Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in DISCRETE MATHEMATICS

www.elsevier.com/locate/endm

Almost-Fisher families

Shagnik Das [1]

Department of Mathematics
Freie Universität
Berlin, Germany

Benny Sudakov [2,3]

Department of Mathematics
ETH
Zürich, Switzerland

Pedro Vieira [4]

Department of Mathematics
ETH
Zürich, Switzerland

Abstract

A classic theorem in combinatorial design theory is Fisher's inequality, which states that a family $\mathcal{F}$ of subsets of $[n] = \{1, 2, \ldots, n\}$ with all pairwise intersections of size λ can have at most n non-empty sets. One may weaken the condition by requiring that for every set in $\mathcal{F}$, all but at most k of its pairwise intersections have size λ. We call such families k-almost λ-Fisher. Vu was the first to study the maximum size of such families, proving that for $k = 1$ the largest family has $2n - 2$ sets, and characterising when equality is attained. We substantially refine his result, showing how the size of the maximum family depends on λ. In particular we prove that for small λ one essentially recovers Fisher's bound. We also solve the next open case of $k = 2$ and obtain the first non-trivial upper bound for general k.

Keywords: Fisher's inequality, Fisher families, almost-Fisher, extremal set theory, design theory, restricted intersections

1 Introduction

1.1 Restricted intersections

Extremal set theory is a rapidly developing area of combinatorics and has enjoyed tremendous growth in the past few decades. No doubt this is fuelled by its deep connections to other areas; extremal set theory both employs methods from and enjoys applications to diverse fields such as algebra, geometry and coding theory.

Many problems in extremal set theory are concerned with the pairwise intersections between sets in a family. Typically one is interested in studying families that are (in some sense) extremal under certain restrictions on possible intersection sizes. A natural setting to consider is to require that all pairwise intersections between sets in a family are of the same size and ask how large such a family can be. We call a family of sets a λ-*Fisher family* if any two distinct sets intersect in λ elements. The foundational result in this direction is Fisher's inequality [3], bounding the size of a λ-Fisher family. Fisher's original result dealt with more restrictive designs, and was extended to uniform λ-Fisher families by Bose [1]. The following non-uniform version was proven by Majumdar [10] and rediscovered by Isbell [8].

Theorem 1.1 *A λ-Fisher family over $[n]$ can have at most n non-empty sets.*

Note that the non-empty condition is necessary, as when $\lambda = 0$, one may take the empty set in addition to the n singletons. However, when $\lambda > 0$, a λ-Fisher family cannot contain the empty set. The classification of extremal constructions (i.e. λ-Fisher families over $[n]$ of size n) remains one of the most important problems in combinatorial design theory. In the case $\lambda = 1$, this reduces to the famous de Bruijn–Erdős theorem [2], for which the extremal constructions are known precisely.

[1] Email: `shagnik@mi.fu-berlin.de`
[2] Email: `benjamin.sudakov@math.ethz.ch`
[3] Research supported in part by SNSF grant 200021-149111.
[4] Email: `pedro.vieira@math.ethz.ch`

Theorem 1.1 has inspired a great deal of research, having been extended in numerous directions by several renowned mathematicians. Ray-Chaudhuri and Wilson [12] provided bounds on families where s different intersection sizes are allowed, and Frankl and Wilson [4] considered the problem where the sizes of sets and intersections are taken modulo a prime p. These results have proven remarkably useful in the field of discrete geometry. Another extension that has attracted plenty of attention in recent years is to restrict the sizes of k-wise intersections instead of pairwise intersections. This problem was first raised by Sós [13], and resolved in the uniform setting by Füredi [5]. Vu [15] studied the problem for intersections modulo 2, and Grolmusz and Sudakov [7] extended this to systems modulo an arbitrary prime. Exact results in this setting were obtained by Szabó and Vu [14]. In the non-modular setting, asymptotically sharp results were provided by Füredi and Sudakov [6].

1.2 Almost-Fisher families

We seek a different extension of Fisher's inequality. As we have seen, requiring that *all* pairwise intersections in a set family $\mathcal{F}$ have size λ severely restricts the size of $\mathcal{F}$. One might hope to find larger families by relaxing the condition and allowing a few 'bad' intersections to appear. The following weakened version of λ-Fisher families was first introduced by Vu [15]. We call a family $\mathcal{F}$ of sets *k-almost λ-Fisher* if for every set $F \in \mathcal{F}$, there are at most k other sets $F' \in \mathcal{F} \setminus \{F\}$ for which $|F \cap F'| \neq \lambda$. In particular, note that when $k = 0$ this reduces to the λ-Fisher families defined previously. We are interested in bounding the size of a k-almost λ-Fisher family over $[n]$, and denote the largest possible size by

$$f(n, k, \lambda) = \max \left\{ |\mathcal{F}| : \mathcal{F} \subset 2^{[n]} \text{ is a } k\text{-almost } \lambda\text{-Fisher family} \right\}.$$

Most results regarding restricted intersections are proven by linear algebraic methods, using the restrictions on the system to build a linearly independent set of vectors in an appropriate vector space. Dimensional arguments then provide the required bound on the size of the set family. Allowing some intersections of different sizes destroys the linear independence of these vectors. However, by marrying the algebraic arguments with some graph theoretic considerations, we can still recover some bounds on $f(n, k, \lambda)$.

Given a k-almost λ-Fisher family $\mathcal{F}$, we can define an auxiliary graph $G = G(\mathcal{F}) = (V, E)$, where $V = \mathcal{F}$ and $\{F, F'\} \in E$ if and only if $|F \cap F'| \neq \lambda$. Since every set in $\mathcal{F}$ can have at most k pairwise intersections not equal to λ, it follows that the maximum degree of G is at most k. Moreover, an independent set in G corresponds to a λ-Fisher subfamily of $\mathcal{F}$. Since any such family can have at most n non-empty sets, and any graph with m vertices and maximum degree Δ has an independent set of size at least $m/(\Delta+1)$, we arrive at the following upper bound for $f(n,k,\lambda)$, first given in [15].

Proposition 1.2 *We have $f(n,k,\lambda) \leq (k+1)n+1$ for any positive integer n and non-negative integers k and λ. Moreover, if $\lambda \neq 0$, we can improve this bound to $f(n,k,\lambda) \leq (k+1)n$.*

Vu [15] showed that this essentially gives the correct bound when $k = 1$, and was further able to prove that the extremal constructions arise from Hadamard matrices.

Theorem 1.3 *For $n \geq 3$ and for any non-negative λ, $f(n,1,\lambda) \leq 2(n-1)$. Moreover, if $n \geq 4$ and equality holds, then $\lambda = n/4$ and a Hadamard matrix of order n exists.*

For large values of k, however, Vu noted that Proposition 1.2 appears to be far from the truth, and asked to determine the correct behavior of $f(n,k,\lambda)$.

2 Our results

In this paper we continue the study of k-almost λ-Fisher families, approaching the problem of determining $f(n,k,\lambda)$ from a few different directions. In doing so, we are able to substantially refine Vu's result, while also obtaining some evidence that $nk/4$ might be the correct asymptotic behavior for large k.

As a warm up, we start with the case $\lambda = 0$, and for brevity call a k-almost 0-Fisher family a k-almost disjoint family. We obtain the following bound on $f(n,k,0)$, and show for every k that this is tight for infinitely many values of n.

Theorem 2.1 *For any positive integers n and k, we have*

$$f(n,k,0) \leq \frac{n}{k}\left\lfloor \frac{k^2}{4} \right\rfloor + n + 1.$$

In particular, this shows that as k grows, the largest k-almost disjoint family has size asymptotically $nk/4$, matching a construction of Vu (for which $\lambda = 2$).

We next turn our attention to the case $k = 1$. Recall that Vu proved $f(n,1,\lambda) \leq 2n-2$, attainable only if $\lambda = n/4$. It is very natural to ask what happens for other values of λ, i.e., to study the dependence of the function $f(n,1,\lambda)$ on the parameter λ. Here we practically resolve this question, obtaining the following essentially tight result.

Theorem 2.2 *For integers $n \geq 1$ and $\lambda \geq 0$, we have*

$$f(n,1,\lambda) \leq \max\left\{n+2, 8\min\left\{\lambda, \frac{n-\lambda}{3}\right\} + o(\lambda)\right\}.$$

Note that $f(n,1,\lambda)$ is only close to $2n$ when λ is close to $n/4$, providing stability for Theorem 1.3. Moreover, if $\lambda < n/8$ or $\lambda > 5n/8$, then allowing one non-λ intersection per set provides almost no gain compared to Fisher's inequality.

We provide further evidence of the Hadamard construction being atypically large by extending Vu's methods and showing that it is the best possible even when we allow two bad intersections per set. Once again, we provide stability by showing that $f(n,2,\lambda)$ is much smaller than $2n$ when λ is far from $n/4$. We also show that when $\lambda = o(n)$, $f(n,2,\lambda) = \left(\frac{3}{2} + o(1)\right)n$, which is asymptotically the bound obtained when $\lambda = 0$ in Theorem 2.1. This suggests that perhaps the $\lambda = 0$ case exemplifies the true behaviour of the function $f(n,k,\lambda)$ for large k, implying that $nk/4$ is the correct bound.

Theorem 2.3 *For n sufficiently large and $0 \leq \lambda \leq n$, we have the bounds*

(i) $f(n,2,\lambda) \leq 2n-2$.

(ii) $f(n,2,\lambda) \leq \frac{1}{3}\left(5n + 4\min\left\{\lambda, \frac{n-\lambda}{3}\right\} + 7\right)$.

(iii) $f(n,2,\lambda) \leq \left(\frac{3}{2} + o(1)\right)n$ *when* $\lambda = o(n)$.

Finally, using our results together with a theorem of Lovász [9] we obtain the first non-trivial upper bound for general k.

Corollary 2.4 *For $k \geq 1$, we have $f(n,k,\lambda) \leq (2n-2)\left\lceil \frac{k+1}{3} \right\rceil$. Moreover, if $\lambda = o(n)$, then $f(n,k,\lambda) \leq \left(\frac{3}{2} + o(1)\right)n\left\lceil \frac{k+1}{3} \right\rceil$.*

3 Proof ideas

3.1 Theorem 2.1

A key ingredient in our proof of Theorem 2.1 is to note that if $\mathcal{F}$ is a k-almost disjoint family, if $F \in \mathcal{F}$ and if $F' \notin \mathcal{F}$ is a proper subset of F then the family $\mathcal{F}' = (\mathcal{F} \cup \{F'\}) \setminus \{F\}$ is a k-almost disjoint family of the same size as $\mathcal{F}$ but whose sets are in some sense smaller. This shows that in order to bound $f(n,k,0)$ one can consider only k-almost disjoint families $\mathcal{F}$ which are downwards closed (i.e. such that if $F \in \mathcal{F}$ and $F' \subseteq F$ then $F' \in \mathcal{F}$). This in turn suggests that we should take our sets as small as possible. Indeed, a simple induction on the number of sets in $\mathcal{F}$ of size at least 4 allows one to reduce the problem to the case when $\mathcal{F}$ only has sets of size at most 3.

Furthermore, using shifting arguments we can further restrict our attention to the case when $\mathcal{F}$ only has sets of size at most 2. In this simplified setting a careful analysis then shows that the theorem holds.

3.2 Theorem 2.2

The proof of Theorem 2.2 has three main ingredients: linear algebra, structural analysis of 1-almost λ-Fisher families and a bound from coding theory due to Plotkin [11]. The linear algebra part mostly concerns intersection matrices. Given a family $\mathcal{F} = \{F_1, \dots, F_m\}$ we define its *intersection matrix* $M(\mathcal{F})$ to be the $m \times m$ matrix with $M(\mathcal{F})_{i,j} = |F_i \cap F_j| - \lambda$. Most of our algebraic arguments rely on estimating the rank of the intersection matrix. Indeed, it is not hard to see that if $\mathcal{F} \subseteq 2^{[n]}$ then $M(\mathcal{F})$ has rank at most $n+1$. Moreover, if $\mathcal{F}$ is a 1-almost λ-Fisher family then, up to relabeling the sets in $\mathcal{F}$, $M(\mathcal{F})$ is a block diagonal matrix all of whose blocks have size 1 or 2, the latter corresponding to pairs $F, F' \in \mathcal{F}$ such that $|F \cap F'| \neq \lambda$. Thus, one can estimate the rank of $M(\mathcal{F})$ based on the ranks of its blocks. If all but at most one of these blocks have full rank then one can deduce that $|\mathcal{F}| \leq n+2$. If on the other hand there are at least 2 blocks which do not have full rank then one can deduce structural information about the sets in $\mathcal{F}$. Indeed, in that case, most pairs of sets in $\mathcal{F}$ must have large symmetric difference. A bound due to Plotkin [11] then implies that $\mathcal{F}$ cannot be too large.

3.3 Theorem 2.3

The proof of Theorem 2.3 follows similar lines as the one of Theorem 2.2. The major difference lies in the fact that, although the intersection matrix $M(\mathcal{F})$ is still a block diagonal matrix, if $\mathcal{F}$ is 2-almost λ-Fisher family then the blocks of $M(\mathcal{F})$ can have arbitrarily large size. In fact, one can easily see that these blocks correspond to the connected components of the auxiliary graph $G(\mathcal{F})$ (defined in section 1.2), and thus, since $G(\mathcal{F})$ has maximum degree 2, correspond to paths or cycles. Estimating the ranks of these blocks is still possible but a deeper analysis is needed in order to obtain some structural information about the sets in $\mathcal{F}$.

3.4 Corollary 2.4

Finally Corollary 2.4 follows immediately from a theorem of Lovász [9] and Theorem 2.3. Indeed, we know that the auxiliary graph $G(\mathcal{F})$ of a k-almost λ-Fisher family $\mathcal{F}$ has maximum degree at most k. Lovász's theorem implies that the vertex set of $G(\mathcal{F})$ can be partitioned into $\lceil \frac{k+1}{3} \rceil$ parts, each inducing a graph of maximum degree at most 2. However, a subset of vertices of $G(\mathcal{F})$ inducing a graph of maximum degree at most 2 is readily seen to correspond to a 2-almost λ-Fisher subfamily $\mathcal{F}' \subseteq \mathcal{F}$. The corollary follows by using Theorem 2.3 to bound the sizes of the subfamilies $\mathcal{F}'$.

References

[1] Bose, R. C., *A note on Fisher's inequality for balanced incomplete block designs*, Ann. Math. Stat. **20.4** (1949), 619–620.

[2] de Bruijn, N. G., and P. Erdős, *A combinatorial problem*, K. N. A. W. Verhan. **49.49** (1946), 758–764.

[3] Fisher, R. A., *An examination of the different possible solutions of a problem in incomplete blocks*, Ann. Eugenic **10** (1940), 52–75.

[4] Frankl, P., and R. M. Wilson, *Intersection theorems with geometric consequences*, Combinatorica **1** (1981), 357–368.

[5] Füredi, Z., *On finite set-systems whose every intersection is a kernel of a star*, Discrete Math. **47** (1983), 129–132.

[6] Füredi, Z., and B. Sudakov, *Extremal set systems with restricted k-wise intersections*, J. Comb. Theory A **105** (2004), 143–159.

[7] Grolmusz, V., and B. Sudakov, *On k-wise set-intersections and k-wise hamming-distances*, J. Comb. Theory A **99** (2002), 180–190.

[8] Isbell, J. R., *An inequality for incidence matrices*, Proc. Amer. Math. Soc. **10** (1959), 216–218.

[9] Lovász, L., *On decomposition of graphs*, Studia Sci. Math. Hungarica **1** (1966), 237–238.

[10] Majumdar, K. N., *On some theorems in combinatorics, relating to incomplete block designs*, Ann. Math. Stat. **24** (1953), 377–389.

[11] Plotkin, M., *Binary codes with specified minimum distance*, IRE T. Inform. Theor. **6** (1960), 445–450.

[12] Ray-Chaudhuri, D. K., and R. M. Wilson, *On t-designs*, Osaka J. Math. **12** (1975), 737–744.

[13] Sós, V. T., *Remarks on the connection of graph theory, finite geometry and block designs*, "Colloq. Inter. sulle Teorie Combinatorie (Roma, 1973), Tomo II", 223–233. Atti dei Convegni Lincei, No. 17, Accad. Naz. Lincei, Rome, 1976.

[14] Szabó, T., and V. H. Vu, *Exact k-wise intersection theorems*, Graph. Combinator. **21.2** (2005), 247–261.

[15] Vu, V. H., *Extremal set systems with weakly restricted intersections*, Combinatorica **19.4** (1999), 567–587.