Strong qualitative independence*†

Gyula O.H. Katona
Alfréd Rényi Institute of Mathematics, HAS,
Budapest P.O.B. 127 H-1364 HUNGARY
ohkatona@renyi.hu

January 26, 2012

*AMS Subject classification Primary 90B40, Secondary 05D05, 94A99, 94B99.
Keywords: search, qualitatively independent sets, Hadamard matrix, code
†The work was supported by the Hungarian National Foundation for Scientific Research grant numbers T029255.

Abstract

The subsets A, B of the n-element X are said to be s-strongly separating if the two sets divide X into 4 sets of size at least s. The maximum number $h(n, s)$ of pairwise s-strongly separating subsets was asymptotically determined by Frankl [8] for fixed s and large n. A new proof is given. Also, estimates for $h(n, cn)$ are found where c is a small constant.

1 Introduction

Let X be a finite set of n elements. The following notion was introduced by Marczewski [17]. The subsets $A, B \subset X$ are called *qualitatively independent* if they divide X into four non-empty parts that is $A \cap B, A \cap \overline{B}, \overline{A} \cap B, \overline{A} \cap \overline{B}$ are all non-empty. The significance of this notion in search theory lies in the consequence that after knowing if an unknown x is in A or not we cannot decide the same question for B, independenly on the answer for the first question. The family $\mathcal{F} \subset 2^X$ is called *independent* if their members are pairwise qualitatively independent.

Rényi asked [23] the question what is the maximum size of an independent family in an n-element set. The answer was found by the present author ([12], see also [23]), and independently by Brace and Daykin [5], Bollobás [3] and Kleitman and Spencer [16] (see also Schönheim [24]).

Theorem 1.1 *The maximum number of pairwise qualitatively independent set in an n-element sets is*

$$\binom{n-1}{\lceil \frac{n}{2} \rceil}.$$

If the family $\mathcal{F}$ is independent then $A, B \in \mathcal{F}, A \neq B$ implies $A \not\subset B$ that is $\mathcal{F}$ is *inclusion-free.* The maximum size of an inclusion-free family is determined by a classical theorem of Sperner [25].

Theorem 1.2 *If $\mathcal{F}$ is an inclusion-free family in an n-element set then*

$$|\mathcal{F}| \leq \binom{n}{\lfloor \frac{n}{2} \rfloor}.$$

The following sharpening of Sperner's theorem is also well-known as the YBLM-inequality (earlier LYM) (see [27], [4], [13], [18]). It involves the sizes of the "levels" of the family. If $\mathcal{F}$ is a family, define $f_i(\mathcal{F}) = |\{F : F \in \mathcal{F}, |F| = i\}|$.

Theorem 1.3 *If $\mathcal{F}$ is an inclusion-free family in an n-element set then*

$$\sum_{i=0}^{n} \frac{f_i(\mathcal{F})}{\binom{n}{i}} \leq 1$$

holds.

Kleitman and Spencer [16] introduced the notion of k-indpendence. A family $\mathcal{F} \subset 2^X$ is called k-independent iff

$$\cap_{i=1}^{k} A_i^{\varepsilon_i} \neq \emptyset$$

holds for any choice of distinct $A_1, \dots, A_k \in \mathcal{F}$ and $\varepsilon_i = 0, 1$ where $A^0 = A, A^1 = \overline{A}$ that is when any k members divide X into 2^k non-empty parts. Let $f(n, k)$ denote the maximum size of a k-independent family. They proved the following theorem.

Theorem 1.4 (Kleitman and Spencer [16]).

$$2^{c_1 2^{-k} k^{-1} n} \leq f(n, k) \leq 2^{c_2 2^{-k} n}.$$

See also [1].

A set A can be considered as a partition $(A, \overline{A})$. A generalization in this direction is to consider partitions into r parts that is *r-partitions*. Two r-partitions are called *qualitative independent* if all the r^2 intersections of the classes are non-empty. The maximum number of pairwise qualitatively independent r-partitions is denoted by $g(n, r)$. As this is exponential in n and an exact formula for $g(n, r)$ is hopeless, it is sufficient to consider the exponent:

$$q_r = \limsup_{n \to \infty} \frac{1}{n} \log g(n, r).$$

After some preliminary results ([19], [20], [21], [22], [15]) Gargano, Körner and Vaccaro have determined the exact exponent.

Theorem 1.5 (Gargano, Körner and Vaccaro [9]).

$$q_r = \frac{2}{r} \quad (2 \leq r).$$

For a recent interesting result concerning qualitative independence see [14] .

A family $\mathcal{S} \subseteq 2^X$ is *s-strongly separating* iff all four intersections $A \cap B, A \cap \overline{B}, \overline{A} \cap B, \overline{A} \cap \overline{B}$ are of size at least s for any two distinct members $A, B \in \mathcal{S}$. The maximum size of an s-strongly separating family is denoted by $h(n, s)$. The determination of $h(n, s)$ was asked in [12]. It has been asymptotically answered by Frankl for fixed s.

Theorem 1.6 [8]

$$d_1(s)\frac{2^n}{n^{s-\frac{1}{2}}} \le h(n,s) \le d_2(s)\frac{2^n}{n^{s-\frac{1}{2}}}.$$

where

$$d_1(s) = \sqrt{\frac{2}{\pi}}\frac{1}{2^s} - \varepsilon$$

and

$$d_2(s) = \sqrt{\frac{2}{\pi}}2^{s-2}(s-1)! + \varepsilon.$$

The author returned to this question now because Benjamin Weiss [26] suggested the following related problem. Find the maximum number of sequences of length n over the alphabet $\{0,1,2\}$ such that if $a = (a_1, \dots, a_n), b = (b_1, \dots, b_n)$ are two such sequences then every pair (c,d) $c,d \in \{0,1,2\}$ occurs either 0 times or at least twice among the pairs (a_i, b_i) $(1 \le i \le n)$. The analogous problem for $\{0,1\}$ is just $h(n,2)$.

However we only found a new proof for Frankl's theorem and some estimates for the case when $s = cn$ where c is a small constant.

In the new proof of the upper estimate (in Theorem 1.6) it is sufficient to use the following weaker condition. A family $\mathcal{F}$ is called *s-diffbounded* iff $|A-B| \ge s$ holds for any two members of $\mathcal{F}$. Let us mention that a family is 1-diffbounded iff $A \not\subset B$ for any two distinct members, in other words, iff the family is *inclusion-free.* It is obvious that if $\mathcal{S}$ is *s-strongly separating* then it is *s-diffbounded.* The following theorem is a generalization of Theorem 1.3.

Theorem 1.7 *Let s be a positive integer. If $\mathcal{F}$ is an s-diffbounded family in an n-element set then*

$$\sum_{i=s-1}^{n} f_i(\mathcal{F})\frac{\binom{i}{s-1}}{\binom{n}{i-s+1}} \le 1. \tag{1.1}$$

Christian Bey [2] called the author's attention to the fact that Theorem 1.7 is a special case of Theorem 4 of [28] ($i = t$ in its notation).

Corollary 1.8 *Let s be a positive integer. If $\mathcal{F}$ is an s-diffbounded family in an n-element set then*

$$|\mathcal{F}| \le d_3(s)\frac{2^n}{n^{s-\frac{1}{2}}}$$

where

$$d_3(s) = \left(\sqrt{\frac{2}{\pi}}2^{s-1}(s-1)! + \varepsilon\right).$$

If $\mathcal{S}$ is s-strongly separating then the family

$$\mathcal{F} = \{S, \overline{S} : \ S \in \mathcal{S}\}$$

is s-diffbounded. Here $|\mathcal{S}| \le \frac{1}{2}|\mathcal{F}|$ holds and this yields the following improvement of Corollary 1.8.

Corollary 1.9 *Let s be a positive integer. If $\mathcal{S}$ is an s-strongly separating family in an n-element set then*

$$|\mathcal{S}| \le \left(\sqrt{\frac{2}{\pi}}2^{s-2}(s-1)! + \varepsilon\right) \frac{2^n}{n^{s-\frac{1}{2}}}.$$

In Section 3 the case when $s = cn$ is considered. We have a non-trivial upper estimate only when c is small.

Theorem 1.10

$$h(n, cn) \le 2^{n(-\frac{1}{2}\log c - 1.099) + o(n)}$$

where $c \le 0.099$.

2 The case of fixed s

Proof of Theorem 1.7

Define the r-*shadow* of a family $\mathcal{F}$ in the following way.

$$\sigma_r(\mathcal{F}) = \{A - B : B \subset A \in \mathcal{F}, |B| = r\}.$$

Let $A, B \in \mathcal{F}$ be distinct members and choose an $s-1$-element subset C of A. Here $|A - B| \ge s$ implies $A - C \not\subseteq B$. If D is an $s-1$-element subset of B then $A - C \not\subseteq B - D$ follows. This is obviously true if $A = B$ but $C \ne D$. Therefore $\sigma_{s-1}(\mathcal{F})$ is inclusion-free.

Determine the sizes of the levels of $\sigma_{s-1}(\mathcal{F})$. The sets on level $j = i - s + 1$ $(s - 1 \le i \le n)$ are obtained from sets on level i in $\mathcal{F}$:

$$f_j(\sigma_{s-1}(\mathcal{F})) = f_i(\mathcal{F})\binom{i}{s-1}.$$

The YBLM-inequality for the family $\sigma_{s-1}(\mathcal{F})$ is exactly (1.1). ∎

Proof of Corollary 1.8
Theorem 1.7 will be used. Find the minimum of the ratio

$$\frac{\binom{i}{s-1}}{\binom{n}{i-s+1}}$$

in the terms of (1.1). The above ratio is equal to

$$\frac{\binom{n+s-1}{s-1}}{\binom{n+s-1}{i}}.$$

Its minimum is

$$\frac{\binom{n+s-1}{s-1}}{\binom{n+s-1}{\lfloor \frac{n+s-1}{2} \rfloor}}.$$

Therefore (1.1) implies

$$|\mathcal{F}| = \sum_{i=s-1}^{n} f_i(\mathcal{F}) \leq \frac{\binom{n+s-1}{\lfloor \frac{n+s-1}{2} \rfloor}}{\binom{n+s-1}{s-1}}.$$

The application of the Stirling formula leads to the desired result. ∎

It is easy to see that

$$d_3(s) = \sqrt{\frac{2}{\pi}} 2^{s-1}(s-1)! + \varepsilon$$

for large enough n.

Now *Corollary 1.9* easily follows. ∎

In the proof of the lower estimate of Theorem 1.6 the following result will be used from coding theory. A *codeword* is a 0,1 sequence of length n. We say that the codeword has *weight* w if the number of 1s is w. The Hamming distance of two codewords is the number of different digits. Let $A(n, 2\delta, w)$ is the maximum number of codewords of weight w with pairwise Hamming distance at least 2δ.

Theorem 2.1 (Graham and Sloane [10]) *Let $q \geq n$ be a prime power. Then*

$$\frac{\binom{n}{w}}{q^{\delta-1}} \leq A(n, 2\delta, w).$$

Proof of of Theorem 1.6
The upper estimate is a consequence of Corollary 1.9.

The construction of the lower estimate is based on Theorem 2.1. Use the theorem for $n-s, 2s, \lfloor \frac{n}{2} \rfloor - s$. The codewords define, in an obvious way, subsets of an $n-s$-element set Y with pairwise differences at least s. Denote this family by $\mathcal{F}'$. Add a disjoint set U to Y, where $|U| = s$. The extended underlying set is $X = Y \cup U$. Define

$$\mathcal{F} = \{F \cup U : \ F \in \mathcal{F}'\},$$

that is U is contained in all the sets. Since the difference of any two members of $\mathcal{F}'$ have size at least s, the same holds for any two members of $\mathcal{F}$, that is, $\mathcal{F}$ satisfies the conditions of the theorem. By Theorem 2.1 we have

$$\frac{\binom{n-s}{\lfloor \frac{n}{2} \rfloor - s}}{q^{s-1}} \leq |\mathcal{F}|. \tag{2.1}$$

By the prime number theorem there is a prime q satisfying $n \leq q \leq n + o(n)$. This fact, (2.1) and the Stirling formula proves the lower estimate with

$$d_1(s) = \sqrt{\frac{2}{\pi}} \frac{1}{2^s} - \varepsilon$$

for large enough n. ∎

3 The case when $s = cn$

The other end.

In the previous section the case when s is much smaller than n was cosidered. Suppose now that n is divisible by 4 and $s = \frac{n}{4}$. Let $\mathcal{S}$ be an $\frac{n}{4}$-strongly separating family. Then $A, B \in \mathcal{S}, (A \neq B)$ divide X into four parts of $\frac{n}{4}$ elements each. Associate a vector with coordinates $1, -1$ with a member A of $\mathcal{S}$ writing 1 in the ith position iff the ith element of X is in A. Denote the vectors obtained in this way from the members of $\mathcal{S}$ by $v_1, v_2, \ldots v_m$. It is easy to see that the inner product $v_i v_j$ is 0 for $1 \leq i < j \leq m$. Let v_0 have 1s in each coordinate. Then $v_0 v_i = 0$ also holds $(1 \leq i \leq m)$. That is, $v_0, v_1, \ldots, v_m$ are pairwise orthogonal vetors in an n-dimensional space. Then $m \leq n-1$ follows. The following little statement is obtained.

Proposition 3.1 *Let n divisible by 4. Then*

$$h\left(n, \frac{n}{4}\right) \leq n - 1$$

with equality iff there is a Hadamard matrix.

Consider now the case when $n = 4s + t$ with a small positive integer t. Then $|v_i v_j| \leq t$ must hold for all $0 \leq i < j \leq m$. Since $|v_i| = n$ $(0 \leq i \leq m)$ therefore the normed version of v_i is $u_i = \frac{v_i}{|v_i|} = \frac{v_i}{\sqrt{n}}$. Their lengths are 1, on the other hand $|u_i u_j| \leq \frac{t}{n}$ holds for $0 \leq i < j \leq m$. We arrived to a geometric problem: what is the maximum number m of points on the surface of the n-dimensional unit ball B^n (it is called the $n-1$-dimensional spherical space $\mathbb{S}^{n-1}$) if the angle φ between any two vectors determined by these points satisfies $|\cos\varphi| \leq \frac{t}{n}$.

A slightly different version of this problem is widely studied: what is the maximum number $M(d, \varphi)$ of points on the surface of the $d = (n-1)$-dimensional spherical space $\mathbb{S}^d$ if the angle between any two vectors determined by these points is at least φ. For good surveys see [7] and [6] .

In our case the vectors $u_0, u_1, \ldots, u_m, -u_0, -u_1, \ldots, -u_m$ possess the property that their pairwise angle is at most $\arccos\frac{t}{d+1}$. Denote the maximum number of points $P_1, \ldots, P_m$ in the d-dimensional spherical space such that the angle between any two of these points and their opposites is at least φ by $N(d, \varphi)$. The inequality $2N(d, \varphi) \leq M(d, \varphi)$ is obvious. It does not always hold with equality as the case $d = 2, \varphi = \arccos\frac{1}{3}$ shows, since $M(2, \arccos\frac{1}{3}) = 9$ while $N(2, \arccos\frac{1}{3}) = 4$. We have proved the following lemma.

Lemma 3.2

$$h(4s+t, s) \leq N(4s+t-1, \arccos\frac{t}{4s+t})-1 \leq \frac{1}{2}M(4s+t-1, \arccos\frac{t}{4s+t})-1.$$

In order to obtain good estimates on $h(4s + t, s)$ for fixed t and large s one should study the corresponding values of $M(d, \varphi)$ and $N(d, \varphi)$. E.g. it seems to be not difficult to prove the inequality $h(4s+1, s) \leq 4s+1$, but $h(5, 1) = 4$ by Theorem 1.1 showing that the equality does not always hold.

The middle case.

Here we want to find $h(n, cn)$ where c is a constant and n is large. Our only result is Theorem 1.10.

Proof of Theorem 1.10. This upper estimate is a trivial combination of Lemma 3.2 and a theorem of Kabatjanskiĭ and Levenštein [11] claiming

$$M(d,\varphi) \le (1-\cos\varphi)^{-\frac{d}{2}} 2^{-0.099d++o(d)} \quad (\varphi \le 62.9974\ldots^\circ).$$

∎

We have no reasonable lower estimate. Theorem 2.1 does not help, since its estimate tends to 0 when $s = cn$ and n tends to ∞.

4 Open problems

Restriction with the pairwise entropy.

Let us repeat a problem posed in [12]. If $p = (p_1, \ldots, p_m)$ is a probability distribution then its *entropy* is

$$H(p) = \sum_{i=1}^{m} -p_i \log p_i.$$

Define the *entropy* of a pair of sets $A, B \subset X$ as the entropy of the probability distribution

$$\left(\frac{|A \cap B|}{|X|}, \frac{|\overline{A} \cap B|}{|X|}, \frac{|A \cap \overline{B}|}{|X|}, \frac{|\overline{A} \cap \overline{B}|}{|X|} \right).$$

What is the maximum number of subsets $A_1, \ldots, A_m$ of an n-element set X if the entropy of any pair A_i, A_j $(i \ne j)$ is at least ρ? The case $\rho = 4$ is solved by Proposition 3.1.

Combining the problems.

The combinations of the problems of Theorems 1.4, 1.5 and 1.6 give rise to many new problems.

References

[1] Alon, Noga, Explicit construction of exponential sized families of k-independent sets, *Discrete Mat.* **58**(1986) 191-193.

[2] Bey, Christian, personal communication.

[3] BOLLOBÁS, B., Sperner systems consisting of pairs of complementary subsets, *J. Combin. Theory Ser A* **15**(1973) 363-366.

[4] BOLLOBÁS, B., On generalized graphs, *Acta Math. Hungar* **16**(1965) 447-452.

[5] BRACE, A. AND DAYKIN, D.E. Sperner-type theorems for finite sets, in D.J.A. Welsch and D.R. Woodall, editors, *Combinatorics, (Proc. Conf. Combinatorial Math., Math. Inst., Oxford, 1972*, pp. 18-37, Inst. Math. Appl., Southend-on-Sea (1972).

[6] CONWAY, J.H. AND SLOANE, N.J.A. *Sphere Packings, Lattices and Groups*, 2nd edition. Springer Verlag, New York, 1993.

[7] FEJES TÓTH, G., Spherical Space, in *Handbook of Computational Geometry* pp. 32-34.

[8] FRANKL, P, An extremal problem of coding type, *Ars Combin* **1**(1976) 53–55.

[9] GARGANO, L., KÖRNER, J. AND VACCARO, U., Sperner capacities, *Graphs and Combin.* **9**(1993) 31-56.

[10] GRAHAM, R.L. AND SLOANE, N.J.A., Lower bounds for Constant Weight Codes, *IEEE Trans. Inform. Theory* **26**(1980) 37-43.

[11] KABATJANSKIĬ, G.A., LEVENŠTEIN, V.I., Bounds for packings on a sphere and in space, *Problemy Peredachy Inf.* **14**(1)(1978) 3-25 and *Problems of Inf. Trans.* **14**(1)(1978) 1-17.

[12] KATONA, G.O.H., Two applications (for search theory and truth functions) of Spener type theorems, *Periodica Math. Hungar.* **3**(1973) 19-26.

[13] LUBELL, D., A short proof of Sperner's lemma, *J. combin. Theory* **1**(1966) 299.

[14] KÖRNER, J. AND MONTI, A., Delta -Systmes and Qualitative (In)Dependence, *J. Combin. Theory* **99**(2002) 75–84.

[15] KÖRNER, J. AND SIMONYI, G., A Sperner-type theorem and qualitative independence, *J. Comb. Theory Ser A*, **59**)(1992) 90–103

[16] Kleitman, D. and Spencer, J., Families of k-indpendent sets, *Discrete Math* **6**(1973) 255-262.

[17] Marczewski, E., Indépendance d'ensembles et prolongement de mesures, *Colloq. Math.* **1**(1948) 122-132.

[18] Mešalkin, L.D., A generalization of Sperner's theoremon the number of subsets of a finite set, *Teor. Verojatnost. i Primen.* **8**(1963) 219-220 (in Russian with German summary).

[19] Poljak, S., Pultr, A. and Rödl, V., On the dimension of Kneser graphs, in *Algebraic Methods in Graph Theory, Szeged, Hungary,* Coll. Soc. Math. J. Bolyai, pp. 631-646.

[20] Poljak, S., Pultr, A. and Rödl, V., On qualitatively independent partitions and related problems, *Discrete Appl. Math.* **6**(1983) 193-205.

[21] Poljak, S. and Rödl, V., Orthogonal partitions and covering of graphs, *Czech Math. J.* **30**(1980) 475-485.

[22] Poljak, S., Tuza, Z., On the maximu number of qualitatively independent partititons, *J. Combin. Theory Ser A* **51**(1989) 111-116.

[23] Rényi, A., *Foundations of Probability,*Wiley, New York, 1971.

[24] Schönheim, J. On a problem of Purdy related to Sperner systems, *Canad. Math. Bull.* **17**(1974) 135-136.

[25] Sperner, E., Ein Satz über Untermengen einer endlichen Menge, *Math. Z.* **27**(1928) 544-548.

[26] Weiss, Benjamin, personal communication.

[27] Yamamoto, K., Logarithmic order of free distributive lattices, *J. Math. Soc. Japan* **6**(1954) 347-357.

[28] Zhang, Zhen, LYM-TYpe Inequalities for tEC/AUED Codes, *IEEE Trans. IT* **39**(1993) 232-238.