and $a_2 + b$ are both even. By Lemma 4.3 we can suppose that $b > 0$ (otherwise $d = k$ and this is clearly not minimal). We now use the result [6, Theorem 3.11], which states that if $b, d > 0$ then we have:

$$\mathrm{split}(a_1 + 1, b - 1, a_2 + 1, d - 1) \leq \mathrm{split}(a_1, b, a_2, d).$$

Note that the arrangement on the left-hand side again satisfies the hypotheses of Lemma 4.3. Thus we can repeat the process inductively to obtain the desired conclusion. $\qquad\square$

The next result is the key, as it allows one to move 2 points from one of the sets to the other without increasing the number of splitters, provided it would make the regions more balanced in size.

Lemma 4.6. *Assume that $a_1 + b$ and $a_2 + b$ are even. If $2 \leq a_1 \leq a_2$ then*

$$\mathrm{split}(a_1, b, a_2, 0) \leq \mathrm{split}(a_1 - 2, b, a_2 + 2, 0).$$

Proof. Let $\mathcal{B} = \{B_1, B_2\}$ be a family with arrangement $(a_1, b, a_2, 0)$ and $\mathcal{B}' = \{B_1', B_2'\}$ be a family with arrangement $(a_1 - 2, b, a_2 + 2, 0)$. We can again assume $\mathcal{B}'$ is constructed from $\mathcal{B}$ in the natural way, that is, fix $x, y \in B_1 \backslash B_2$ and let $B_1' = B_1 \setminus \{x, y\}$ and $B_2' = B_2 \cup \{x, y\}$. We wish to show that there exists an injection from the splitters of $\mathcal{B}$ to the splitters of $\mathcal{B}'$. Note immediately that if A is a splitter of $\mathcal{B}$ that contains either x or y, but not both, then A also splits $\mathcal{B}'$. Thus it remains to show that there are fewer splitters of $\mathcal{B}$ that contain (omit) both x, y than splitters of $\mathcal{B}'$ that contain (omit) both x, y.

We address the "omit" case with the "contain" case being similar. Let S be the number of splitters of $\mathcal{B}$ that omit x, y, and S' be the number of splitters of $\mathcal{B}'$ that omit x, y. We wish to show that $S \leq S'$. To proceed, let us first assume that b is even. Let $t_i = \frac{a_i + b}{2}$, the *target* number of elements of B_i for a splitter A of $\mathcal{B}$. We calculate:

$$S' - S = \sum_{i=0}^{b} \binom{a_1 - 2}{t_1 - i - 1}\binom{b}{i}\binom{a_2}{t_2 - i + 1} - \binom{a_1 - 2}{t_1 - i}\binom{b}{i}\binom{a_2}{t_2 - i}$$

$$= \sum_{i=0}^{b/2} \binom{a_1 - 2}{t_1 - i - 1}\binom{b}{i}\binom{a_2}{t_2 - i + 1} - \binom{a_1 - 2}{t_1 - i}\binom{b}{i}\binom{a_2}{t_2 - i}$$

$$+ \binom{a_1 - 2}{t_1 - (b+1-i)}\binom{b}{b+1-i}\binom{a_2}{t_2 - (b+1-i)+1}$$

$$- \binom{a_1 - 2}{t_1 - (b+1-i)}\binom{b}{b+1-i}\binom{a_2}{t_2 - (b+1-i)}$$

$$= \sum_{i=0}^{b/2} \binom{b}{i}\left(\binom{a_1 - 2}{t_1 - i - 1}\binom{a_2}{t_2 - i + 1} - \binom{a_1 - 2}{t_1 - i}\binom{a_2}{t_2 - i}\right)$$

$$+ \binom{b}{i-1}\left(\binom{a_1 - 2}{t_1 - i}\binom{a_2}{t_2 - i} - \binom{a_1 - 2}{t_1 - i - 1}\binom{a_2}{t_2 - i + 1}\right)$$

$$= \sum_{i=0}^{b/2} \left(\binom{b}{i} - \binom{b}{i-1}\right)\left(\binom{a_1 - 2}{t_1 - i - 1}\binom{a_2}{t_2 - i + 1} - \binom{a_1 - 2}{t_1 - i}\binom{a_2}{t_2 - i}\right)$$

The second line above was the key step, wherein we paired term i with term $j = b + 1 - i$. In the third equality we reflect five binomial coefficients and observe that $a_i - t_i = t_i - b$.

Clearly, $\binom{b}{i} - \binom{b}{i-1}$ is positive for each $i \leq b/2$. Thus it is enough to show that for all $i \leq b/2$ we have

$$\binom{a_1 - 2}{t_1 - i - 1}\binom{a_2}{t_2 - i + 1} - \binom{a_1 - 2}{t_1 - i}\binom{a_2}{t_2 - i} \geq 0. \tag{1}$$

Note that if the second term of inequality (1) is zero, we are done. On the other hand if the second term of inequality (1) is nonzero, then we claim that the first term is nonzero as well. To see this, the indices where the first term is nonzero are $[t_1 - a_1 + 1, t_1 - 1] \cap [t_2 - a_2 + 1, t_2 + 1] = [t_1 - a_1 + 1, t_1 - 1]$, and the indices where the second term is nonzero are $[t_1 - a_1 + 2, t_1] \cap [t_2 - a_2, t_2] = [t_1 - a_1 + 2, t_1]$. (We are using here that $a_1 \leq a_2$, $t_1 \leq t_2$, and $t_2 - a_2 \leq t_1 - a_1$.) The only index in the latter set but not the former is $i = t_1$, but the hypothesis $2 \leq a_1$ implies $b/2 < t_1$, and so this index is not in the sum.

Now for indices i such that the terms of inequality (1) are nonzero, we have:

$$\frac{\binom{a_1 - 2}{t_1 - i - 1}\binom{a_2}{t_2 - i + 1}}{\binom{a_1 - 2}{t_1 - i}\binom{a_2}{t_2 - i}} = \frac{(t_1 - i)(a_2 - t_2 + i)}{(a_1 - t_1 + i - 1)(t_2 - i + 1)} = \frac{(t_1 - i)(t_2 - b + i)}{(t_1 - b + i - 1)(t_2 - i + 1)}. \tag{2}$$

It suffices to show that the last quantity in equation (2) is > 1. Recall that the Mediant Inequality states that if $A, B, C, D > 0$ and $\frac{A}{B} > \frac{C}{D}$ then $\frac{A}{B} > \frac{A+C}{B+D} > \frac{C}{D}$. Since $t_1 \leq t_2$ we can write:

$$\frac{t_2 - b + i}{t_1 - b + i - 1} > 1 = \frac{b - 2i + 1}{b - 2i + 1}.$$

Since we are assuming the numerator and denominator of the left-hand side are positive, and since $b - 2i + 1$ is positive as well, the Mediant Identity implies that

$$\frac{t_2 - b + i}{t_1 - b + i - 1} > \frac{t_2 - i + 1}{t_1 - i} > \frac{b - 2i + 1}{b - 2i + 1}.$$

This implies that the last quantity in Equation (2) is greater than 1, as desired.

If b is odd then the calculation of $S' - S$ is similar. We again pair the i and $b + 1 - i$ terms. Of course the $i = (b + 1)/2$ term has nothing to pair with, but fortunately it cancels out completely. To see this note that the left and right terms are reflections of one another, since

$$(t_1 - (b+1)/2 - 1) + (t_1 - (b+1)/2) = a_1 - 2, \text{ and}$$
$$(t_2 - (b+1)/2 + 1) + (t_2 - (b+1)/2) = a_2.$$

Thus in the second line we can take the sum from $i = 1$ up to $i = (b - 1)/2$. The rest of the proof is the same as before. □

The combination of Lemmas 4.3 and 4.6 means that (when $d = 0$) moving two points from any region to another region with fewer points does not increase the number of splitters. We are now ready to prove our main result.

Proof of Theorem 4.2. We will give the proof when $k = 3m$, the remaining cases are similar. By the lemmas together, it suffices to show that for any even $x, y \geq 0$ we have:

$$\text{split}(m, m, m, 0) \leq \text{split}(m + x, m - x + y, m - y, 0) .$$

We show this in two steps, with each step consisting of several applications of Lemma 4.6 and 4.3. Assuming $x \leq y$, first use the two lemmas to achieve:

$$\text{split}(m, m, m, 0) \leq \text{split}(m + x, m, m - x, 0).$$

We then use the two lemmas again to achieve:

$$\text{split}(m + x, m, m - x, 0) \leq \text{split}(m + x, m - x + y, m - y, 0).$$

We refer to Figure 4 for a visualization of these two inequalities. The steps are similar if $y \leq x$. This concludes the proof that $\text{split}(m, m, m, 0)$ is minimal.

For the last statement, observe that $\text{split}(k/3, k/3, k/3, 0)$ is exactly equal to $\sum_i \binom{k/3}{i}^3$. These are *Franel numbers*, and it is not hard to see this expression is asymptotic to $2^k/k$. (See for instance [9].) □

While we kept it out of the statement for cleanliness, we record here that the proof of Theorem 4.2 may be used to classify all minimum arrangements. First observe that Lemma 4.3 implies there are two additional minimum arrangements in the cases when $k \equiv 1, 2 \pmod 3$. That these are exhaustive may be seen from the strict inequality obtained in the end of the proof of Lemma 4.6.

Before leaving the case when there are $n = 2$ sets, we mention an approximate formula for the value of $\text{split}(a_1, b, a_2, d)$. Recall that the de Moivre–Laplace Theorem [10, Section 7.3] gives the approximation

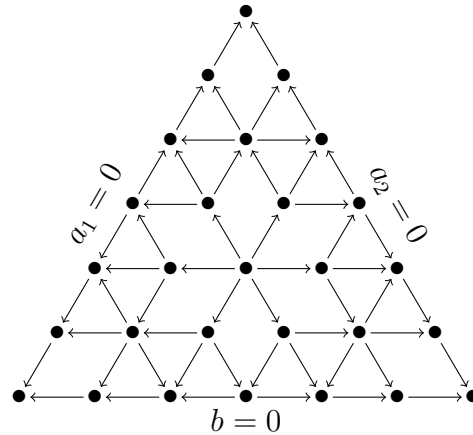$$\binom{n}{k} \approx 2^n \sqrt{\frac{2}{\pi n}} e^{-\frac{2}{n}\left(k - \frac{n}{2}\right)^2}.$$

Figure 4: The lattice of arrangements $(a_1, b, a_2, 0)$ with $k = 12$ and $a_1 + b, a_2 + b$ even. An arrow $\alpha \to \beta$ denotes α has fewer or equal splitters than β. The proof of Theorem 4.2 uses the fact that one can follow the arrows from the centroid to any other point using two straight-line steps.

Thus when $a_1 \equiv b \equiv a_2 \pmod 2$ we can approximate the number of splitters by

$$\text{split}(a_1, b, a_2, d) = 2^d \sum_{i=0}^{b} \binom{b}{i} \binom{a_1}{t_1 - i} \binom{a_2}{t_2 - i}$$

$$\approx 2^{a_1+b+a_2+d} \sqrt{\frac{8}{\pi^3 a_1 b a_2}} \int_{-\infty}^{\infty} e^{-\frac{2}{b}\left(x - \frac{b}{2}\right)^2 - \frac{2}{a_1}\left(\frac{b}{2} - x\right)^2 - \frac{2}{a_2}\left(\frac{b}{2} - x\right)^2} dx$$

$$= \frac{2^{k+1}}{\pi \sqrt{a_1 a_2 + a_1 b + a_2 b}}.$$

It is not difficult to calculate that the latter expression has its minimum when $a_1 = b = a_2 = k/3$ and $d = 0$. However without very tight control over the error in the approximation, it would not be possible to use this information to replace the proof of the previous theorem.

We close this section by briefly considering the case when there are $n = 3$ sets. The formula for the number of splitters when $n = 3$ is considerably more complex than the $n = 2$ case. Using an exhaustive search elaborated in Appendix 5, we found that if N_k denotes the minimum number of splitters of a splittable 3-set family, then N_k appears to obey $N_6 = 4$ and

$$\frac{N_{k+1}}{N_k} = \begin{cases} 2 - \frac{1}{\lfloor k/6 \rfloor + 1}, & k \text{ even} \\ 2, & k \text{ odd} \end{cases}$$

It is not difficult to show that the recurrence above implies that N_k is asymptotic to $2^k/k^{3/2}$. Together with the known asymptotics for the case $n = 2$, this supports the following.

Conjecture. The minimum number of splitters of a splittable n-set family is asymptotic to
$$2^k/k^{n/2} .$$

5   The splitting game

In this section we again look for arrangements that are in some sense the least splittable. This time, rather than considering the number of splitters an arrangement has, we instead consider whether the arrangement can be split under adversarial conditions. To do this, we adopt a game theoretic perspective on the notion of set splitting.

The *splitting game* is played by players Split and Skew on a game board $(k, \mathcal{B})$, where k is a positive integer and $\mathcal{B}$ is a family of subsets of $[k]$. An instance $(k, \mathcal{B}, t)$ of the game consists of a game board together with $t \in \{\text{Split}, \text{Skew}\}$ indicating which player goes first. The players alternately claim an element from $[k]$, without repetition, until all the elements have been claimed. Split wins the game if the set of elements Split claimed splits $\mathcal{B}$ (*i.e.*, splits every $B \in \mathcal{B}$), and Skew wins otherwise.

The splitting game lies in the general class of games known as Maker–Breaker games. Introduced by Paul Erdős and John Selfridge, such games consist of two players choosing objects with Maker trying to occupy some winning arrangement and Breaker trying to prevent Maker's success [2]. Since such games are finite and of perfect information, Zermelo's theorem implies that in each instance of the game, one of the two players must have a winning strategy.

It is clear that if Split has a winning strategy in $(k, \mathcal{B}, t)$, then $\mathcal{B}$ is splittable. However, the converse is not true. For example, consider a game on $k = 3$ consisting of the sets $\mathcal{B} = \{\{1, 2\}, \{2, 3\}\}$. Note that $\mathcal{B}$ is splittable with two splitters: $\{1, 3\}$ and $\{2\}$. Therefore, in the game $(3, \mathcal{B}, \text{Skew})$, if Skew first claims 2, then Split can claim 1 or 3 but not both, resulting in a victory for Skew.

We begin with some general observations on the splitting game.

Lemma 5.1 (Reduction lemma). *Let* $(k, \mathcal{B}), (k', \mathcal{B}')$ *be game boards with* $|\mathcal{B}| = |\mathcal{B}'|$, *and suppose* $|R| \equiv |R'| \pmod 2$ *for corresponding Venn regions* R, R' *of* $\mathcal{B}, \mathcal{B}'$. *Then Player* p *has a winning strategy in* $(k, \mathcal{B}, t)$ *if and only if* p *has a winning strategy in* $(k', \mathcal{B}', t)$.

Proof. It is enough to consider the case when $\mathcal{B}'$ is formed by adding a pair of points to some Venn region of $\mathcal{B}$. One can then use this case repeatedly, adding or removing pairs of points, to obtain the full statement. So assume that $k' = k + 2$, and that $\mathcal{B}'$ is formed by adding the points $k + 1, k + 2$ to some subset of the sets in $\mathcal{B}$.

Let σ be a winning strategy for Player p in $(k, \mathcal{B}, t)$ and let Player q be the other player. We define a strategy for p in $(k', \mathcal{B}', t)$ as follows. If q claims one of $k + 1$ or $k + 2$, then p claims the other immediately. Otherwise p plays according to σ. If every element of $[k]$ is claimed before this happens, and it is p's turn to move, then p takes either one, and q is forced to take the other.

Let A be the set of elements claimed by Player p after a run where p played according to this strategy. Since p followed σ, considering only turns involving $[k]$, A splits or skews $\mathcal{B}$ as desired by p. Moreover since A contains exactly one of $k + 1$ and $k + 2$, it follows that A remains a splitter or a skewer, as the case may be, of $\mathcal{B}'$.

In any case, p wins $(k', \mathcal{B}', t)$.

Conversely, if Player p does not have a winning strategy in $(k, \mathcal{B}, t)$, then by Zermelo's Theorem, q must have a winning strategy. By the forward implication, q then has a winning strategy in $(k', \mathcal{B}', t)$. Thus p does not have winning strategy in $(k', \mathcal{B}', t)$ if p does not have a winning strategy in $(k, \mathcal{B}, t)$. $\qquad\square$

For the next result, if $\mathcal{B}$ is a family of subsets of $[k]$, we say that Split has a *pairing strategy* for $\mathcal{B}$ if there exists a set of disjoint pairs $\mathcal{P} = \{P_1, \ldots, P_r\}$ of elements of $[k]$ such that every splitter of $\mathcal{P}$ splits $\mathcal{B}$. Similarly we say that Skew has a pairing strategy if every splitter of $\mathcal{P}$ does not split $\mathcal{B}$; that is, if $\mathcal{P} \cup \mathcal{B}$ is unsplittable.

Lemma 5.2 (Pairing lemma). *Let $\mathcal{B}$ be a family of subsets of $[k]$. If Player p has a pairing strategy for $\mathcal{B}$, then p has a winning strategy for $(\ell, \mathcal{B}, t)$ for either value of t and for each $\ell \geq k$.*

Proof. Suppose Split has a pairing strategy given by $\mathcal{P} = \{P_1, \ldots, P_r\}$, and consider the side game $(\ell, \mathcal{P}, t)$. By Lemma 5.1, Split has a winning strategy for $(\ell, \mathcal{P}, t)$ if and only if Split has a winning strategy in a game in which all sets are empty, which is clearly the case. Since any set that splits $\mathcal{P}$ also splits $\mathcal{B}$, this yields a winning strategy for $(\ell, \mathcal{B}, t)$.

If Skew has a pairing strategy, we can use the same argument, but note that Skew will play the role of Split in the side game. $\qquad\square$

In the rest of the section, we carry out case studies, analyzing which player wins in several special cases of the splitting game. We begin with a very trivial such case.

Theorem 5.3. *Suppose that $(k, \mathcal{B})$ has at most two odd-sized Venn regions of nonzero multiplicity. Then for either value of t, Split has a winning strategy in $(k, \mathcal{B}, t)$.*

Proof. By Lemma 5.1, the game board $(k, \mathcal{B})$ is equivalent to a game board $(\ell, \mathcal{C})$ in which $\left|\bigcup_{C \in \mathcal{C}} C\right| \leq 2$. If $\left|\bigcup_{C \in \mathcal{C}} C\right| \leq 1$, then every subset of $[\ell]$ splits $\mathcal{C}$, so any run of the splitting game results in a Split win. If $\left|\bigcup_{C \in \mathcal{C}} C\right| = 2$, then Split has a pairing strategy given by $\left\{\bigcup_{C \in \mathcal{C}} C\right\}$. $\qquad\square$

Next we proceed to study games where the number of sets $n = |\mathcal{B}|$ is small. By Lemma 5.1, one can reduce any game to one in which each Venn region has at most one element. Thus for fixed n there are just finitely many n-set game boards after reduction, and for small n we can exhaustively check who has a winning strategy in each case.

Theorem 5.4. *For all games with $|\mathcal{B}| \leq 3$, the winning player and corresponding winning strategy are known and identified in Appendix 5.*

This case study gives some insight into the question of which player wins more often. Let $P(n)$ be the proportion of n-set boards $(k, \mathcal{B})$ (with at most one element in each Venn region) for which Split has a winning strategy for $(k, \mathcal{B}, \text{Split})$. From

our data, we have $P(1) = 1$, $P(2) = 7/8$, and $P(3) = 65/128$. Moreover, using the fact that games won by Skew are closed under adding sets, it is not difficult to show that $P(n)$ is a decreasing function of n, and in fact, we have the following result.

Theorem 5.5. *With $P(n)$ as described above, $P(n) \to 0$ as $n \to \infty$.*

Proof. Let $N \in \{0, 1\}^{2^n}$ be chosen uniformly at random. Construct a game board $(k, \mathcal{B})$ on n sets from N by identifying each of the 2^n Venn regions of $\mathcal{B}$ with a coordinate of N, and placing 0 or 1 points in each region, according to the corresponding coordinate of N. Then $P(n)$ is precisely the probability that Split has a winnning strategy for $(k, \mathcal{B}, \text{Split})$.

Our strategy is to find subfamilies of $\mathcal{B}$ of size 2, with high probability, for which Skew has a winning strategy. Let $t = \lfloor n/2 \rfloor$, and let $\mathcal{B}_1, \ldots, \mathcal{B}_t$ be disjoint subfamilies of $\mathcal{B}$ on 2 sets. Let S_i be the event that Split has a winning strategy for $(k, \mathcal{B}_i, \text{Split})$. Now consider a single Venn region R of $\mathcal{B}_i$, and let X be a random variable denoting $|R|$. Then X is binomially distributed with $p = 1/2$, so we have $\Pr(X \text{ is even}) = \Pr(X \text{ is odd}) = 1/2$. It follows from Lemma 5.1 and Appendix 5 that $\Pr(S_i) = P(2) = 7/8$.

We claim that $S_1, \ldots, S_t$ are mutually independent. It suffices to show that for any nonempty subset $\{i_1, \ldots, i_s\}$ of $[t-1]$, we have $\Pr(S_t \mid S_{i_1} \cap \cdots \cap S_{i_s}) = \Pr(S_t)$. Now recall that we are assigning 0 or 1 points to each Venn region of the total configuration. Imagine that we assign the regions contained in sets of $\mathcal{B}_1 \ldots \mathcal{B}_{t-1}$ first, thus determining the occurrence of $S_{i_1} \cap \cdots \cap S_{i_s}$. Note that each of the three interior Venn regions of B_t has at least one Venn region of $\mathcal{B}$ unassigned. Therefore, the number of additional points assigned will follow a nontrivial binomial distribution, so the probability that an even (or odd) number of additional points will be assigned is $1/2$. Therefore, the initial assignment has no bearing on the probability of S_t, and we have mutual independence.

Now, since $S_1 \cap \cdots \cap S_t$ must occur in order for Split to have a winning strategy for $(k, \mathcal{B}, \text{Split})$, we have

$$P(n) \leq \Pr(S_1 \cap \cdots \cap S_t) = (7/8)^t \to 0 \text{ as } n \to \infty.$$

This completes the proof. □

It is also worth remarking that while we were able to give a pairing strategy for every 3-set game, there is not always a pairing strategy. For an example with $n = 6$, see the first statement of Theorem 5.7 below. In this case we can show that Player II has a winning strategy, so Lemma 5.2 implies there cannot be a pairing strategy. For an example with $n = 5$, we can also show Player II has a winning strategy in the game on $\mathbb{Z}/5\mathbb{Z}$ whose board is described by $B_i = \{i, i+1, i+2\}$, $i = 1, \ldots, 5$. We do not know whether there is such an example using just four sets.

As there is significant literature on maker-breaker tic-tac-toe and its variants, we next study splitting tic-tac-toe. Here given m, n we let $k = mn$ and think of $[k]$ as an $m \times n$ grid. We then let $\mathcal{B}$ consist of the rows and columns of the grid.

Observe that $\mathcal{B}$ is splittable by a checkerboard pattern. For comparison, in 3×3 maker-breaker tic-tac-toe, Maker tries to make a line and Breaker tries to prevent Maker from making a line. In 3×3 splitting tic-tac-toe, Skew tries to make a line or cause Split to make a line.

Because of the special status of the 3×3 board, we first consider the "classic" variant of 3×3 splitting tic-tac-toe where $\mathcal{B}$ includes the diagonals as well as the rows and columns.

Proposition 5.6. *In "classic" 3×3 splitting tic-tac-toe, Skew has a winning strategy.*

Proof. If Skew is Player I, begin by claiming the upper-left corner. Split is then forced to claim the middle, or else Skew can win traditional tic-tac-toe. Skew then claims the middle-top and middle-left, forcing Split to form a diagonal line.

If Skew is Player II, we consider the three cases depending on which element Split claims first. If Split claims the middle, then by claiming a corner, Skew can threaten to form three in a row no matter where Split goes next. Skew can take advantage of this by choosing one of the squares adjacent to its first move, forcing Split to take the final corner in that row or column. Thus, along this diagonal, either Split has already formed a full diagonal or there is one open square which Skew can force Split to claim as Split must make the last move.

Next suppose Split claims the upper middle. Skew can then go in the middle and no matter what Split does next, Skew can occupy one of the two bottom corners. Thus, Skew threatens to form a diagonal so Split must prevent this by taking one of the two upper corners depending on where Skew went. Again, by avoiding the final spot in this row, Skew can force Split to take it, resulting in a full row for Split and a Skew win.

Finally suppose Split claims the upper left. Skew again responds by selecting the middle. No matter where Split goes next, Skew can force Split to claim one of the two free squares adjacent to their first one. As before, Skew can simply avoid the third square in this row or column, eventually forcing Split to form three in a row.

This covers all cases up to symmetry, so Skew has a winning strategy. $\qquad\square$

We now return to the version of splitting tic-tac-toe without the diagonals.

Theorem 5.7. *Consider the $m \times n$ splitting tic-tac-toe game, with $m, n \neq 1$.*

- *If $m = n = 3$, Player II has a winning strategy.*

- *Otherwise, Skew has a winning strategy.*

The proof consists of a number of cases and is carried out in Appendix 5.

We conclude with the generalization of splitting tic-tac-toe to higher dimensions. Specifically, we play on a d-dimensional grid and let $\mathcal{B}$ consist of the maximal axis-parallel lines.

Theorem 5.8. *Consider the $n_1 \times \cdots \times n_d$ game described above. Assume that $n_i \neq 1$ for all i (this would reduce to a lower dimensional game).*

○ *If $d = 2$ and $n_1 = n_2 = 3$, then Player II has a winning strategy.*

○ *In all other cases, Skew has a winning strategy.*

Proof. The first statement is part of the previous theorem.

For the second statement, unless $n_i = 3$ for all i, the previous theorem implies there is a two-dimensional sub-board where Skew has a winning strategy.

If $n_i = 3$ for all i, regard the board as a union of parallel 3×3 sub-boards. First suppose Skew is player I. After the first move, we can assume that Player II moves on a different sub-board (otherwise we could cut the boards in a different way). Then Skew gets to make a second move in the same sub-board, giving them a one move lead. It is not difficult to see that Skew can now make a line in this sub-board. For example Skew's first four moves can form a 2×2 corner of the sub-board, forcing Split to play in other corners. After this Skew is threatening two lines and Split cannot block both.

Next suppose that Skew is Player II. After Split makes the first move, Skew can play on the same sub-board. If Split continues to play on the same sub-board, then Skew wins because Skew is Player II in a 3×3 game. If Split plays outside the sub-board, Skew can copycat and also play outside the sub-board. Since there are an even number of elements in the complement of the sub-board, Skew can continue to use the winning strategy on the sub-board. □

Appendices
A 3-set families with the minimum number of splitters

Using a computer search, we found all types of splittable 3-set families on all $k \leq 60$. The code and its output may be found in [15].

For each of these k there are several distinct families with the fewest number of splitters. However, the various solutions all look similar to one another (with some Venn regions permuted or single point moved somewhat). Every solution has either exactly two Venn regions with 1 point or exactly one Venn region with 2 points. The solution sets repeat in pattern every 6 values of k.

In Figure 5 we give one of the solutions for each equivalence class of k modulo 6. We conjecture that the pattern continues for all k.

The pattern here has been verified for $k \leq 60$.

In Figure 6, we give some computational results for the minimum number of splitters of 3-set families for small values of k.
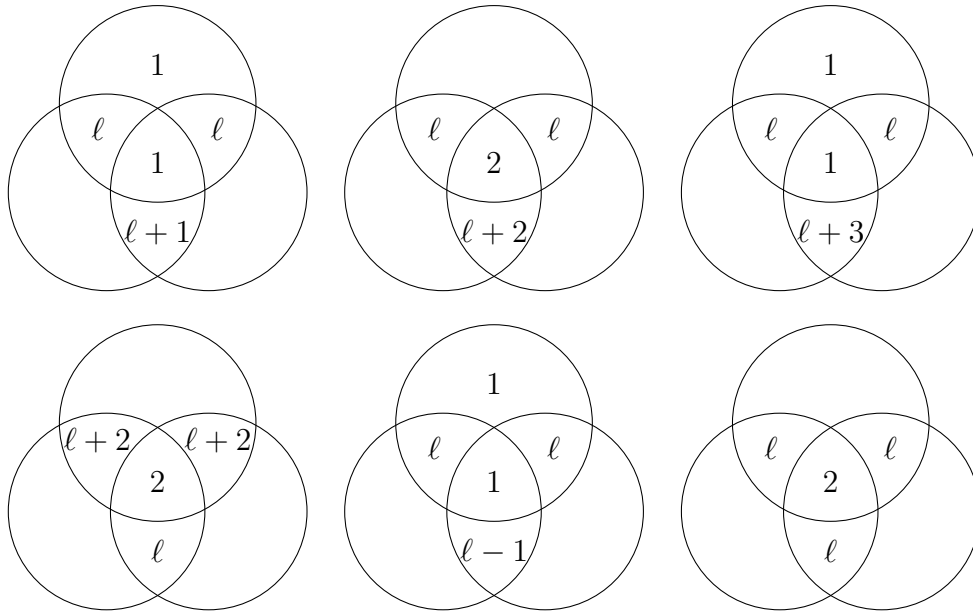
Figure 5: Conjectural 3-set families with the minimum number of splitters. From left to right, top to bottom, the congruence classes $0, \ldots, 5$ of k modulo 6. (In each case, ℓ is an odd integer.)

k	# of splitters in minimum arrangement
≤ 5	2
6	4
7	6
8	12
9	18
10	36
11	54
12	108
13	180
14	360
15	600
16	1200
17	2000
18	4000
19	7000
20	14000

Figure 6: Minimum number of splitters for 3-families on $[k]$.

B Two and three set games

In this section we catalog the winner and winning strategy for all instances of the splitting game on two or three sets. By Lemma 5.1, we may suppose there are 0 or 1 elements in each Venn region. We use an empty region to represent 0 elements and a • to represent 1 element. Since we will see that each of these games may be won using a (possibly trivial) pairing strategy, the winner is independent of which player goes first. It is also independent of whether or not there are points in the Venn region of multiplicity 0; we use a ⋆ to suggest either case.

For two set games, there are 8 arrangements, not considering the multiplicity 0 region. Of these, 7 are won by Split and 1 by Skew. The 8 arrangements are divided into 5 types up to symmetry. Of the 5 types, 4 are won by Split and 1 by Skew (see Figure 7).
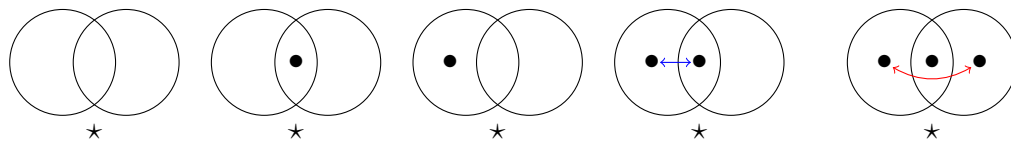

Figure 7: Types of two set games, with pairing strategies shown. The first four arrangements are won by Split. The final arrangement is won by Skew.

For 3-set games, there are 128 arrangements. Of these, 65 are won by Split and 63 are won by Skew. The arrangements fall into 40 types up to symmetry. By Theorem 5.3, Split has a winning strategy for any game with ≤ 2 occurrences of •, which accounts for 10 of the types. The remaining 30 types are shown below. Figure 8 depicts games won by Split, and Figure 9 depicts games won by Skew.
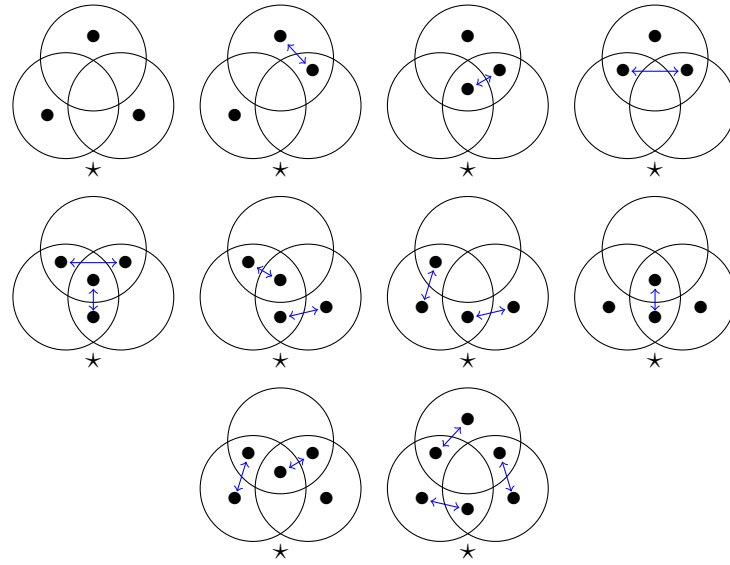
Figure 8: Three-set games won by Split with ≥ 3 instances of $\bullet$, with pairing strategies shown.

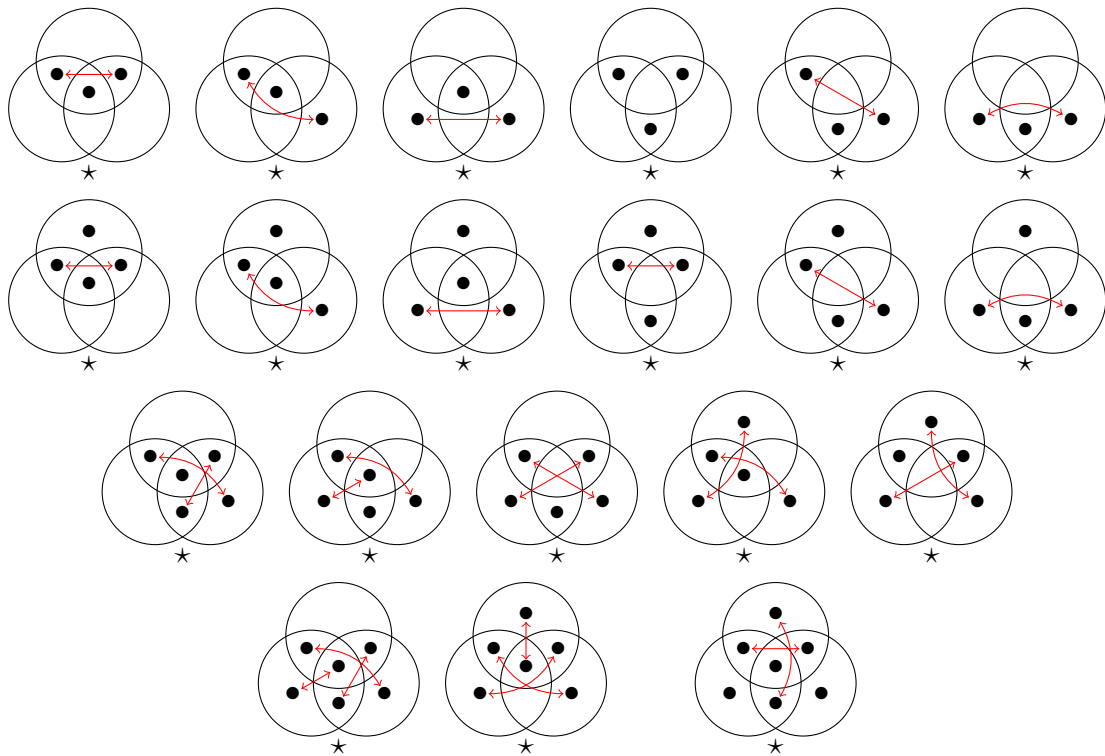Figure 9: Three-set games won by Skew, with pairing strategies shown.

C 2-dimensional splitting tic-tac-toe

In this section we carry out the details of the proof of Theorem 5.7. Recall the theorem states that in the $m \times n$ splitting tic-tac-toe game with $m, n \neq 1$:

- If $m = n = 3$, Player II has a winning strategy.

- Otherwise, Skew has a winning strategy.

For the first item, consider first the case where Skew is Player I. The board is completely symmetric with respect to the squares, so assume without loss of generality that Skew takes the upper left square. Then Split's strategy is to take the top middle. Up to symmetry, Skew can now take the top right, the left middle, the center, or the right middle.

Suppose Skew takes the top right. Then split takes the left middle. Then if Skew takes the bottom right, then Split takes the right middle. Then as long as Split takes something other than the middle on his remaining turn, Split wins. On the other hand if Skew takes something other than the bottom right on his third turn, then Split takes the bottom right. Now Split has one square from each row and each column, so Skew cannot claim an entire row, and neither can Split, as Split has just one move left. Thus Split wins in this case.

Now suppose Skew takes the left middle. Then Split takes the bottom left. As noted before, if Split has claimed a square from every row and column, Split wins. Thus Skew is forced to take the right middle. Then Split takes the center. Then Split can win by splitting the pair of remaining squares on the right.

Now suppose Skew takes the center. Then Split takes the right middle. Then, so as to prevent Split from taking a square from each row or column, Skew must take the bottom right. Then Split can win by simultaneously splitting the pair of remaining squares on the bottom and the pair of remaining squares on the right.

Finally, consider the case where Skew goes in the right middle. Then Split takes the left middle. Then Skew is forced to take the bottom right, as in the other cases. Then Split takes the top right. Now Split can win by splitting the pair of remaining squares on the bottom.

It remains to show that Skew has a winning strategy in the first item when Split is Player I. Without loss of generality, Split begins by taking the top left. Then Skew takes the center. Split cannot take a square from the top or the left since Split has the last move, and Skew can force Split to eventually take three in a row. Thus Split has two remaining options, up to symmetry: Split can take the right middle or the bottom right. In either case, after Split makes a move, then Skew takes the bottom middle and forces Split to take the top middle. Then Skew can force Split to eventually take the top right, so Skew wins.

For the second item, if m, n are both even, notice that any row and column forms a two-set family with an odd number of elements in each Venn region. By our previous work on two-set games, Skew has a winning strategy on just these two sets.

If $m = 2$, Skew uses a pairing strategy with the pairs $a_{2,j}, a_{1,j+1}$ (cyclically).

For the remaining cases, Skew has a winning strategy by "getting ahead" in one of the rows or columns. If at any point, the number of squares claimed by Split and Skew in an even row (column) differ by at least two, or in an odd row (column) by at least three, then Skew can skew the row (column) by claiming squares from that row (column) on each turn until they run out.

If m is even and n is odd, consider first when Skew is Player I. Skew takes the top left corner. Since m is even, Split cannot allow Skew to take two squares from any column before Split takes any, so Split is forced to go in column 1. Then Skew takes the second square in row 1. Now if Split takes a square from row 1, then Skew takes the second square in column 2, which puts Skew two ahead in column 2. If Split takes a square from column 2, then Skew takes a square from row 1, which puts Skew three ahead in row 1. In either case, Skew wins.

Now consider when m is even and n is odd and Split is Player I. Without loss of generality, Split goes in the bottom right corner. Then Skew can follow the same strategy as before without Split's first move coming into play at all.

If $m, n \geq 5$ are odd and Skew is Player I, Skew takes the top left. Assume without loss of generality that Split does not go in column 1 or column 2. Then Skew takes a square from column 1 in an empty row. Then Split is forced to go in column 1. Skew continues taking squares from column 1 in empty rows until Split takes the last square from column 1. Now column 2 is empty, and there are at least two rows in which only Skew has claimed a square. Skew now takes squares from those rows in column 2. Each time, Split is forced to take a square from the same row. Then Skew is free to take a third square from column 2, where Split has taken none, and Skew wins.

If $m, n \geq 5$ are odd and Split is Player I, assume without loss of generality that Split takes the bottom right square. Then Skew plays as described above, but in the case where Split takes a square not in row 1 in his second move, Skew takes a square from column 1 from the same row, then continues as described. The result is the same, since in either case, it will be Skew's turn with column 2 empty and at least two rows with only Skew having claimed a square.

If m, n are odd and exactly one of them is 3, more explanation is required. Assume $m \geq 5$ and $n = 3$ without loss of generality, and first consider the case where Skew is Player I. Skew takes the top left. If Split does not go in column 1, then Skew can play as in the 5×5 game. If Split goes in column 1, then Skew takes the middle square of row 1. Then Split is forced to take the right square of row 1. Then Skew takes squares from column 2 from empty rows until all have been taken, and Split is forced to also take squares from column 2. When Split takes the last square from column 2, there are at least two rows in which only Skew has taken a square. Skew then takes the left square from each of these rows, and Split is forced to take the right square of that row each time. Now Skew takes another square from column 1, which puts Skew three ahead, and Skew wins.

For the last case, assume $m \geq 5$ is odd and $n = 3$ and Split is Player I. Assume

without loss of generality that Split takes the bottom right square. Then Skew takes the top left. If Split doesn't go in go in column 1, then Skew takes squares from column 1, not row m, until they run out. Split is forced to take squares from column 1 as well, including the left square in row m. Now since there are an odd number of squares, and Split is Player I, Skew can force Split to eventually take the third square in row m, and Skew wins. If Split instead goes in column 1 for his second move, then Skew takes top middle. Split is forced to take top right. Then Skew takes squares from column 2, not row m until they run out, and Split is eventually forced to take bottom middle. Then, since Split has the last move, Skew can force Split to eventually take bottom left, and Skew wins.

This completes the proof.

Acknowledgements

References

[1] N. Alon, E.xi E. Bergmann, D. Coppersmith and A. M. Odlyzko, Balancing sets of vectors, *IEEE Trans. Inform. Theory* 34 (1) (1988), 128–130.

[2] J. Beck, *Combinatorial games*, vol. 114 of *Encyclopedia of Mathematics and its Applications*, Cambridge University Press, Cambridge, 2011; Tic-tac-toe theory, Paperback edition of the 2008 original.

[3] P. Bernstein, C. Bortner, S. Coskey, S. Li and C. Simpson, The set splittability problem, *Australas. J. Combin.* 75 (2019), 190–209.

[4] M. Charikar, A. Newman and A. Nikolov, Tight hardness results for minimizing discrepancy, In *Proc. Twenty-Second Annual ACM-SIAM Symp. on Discrete Algorithms*, pp. 1607–1614; SIAM, Philadelphia, PA, 2011.

[5] B. Chazelle, *The discrepancy method*, Cambridge University Press, Cambridge, 2000; Randomness and complexity.

[6] D. Condon, S. Coskey, L. Serafin and C. Stockdale, On generalizations of separating and splitting families, *Electron. J. Combin.* 23 (3) (2016), #P3.36, 19 pp.

[7] 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) (2007), 18–37.

[8] H. Enomoto, P. Frankl, N. Ito and K. Nomura, Codes with given distances, *Graphs Combin.* 3 (1) (1987), 25–38.

[9] J. D. Farmer and S. C. Leth, An asymptotic formula for powers of binomial coefficients, *Math. Gazette* 89 (516) (2005), 385–391.

[10] W. Feller, *An introduction to probability theory and its applications*, Vol. I, Third ed., John Wiley & Sons, Inc., New York-London-Sydney, 1968.

[11] B. Frederickson, Splitting families code, 2018, `https://bitbucket.org/brycefred/maximal-splitting-families/src/master/`.

[12] X. Li and H. Shen, Constructions and bounds for splitting and separating systems, *Ars Combin.* 123 (2015), 381–406.

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

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

[15] H.-T. Yan, Splittable families code, 2018, `https://drive.google.com/drive/folders/1-greIn1J6KMUz4S4PnoF0d0L2PbgSjFd?usp=sharing`.

[16] C. Yost-Wolff, Personal communication, 2018.