Probabilistic Methods in Extremal Finite Set Theory

Noga Alon *
Department of Mathematics
Raymond and Beverly Sackler Faculty of Exact Sciences
Tel Aviv University, Tel Aviv, Israel

Abstract

There are many known applications of the Probabilistic Method in Extremal Finite Set Theory. In this paper we describe several examples, demonstrating some of the techniques used and illustrating some of the typical results obtained. This is partly a survey paper, but it also contains various new results.

*Research supported in part by a United States Israel BSF Grant and by a Bergmann Memorial Grant

The Probabilistic Method is a powerful tool in tackling many problems in Combinatorics. Roughly speaking, the method works as follows: Trying to prove that a combinatorial structure (or a substructure of a given one) with certain desired properties exists, one defines an appropriate probability space of structures and then shows that the desired properties hold in this space with positive probability.

Extremal Finite Set Theory is one of the most rapidly developing areas in Combinatorics, which has applications in various other branches of Mathematics and Computer Science including Discrete Geometry, Functional Analysis, Probability Theory and Circuit Complexity. There are numerous known applications of probabilistic arguments in Extremal Finite Set Theory. These include many examples of probabilistic proofs of existence of families of sets with certain properties, as well as proofs of results whose statement often does not seem to suggest any connection to probability. In this paper we describe examples of both types.

It is worth mentioning that our choice of the examples described here is limited by the length of this paper and obviously reflects a personal choice as well. It is impossible to even mention all the beautiful known probabilistic proofs in the area. Many additional interesting examples and important techniques can be found in the excellent survey [16], as well as in the nice books [7] and [24]. Here we first present simple known probabilistic proofs of three basic results, and then describe several examples of random constructions supplying proofs of existence of certain set systems. Afterwards we describe a few more recent results, and conclude with a section containing various applications of the properties of the entropy function of a random variable to Extremal Finite Set Theory.

1 Probabilistic proofs of three basic results

Each of the three results presented in this section has numerous generalizations, extensions and applications. Their simple probabilistic proofs demonstrate nicely the basic application of probabilistic arguments.

The first result is Sperner's Theorem [25], considered by many researchers to be the starting point of Extremal Finite Set Theory. Recall that a family $\mathcal{F}$ of subsets of $\{1, \ldots, n\}$ is called an *antichain* if no set of $\mathcal{F}$ is contained in another.

Theorem 1.1 *Let $\mathcal{F}$ be an antichain. Then*

$$\sum_{A\in\mathcal{F}} \frac{1}{\binom{n}{|A|}} \leq 1$$

Proof Let σ be a uniformly chosen permutation of $\{1,\ldots,n\}$ and set

$$\mathcal{C}_\sigma = \{\{\sigma(j) : 1 \leq j \leq i\} : 0 \leq i \leq n\}$$

(The cases $i = 0, n$ give $\emptyset, \{1,\ldots,n\} \in \mathcal{C}_\sigma$, respectively.) Define a random variable

$$X = |\mathcal{F} \cap \mathcal{C}_\sigma|$$

Clearly

$$X = \sum_{A\in\mathcal{F}} X_A$$

where X_A is the indicator random variable for $A \in \mathcal{C}$. Thus

$$E(X_A) = \Pr(A \in \mathcal{C}_\sigma) = \frac{1}{\binom{n}{|A|}}$$

since $\mathcal{C}_\sigma$ contains precisely one set of size $|A|$, which is distributed uniformly among the $|A|$-sets. By linearity of expectation

$$E(X) = \sum_{A\in\mathcal{F}} \frac{1}{\binom{n}{|A|}}$$

For *any* σ, $\mathcal{C}_\sigma$ forms a chain - every pair of sets is comparable. Since $\mathcal{F}$ is an antichain we *must* have $X = |\mathcal{F} \cap \mathcal{C}_\sigma| \leq 1$. Thus $E(X) \leq 1$. □

Corollary 1.2 ([25]) *Let $\mathcal{F}$ be an antichain. Then*

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

Proof The function $\binom{n}{x}$ is maximized at $x = \lfloor n/2 \rfloor$ so that

$$1 \geq \sum_{A\in\mathcal{F}} \frac{1}{\binom{n}{|A|}} \geq \frac{|\mathcal{F}|}{\binom{n}{\lfloor n/2 \rfloor}}.$$

□

The second basic result presented here is the Erdős Ko Rado Theorem [13]. A family $\mathcal{F}$ of sets is called *intersecting* if for every $A, B \in \mathcal{F}$ $A \cap B \neq \emptyset$.

Theorem 1.3 ([13]) *Suppose $n \geq 2k$ and let $\mathcal{F}$ be an intersecting family of k-element subsets of an n-set. Then $|\mathcal{F}| \leq \binom{n-1}{k-1}$.*

Observe that the result is sharp, as shown by taking the family of all k-sets containing a particular point. The short proof given here is due to G. Katona (1972). Let $\mathcal{F}$ be an intersecting family of subsets of $\{0, \ldots, n-1\}$.

Lemma 1.4 *For $0 \leq s \leq n-1$ set $A_s = \{s, s+1, \ldots, s+k-1\}$ where addition is modulo n. Then $\mathcal{F}$ contains at most k of the sets A_s.*

Proof By symmetry we can suppose that $A_0 \in \mathcal{F}$. The only sets A_s that intersect A_0 other than A_0 itself are the $2k-2$ sets A_s with $-(k-1) \leq s \leq k-1$, $s \neq 0$ (where the indices are taken modulo n). These sets can be partitioned into $k-1$ pairs of disjoint sets, A_i, A_{i+k}, where $-k \leq i \leq -1$. Since $\mathcal{F}$ can contain at most one set of each such pair the assertion of the lemma follows. □

Now we prove the Erdős Ko Rado Theorem. Let a permutation σ of $\{0, \ldots, n-1\}$ and $i \in \{0, \ldots, n-1\}$ be chosen randomly, uniformly and independently and set $A = \{\sigma(i), \sigma(i+1), \ldots, \sigma(i+k-1)\}$, addition again modulo n. Conditioning on any choice of σ the Lemma gives $\Pr(A \in \mathcal{F}) \leq k/n$. Hence $\Pr(A \in \mathcal{F}) \leq k/n$. But A is uniformly chosen from all k-sets so

$$\frac{k}{n} \geq \Pr(A \in \mathcal{F}) = \frac{|\mathcal{F}|}{\binom{n}{k}},$$

implying that

$$|\mathcal{F}| \leq \frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}.$$

□

We conclude this section with a theorem of Bollobás [6]. The proof presented here is due to Jaeger and Payan [17] and Katona [19].

Let $\mathcal{F} = \{(A_i, B_i)\}_{i=1}^h$ be a family of pairs of subsets of an arbitrary set. We call $\mathcal{F}$ a (k,l)-*system* if $|A_i| = k$ and $|B_i| = l$ for all $1 \leq i \leq k$, $A_i \cap B_i = \emptyset$ and $A_i \cap B_j \neq \emptyset$ for all $1 \leq i, j \leq k$.

Theorem 1.5 *If $\mathcal{F} = \{(A_i, B_i)\}_{i=1}^h$ is a (k,l)-system then $h \leq \binom{k+l}{k}$.*

Proof Put $X = \cup_{i=1}^h (A_i \cup B_i)$ and consider a random order π of X. For each i, $1 \leq i \leq k$, let X_i be the event that all the elements of A_i precede all those of B_i in this order. Clearly

$$Pr(X_i) = 1/\binom{k+l}{k}.$$

It is also easy to check that the events X_i are pairwise disjoint. Indeed, assume this is false and let π be an order in which all the elements of A_i precede those of B_i and all the elements of A_j precede those of B_j. Without loss of generality we may assume that the last element of A_i does not appear after the last element of A_j. But in this case, all elements of A_i precede all those of B_j, contradicting the fact that $A_i \cap B_j \neq \emptyset$. Therefore, all the events X_i are pairwise disjoint, as claimed. It follows that

$$1 \geq Pr(\cup_{i=1}^{h} X_i) = \sum_{i=1}^{h} Pr(X_i) = h \cdot 1 \binom{k+l}{k},$$

completing the proof. □

Theorem 1.5 is sharp, as shown by the family $\mathcal{F} = \{(A, X \setminus A) : A \subset X, |A| = k\}$, where $X = \{1, 2, \ldots, k+l\}$.

2 Random constructions

Probabilistic arguments are very useful in proofs of existence of combinatorial structures satisfying a set of required properties. Existence proofs of this type are sometimes called *random constructions*. Here is a typical, simple example. A collection $\mathcal{F}$ of subsets of $N = \{1, 2, \ldots, n\}$ is *k-independent* if for every k distinct sets $F_1, F_2, \ldots, F_k$ of $\mathcal{F}$, all the 2^k intersections $\cap_{i=1}^{k} G_i$ are nonempty, where each G_i is either F_i or its complement $N \setminus F_i$. Kleitman and Spencer [21] considered the problem of estimating the maximum possible cardinality of a k-independent family of subsets of an n-set. Their lower bound is proved by a random construction.

Theorem 2.1 *If*

$$\binom{m}{k} 2^k (1 - 2^{-k})^n < 1 \tag{1}$$

then there is a k-independent family of subsets of $N = \{1, \ldots, n\}$ *whose cardinality is at least* m.

Proof Suppose (1) holds, and let $v_1, v_2, \ldots, v_m$ be a sequence of m randomly chosen binary vectors of length n, where each coordinate of each of the vectors v_i is chosen, randomly and independently, to be either 0 or 1 with equal probability. For each fixed set I of k distinct indices $1 \leq i_1 < i_2 < \ldots i_k \leq m$ and each fixed value of $\epsilon = (\epsilon_1, \ldots \epsilon_k) \in \{0,1\}^k$ let $A(I, \epsilon)$ denote the event that there is no coordinate j, $1 \leq j \leq n$ such that $v_{i_l}(j) = \epsilon_l$ for all $1 \leq l \leq k$. Clearly $Pr(A(I, \epsilon)) = (1 - 2^{-k})^n$.

Therefore, by (1), with positive probability none of the events $A(I,\epsilon)$ holds. Let $v_1,\ldots,v_m$ be a fixed sequence for which none of these events holds, and let F_i be the subset of N whose characteristic function is v_i. One can easily check that the family $\mathcal{F}=\{F_1,\ldots F_m\}$ is k-independent, completing the proof. $\Box$

We note that by using the fact that $(1-2^{-k})^n \leq e^{-n/2^k}$, where $e=2.718...$ is the base of the natural logarithm, and by a special simple construction for $k\leq 2$ one can derive from the last theorem that for every n and k there is a k-independent family of subsets of N whose cardinality is at least $\lfloor e^{n/(k2^k)}\rfloor$. There is no known explicit construction of such a large k-independent family, although there are known explicit constructions of such families of size $2^{c_k n}$, where $c_k>0$ is a constant depending only on k. See [2] for more details.

The second random construction we describe, due to Erdős and Fűredi [12], is similar, but has an interesting application in Combinatorial Geometry.

Proposition 2.2 *For every $n\geq 1$ there is a family $\mathcal{F}$ of m subsets of $N=\{1,\ldots,n\}$, where $m=\lfloor \frac{1}{2}(\frac{2}{\sqrt{3}})^n\rfloor$, such that there are no three distinct members A,B and C of $\mathcal{F}$ satisfying*

$$A\cap B\subset C\subset A\cup B. \tag{2}$$

Proof Define $m=\lfloor \frac{1}{2}(\frac{2}{\sqrt{3}})^n\rfloor$, and choose, randomly and independently, $2m$ 0,1-vectors of length n, where each coordinate of each of the vectors independently is chosen to be either 0 or 1 with equal probability. Each vector is the characteristic vector of a corresponding subset of N. For every fixed triple a,b and c of the chosen vectors, the probability that the corresponding sets satisfy equation (2) is precisely $(3/4)^n$. This is because (2) simply means that for each i, $1\leq i\leq n$, neither $a_i=b_i=0,c_i=1$ nor $a_i=b_i=1,c_i=0$ hold. Therefore, the probability that for three fixed indices i,j and k, the sets A,B,C corresponding to the chosen vectors a,b,c respectively satisfy (2) is $(3/4)^n$. Since there are $\binom{2m}{3}3$ possible triples as above, the expected number of triples A,B,C that satisfy (2) is

$$\binom{2m}{3}3(3/4)^n\leq m,$$

where the last inequality follows from the choice of m. Thus there is a choice of a family X of $2m$ subsets of N in which the number of triples A,B,C satisfying (2) is at most m. By deleting one set from each such triple we obtain a family $\mathcal{F}$ of at least $2m-m=m$ subsets of N satisfying

the assertion of the proposition. Notice that the members of $\mathcal{F}$ are all distinct since (2) is trivially satisfied if $A = C$. This completes the proof. $\Box$

There are several striking examples, in different areas of Combinatorics, where the probabilistic method supplies simple counter- examples to long-standing conjectures. The following consequence of the last proposition is such an example.

Theorem 2.3 ([12]) *For every $n \geq 1$ there is a set of at least $\lfloor \frac{1}{2}(\frac{2}{\sqrt{3}})^n \rfloor$ points in the n-dimensional Euclidean space R^n, such that all angles determined by three points from the set are strictly less than $\pi/2$.*

This theorem disproves an old conjecture of Danzer and Grűnbaum [11], that the maximum cardinality of such a set is at most $2n - 1$. We note that as proved by Danzer and Grűnbaum the maximum cardinality of a set of points in R^n in which all angles are at most $\pi/2$ is 2^n.

Proof of Theorem 2.3 We select the points of a set X in R^n from the vertices of the n-dimensional cube. As before, we view the vertices of the cube, which are $0,1$-vectors of length n, as the characteristic vectors of subsets of an n-element set. A simple consequence of Pythagoras' Theorem gives that the three vertices a, b and c of the n-cube, corresponding to the sets A, B and C, respectively, determine a right angle at c if and only if (2) holds. As the angles determined by triples of points of the n-cube are always at most $\pi/2$, the result follows immediately from Proposition 2.2. $\Box$

Our final example for a random construction is more complicated than the previous ones and is only sketched here. let $\mathcal{F}$ and $\mathcal{H}$ be two families of subsets of $N = \{1, \ldots, n\}$. We say that $F \in \mathcal{F}$ *hits* $H \in \mathcal{H}$ if $|F \cap H| = 1$. We say that the family $\mathcal{F}$ *hits* the family $\mathcal{H}$ if for every $H \in \mathcal{H}$ there is an $F \in \mathcal{F}$ that hits it. Let $t(\mathcal{H})$ be the minimum cardinality of a family $\mathcal{F}$ that hits $\mathcal{H}$. Finally, let $t(n)$ denote the maximum possible value of $t(\mathcal{H})$, where the maximum is taken over all families $\mathcal{H}$ of n subsets of n. Our objective is to estimate $t(n)$.

This problem, considered in [1], is motivated by the study of a certain communication network, as we briefly describe below. A *radio network* is a synchronous network of processors that communicate by transmitting messages to their neighbors. A processor receives a message in a given step iff it is silent in this step and precisely one of its neighbors transmits. This model is discussed in various papers, see, e.g., [8], [1] and their references. Suppose that a processor p in this model

has a message which it has to broadcast to all the other processors in the network. Suppose also, for simplicity, that p has a common neighbor with each other processor in the network. In one step p can transmit the message to all its neighbors, and then these have to transmit it to the other processors. It is not difficult to check that the problem of completing all these transmissions in a minimum total number of steps is closely related to the problem of determining $t(\mathcal{H})$ for an appropriately defined family of sets $\mathcal{H}$. In particular, $t(n)$ is equal, up to a constant factor, to the maximum possible number of steps required to complete the broadcast task in an n-processors network of the above type.

In [8] it is shown that there exists a positive constant c such that $t(n) \leq c(\log_2 n)^2$. As shown in [1] this is sharp, up to a constant factor:

Theorem 2.4 *There are two positive constants b and c such that for every n,*

$$b(\log_2 n)^2 \leq t(n) \leq c(\log n)^2.$$

The lower bound is proved in [1] by a random construction. Since the function $(\log_2 n)^2$ changes only by a constant factor when n is replaced by any polynomial in n it suffices to prove the existence of a polynomial size family $\mathcal{H}$ for which $t(\mathcal{H})$ is at least, say, $(\log_2 n)^2/100$. Construct a family $\mathcal{H}$ composed of $0.2 \log n$ subfamilies $\mathcal{H}_l$, each of cardinality n^7. For each l, $0.4 \log n \leq l \leq 0.6 \log n$, let $\mathcal{H}_l$ be a random family of n^7 (not necessarily distinct) subsets H of N chosen as follows: for each $i \in N$, randomly and independently, $Pr(i \in H) = 2^{-l}$. It can be shown that for such an $\mathcal{H}$, the probability that $t(\mathcal{H}) \leq (\log_2 n)^2/100$ is (much) less than 1, establishing the lower bound in theorem 2.4. The detailed proof, given in [1], which is rather complicated and includes various combinatorial arguments and an application of the FKG-Inequality (cf., e.g., [7]), is omitted.

3 Two additional examples

The probabilistic method is most striking when it is applied to prove theorems whose statement does not seem to suggest at all the need of probability. Most of the examples given in the previous sections are simple instances of such statements. In this section we describe two slightly more complicated examples.

The first result solves a conjecture of Daykin and Erdős. Let $\mathcal{F}$ be a family of m distinct subsets of $X = \{1, 2, \ldots, n\}$. Let $d(\mathcal{F})$ denote the number of disjoint pairs in F, i.e.,

$$d(\mathcal{F}) = |\{(F, F') : F, F' \in \mathcal{F}, F \cap F' = \emptyset\}|.$$

Daykin and Erdős conjectured that if $m = 2^{(\frac{1}{2}+\delta)n}$, then, for every fixed $\delta > 0$, $d(\mathcal{F}) = o(m^2)$, as n tends to infinity. More generally, Erdős conjectured that if $\mathcal{F}$ is a family of m subsets of X, and $m = 2^{(1/(k+1)+\delta)n}$, where $\delta > 0$, then

$$d(\mathcal{F}) \leq (1 - \frac{1}{k})\binom{m}{2} + o(m^2)$$

as n tends to infinity.

The more general conjecture is proved in [3]. Since the proof for the general case is somewhat complicated, we describe here only that of the special case $k = 1$, mentioned above. We prove a stronger result, as follows.

Theorem 3.1 ([3]) *Let $\mathcal{F}$ be a family of $m = 2^{(\frac{1}{2}+\delta)n}$ subsets of $X = \{1, 2, \ldots, n\}$, where $\delta > 0$. Then*

$$d(\mathcal{F}) < m^{2-\frac{\delta^2}{2}} \ . \tag{3}$$

Proof Suppose (3) is false and pick independently t members $A_1, A_2, \ldots, A_t$ of $\mathcal{F}$ with repetitions at random, where t is a large positive integer, to be chosen later. We will show that with positive probability $|A_1 \cup A_2 \cup \ldots \cup A_t| > n/2$ and still this union has an empty intersection with more than $2^{n/2}$ distinct subsets of X. This contradiction will establish (3).

In fact

$$\begin{aligned} & Pr(|A_1 \cup A_2 \cup \ldots \cup A_t| \leq n/2) \\ \leq & \sum_{S \subset X, |S| \leq n/2} Pr(A_i \subset S, i = 1, \ldots, t) \\ & \leq 2^n (2^{n/2}/2^{((1/2)+\delta)n})^t = 2^{n(1-\delta t)}. \end{aligned} \tag{4}$$

Define

$$v(B) = |\{A \in \mathcal{F} : B \cap A = \emptyset\}|.$$

Clearly

$$\sum_{B \in \mathcal{F}} v(B) = 2d(\mathcal{F}) \geq 2m^{2-\delta^2/2}.$$

Let Y be the random variable whose value is the number of members $B \in \mathcal{F}$ which are disjoint to all the A_i $(1 \leq i \leq t)$. By the convexity of the function z^t the expected value of Y satisfies

$$E(Y) = \sum_{B \in \mathcal{F}} (v(B)/m)^t = \frac{1}{m^t} \cdot m \left(\frac{\sum v(B)^t}{m} \right) \geq \frac{1}{m^t} \cdot m \left(\frac{2d(\mathcal{F})}{m} \right)^t > 2m^{1-t\delta^2/2} .$$

Since $Y \leq m$ we conclude that

$$Pr(Y \geq m^{1-t\delta^2/2}) \geq m^{-t\delta^2/2}. \tag{5}$$

One can check that for $t = \lfloor 1+1/(\delta - \delta^2/4 - \delta^3/2) \rfloor$, $m^{1-t\delta^2/2} > 2^{n/2}$ and the right-hand side of (5) is greater than the right-hand side of (4). Thus, with positive probability, $|A_1 \cup A_2 \cup \ldots \cup A_t| > n/2$ and still this union has an empty intersection with more than $2^{n/2}$ members of $\mathcal{F}$. This contradiction implies inequality (3). $\Box$

The second example we describe in this section is a very recent result of the author, and it is very likely that the estimate here can be still improved. The problem we consider was raised by Fiat and Naor [14], who were motivated by the study of a method for distributing keys in a certain multi-user crypto-system. The objective is, roughly, to distribute keys among n users, so that each one receives a small set of keys and so that for every pair of users and every set of m other users, the pair would have at least one key in common among the keys that none of the other m have. Formulated as a problem on set-systems, the problem is the following.

Let $\mathcal{F} = \{F_1, \ldots, F_n\}$ be a family of finite sets, and define $N = \{1, 2, \ldots, n\}$. We say that $\mathcal{F}$ has *property* $(2, m)$ if there are no two distinct indices $i, j \in N$ and a subset M of N such that $|M| = m$, $i, j \notin M$ and

$$F_i \cap F_j \subset \cup_{k \in M} F_k.$$

Let $c(\mathcal{F})$ denote the minimum cardinality of a member of $\mathcal{F}$ and let $C(\mathcal{F})$ denote the maximum cardinality of a member of $\mathcal{F}$. For $n \geq m+2 \geq 3$, let $c(m, n)$ denote the minimum possible value of $c(\mathcal{F})$, where the minimum is taken over all families $\mathcal{F}$ of n sets that have property $(2, m)$. Similarly, $C(m, n)$ denotes the minimum possible value of $C(\mathcal{F})$, as $\mathcal{F}$ ranges over all families as above. Our objective is to estimate the two functions $c(m, n)$ and $C(m, n)$. As suggested by the notation, one can consider families that satisfy the naturally defined more general property (k, m), and indeed the results we prove below for property $(2, m)$ can be extended to this more general case. For simplicity we only deal here with property $(2, m)$.

As observed by Fiat and Naor, a simple random construction shows that there exists an absolute constant a such that for all admissible m and n

$$c(m,n) \leq C(m,n) \leq am^2 \log_2 n.$$

To see this, take a set X of $\frac{a}{2}m^3 \log_2 n$ elements and let $F_1, \ldots F_{2n}$ be $2n$ random subsets of X, where for each set F_i and for each element $x \in X$ independently, $Pr(x \in F_i) = 1/m$. An easy calculation shows that for a sufficiently large a (say $a = 20$), there is a positive probability that the collection of all the above $2n$ sets has property $(2, m)$ and at least half of the sets have cardinality at most $am^2 \log_2 n$. By letting $\mathcal{F}$ be a collection of n of the small sets the above inequality follows. We omit the detailed calculation.

Another simple observation is the fact that for all admissible m and n

$$C(m,n) \geq c(m,n) \geq m \log_2((n-1)/m).$$

This can be shown as follows. Let $\mathcal{F} = \{F_1, \ldots F_n\}$ be a family satisfying the property $(2, m)$ and suppose, without loss of generality, that $c(m,n) = c(\mathcal{F}) = |F_n| = c$. For each $1 \leq i \leq n-1$, put $G_i = F_i \cap F_n$. We claim that there is no union of m sets G_i which is contained in a union of another collection of m of the sets G_i. Indeed, suppose this is false and suppose that S_1 and S_2 are two distinct subsets of cardinality m of $\{1, \ldots, n-1\}$ and that $\cup_{s_1 \in S_1} G_{s_1} \subset \cup_{s_2 \in S_2} G_{s_2}$. Choose arbitrarily an index $j \in S_1 \setminus S_2$. Then

$$F_j \cap F_n = G_j \subset \cup_{s_1 \in S_1} G_{s_1} \subset \cup_{s_2 \in S_2} G_{s_2} \subset \cup_{s_2 \in S_2} F_{s_2},$$

contradicting the assumption that $\mathcal{F}$ has property $(2, m)$. Thus the claim is true and there is an antichain of $\binom{n-1}{m}$ subsets of F_n, implying that

$$\binom{n-1}{m} \leq \binom{c}{\lfloor c/2 \rfloor} \leq 2^c.$$

Therefore $c \geq m \log_2((n-1)/m)$, as needed.

It is easy to see that for all admissible m and n, $c(m,n) \leq C(m,n) \leq n-1$. This is because the family $\mathcal{F} = \{F_1, \ldots, F_n\}$ in which F_i is the set of unordered pairs given by $F_i = \{\{i,j\} : 1 \leq j \leq n, j \neq i\}$ has property $(2, m)$ for all $m \leq n-2$. In fact, it is easy to see that $c(n, n-2) = n-1$, since every pair of sets in a collection of n sets that has property $(2, n-2)$ must have a common element that does not belong to any other set in the collection besides these two.

In case $n >> m >> \log n$ the gap between the upper and lower bounds given above is rather large. A better lower bound for these cases is given in the following proposition.

Proposition 3.2 *For every $n \geq m + 2 \geq 3$,*

$$C(m,n) \geq c(m,n) \geq Min\{\frac{n}{2}, \frac{m^2}{8}\}.$$

The multiplicative constants in the above proposition can be improved, but we are only interested here in the asymptotic behaviour of the corresponding functions, up to a constant factor.

Proof Let $\mathcal{F}$ be a family of sets satisfying property $(2,m)$ and let F be a member of $\mathcal{F}$. We must show that

$$|F| \geq Min\{\frac{n}{2}, \frac{m^2}{8}\}.$$

Let x_i denote the number of points in F that belong to i members of $\mathcal{F}$ (including F). If $x_2 \geq n/2$ the desired result holds and hence we may assume that $x_2 \leq \frac{n}{2} - 1$. Therefore, there are at least $n/2$ members G of $\mathcal{F}$ such that any point in the intersection $F \cap G$ belongs to at least 3 members of $\mathcal{F}$. We call these members of $\mathcal{F}$ *F-good*.

Let G be a randomly chosen F-good member of $\mathcal{F}$, chosen uniformly among all F-good members of $\mathcal{F}$. Define $p = \frac{m}{2(n-2)} (\leq 1/2)$ and let $\mathcal{S}$ be a random collection of members of $\mathcal{F}$ obtained by choosing each member of $\mathcal{F}$ other than F and G, randomly and independently, to be a member of $\mathcal{S}$ with probability p. Let Y_1 be the random variable whose value is the number of members of $\mathcal{S}$, and let Y_2 be the random variable whose value is the number of points in $F \cap G$ which do not belong to $\cup_{S \in \mathcal{S}} S$. Define, also $Y = Y_1 + Y_2$.

We claim that the random variable Y is always greater than m. Indeed, suppose this is false and for some choice of G and $\mathcal{S}$ $Y = Y_1 + Y_2 \leq m$. For each of the Y_2 points that contribute to Y_2 (i.e., that lie in $F \cap G$ but not in $\cup_{S \in \mathcal{S}} S$) choose arbitrarily a set T other than F and G that contains this point. (There is always such a set since G is F-good). Let $\mathcal{T}$ be the collection of all these sets T. Then $|\mathcal{T}| \leq Y_2$ and hence $|\mathcal{S} \cup \mathcal{T}| \leq m$. Moreover,

$$F \cap G \subset \cup_{S \in \mathcal{S}} S \cup_{T \in \mathcal{T}} T,$$

contradicting the assumption that $\mathcal{F}$ has property $(2,m)$. Thus $Y > m$, as claimed.

It follows that the expectation of Y is greater than m. Since the expectation of Y_1 is $p(n-2) = m/2$ this implies that $E(Y_2) > m/2$. We next obtain an upper bound for the expectation of Y_2.

Let q be a point of F that belongs to i members of $\mathcal{F}$. The probability that q contributes to Y_2 is the probability that it belongs to G and does not belong to all the members of $\mathcal{S}$. This probability is 0 if $i \leq 2$, since G is chosen among the F-good members of $\mathcal{F}$ and these do not contain such a point q. In case $i > 2$, the probability that q belongs to G is at most $\frac{i-1}{n/2}$, since there are at least $n/2$ F-good sets in $\mathcal{F} \setminus \{F\}$, and $i-1$ of them contain q. Given the choice of G that contains q, the probability that none of the sets in $\mathcal{S}$ contains q is $(1-p)^{i-2}$, since there are precisely $i-2$ members of $\mathcal{F}$ other than F and G that contain q. It follows that the probability that q contributes to Y_2 is at most

$$\frac{2(i-1)}{n}(1-p)^{i-2} \leq \frac{4(i-1)}{n}(1-p)^{i-1} \leq \frac{4(i-1)}{n}e^{-p(i-1)},$$

where here we used the fact that $1-p \geq 1/2$.

Linearity of expectation now implies that

$$E(Y_2) \leq \sum_{i>2} x_i \frac{4(i-1)}{n}e^{-p(i-1)} = \frac{4}{n}\sum_{i>2} x_i(i-1)e^{-p(i-1)}.$$

A simple computation shows that the maximum of the function $g(z) = ze^{-pz}$ for $z \geq 0$ is attained at $z = 1/p$, and this maximum is $1/(ep)$. Therefore,

$$E(Y) \leq \frac{4}{n}\sum_{i>2} x_i \frac{1}{ep} = \frac{4}{enp}\sum_{i>2} x_i \leq \frac{4}{enp}|F|$$

$$= \frac{8(n-2)}{emn}|F| \leq \frac{8}{em}|F| \leq \frac{4}{m}|F|.$$

Since $E(Y_2) > m/2$ this implies that $|F| > m^2/8$, completing the proof. □

4 Some applications of the properties of the entropy function

Let X be a random variable taking values in some range S, and let p_x denote the probability that the value of X is x. The *binary entropy* of X, denoted by $H(X)$ is defined by

$$H(X) = \sum_{x \in S} -p_x \log_2 p_x.$$

The following well known fact is simple but useful.

Proposition 4.1 *Let $X = (X_1, \ldots, X_n)$ be a random variable taking values in the set $S = S_1 \times S_2 \times \ldots \times S_n$, where each of the coordinates X_i of X is a random variable taking values in S_i. Then*

$$H(X) \leq \sum_{i=1}^{n} H(X_i).$$

Proof Let $p(x_1, \ldots, x_n)$ denote the probability that $X = (x_1, \ldots, x_n)$ and let $p(i : x_i)$ denote the probability that $X_i = x_i$. Since the summation of $p(x_1, \ldots, x_n)$ over all $x_j \in S_j, j \neq i$ is precisely $p(i : x_i)$ it follows from the definition of the entropy function that

$$H(X) - \sum_{i=1}^{n} H(X_i) = \sum_{i=1}^{n} \sum_{x_i \in S_i} -p(x_1, \ldots, x_n) \log_2 \frac{p(x_1, \ldots, x_n)}{p(1 : x_1)p(2 : x_2) \ldots p(n : x_n)}$$

$$= \sum_{i=1}^{n} \sum_{x_i \in S_i} p(1 : x_1)p(2 : x_2) \ldots p(n : x_n) f(\frac{p(x_1, \ldots, x_n)}{p(1 : x_1)p(2 : x_2) \ldots p(n : x_n)}),$$

where here $f(z) = -z \log_2 z$. By the convexity of the function $f(z)$ and by Jensen Inequality the last quantity is at most $-x \log_2 x$ where

$$x = \sum_{i=1}^{n} \sum_{x_i \in S_i} p(1 : x_1)p(2 : x_2) \ldots p(n : x_n)(\frac{p(x_1, \ldots, x_n)}{p(1 : x_1)p(2 : x_2) \ldots p(n : x_n)}) = 1.$$

Therefore $H(X) - \sum_{i=1}^{n} H(X_i) \leq -1 \log_2 1 = 0$, completing the proof. □

The above proposition is used in [20] to derive several interesting applications in Extremal Finite Set Theory, including an upper estimate for the maximum possible cardinality of a family of k-sets in which the intersection of no two is contained in a third. The basic idea in [20] can be illustrated by the following simple corollary of Proposition 4.1.

Corollary 4.2 *Let $\mathcal{F}$ be a family of subsets of $\{1, 2, \ldots, n\}$ and let p_i denote the fraction of sets in $\mathcal{F}$ that contain i. Then*

$$|\mathcal{F}| \leq 2^{\sum_{i=1}^{n} H(p_i)},$$

where $H(y) = -y \log_2 y - (1 - y) \log_2 (1 - y)$.

Proof Associate each set $F \in \mathcal{F}$ with its characteristic vector $v(F)$, which is a binary vector of length n. Let $X = (X_1, \ldots, X_n)$ be the random variable taking values in $\{0, 1\}^n$, where $Pr(X = v(F)) = 1/|\mathcal{F}|$ for all $F \in \mathcal{F}$. Clearly $H(X) = |\mathcal{F}|(-\frac{1}{|\mathcal{F}|} \log \frac{1}{|\mathcal{F}|}) = \log |\mathcal{F}|$, and since here $H(X_i) = H(p_i)$ for all $1 \leq i \leq n$, the result follows from Proposition 4.1. □

As observed by Peter Frankl [15], the last corollary supplies a quick proof for the well known estimate, (that follows, e.g., from the results in [9]), that for every integer n and for every real $0 < p \leq 0.5$, $\sum_{i \leq np} \binom{n}{i} \leq 2^{nH(p)}$. Indeed, let $\mathcal{F}$ be the family of all subsets of cardinality at most pn of $\{1, 2, \ldots, n\}$. If p_i is the fraction of subsets of $\mathcal{F}$ that contain i, then clearly $p_i \leq p$ for all i. Since the function $H(p)$ is increasing for $0 \leq p \leq 0.5$ this, together with Corollary 4.2 implies that

$$\sum_{i \leq np} \binom{n}{i} = |\mathcal{F}| \leq 2^{\sum_{i=1}^{n} H(p_i)} \leq 2^{nH(p)},$$

as needed.

An interesting extension of Proposition 4.1 is proved in [10]. As in that proposition, let $X = (X_1, \ldots, X_n)$ be a random variable taking values in the set $S = S_1 \times S_2 \times \ldots \times S_n$, where each X_i is a random variable taking values in S_i. For a subset I of $\{1, 2, \ldots, n\}$, let $X(I)$ denote the random variable $(X_i)_{i \in I}$. With these notations, the following proposition is proved in [10] for the case $S_i = \{0, 1\}$ for all i. The proof for the general case is analogous, and we omit it.

Proposition 4.3 *Let $X = (X_1, \ldots, X_n)$ and S be as above. If $\mathcal{G}$ is a family of subsets of $\{1, \ldots, n\}$ and each $i \in \{1, \ldots, n\}$ belongs to at least k members of $\mathcal{G}$ then*

$$kH(X) \leq \sum_{G \in \mathcal{G}} H(X(G)).$$

□

Corollary 4.4 ([10]) *Let S be a finite set, and let $\mathcal{F}$ be a family of subsets of S. Let $\mathcal{G} = \{G_1, G_2, \ldots G_m\}$ be a collection of subsets of S, and suppose that each element of S belongs to at least k members of $\mathcal{G}$. For each $1 \leq i \leq m$ define $\mathcal{F}_i = \{F \cap G_i : F \in \mathcal{F}\}$. Then*

$$|\mathcal{F}|^k \leq \prod_{i=1}^{m} |\mathcal{F}_i|.$$

Proof Suppose $S = \{1, \ldots, n\}$, and define $S_i = \{0, 1\}$ for $1 \leq i \leq n$. Let $X = (X_1, \ldots, X_n)$ be the random variable taking values in S, where for each $F \in \mathcal{F}$ X is equal to the characteristic vector of F with probability $1/|\mathcal{F}|$. By Proposition 4.3

$$kH(X) \leq \sum_{i=1}^{m} H(X(G_i)).$$

But $H(X) = \log_2 |\mathcal{F}|$, whereas $H(X(G_i)) \leq \log_2 |\mathcal{F}_i|$, implying the desired result. □

A special case of the last corollary is proved in [23], in a different method.

Another application of Proposition 4.3 is given in [4]. The *d-dimensional grid* $G_{n,d}$ is the graph formed by the product of d n-vertex paths. It has $N = n^d$ vertices and $dn^{d-1}(n-1)$ edges. For a spanning tree T of $G = G_{n,d}$, let $V(T)$ denote the average distance in T between u and v, where the average is taken over all edges uv of G. Motivated by the study of a certain game on trees, it is proved in [4] that there exists an absolute constant $c > 0$ such that for every $n, d \geq 2$, there is a spanning tree T of G such that $V(T) \leq c \log_2 N$. Moreover, there exists another absolute constant $c' > 0$, such that for every spanning tree T of G, $V(T) \geq c' \log_2 N$. The proof of the lower bound relies on Proposition 4.3, together with some additional ideas. The somewhat lengthy details can be found in [4].

For a family of subsets $\mathcal{A}$ of $S = \{1, 2, \ldots, n\}$, define the *average distance in* $\mathcal{A}$, denoted $dist(\mathcal{A})$, by

$$dist(\mathcal{A}) = \frac{1}{|\mathcal{A}|^2} \sum_{A \in \mathcal{A}} \sum_{B \in \mathcal{A}} d(A, B),$$

where $d(A, B)$ is the cardinality of the symmetric difference of A and B (i.e., the cardinality of $(A \setminus B) \cup (B \setminus A)$). The authors of [5] proved that for every family $\mathcal{A}$ as above,

$$dist(\mathcal{A}) \geq \frac{n+1}{2} - \frac{2^{n-1}}{|\mathcal{A}|}.$$

This result is sharp for $|\mathcal{A}| = 2^{n-1}$ (and, of course, for $|\mathcal{A}| = 2^n$), but it is very far from being sharp for smaller values of $|\mathcal{A}|$.

The following result supplies a better estimate for smaller values of $|\mathcal{A}|$.

Proposition 4.5 *For every family $\mathcal{A}$ of $|\mathcal{A}| = a$ subsets of $\{1, \ldots, n\}$:*

$$dist(\mathcal{A}) \geq \frac{n}{2} - \log_e(\frac{2^n}{a}).$$

We note that it is easy to show that if $a = |\mathcal{A}| \geq \sum_{i=1}^{t} \binom{n}{i}$ then $dist(\mathcal{A}) \geq \Omega(t)$, which is better than the above estimate for very small values of a. The above proposition is useful to estimate how close to $n/2$ $dist(\mathcal{A})$ must be when $|\mathcal{A}| = a = 2^{(1-o(1))n}$.

The above proposition will be derived from Corollary 4.2, together with the following simple technical lemma.

Lemma 4.6 *For every real x, $0 < x < 1$,*

$$2x(1-x) \geq -x\log_e x - (1-x)\log_e(1-x) + 1/2 - \log_e 2.$$

Prof Define $f(x) = 2x(1-x) + x\log_e x + (1-x)\log_e(1-x) - 1/2 + \log_e 2$. Then $f'(x) = -4(x - \frac{1}{2}) + \log_e(x/(1-x))$, $f''(x) = -4 + \frac{1}{x(1-x)}$. Therefore, $f(1/2) = f'(1/2) = 0$ and $f''(x) \geq 0$ for all $0 < x < 1$. This easily implies that $f(x) \geq 0$ for all $0 < x < 1$, as needed. □

Proof of Proposition 4.5 Let $\mathcal{A}$ be a family of a subsets of $\{1, \ldots, n\}$, and let p_i denote the fraction of subsets of $\mathcal{A}$ that contain i, $(1 \leq i \leq n)$. By Corollary 4.2

$$\sum_{i=1}^{n} H(p_i) \geq \log_2 a.$$

Define $H_e(z) = -z\log_e z - (1-z)\log_e(1-z)$ and observe that by the last inequality

$$\sum_{i=1}^{n} H_e(p_i) \geq \log_e a.$$

Clearly, the number of ordered pairs (A, B) of subsets of $\mathcal{A}$ for which i is in the symmetric difference of A and B is precisely $2|\mathcal{A}|p_i|\mathcal{A}|(1-p_i)$. Therefore, by Lemma 4.6,

$$dist(\mathcal{A}) = \sum_{i=1}^{n} 2p_i(1-p_i) \geq \sum_{i=1}^{n} H_e(p_i) + n/2 - n\log_e 2$$

$$\geq \log_e a + n/2 - n\log_e 2 = \frac{n}{2} - \log_e(\frac{2^n}{a}),$$

completing the proof. □

Combining Proposition 4.5 with the convexity of the function $-z\log_e z$ we obtain the following.

Corollary 4.7 *If $\mathcal{A}_1, \ldots, \mathcal{A}_k$ is a partition of all 2^n subsets of an n-set into k pairwise disjoint sets, then*

$$\sum_{i=1}^{k} \frac{|\mathcal{A}_i|}{2^n} dist(\mathcal{A}_i) \geq \frac{n}{2} - \log_e k.$$

This improves, for all $k \geq 4$, the lower bound of $(n+1-k)/2$ derived for the above quantity in [5].

The final result we mention here deals with vectors over Z_k, (and not with binary vectors, which correspond naturally to subsets of a set). We describe it here, since the basic ideas used in its proof are similar to the ones described in this section. This result answers a question of Lapidot and Shamir, and although it may look somewhat artificial it is naturally suggested, as shown in [22], in the study of the possibility to parallelize certain two-prover zero-knowledge protocols.

Let $Z_k^n \times Z_k^n$ denote the set of all ordered pairs (u, v), where u and v are vectors of length n over $Z_k = \{0, \ldots, k-1\}$. We say that a subset $\mathcal{F}$ of $Z_k^n \times Z_k^n$ has *property P* if there are $u_0, \ldots, u_{k-1}$ and $v_0, \ldots v_{k-1}$ in Z_k^n, such that $(u_i, v_j) \in \mathcal{F}$ for all $0 \leq i, j \leq k-1$, and there is a coordinate l such that the value of u_i as well as that of v_i in that coordinate is i for all $0 \leq i \leq k-1$. Let $f(n, k)$ denote the maximum possible cardinality of a family $\mathcal{F} \subset Z_k^n \times Z_k^n$ which does not have property P.

The following theorem supplies an estimate for $f(n, k)$.

Theorem 4.8 *There are two absolute positive constants c_1 and c_2 such that for every $k \geq 2$ and for every $n \geq 1$,*

$$k^{2n(1-c_1/(k^2 \log_2 k))} \leq f(n,k) \leq k^{2n(1-c_2/(k^2(\log_2 k)^2))}.$$

The lower bound is very simple, but the proof of the upper bound is more complicated and relies on several probabilistic arguments some of which depend on the properties of the entropy function.

References

[1] N. Alon, A. Bar-Noy, N. Linial and D. Peleg, *A lower bound for radio broadcast*, J. Comp. Sys. Sci., to appear.

[2] N. Alon, *Explicit construction of exponential sized families of k-independent sets*, Discrete Math. 58 (1986), 191-193.

[3] N. Alon and P. Frankl, *The maximum number of disjoint pairs in a family of subsets*, Graphs and Combinatorics 1 (1985), 13-21.

[4] N. Alon, R. M. Karp and D. West, *A graph-theoretic game and its application to the k-server problem*, in preparation.

[5] I. Althőfer and T. Sillke, *An average distance inequality for subsets of the cube*, to appear.

[6] B. Bollobás, *On generalized graphs*, Acta Math. Acad. Sci. Hungar. 16 (1965), 447-452.

[7] B. Bollobás, *Combinatorics*, Cambridge University Press, Cambridge, 1986.

[8] R. Bar-Yehuda, O. Goldreich and A. Itai, *Broadcast in radio networks; an exponential gap between determinism and randomization*, Proc. 4^{th} ACM Symp. on Principles of Distributed Computing, 1986, 98-107.

[9] H. Chernoff, *A measure of the asymptotic efficiency for tests of a hypothesis based on the sum of observations*, Ann. Math. Stat. 23 (1952), 493-509

[10] F. R. K. Chung, P. Frankl, R. L. Graham and J. B. Shearer, *Some intersection theorems for ordered sets and graphs*, J. Combinatorial Theory, Ser. A 43 (1986), 23-37.

[11] Danzer and Grűnbaum *Űber zwei Probleme bezűglich konvexer Kőrper von P. Erdős und von V. L. Klee*, Math. Z. 79 (1962), 95-99.

[12] P. Erdős and Z. Fűredi, *The greatest angle among n points in the d-dimensional Euclidean space*, Annals of Discrete Math. 17 (1983), 275-283.

[13] P. Erdős, C. Ko and R. Rado, *Intersection theorems for systems of finite sets*, Quart. J. Math. Oxford (2) 12 (1961), 313-320.

[14] A. Fiat and M. Naor, Private communication.

[15] P. Frankl, Private communication.

[16] Z. Fűredi, *Matchings and covers in hypergraphs*, Graphs and Combinatorics 4 (1988), 115-206.

[17] F. Jaeger and C. Payan, *Nombre maximal d'arétes d'un hypergrphe critique de rang h*, C. R. Acad. Sci. Paris 273 (1971), 221-223.

[18] G. O. H. Katona, *A simple proof of the Erdős Ko Rado Theorem*, J. Combinatorial Theory Ser. 13 (1972), 183-184.

[19] G. O. H. Katona, *Solution of a problem of Ehrenfeucht and Mycielski*, J. Combinatorial Theory Ser. A 17 (1974), 265-266.

[20] D. J. Kleitman, J. B. Shearer and D. Sturtevant, *Intersection of k-element sets*, Combinatorica 1 (1981), 381-384.

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

[22] D. Lapidot and A. Shamir, *Parallel two-prover zero-knowledge protocols*, to appear.

[23] L. H. Loomis and H. Whitney, *An inequality related to the isoperimetric inequality*, Bull. Amer. Math. Soc. 55 (1949), 961-962.

[24] J. Spencer, *Ten Lectures on the Probabilistic Method*, SIAM, Philadelphia, 1987.

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