On K-wise Independent Distributions and Boolean Functions

Itai Benjamini * Ori Gurel-Gurevich † Ron Peled ‡

August 7, 2007

Abstract

We pursue a systematic study of the following problem. Let $f : \{0,1\}^n \to \{0,1\}$ be a (usually monotone) boolean function whose behaviour is well understood when the input bits are identically independently distributed. What can be said about the behaviour of the function when the input bits are not completely independent, but only k-wise independent, i.e. every subset of k bits is independent? more precisely, how high should k be so that any k-wise independent distribution "fools" the function, i.e. causes it to behave nearly the same as when the bits are completely independent?

In this paper, we are mainly interested in asymptotic results about monotone functions which exhibit sharp thresholds, i.e. there is a critical probability, p_c, such that $P(f = 1)$ under the completely independent distribution with marginal p, makes a sharp transition, from being close to 0 to being close to 1, in the vicinity of p_c. For such (sequences of) functions we define 2 notions of "fooling": K_1 is the independence needed in order to force the existence of the sharp threshold (which must then be at p_c). K_2 is the independence needed to "fool" the function at p_c.

In order to answer these questions, we explore the extremal properties of k-wise independent distributions and provide ways of constructing such distributions. These constructions are connected to linear error correcting codes.

We also utilize duality theory and show that for the function f to behave (almost) the same under all k-wise independent inputs is equivalent to the function f being well approximated by a real polynomial in a certain fashion. This type of approximation is stronger than approximation in L_1.

We analyze several well known boolean functions (including AND, Majority, Tribes and Percolation among others), some of which turn out to have surprising properties with respect to these questions.

In some of our results we use tools from the theory of the classical moment problem, seemingly for the first time in this subject, to shed light on these questions.

*Weizmann Institute, Rehovot, 76100, Israel. itai.benjamini@weizmann.ac.il

†Weizmann Institute, Rehovot, 76100, Israel. ori.gurel-gurevich@weizmann.ac.il

‡UC Berkeley. peledron@stat.berkeley.edu

1 Introduction

Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function whose behaviour is well understood when the input bits are independent and identically distributed, with probability p for each bit to be 1. As an example we may consider the majority function, Maj, whose output is the bit which occurs more in the input (suppose that n is odd). When $p = 1/2$ we know that the output is also distributed uniformly. When $p < 1/2$ the output tends to be 0. More precisely, if $p < 1/2$ is constant, the probability of $\mathrm{Maj} = 1$ decays exponentially fast with n.

Suppose, however, that the input bits are not truly IID . For example, they might be the result of a derandomization procedure. A reasonable, but weaker assumption would be that the probability of each bit to be 1 is still p, and that they are k-wise independent, i.e. the distribution of any k of the bits is independent.

Under this assumption, what can be said about the distribution of f? For fixed p, which k (as a function of n) is enough to guarantee the same asymptotic behaviour? Majority turns out to be relatively easy to analyze: $k = 2$ is enough to guarantee that for any fixed $p < 1/2$, the probability of $\mathrm{Maj} = 1$ tends to 0 (though only polynomially fast), while for $p = 1/2$, we have $P(\mathrm{Maj} = 1)$ guaranteed to tend to $1/2$ if and only if $k = \omega(1)$ ("Guarantee" here means that Maj behaves as prescribed under any k-wise independent distribution). In fact, for $p = 1/2$ we have more precise results, that $|P(\mathrm{Maj} = 1) - 1/2| \leq O(1/\sqrt{k})$ under any k-wise independent distribution. As can been seen, the k needed to "fool" majority at $p \neq p_c$ (which we denote K_1) is much smaller then the k needed to "fool" majority at p_c (which we denote K_2). This phenomenon is shared by the other functions we explore, and we provide a partial explanation. Other functions exhibit much more complex behaviour and the required analysis is accordingly complex. We pursue a systematic study of the above question.

k-wise independent distributions are often used in computer science for derandomization of algorithms. This was initiated by the papers [2], [13], [25], [31] and further developed in [10], [36], [32], [41], [26], [27] and others (see [33] for a survey). For derandomization one checks that the algorithm still behaves (about) the same on a particular k-wise independent input as in the completely independent case. The question we ask is of the same flavor, for a given boolean function f, we ask how much independence is required for it to behave about the same on all k-wise independent inputs (including the completely independent one).

Typically, k-wise independent distributions are constructed by sampling a uniform point of a small sample space, which is usually also a linear subspace ([24], [23], [34]). In this work, like in the works of [26], [27], we do not impose this restriction and consider general k-wise independent distributions. Still, our work is of interest even for the reader only interested in the more restrictive model since, on the one hand, anything we show is impossible would still be impossible in that model and on the other hand almost all of our constructions are of the linear subspace type. Interestingly, in section 4.8 we give an example where the general and more restrictive case give asymptotically different results, i.e., that the general distribution case is richer, not just up to constants, in what can be achieved with it.

The tools we use include the duality of linear programming, in section 3.1, used to show an *equivalence* between our question and the question of approximating the function f by a real polynomial in a certain "sandwich L_1" approximation (stronger than ordinary L_1 approximation). This connects our results to the subject of approximation of boolean functions, used for example in learning theory (e.g. [29], [38], [5], [40]).

In section 3.2 we recall a theorem about weak convergence of distributions, later used to give sharp bounds on K_2 very easily. In section 3.3 we introduce a tool from the Theory of the Classical Moment Problem (TCMP), seemingly for the first time in this context. In sections 4.8 and 4.9 we use it to prove bounds on the maximal and minimal probabilities of all bits to

be 1 under a k-wise independent distribution, in a simple way. We then observe that if $p = \frac{1}{q}$ for a prime-power q, then an upper bound on this maximal probability translates to a lower bound on the size of a symmetric sample space for k-wise independent $GF(q)$-valued random variables, we apply our upper bound to obtain new lower bounds for such sample spaces. For the binary case $q = 2$ our bound equals the well-known bound of [2], [13].

In section 4 we explore K_1 and K_2 for various boolean functions, and also prove some general theorems. In section 4.6 we present a novel construction of a distribution (of the linear subspace type) designed to change the behaviour of a particular function. We use a variation of the $(u \mid u + v)$ construction of error-correcting codes [34] and we would like to emphasize the technique used there. We think there is a shortage of ways to construct k-wise independent distributions with specified properties and that this technique will be useful for changing the behaviour of other functions as well.

The approach in this paper is a little different than that usually taken in pseudo-random generators (see [39]). There one seeks a distribution under which all functions from a certain complexity class behave the same as on fully independent bits. In contrast, we start with a function f and wish to show that it behaves the same on all k-wise independent inputs. Still, one may expect this to hold if the function f is "simple enough". Indeed, a conjecture of Linial and Nisan [30] makes this precise when f is a function from the class AC0. In section 4.3 we recall the precise conjecture and make some modest progress towards confirming it.

There are other notions of "simple functions". Another such notion is that the function be noise stable [7], i.e., having most of its Fourier mass on constant level coefficients. In section 4.7 we show a connection between the Fourier spectrum and the behaviour on k-wise independent inputs, but surprisingly show that a noise stable function can behave very differently on k-wise independent inputs than on fully independent inputs even when k grows fast with n.

There is also a lot of interest in *almost* k-wise independent distributions ([37],[4],[3],[6],[42],[16]), though our questions can equally be formulated for that case, in our work we concentrate only on perfect k-wise independence, this is both because it seems the analysis is simpler for perfect k-wise independence and they could serve as a starting point for further research and because we think the perfect k-wise independent case is interesting on its own.

2 Basic definitions and properties

We begin with a definition

Definition 1 *Let $\mathcal{A}(n, k, p)$ be the set of all k-wise independent distributions $\mathbb{Q}$ on n bits $(X_1, \ldots, X_n)$ with $\mathbb{Q}(X_i = 1) = p$ for all i.*

Also denote by $\mathbb{P}_p$ the fully independent distribution on n bits, each with probability p to be 1.

In most of the sequel we will be concerned with understanding

$$\max_{\mathbb{Q} \in \mathcal{A}(n,k,p)} \mathbb{Q}(f = 1) \qquad \text{and} \qquad \min_{\mathbb{Q} \in \mathcal{A}(n,k,p)} \mathbb{Q}(f = 1) \tag{1}$$

for a boolean function $f : \{0,1\}^n \to \{0,1\}$ and given n, k and p. We first note that $\mathcal{A}(n, k, p)$ is a convex set, since, if the distribution on a subset of k bits is independent with marginal p in both $\mathbb{Q}_1$ and $\mathbb{Q}_2$ then it is so also in $\alpha\mathbb{Q}_1 + (1 - \alpha)\mathbb{Q}_2$.

This implies that the extremal values in (1) are attained at extreme points of $\mathcal{A}(n, k, p)$, hence if we could only find all these extreme points we could then find the values (1) for all f. Unfortunately saying anything about these extreme points appears to be very difficult and so in the sequel we will need to resort to special methods for each function f considered.

For later reference, we identify the two extreme points of $\mathcal{A}(n, n-1, \frac{1}{2})$. XOR0 is the distribution on $(X_1, \ldots, X_n)$ having $\{X_i\}_{i=1}^{n-1}$ IID and $X_n \equiv \sum_{i=1}^{n-1} X_i \bmod 2$, and XOR1 is the same with $X_n \equiv 1 + \sum_{i=1}^{n-1} X_i \bmod 2$.

We next define precisely what we mean by "k large enough so that f behaves on all k-wise independent inputs the same as on the fully independent input".

Definition 2 $\varepsilon^f(k,p) = \max_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f=1) - \min_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f=1)$
$k^f(\varepsilon,p)$ *is the minimal* k *such that* $\varepsilon^f(k,p) < \varepsilon$.

We will be mostly interested in asymptotic (in n) results. Let $f_n : \{0,1\}^n \to \{0,1\}$ be a sequence of monotone boolean functions. Assume that the functions have a sharp threshold, i.e. there is a p_c such that $\lim_{n\to\infty} \mathbb{P}_p(f=1)$ is 0 if $p < p_c$, 1 if $p > p_c$.

For example, any sequence of balanced monotone transitive functions has a sharp threshold, as is proved by Friedgut and Kalai [17].

Definition 3 K_1 *is the class of functions* $k(n)$, *such that* $\varepsilon(k,p) \to 0$ *for any* $p \neq p_c$.
K_2 *is the class of functions* $k(n)$, *such that* $\varepsilon(k,p_c) \to 0$.

In other words, K_1-wise independence is enough to guarantee the existence of sharp threshold (which is then necessarily at p_c), while K_2-wise independence is enough to guarantee that f behaves as if the bits were completely independent, when $p = p_c$.

Notice that while K_1 and K_2 are classes of functions, we occasionally abuse the formal notation, and write, as above, K_1-wise independence. Similarly, we write $K_1 > k(n)$ to indicate that $k(n)$ does not belong to K_1, or $K_2 < \omega(1)$ to indicate $K_2 \supset \omega(1)$, etc.

It is not a-priori clear whether $K_1 \leq K_2$ or vice versa (or neither). Consult the appendix for a partial result. In all the examples we encountered K_2 is at least $\omega(K_1)$.

3 General tools

In this section we discuss some general tools for finding K_1 and K_2 as defined in the previous section.

3.1 Duality - Approximation by polynomials

We note that the values (1) are the solution to a simple linear program. What is the dual of this program? We observe

Proposition 4 *For any* $f : \{0,1\}^n \to \{0,1\}$, *any* k *and any* $0 < p < 1$.

$$\begin{aligned} \max_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f=1) &= \min_{P\in P_k^+(f)} \mathbb{E}_{\mathbb{P}_p} P(X_1,\ldots,X_n) \\ \min_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f=1) &= \max_{P\in P_k^-(f)} \mathbb{E}_{\mathbb{P}_p} P(X_1,\ldots,X_n) \end{aligned} \tag{2}$$

where $P_k^+(f)$ *is the set of all real polynomials* $P : \mathbb{R}^n \to \mathbb{R}$ *of degree not more than* k *satisfying* $P \geq f$ *on all points of the boolean cube.* P_k^- *is defined analogously with* $P \leq f$.

The proof is simple using linear programming duality. We deduce that $\varepsilon^f(k,p) < \varepsilon$ is equivalent to having two polynomials $P^+ \geq f$ and $P^- \leq f$ of degree not more than k with $\mathbb{E}_{\mathbb{P}_p}(P^+ - P^-) < \varepsilon$. We call this type of approximation of f a *"sandwich* L_1*" approximation.* In section 4.7 we show that it is strictly stronger than L_1 approximation (by real polynomials of degree not more than k). Whether it is stronger than L_2 approximation is one of our main open questions.

3.2 Distributions determined by their moments

Definition 5 *We say that a real random variable X has* distribution determined by its moments *if any random variable Y satisfying $\mathbb{E}X^m = \mathbb{E}Y^m$ for all integer $m \geq 1$ has the same distribution as X.*

We shall often use the following principle

Proposition 6 *Suppose a sequence of RV's $\{X_n\}_n$ satisfies for all m, $\mathbb{E}X_n^m \to \mathbb{E}X^m$ for some RV X whose distribution is determined by its moments. Then $X_n \to X$ in the weak sense.*

For the proof, see [15], section 2.3 . We remark that a distribution is determined by its moments whenever these do not grow too fast. The best criterion is called Carleman's condition (see [15]). But for our purposes it will mostly be enough to know that the Normal and Poisson distributions are determined by their moments.

3.3 Bounds from the classical moment problem

Given a real sequence $\mathcal{S} := \{s_m\}_{m=0}^k$, with k even and $s_0 = 1$, let

$$\mathcal{A}_{\mathcal{S}} = \{\mathbb{Q} \mid \mathbb{Q} \text{ a probability distribution on } \mathbb{R},\ s_m = \mathbb{E}_Q(X^m) \text{ for } 0 \leq m \leq k\} \tag{3}$$

be all probability distributions with these first k moments (X is a random variable distributed according to $\mathbb{Q}$). In the theory of the classical moment problem [1], [28], based on $\mathcal{S}$ a certain sequence of real functions ρ_m is defined and the following theorem is proved

Theorem 7 *[1, 2.5.2 and 2.5.4] For any x and any $\mathbb{Q}_1, \mathbb{Q}_2 \in \mathcal{A}_{\mathcal{S}}$*

$$|\mathbb{Q}_1(X \leq x) - \mathbb{Q}_2(X < x)| \leq \rho_{\frac{k}{2}}(x) \tag{4}$$

and in particular by taking $\mathbb{Q}_1 = \mathbb{Q}_2$ we get $\max_{\mathbb{Q} \in \mathcal{A}_{\mathcal{S}}} \mathbb{Q}(X = x) \leq \rho_{\frac{k}{2}}(x)$.

For brevity we do not give the general definitions of ρ_m here but differ them to the appendix. In the case of interest for us $s_m := \mathbb{E}(X^m)$ when $X \sim \mathrm{Bin}(n, p)$ and then $\rho_m(x) := (\sum_{j=0}^m P_j^2(x))^{-1}$ where $\{P_j\}_j$ are the (normalized) Krawtchouk polynomials (see [43]). These polynomials are very well known and from them we easily deduce the following (see appendix for a proof)

$$\rho_m(n) = \frac{p^n}{\mathbb{P}(\mathrm{Bin}(n, 1-p) \leq m)} \tag{5}$$

$$\rho_m(\frac{n}{2}) \leq \frac{2}{\sqrt{m}} \qquad \text{for } p = \frac{1}{2}, \text{ even } n \text{ and even } m \leq \frac{n}{2} \tag{6}$$

In many cases, the theory also has constructions achieving the bound of theorem 7. However, these are not necessarily supported by the integers, which we require. It might be that they can be suitably modified to give sharp results in our cases.

4 Boolean functions

In this section we investigate K_1 and K_2 for several boolean functions, and also present some general theorems. We start with a simple, but already non-trivial example.

4.1 Majority

Let Maj_n be the majority function on n bits (for odd n). Let $S_n = \sum_{i=1}^n x_i$, where x_i are the input bits. Let $\overline{S_n} = (2S_n - n)/\sqrt{n}$. The central limit theorem implies that under $\mathbb{P}_{1/2}$, $\overline{S_n} \to N(0,1)$. Identifying K_1 is easy

Theorem 8 *$K_1(\mathrm{Maj}) = 2$.*

Proof. Obviously, $k = 1$ is not in K_1. However, for $\mathbb{Q} \in \mathcal{A}(n, 2, p)$ we have $\mathbb{E}_{\mathbb{Q}}(S_n) = np$ and $\mathrm{Var}_{\mathbb{Q}}(S_n) = np(1-p)$. If, WLOG, $p < 1/2$ then by Chebyshev's inequality

$$\mathbb{Q}(S_n > n/2) \le \mathbb{Q}((S_n - np) > n(1/2 - p)) \le \frac{np(1-p)}{(n(1/2-p))^2} = O(\frac{1}{n}) \to 0 \quad \blacksquare$$

Identifying K_2 is harder. The ideas of section 3.2 give the following

Proposition 9 *$K_2(\mathrm{Maj}) \le \omega(1)$*

Proof. Consider the distribution of S_n under some $\mathbb{Q} \in \mathcal{A}(n, k, 1/2)$. Obviously, $E_{\mathbb{Q}}(S_n^l) = E_{\mathbb{P}_{1/2}}(S_n^l)$ for any $l \le k$. The same holds for $\overline{S_n}$ as it is a linear function of S_n. Therefore, $E_{\mathbb{Q}_n}(\overline{S_n}^l) \to s_l$ where $s_l = \mathbb{E}(N(0,1)^l)$ is the l-th moment of a standard normal distribution. The normal distribution is determined by its moments. Hence, if $k(n) \in \omega(1)$ and $\mathbb{Q}_n \in \mathcal{A}(n, k(n), 1/2)$ then $\overline{S_n} \to N(0,1)$ weakly by proposition 6. In particular, $\mathbb{Q}_n(\mathrm{Maj}_n = 1) = \mathbb{Q}_n(\overline{S_n} > 0) \to 1/2$. $\blacksquare$

In fact for Maj we can be much more specific.

Theorem 10 *There exists a $C > 0$ such that for any even $2 \le k < n$*

$$\frac{C}{\sqrt{k \log k}} \le \max_{Q \in \mathcal{A}(n,k,\frac{1}{2})} |\mathbb{Q}(\mathrm{Maj}_n = 1) - \frac{1}{2}| \le \frac{2\sqrt{2}}{\sqrt{k}} \tag{7}$$

And when $\mathbb{Q}_0 \in \mathcal{A}(n, n-1, \frac{1}{2})$ is the XOR0 distribution we have $|\mathbb{Q}_0(\mathrm{Maj}_n = 1) - \frac{1}{2}| \ge \frac{1}{3\sqrt{n}}$.

The theorem implies that $K_2 = \omega(1)$, but is much stronger in that it bounds $\varepsilon^{\mathrm{Maj}}(k, \frac{1}{2})$.

Proof. The claim about XOR0 is easy to verify directly. The lower bound comes from a direct construction sketched in the appendix. The upper bound is actually known in the context of error-correcting codes [34, Ch. 9, thm. 23] and it appears the proof there also works in our case. But we point out that a very simple proof of it can be obtained just by applying theorem 7 and (6) to the distribution of S_n and this proof even improves a little on the constant. $\blacksquare$

4.2 Tribes

Let m be an integer and let $n = m2^m$ and let $m(n)$ be its inverse function. Tribes_n is the following function: Let the input bits be divided into 2^m sets of size m each, called *tribes*. Let y_i be the AND of the bits in the i-th tribe. Then Tribes_n is the OR of the y_i's. Let $S_n = \sum_{0 \le i < 2^m} y_i$. Then $\mathrm{Tribes}_n = 0$ iff $S_n = 0$. Under $\mathbb{P}_{\frac{1}{2}}$, $S_n \to Poisson(1)$. It is easily checked that Tribes is a sequence of monotone functions with sharp threshold at $p_c = 1/2$ and $\mathbb{P}_{\frac{1}{2}}(\mathrm{Tribes}_n = 1) \to 1 - 1/e$.

Theorem 11 *For some $C > 0$, $Cm(n) \le K_1(\mathrm{Tribes}) \le 2m(n)$*

Proof. The proof is similar to that of proposition 8. First, notice that for $\mathbb{Q} \in \mathcal{A}(n, m(n), p)$ we have $\mathbb{Q}(y_i = 1) = p^m$. For $p < 1/2$, a union bound now yields $\mathbb{Q}(\max_{0 \le i < 2^m} y_i = 1) \le (2p)^m \to 0$. If $\mathbb{Q} \in \mathcal{A}(n, 2m(n), p)$ then the y_is are pairwise independent. Using Chebyshev's inequality on S_n yields the desired result for $p > 1/2$.

For the lower bound we use equation 14 to produce a $\mathbb{Q} \in \mathcal{A}(n, Cm(n), p)$ such that the probability of all bits of any tribe are 1 is 0. ■

Theorem 12 $K_2(\text{Tribes}) \le \omega(m(n)) = \omega(\log(n))$

Proof. This is like the proof of 9. There is no need to normalize S_n as it tends to Poisson(1) as is. Again, we only need to check that Poisson distribution satisfies Carleman's condition. ■

A more refined results, like those for Maj can be reached using Theorem 7.

Theorem 13 $\varepsilon^{\text{Tribes}_n}(km(n), 1/2) \le \frac{2}{(k/2)!}$

4.3 AC^0 functions

AC^0 is the class of functions computable by boolean circuits using Not gates, a polynomial number of AND and OR gates (with unlimited fan-in) and of bounded depth. Tribes is a notable example of an AC^0 function of depth 2. Linial and Nisan ([30]) conjectured that any boolean circuit of depth d and size s has $K_2 \supset \omega(\log^{d-1} s)$.

We prove a very special case of this conjecture. Let $n = 2^{2m}$ and let the input bits be divided into disjoint sets, A_i, consisting of m bits each. A function is *paired* if it is the OR of AND gates, each operating on the bits in exactly 2 of the A_i's. A paired function is, in particular, an AC^0 function of depth 2.

Theorem 14 *If f is paired then $K_2(f) \le \omega(\log n)$*

proof sketch. Let $S(f)$ be the number of satisfied AND gates in f. The crux of the proof is to trim f by removing some of the AND gates to produce a function f', which is (a) very close to f under any $\omega(\log n)$-wise independent distribution, and (b) $S(f')$ under $\mathbb{P}_p$ tends to a RV which is determined by its moments. ■

4.4 Majority of majorities

Let m be an odd integer and let $n = m^2$. Maj^2 is the following function: divide the input bits into m disjoint sets of size m. Let y_i be the majority of the i-th set, then Maj^2 is the majority of the y_i's.

Theorem 15 $K_1(\text{Maj}^2) = 2$

Proof. The proof of 8 yields $\mathbb{Q}(y_i = 1) \le 1/(m(1-2p)^2)$ for any $\mathbb{Q} \in \mathcal{A}(n, k, p)$, when $p < 1/2$. Therefore $E_{\mathbb{Q}}(\sum_i y_i) \le 1/(1-2p)^2$. The y_i's are not pairwise independent, but Markov's inequality is enough: $\mathbb{Q}(\text{Maj}_n^2 = 1) \le 2/(m(1-2p)^2) \to 0$. ■

Notice that this proof applies also to i-levels majority, Maj^i, defined similarly on m^i bits.

Theorem 16 $\sqrt{n} \le K_2 \le \omega(\sqrt{n})$

Proof. To show that any function $k \in \omega(\sqrt{n})$ belongs to $K_2(\text{Maj}^2)$ notice that if $\mathbb{Q} \in \mathcal{A}(n, k, 1/2)$ then the distribution generated on the y_i's belongs to $\mathcal{A}(n, k/m, 1/2)$. Since the y_i's enter majority to produce the output, it is enough, by theorem 9 to have $k/m = \omega(1)$ in order for the output to tend to 1/2.

To show that $k = m-1$ is not in K_2, let $\mathbb{Q}$ be the following distribution: $\mathbb{Q}$ is $XOR0$ on each A_i and completely independent on different A_i's. Obviously, $\mathbb{Q} \in \mathcal{A}(n, m-1, 1/2)$. By theorem 10, $\mathbb{Q}(y_i = 1) \geq 1/2 + 1/3\sqrt{m}$ (assume WLOG that $(n+1)/2$ is even). Let $S_n = \sum_{i=0}^{m-1} y_i$ and $\overline{S_n} = (2S_n - m)/\sqrt{m}$. Since the y_i's are independent we have that $\overline{S_n} \to N(a, 1)$ where $a = \lim(2\mathbb{Q}(y_i = 1) - 1)\sqrt{m} \geq 2/3$. Obviously, $\mathbb{Q}(\mathrm{Maj}^2 = 1) = \mathbb{Q}(\overline{S_n} > 0)$ is bounded away from $1/2$. ■

The surprising fact here is the lower bound of $\sqrt{n}$. First it shows an example where K_2 is much larger then $\omega(K_1)$. Second, it demonstrates that L_2 approximation does not imply "Sandwich L_1" approximation (see section 4.7).

4.5 Composition of functions

Maj^2 is a simple example of composition of functions. What can we say about compositions in general?

Let $n = ml$ and let $f = g(h_1, .., h_m)$ where the h_i's receive disjoint sets A_i of l bits each. Assume that h_i's are balanced with respect to p_c and that $p_c(g) = 1/2$.

Theorem 17 *For $\varepsilon \leq \frac{1}{2m}$, $k^f(4m\varepsilon, p_c) \leq \sum_i k^{h_i}(\varepsilon, p_c)$*

Proof. $g(y_1, .., y_m)$ can be expressed as a sum of monomials of the form $\prod y_i \prod(1 - y_j)$, each involving all of the y's. We take the upper and lower "sandwich L_1" approximating polynomials of each h_i (which have degree $k^{h_i}(\varepsilon, p_c)$) and plug the upper in place of any y_i and one minus the lower in place of any $(1 - y_j)$. This produces a polynomial of degree $k = \sum_i k^{h_i}(\varepsilon, p_c)$ which bounds f from above. The error of each monomial, when the distribution is k-wise independent is at most $(1/2 + \varepsilon)^m - 1/2^m \leq m\varepsilon/2^{m-2}$ for $\varepsilon \leq \frac{1}{2m}$. Summing over the monomials we have an error of no more then $4m\varepsilon$. The lower bound is similar. ■

This is a very general bound - we did not put any restriction on g, it can even be nonmonotone. For example, K_2 for the XOR of two (or boundedly many) majorities is still $\omega(1)$.

For $0 < a < 1$, define Maj_a^2 to be the majority of n^a majorities of n^{1-a} bits each. It is easy to see that $K_2 \leq \omega(n^{1-a})$. Theorem 17 gives a bound of $K_2 \leq \omega(n^{3a})$. However, using finer properties of the "sandwich L_1" approximating polynomials of Maj, we can do better.

Theorem 18 $K_2(\mathrm{Maj}_a^2) = \omega(n^{\min(a,1-a)})$

Proof. We use the approximating polynomials of the upper Maj function (the "g") instead of the generic polynomial of theorem 17. These are not only of bounded degree, but also have small coefficients. This implies that the resulting polynomial is of degree $O(m)$ and produces an error of $O(n^{a/2}\varepsilon)$, where m is the degree of the approximating polynomial of the lower Maj functions and ε is their error. Taking $m = n^a$ gives $\varepsilon = 1/\sqrt{n^a} = n^{-a/2}$, as required. ■

4.6 Percolation

Another very interesting example to consider is that of percolation. Briefly, *percolation* on a graph $G = (V, E)$ is a distribution on $\{0, 1\}^V$, where we identify the bits with the states $\{\texttt{open}, \texttt{close}\}$. We refer the reader to [20] for details of the theory of percolation. We denote the set of all k-wise independent percolation with marginal probability p for every vertex to be open by $\mathcal{A}(G, k, p)$. When G is infinite, we are interested in the probability of existence of an infinite cluster of open vertices. This event is a boolean function on infinitely many bits.

Theorem 19 *For $G = \mathbb{Z}^d$ or $G = \mathbb{T}^d$ (the d-ary tree), for any $0 < p < 1$ and any k there exist a $\mathbb{Q} \in \mathcal{A}(G, k, p)$ such that there is an infinite open cluster $\mathbb{Q}$-almost surely, and another such $\mathbb{Q}$ with no infinite open cluster $\mathbb{Q}$-almost surely.*

The positive part of this theorem follows from the following 2 theorems about finite versions of percolation. Let f be the function indicating an open crossing of the $n \times n$ grid.

Theorem 20 $2^{\sqrt{\log\log n}} = (\log n)^{1/\sqrt{\log\log n}} \leq K_1(f) \leq \omega(\log n)$

For the tree case, we need to diverge slightly from the boolean valued setting. Let f be the number of open paths from the root to the leaves of $\mathbb{T}_n^d$, the n-levels d-ary tree.

Theorem 21 *For any p, for $k = C \log n$, there is a $\mathbb{Q} \in \mathcal{A}(\mathbb{T}_n^d, k, p)$ such that $E_{\mathbb{Q}}(f) \geq 2$.*

To this end, we present a way of combining k-wise independent distributions to "amplify" the amount of independence, inspired by the $(u \mid u+v)$ lemma of error-correcting codes. Let $\mathbb{Z}_r$ be the cyclic group of size r. Let $\mathcal{A}^r(n,k)$ be the set of all k-wise independent distributions on vectors $(X_1, \ldots, X_n) \in \mathbb{Z}_r^n$ with each X_i uniform in $\mathbb{Z}_r$. Define $A^r(G,k)$ similarly.

Lemma 22 *Fix $m \geq 1$. Let $X := (X_1, \ldots, X_n) \in \mathcal{A}^r(n,k)$. Let $X^i := (X_j^i)_{j=1}^n$ be m IID copies of X. Let also $Y := (Y_1, \ldots, Y_n) \in \mathcal{A}^r(n, 2k+1)$ be a vector independent of all the X's. Then the vector with the following coordinates*

$$\begin{array}{cccc} X_1^1 + Y_1, & X_2^1 + Y_2, & \ldots, & X_n^1 + Y_n, \\ X_1^2 + Y_1, & X_2^2 + Y_2, & \ldots, & X_n^2 + Y_n, \\ \vdots, & \vdots, & \vdots, & \vdots, \\ X_1^m + Y_1, & X_2^m + Y_2, & \ldots, & X_n^m + Y_n \end{array} \tag{8}$$

is in $\mathcal{A}^r(mn, 2k+1)$

Consult the appendix for a proof of a more general result.

Proof. (Sketch, of theorem) We build distributions in $A^r(\mathbb{T}_n^d, k)$ such that when we identify 0 with `open` and the rest with `close`, we get the desired percolation for $p = 1/r$.

The proof goes by induction. For $k = 1$ (i.e. no independence) a suitable distribution is just taking X_i to be identical and n to be large enough.

Assume we have $X \in \mathcal{A}^r(\mathbb{T}_n^d, k)$ such that $E_X(f) \geq 2$. We will construct a suitable $Z \in \mathcal{A}^r(\mathbb{T}_m^d, 2k+1)$ for $m = n + n^2 k \log d$. Let X^i be independent copies of X and let $Y \in \mathcal{A}^q(n, 2k+1)$be such that probability of $Y = 0$ is maximal, which is roughly d^{-nk}, because there are about d^n RVs in Y. Using lemma 22 we now assign the RVs in $X^i + Y^i$ to the vertices of $\mathbb{T}_m^d$ such that each is assigned to a subtree of depth n with root at a level divisible by n. Thus, with probability d^{-nk} we have $Y = 0$ and then the open paths form a Galton-Watson tree with an expectation of $2^{m/n} = 2^{1+nk\log d}$. Thus, the total expectation is $d^{-nk} 2^{1+nk\log d} = 2$. ■

Notice that having an open path from the root to the leaves of $\mathbb{T}_n^d$ is an AC^0 function of depth 2. This is an example of a rather complicated depth 2 function, very different then the paired functions considered in section 4.3. Also, this function does not exhibit a sharp threshold, thus the need for different terminology.

4.7 Fourier transform and K_2

In this section we consider only the case $p = \frac{1}{2}$. The quantity $E_{\mathbb{Q}}(f)$ may be represented using the fourier transform as $\sum \hat{f}(S)\hat{\mathbb{Q}}(S)$. When we consider f as having values of ± 1 we have $\sum \hat{f}^2(S) = \sum f^2(S) = 2^n$. Therefore $\hat{f}^2(S)/2^n$ is a probability measure on all subsets of the bits, called the *Fourier mass*. Now, a distribution is k-wise independent if and only if all of its Fourier coefficients of levels between 1 and k (inclusive) are 0. Therefore, if the fourier mass of f

is supported by the first k levels, then $\mathbb{E}_{\mathbb{Q}}(f) = \hat{f}(\emptyset) = \mathbb{P}_{\frac{1}{2}}(f)$. One might conjecture that if most of the Fourier mass is on the first k levels then $E_{\mathbb{Q}}(f)$ would be small for all $\mathbb{Q} \in \mathcal{A}(n, k, 1/2)$.

In [29] and [22], Linial, Mansour and Nisan with an improvement by Håstad prove that any AC^0 function has its fourier mass concentrated on the first $O(\log^{d-1} s)$ levels (where s is the size and d the depth). Had the above conjecture been true, we would have proved that $K_2 = \omega(\log^{d-1} s)$ immediately for any such AC0 function (see section 4.3).

However, Maj^2 provides a counterexample for this conjecture, as its $K_2 > \sqrt{n}$ while its fourier mass is concentrated on the bounded levels, i.e, for any $\varepsilon > 0$ there exists $C > 0$ such that all but ε of the mass is below level C. This is because Maj^2 is a composition of *noise stable* functions and is therefore noise stable itself (see [7]). Of course, Maj^2 is not an AC^0 function so this conjecture might still be true in that domain.

4.8 Maximal probability that all bits are 1

In this section we investigate the maximal probability that all the bits are 1, i.e, the AND function. At the end of the section two applications of our bounds are given.

Define $M(n, k, p) := \max_{\mathbb{Q} \in \mathcal{A}(n,k,p)} \mathbb{Q}(\text{All bits are 1})$ then

Theorem 23 *For even k*

$$M(n, k, p) \leq \frac{p^n}{\mathbb{P}(\mathrm{Bin}(n, 1-p) \leq \frac{k}{2})} \tag{9}$$

Proof. Fix $\mathbb{Q} \in \mathcal{A}(n, k, p)$, let S count the number of bits which are 1. Since S has the same first k moments as a $\mathrm{Bin}(n, p)$ the result follows immediately from theorem 7 and (5) applied to S. ■

Bound (9) is a powerful bound in that it seems to give good results for most ranges of the parameters. Here are some corollaries

Corollary 24

$$M(n, k, p) \leq 2\sqrt{k} \left(\frac{kp}{2e(1-p)(n - \frac{k}{2})} \right)^{\frac{k}{2}} \qquad \textit{For any } n, \textit{ even } k \textit{ and } p \tag{10}$$

$$M(n, k, p) \leq 10 p^n \qquad k \textit{ even, } n(1-p) \leq \frac{k}{2} \tag{11}$$

We add that it is possible to get a result similar to (10) by letting S count the number of bits which are 1, considering $(S - pn)^k$ and applying Chebyshev's inequality. Still our approach with theorem 7 has the following advantages. First, it is quite simple as the above proof of theorem 23 shows. Second, it gives (10) in all ranges of the parameters n, k and p, estimating $\mathbb{E}(S - pn)^k$ appears to become difficult when k also grows with n, or when np is small. Third, it seems to give slightly better results, the approach with Chebyshev's inequality apparently does not give the factor 2 inside the brackets of (10).

We are also able to obtain *exact* results, for $k = 2, 3$. This is done by adapting the closed-form expressions appearing in Boros and Prekopa [9] to our settings.

Proposition 25 *Let $M := \lfloor (n-1)(1-p) \rfloor$ and $\delta := \{(n-1)(1-p)\}$ (integer and fractional parts respectively). And also $N := \lfloor (n-2)(1-p) \rfloor$ and $\varepsilon := \{(n-2)(1-p)\}$ then*

$$M(n, 2, p) = \frac{p}{M+2} + \frac{\delta^2 - \delta(1+p) + p}{(M+1)(M+2)} \tag{12}$$

$$M(n, 3, p) = \frac{p^2}{N+2} + \frac{p(\varepsilon^2 - \varepsilon(1+p) + p)}{(N+1)(N+2)} = M(n-1, 2, p)p \tag{13}$$

For lower bounds on $M(n,k,p)$ and an exact result for small p, see the appendix.

We present two applications of our bounds. First a definition, for q a prime-power and $k \geq 2$, a matrix $B \in M_{R \times n}(GF(q))$ is an OA(n,k,q), or an *Orthogonal array of strength k with q levels* (see [34] and [23]) if a uniformly chosen row $(X_1, \ldots, X_n)$ of it has k-wise independent entries, each uniform in $GF(q)$. If the rows of B form a linear subspace, then B is called a *linear orthogonal array* and is referred to by its generator matrix $A \in M_{m,n}$ whose rows are a basis for the rows of B. We call A a GOA(n,k,q) for short.

1. The bound (9) can be used to give another proof of the Rao bound (see [23]) on the minimal size of orthogonal arrays over $GF(q)$. To see this, suppose B is an OA(n,k,q) for k even, with R rows. We may assume B contains the all zeroes vector. Consider the distribution $\mathbb{Q} \in \mathcal{A}(n, k, \frac{1}{q})$ obtained by sampling uniformly a row of B and mapping each coordinate to a bit by $0 \mapsto 1$, other elements to 0. We have $\mathbb{Q}(\text{All ones vector}) = \frac{1}{R}$, hence by (9) we now get $R \geq q^n \mathbb{P}(\text{Bin}(n, 1 - \frac{1}{q}) \leq \frac{k}{2})$ which is the Rao bound, or using the less refined (10) we get $R \geq \left(\frac{2e(q-1)(n-\frac{k}{2})}{k} \right)^{\frac{k}{2}} / 2\sqrt{k}$.

   We mention in this context that for $q = 2$, this lower bound is equal to the bound $m(n,k) := \sum_{i=0}^{\frac{k}{2}} \binom{n}{i}$ which also appeared in [2] (in a more general setting) but we note that for $q = 2$ we obtain a somewhat stronger result, the bound (9) is in fact an upper bound for the size of any atom of the distribution (by xoring a constant vector), hence for this case we improve slightly the known results by adding that the maximum atom of the distribution is bounded by $\frac{1}{m(n,k)}$, not just the size of the sample space.

2. Let A be a GOA(n,3,3) with m rows. Since the columns of A are 3-wise linearly independent, a theorem of Meshulam [35] implies that $n = O(\frac{3^m}{m})$. Consider the distribution $\mathbb{Q}$ in $\mathcal{A}(n, 3, \frac{1}{3})$ obtained by sampling a uniform linear combination of the rows of A and mapping to bits by, say, $0 \mapsto 1$ and $1, 2 \mapsto 0$. We have $\mathbb{Q}(\text{All ones vector}) \leq \frac{1}{3^m} = O(\frac{1}{n \log n})$. In contrast, by equation (13) there exist $\mathbb{Q}' \in \mathcal{A}(n, 3, \frac{1}{3})$ with $\mathbb{Q}'(\text{All ones vector}) = \Omega(\frac{1}{n})$. We deduce that there is an asymptotic difference between what distributions obtained from linear orthogonal arrays (by the above method) and general distributions can achieve. This is interesting since most explicit constructions of k-wise independent distributions seem to be based on sampling from linear orthogonal arrays.

4.9 Minimal probability that all bits are 1

Define $m(n,k,p) := \min_{\mathbb{Q} \in \mathcal{A}(n,k,p)} \mathbb{Q}(\text{All bits are 1})$. $m(n,k,p)$ can well be 0, in fact

Proposition 26 *When $k < n$ and $p \leq \frac{1}{2}$ we have $m(n,k,p) = 0$.*

Since for $p = \frac{1}{2}$ we can take the XOR0, or XOR1 distributions according to the parity of n. And for lower p's we can take the AND of this distribution with a fully independent distribution.

For $p \geq \frac{1}{2}$, define $n_c(k,p) := \min\{n \mid m(n,k,p) = 0\}$. Our main result of the section are two sided bounds on $n_c(k,p)$

Theorem 27 *For any $p \geq \frac{1}{2}$*

$$n_c(k,p) \geq \begin{cases} \frac{k}{2(1-p)} + 1 & k \text{ even} \\ \frac{k+1}{2(1-p)} & k \text{ odd} \end{cases} \tag{14}$$

and when $1-p=\frac{1}{q}$ *for a prime-power* q *and* $C>0$ *is a large constant*

$$n_c(k,p) \leq C\frac{k}{1-p}\log(\frac{1}{1-p}) \tag{15}$$

The upper bound is based on the Gilbert-Varshamov bound of error-correcting codes (see [34]) and one extra idea. The lower bound poses the main difficulty and for it we need a different aspect of the TCMP. Fix k and $p \geq \frac{1}{2}$ and let $\mathbb{Q} \in \mathcal{A}(n_c(k,p),k,p)$ satisfy $\mathbb{Q}(\text{All ones vector}) = 0$. Let, as usual, S count the number of 1's. S is supported on $[0,n-1]$ and has the first k moments of a $\text{Bin}(n,p)$. Theorems in the TCMP show this is only possible if n_c satisfies (14). The actual verification is technical and involves calculating determinants. Consult the appendix for more details.

As in the previous section, one can deduce from this a result about orthogonal arrays. Suppose B is an OA(n,k,q) with the property that each row contains the symbol 0. If k is even, then necessarily $n \geq 1+\frac{kq}{2}$ and if k is odd then $n \geq \frac{(k+1)q}{2}$.

5 Open questions

Below we list some of our main open questions:

1. Say anything non-trivial about the extremal points of $\mathcal{A}(n,k,p)$.
2. Is "sandwich L_1" approximation stronger then L_2 approximation?
3. What is $K_2(\text{Maj}^3)$? What is $K_2(\text{Maj}^i)$?
4. What is K_2 for iterated majority of threes?

References

[1] Akhiezer, N. I. (1965) The classical moment problem and some related questions in analysis, Translated by N. Kemmer, Hafner Publishing Co., New York.

[2] Alon N., Babai L. and Itai A. (1986) A Fast and Simple Randomized Parallel Algorithm for the Maximal Independent Set Problem. *J. Algorithms*, **7 (4)**, 567-583.

[3] N. Alon, J. Bruck, J. Naor, M. Naor and R. Roth. Construction of asymptotically good, low-rate error-correcting codes through pseudo-random graphs. IEEE Trans. Info. Theory, 38:509-516, 1992.

[4] N. Alon, O. Goldreich, J. Håstad and R. Peralta, Simple constructions of almost k-wise independent random variables. Random Struct. Algorithms 3 3 (1992), pp. 289-304 (preliminary version in FOCS90) .

[5] J. Aspnes, R. Beigel, M. Furst, S. Rudich: The expressive power of voting polynomials Proc. 23rd ACM Conference STOC, 1991,402-409.

[6] Y. Azar, R. Motwani and J. Naor. Approximating arbitrary probability distributions using small sample spaces. Manuscript, 1990.

[7] I. Benjamini, G. Kalai, O. Schramm, (1999) Noise sensitivity of Boolean functions and applications to percolation, *Publications Mathe'matiques de l'IHE'S*, **90**, 5-43.

[8] I. Benjamini, G. Kozma, D, Romik (2006) Random walks with k-wise independent increments, *Electronic Communications in Probability*, **11**, 100-107

[9] Boros E., Prekopa A. (1989) Closed form two-sided bounds for probabilities that at least r and exactly r out of n events occur, *Mathematics of Operations Research*, **14 (2)**, 317 - 342.

[10] B. Berger and J. Rompel. Simulating ($\log^c n$)-wise independence in NC. Journal of the ACM, 38:1026-1046, 1991.

[11] R. Curto, L. A. Fialkow, Recursiveness, positivity, and truncated moment problems, *Houston J. Math.*, vol 17 (4), 1991.

[12] B. Chor and O. Goldreich, On the power of two-point sampling, Journal of Complexity, vol. 5, pp. 96-106, 1989.

[13] B. Chor, O. Goldreich, J. Hastad, J. Friedman, S. Rudich, and R. Smolensky, "The Bit Extraction Problem or t-Resilient Functions," Proc. 26 th IEEE Symposium on Foundations of Computer Science, 1985, pages 396–407.

[14] Dumer I. and Yekhanin, S. (2004), Long nonbinary codes exceeding the Gilbert-Varshamov bound for any fixed distance, *IEEE Trans. Inform. Theory*, **50 (10)**, 2357–2362.

[15] Durrett R. (1996) Probability: theory and examples, second edition, Duxbury press, Belmont, CA.

[16] G. Even, O. Goldreich, M. Luby, N. Nisan and B. Veličković. Approximations of general independent distributions. In Proc. 24th ACM Symposium on Theory of Computing, pages 10-16, 1992.

[17] Friedgut E. and Kalai G. (1996) Every Monotone Graph Property Has a Sharp Threshold *Proc. Amer. Math. Soc.* **124**, 2993-3002.

[18] M. Fredman, J. Komlos and E. Szemerédi. Storing a sparse table with O(1) worstcase access time. In Proc. 23rd IEEE Symposium on Foundations of Computer Science, pages 165-169, 1982.

[19] Fukuda K. (1999), CDD+ version 0.76a1 (June 8 1999), Institute for Operations Research ETH-Zentrum, CH-8092 Zurich, Switzerland and Department of Mathematics ETFL, CH-1015 Lausanne, Switzerland. http://www.cs.mcgill.ca/ fukuda/soft/cddman/cddman.html.

[20] G.R. Grimmett (1999) Percolation, 2nd Edition, Springer.

[21] O. Häggström (1997) Infinite clusters in dependent automorphism invariant percolation on trees, *Annals of Probability*, **25** no. 3, 1423-1436

[22] Håstad J. (2001), A Slight Sharpening of LMN, *J. Comput. Syst. Sci.* **63 (3)**, 498-508.

[23] Hedayat, A. S. and Sloane, N. J. A. and Stufken, John (1999) Orthogonal arrays, Theory and applications, With a foreword by C. R. Rao, Springer-Verlag, New York.

[24] A. Joffe. On a set of almost deterministic k-independent random variables. Annals of Probability, 2:161-162, 1974.

[25] R. Karp and A. Wigderson. A fast parallel algorithm for the maximum independent set problem. J. ACM, 32: 762-773, 1985.

[26] D. Koller and N. Megiddo. Constructing small sample spaces satisfying given constraints. In Proc. of the 25th Annual ACM Symposium on Theory of Computing, pages 268-277, 1993.

[27] H. Karloff and Y. Mansour. On construction of k-wise independent random variables. In Proc. of the 26th Annual ACM Symposium on Theory of Computing, pages 564-573, 1994.

[28] Kreĭn, M. G. and Nudel'man, A. A. (1977) The Markov moment problem and extremal problems, Ideas and problems of P. L. Čebyšev and A. A. Markov and their further development, Translated from the Russian by D. Louvish, Translations of Mathematical Monographs, Vol. 50, AMS.

[29] Linial N., Mansour Y., and Nisan N.. Constant depth circuits, Fourier transform, and learnability. In 30th Annual Symposium on Foundations of Computer Science, pages 574-579, 1989.

[30] N. Linial and N. Nisan (1990) Approximate inclusion-exclusion, *Combinatorica*, **10**, 349-365.

[31] Luby M. (1985) A simple parallel algorithm for the maximal independent set problem *Proc, of the 17'th annual ACM symposium on Theory of computing*. Providence, Rhode Island, US, 1 - 10.

[32] M. Luby. Removing randomness in parallel computation without a processor penalty. J. Comput. Syst. Sci., 47(2):250-286, 1993.

[33] M. Luby and A. Wigderson. Pairwise independence and derandomization. Technical Report TR-95-035, International Computer Science Institute, Berkeley, California, 1995.

[34] MacWilliams, F. J. and Sloane, N. J. A. (1977) The theory of error-correcting codes, North-Holland Mathematical Library, Vol. 16, North-Holland Publishing Co., Amsterdam.

[35] Meshulam R. (1995), On subsets of finite abelian groups with no 3-term arithmetic progressions. *J. Combin. Theory Series A*, **71 (1)**, 168-172.

[36] R. Motwani, J. Naor and M. Naor. The probabilistic method yields deterministic parallel algorithms. J. Comput. Syst. Sci., 49:478-516, 1994.

[37] J. Naor and M. Naor, Small-bias probability spaces: efficient constructions and applications. SIAM J. Comput. 22 4 (1993), pp. 838-856 (preliminary version in STOC90)

[38] N. Nisan and M. Szegedy. On the degree of boolean functions as real polynomial. In Proceedings of the 24th ACM Symposium on the Theory of Computing, pages 462–467. ACM, New York, 1992.

[39] Nisan N. and Wigderson A. (1994) Hardness vs. randomness, *Journal of Computer and System Sciences*, **49 (2)**, 149 - 167.

[40] O'Donnell R., Servidio R. A. (2003) New degree bounds for polynomial threshold functions, *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, 325 - 334.

[41] L. J. Schulman. Sample spaces uniform on neighborhoods. In Proceedings of the 24th Annual ACM Symposium on Theory of Computing, pages 17-25, 1992.

[42] S. Chari, P. Rohatgi and A. Srinivasan. Improved algorithms via approximations of probability distributions. In Proc. 26th ACM Symposium on Theory of Computing, pages 584-592, 1994.

[43] G. Szegő (1975) Orthogonal polynomials, fourth edition, American Mathematical Society, Colloquium Publications, Vol. XXIII, AMS Providence R.I..

6 Appendix

6.1 K_1 and K_2

It is not a-priori clear whether one of these classes contains the other. Assume that $\lim \mathbb{P}_{p_c}(f = 1)$ exists and denote it by α. We have the following simple result:

Claim 28 *For any $k \in K_2(f)$, for $p < p_c$ we have $\overline{\lim}\ \varepsilon^f(k,p) \leq \alpha$ and for $p > p_c$ we have $\overline{\lim}\ \varepsilon^f(k,p) \leq 1 - \alpha$*

Proof. Obviously, both $\max_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f = 1)$ and $\min_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f = 1)$ are increasing functions of p. The claim now follows immediately from the fact that $\lim \max_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f = 1) = \lim \min_{\mathbb{Q}\in\mathcal{A}(n,k,p)} \mathbb{Q}(f = 1) = \alpha$. ■

So, while we don't know if for $k \in K_2$, $\varepsilon^f(k,p) \to 0$ we do know that it cannot be too large.

6.2 Percolation

Here is the more general result, of which lemma 22 is a corollary (put $l = 1$).

Lemma 29 *(combining distributions) Fix integers $l, m \geq 1$. Suppose for each $1 \leq i \leq m$ we have random vectors $X^i := (X_1^i, \ldots, X_n^i) \in \mathcal{A}^r(n,k)$ and $Y^i := (Y_1^i, \ldots, Y_n^i) \in \mathcal{A}^r(n, lk+l+k)$. Suppose that the X vectors are independent among themselves and independent from the Y vectors, and that the Y vectors are l-wise independent among themselves. Then the vector with the following coordinates*

$$\begin{array}{cccc} X_1^1 + Y_1^1, & X_2^1 + Y_2^1, & \ldots, & X_n^1 + Y_n^1, \\ X_1^2 + Y_1^2, & X_2^2 + Y_2^2, & \ldots, & X_n^2 + Y_n^2, \\ \vdots, & \vdots, & \vdots, & \vdots, \\ X_1^m + Y_1^m, & X_2^m + Y_2^m, & \ldots, & X_n^m + Y_n^m \end{array} \tag{16}$$

is in $\mathcal{A}^r(mn, lk+l+k)$

Proof.

Call the resulting distribution Z, where $Z^i := (Z_1^i, Z_2^i, \ldots, Z_n^i)$ and $Z_j^i := X_j^i + Y_j^i$. Take a set S of at most $lk+l+k$ variables from the vector Z, we need to show they are independent and uniformly distributed. Suppose that a^i of them are from Z^i for each $1 \leq i \leq m$, WLOG we can assume that for each i these are $Z_1^i, \ldots, Z_{a^i}^i$. Consider only the i's for which $a^i \geq k+1$, since $|S| \leq l(k+1)+k$ we can have at most l such i's, WLOG suppose these are $a^1, \ldots, a^t$ for $t \leq l$. Now, fix some values $c_j^i \in \mathbb{Z}_r$ for $1 \leq i \leq m$, $1 \leq j \leq a_i$, define events $A := \{Z_j^i = c_j^i$ for all $1 \leq i \leq t$ and $1 \leq j \leq a_i\}$ and $B := \{Z_j^i = c_j^i$ for all $t+1 \leq i \leq m$ and $1 \leq j \leq a_i\}$. We need to show that $\mathbb{P}(A,B) = r^{-|S|}$. We start with

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{E}\left(\mathbb{P}(A \mid (X^i)_{i=1}^t)\right) = \\ &= \mathbb{E}(\mathbb{P}(Y_j^i = c_j^i - X_j^i \text{ for } 1 \leq i \leq t,\ 1 \leq j \leq a^i \mid (X^i)_{i=1}^t)) = \\ &= \mathbb{E}(r^{-\sum_{i=1}^t a^i}) = r^{-\sum_{i=1}^t a^i} \end{aligned} \tag{17}$$

Where the next to last equality follows since the Y_j^i are uniform and independent from the X's, since they are l-wise independent as vectors (and $l \geq t$), since they are $(lk+l+k)$-wise independent inside each vector and since $a_j^i \leq |S| \leq lk+l+k$.

To finish the lemma we need to show that $\mathbb{P}(B \mid A) = r^{-\sum_{i=t+1}^{m} a^i}$. We will show something stronger, that in fact $\mathbb{P}(B \mid (X^i)_{i=1}^t,\ (Y_i)_{i=1}^m) = r^{-\sum_{i=t+1}^{m} a^i}$. Indeed

$$\begin{aligned}
&\mathbb{P}(B \mid (X^i)_{i=1}^t,\ (Y_i)_{i=1}^m) = \\
&= \mathbb{P}(X_j^i = c_j^i - Y_j^i \text{ for } t+1 \leq i \leq m,\ 1 \leq j \leq a^i \mid (X^i)_{i=1}^t,\ (Y_i)_{i=1}^m) = \\
&= r^{-\sum_{i=t+1}^{m} a^i}
\end{aligned} \tag{18}$$

Where the last equality follows since the X_j^i for $t+1 \leq i \leq m$ are uniform, independent from the Y's and from $(X^i)_{i=1}^t$, since $a^i \leq k$ for each $t+1 \leq i \leq m$ by the definition of t and since $(X_j^i)_{j=1}^n$ are k-wise independent. This finishes the proof of the lemma. ■

6.3 The classical moment problem

Here is the general setup of the classical moment problem ([1], [28]) leading to the bounds of theorem 7. It is followed by a definition of the Krawtchouk polynomials and a proof of (5) and (6).

Consider a real sequence $\mathcal{S} := \{s_m\}_{m=0}^k$, with $s_0 = 1$ (this last condition is convenient for us in order to use probabilistic notation, but it is not necessary for the results of the classical moment problem). Define

$$\mathcal{A}_{\mathcal{S}} = \{\mathbb{Q} \mid \mathbb{Q} \text{ a probability distribution on } \mathbb{R},\ s_m = \mathbb{E}_{\mathbb{Q}}(X^m) \text{ for } 0 \leq m \leq k\} \tag{19}$$

to be all probability distributions with these first k moments (X is a random variable distributed according to $\mathbb{Q}$).

Definition 30 *Given $\mathcal{S} = \{s_m\}_{m=0}^k$ with $s_0 = 1$ and k even, define the* orthogonal polynomials with respect to $\mathcal{S}$, $\{P_m\}_{m=0}^{k/2}$ *as the unique polynomials with the following properties:*

1. *P_m is a polynomial of degree m with positive leading coefficient.*
2. *Defining formally a linear operator T from polynomials of degree k to reals by $T(x^i) := s_i$ for $0 \leq i \leq k$ then $T(P_l(x)P_m(x)) = \delta_{l,m}$.*

Note that the second condition is the same as requiring $E_{\mathbb{Q}}(P_l(X)P_m(X)) = \delta_{l,m}$ for any $\mathbb{Q} \in \mathcal{A}_{\mathcal{S}}$.

We remark that these polynomials cannot be defined for degree larger than n if the sequence $\mathcal{S}$ corresponds to the moments of an atomic distribution with only n atoms.

Define also the function $\rho_n(x) := (\sum_{m=0}^n P_m^2(x))^{-1}$, then we have the following

Theorem 31 *[1, 2.5.2 and 2.5.4] For any x and any $\mathbb{Q}_1, \mathbb{Q}_2 \in \mathcal{A}_{\mathcal{S}}$*

$$|\mathbb{Q}_1(X \leq x) - \mathbb{Q}_2(X < x)| \leq \rho_{\frac{k}{2}}(x) \tag{20}$$

and in particular when $\mathbb{Q}_1 = \mathbb{Q}_2$

$$\max_{\mathbb{Q} \in \mathcal{A}_{\mathcal{S}}} \mathbb{Q}(X = x) \leq \rho_{\frac{k}{2}}(x) \tag{21}$$

We remark that in many cases, the theory also has constructions which achieve these bounds, but we could not use these since in the cases we needed we required the support of the distribution to be on integer points. It is possible, however, that a modification of these constructions can yield a distribution on integer points, this would be very useful to show the sharpness of the bounds in the cases we use.

6.3.1 Krawtchouk polynomials

In our work we utilize the orthogonal polynomials corresponding to the moments of the binomial distribution (i.e., when $s_m = \mathbb{E}(X^m)$ where $X \sim \mathrm{Bin}(n,p)$). These are the well-known Krawtchouk polynomials (see [43]). For given n and p, the m'th polynomial ($0 \le m \le n$) is given by

$$P_m(x) = \binom{n}{m}^{-\frac{1}{2}} (p(1-p))^{-\frac{m}{2}} \sum_{j=0}^{m} (-1)^{m-j} \binom{n-x}{m-j} \binom{x}{j} p^{m-j}(1-p)^j \tag{22}$$

where for real x and integer $b \ge 1$, $\binom{x}{b} := \frac{x(x-1)\cdots(x-b+1)}{b!}$ and $\binom{x}{0} := 1$.

We note that

$$P_m(n) = \binom{n}{m}^{\frac{1}{2}} \left(\frac{1-p}{p}\right)^{\frac{m}{2}} \tag{23}$$

Hence

$$\rho_m(n) = \left(\sum_{j=0}^{m} \binom{n}{j} \left(\frac{1-p}{p}\right)^j \right)^{-1} = \frac{p^n}{\mathbb{P}(\mathrm{Bin}(n,1-p) \le m)} \tag{24}$$

Furthermore, for $p = \frac{1}{2}$

$$P_m(\frac{n}{2}) = \binom{n}{m}^{-\frac{1}{2}} \sum_{j=0}^{m} (-1)^{m-j} \binom{n/2}{m-j} \binom{n/2}{j} \tag{25}$$

but, as is well known, since the sum is the coefficient of z^m in the power series expansion of $f(z) := (1+z)^{\frac{n}{2}}(1-z)^{\frac{n}{2}}$ and since $f(z) = (1-z^2)^{\frac{n}{2}}$ we get by the binomial formula that

$$P_m(\frac{n}{2}) = \begin{cases} 0 & m \text{ odd} \\ \binom{n}{m}^{-\frac{1}{2}} (-1)^{\frac{m}{2}} \binom{n/2}{m/2} & m \text{ even} \end{cases} \tag{26}$$

We then obtain

Lemma 32 *For $p = \frac{1}{2}$, even n and even $m \le \frac{n}{2}$*

$$\rho_m(\frac{n}{2}) \le \frac{2}{\sqrt{m}} \tag{27}$$

Proof. Using (26) we have

$$\rho_m(\frac{n}{2}) = \left(\sum_{j=0}^{\frac{m}{2}} \binom{n}{2j}^{-1} \binom{n/2}{j}^2 \right)^{-1} \tag{28}$$

we recall the well-known inequality that for any integer $a \ge 0$, $a! = \sqrt{2\pi a}\left(\frac{a}{e}\right)^a e^{\lambda_a}$ where $\frac{1}{12a+1} \le \lambda_a \le \frac{1}{12a}$. Using this we notice that

$$\binom{a}{b} = \sqrt{\frac{a}{2\pi b(a-b)}} \frac{a^a}{(a-b)^{a-b} b^b} e^{\lambda_a - \lambda_{a-b} - \lambda_b} \tag{29}$$

Hence after cancelation

$$\binom{n}{2j}^{-1}\binom{n/2}{j}^2 = \sqrt{\frac{n}{\pi(n-2j)j}}e^{2\lambda_{n/2}-2\lambda_j-2\lambda_{\frac{n}{2}-j}+\lambda_{n-2j}+\lambda_{2j}-\lambda_n} \tag{30}$$

so for $\frac{n}{2}, j, (\frac{n}{2}-j) \geq 1$ we get

$$\binom{n}{2j}^{-1}\binom{n/2}{j}^2 \geq \sqrt{\frac{n}{\pi(n-2j)j}}e^{-\frac{5}{12}} \geq \sqrt{\frac{1}{8j}} \tag{31}$$

Plugging back into (28) we get

$$\rho_m(\frac{n}{2}) \leq \left(1+\sum_{j=1}^{\frac{m}{2}}\sqrt{\frac{1}{8j}}\right)^{-1} \leq \left(1+\frac{1}{\sqrt{2}}\left(\sqrt{\frac{m}{2}}-1\right)\right)^{-1} \leq \frac{2}{\sqrt{m}} \tag{32}$$

■

6.4 Majority

We now continue and give a sketch of the proof of the lower bound for theorem 10.

Proof. (sketch of lower bound in theorem 10) Fix an odd n and a $2 \leq k < n$, we would like to construct a distribution $\mathbb{Q} \in \mathcal{A}(n, k, \frac{1}{2})$ such that when we define S to be the number of bits which are 1 when sampling from $\mathbb{Q}$ then

$$|\mathbb{Q}(S \geq \frac{n+1}{2}) - \frac{1}{2}| \geq \frac{C}{\sqrt{k \log k}} \tag{33}$$

for some $C > 0$. We may assume that $k < c\frac{n}{\log n}$ for some small $c > 0$, otherwise the bound follows trivially by taking the distribution XOR0 and using the bound that it satisfies (see theorem 10. Let $M := C\sqrt{\frac{n}{k \log k}}$ be an integer, the idea of the proof is to construct $\mathbb{Q}$ in such a way that with high probability $S \equiv L \bmod M$ for some fixed integer $0 \leq L < M$, and furthermore that on this event S behaves like a $\mathrm{Bin}(n, \frac{1}{2})$ random variable conditioned to be L mod M. Such an S will satisfy (33) for the correct choice of L.

To do this, we consider a distribution $\tilde{\mathbb{Q}}$ on $(X_1, \ldots, X_{k+1}) \in \mathbb{Z}_M^{k+1}$ satisfying that all the X_i are IID uniform in $\mathbb{Z}_M$ except that X_k is chosen so that their sum is always L modulo M. This distribution is of course k-wise independent. the required distribution $\mathbb{Q}$ is a distribution on n bits $(Y_1, \ldots, Y_n)$, we create it from the distribution $\tilde{\mathbb{Q}}$ by dividing the Y's into $k+1$ disjoint groups of bits, each X_i is responsible for the value of one of these groups of bits in the following way, when observing the value of the X variable, we sample as uniformly as is possible a string of bits for the Y variables in its group such that their sum modulo M equals the value of the X variable.

The parameters have been chosen in such a way so that the probability that we do not succeed even at one of the Y groups to have the correct sum modulo M is very small. Hence the distribution $\mathbb{Q}$ thus constructed satisfies the required properties. ■

6.5 More bounds on the maximal probability that all bits are 1

In this section we detail more bounds on the maximial probability that all bits are 1. Recall $M(n, k, p) := \max_{\mathbb{Q} \in \mathcal{A}(n,k,p)} \mathbb{Q}(\text{All bits are 1})$.

In the main text we have shown

Theorem 33 *For even k*

$$M(n,k,p) \leq \frac{p^n}{\mathbb{P}(\mathrm{Bin}(n,1-p) \leq \frac{k}{2})} \tag{34}$$

In particular for even k

$$M(n,k,p) \leq 2\sqrt{k}\left(\frac{kp}{2e(1-p)(n-\frac{k}{2})}\right)^{\frac{k}{2}} \tag{35}$$

and for even k and $n(1-p) \leq \frac{k}{2}$

$$M(n,k,p) \leq 10p^n \tag{36}$$

We now compliment these with lower bounds on $M(n,k,p)$. Both lower bounds come from well-known constructions of linear error-correcting codes. In both we assume $p = \frac{1}{q}$ for either a prime, or a prime-power, q. To get the bounds we first construct the linear code over $GF(q)$, then pass to its dual code, well known to be an orthogonal array. Then sample a line of the orthogonal array uniformly and map to bits using $0 \mapsto 1$ and the rest of the elements mapping to 0. We obtain

Theorem 34 *Using the Gilbert-Varshamov bound, for $p = \frac{1}{q}$ with q a prime power*

$$M(n,k,p) \geq p\left(\frac{p(k-1)}{en}\right)^{k-1} \tag{37}$$

and using BCH codes, when $p = \frac{1}{q}$ with q a prime (not a prime-power!), $k \equiv 1 (mod\ q)$ and $n+1$ is a power of q then

$$M(n,k,p) \geq p\left(\frac{1}{n+1}\right)^{(k-1)(1-p)} \tag{38}$$

We add that there is a gap in the exponent between these lower bounds and our upper bounds, namely the upper bounds have exponent $\frac{k}{2}$ and the lower bounds have, at best, exponent $(k-1)(1-p)$. We do not know to close this gap but remark that it is also present in the theory of error-correcting codes, for a paper discussing this gap for error-correcting codes and the known results there see [14].

We end this section by remarking on one more exact result, for very small p's

Proposition 35 *When $p \leq \frac{1}{n-1}$*

$$M(n,k,p) = p^k \tag{39}$$

This follows quite simply from a direct construction of the distribution. We start by putting probability p^k on the all ones vector, then all the rest of the probabilities of atoms are determined by being k-wise independent with marginal p, we check that for this range of p's all these other probabilities are indeed positive. This is the same as the fact that the weight distribution of an MDS code is determined, see [34].

6.6 More on the minimal probability that all bits are 1

We remark on the proof of theorem 27. The construction of the upper bound on $n_c(k,p)$ goes as follows, we start with an orthogonal array with very good parameters over $GF(q)$ (where now $q = \frac{1}{1-p}$) obtained using the Gilbert-Varshamov bound. We then choose a row uniformly and map each of its coordinates to bits. The mapping is chosen so that in each coordinate exactly

one element of $GF(q)$ is mapped to 0, the rest to 1, but this element is chosen in a greedy fashion to minimize the chance of not having a 0 anywhere. When n is large enough compared to k this idea succeeds in giving a distribution with probability 0 for the all ones vector. This gives the upper bound of the theorem.

As detailed in the main text, the lower bound follows from an existence theorem in the theory of the TCMP, we give this theorem here for easy reference.

Let $X \sim \text{Bin}(n,p)$ and define $s_i := \mathbb{E}(X^i)$. Define the matrices

$$\begin{aligned} A(m) &:= (s_{i+j})_{i,j=0}^{m} \\ B(m) &:= (s_{i+j+1})_{i,j=0}^{m} \\ C(m) &:= (s_{i+j})_{i,j=1}^{m} \end{aligned} \tag{40}$$

then the classical moment problem states (see [1],[28] or [11] which contains a survey)

Proposition 36 *A random variable S with moment sequence $\{s_i\}$ supported on $[a,b]$ exists if and only if*

1. *k is odd and $bA(\frac{k-1}{2}) \geq B(\frac{k-1}{2}) \geq aA(\frac{k-1}{2})$.*
2. *k is even, $A(\frac{k}{2}) \geq 0$ and $(a+b)B(\frac{k}{2}-1) \geq abA(\frac{k}{2}-1) + C(\frac{k}{2})$.*

where as usual, $A \geq B$ means $A - B \geq 0$ means that $A - B$ is non-negative definite.