Approximate Inclusion-Exclusion and orthogonal polynomials

Alex Samorodnitsky *

Abstract

The inclusion-exclusion formula expresses the probability of the union of a family of sets in terms of the probabilities of their intersections:

$$Pr(\cup A_i) = \sum_i Pr(A_i) - \sum_{i<j} Pr(A_i \cap A_j) + ...$$

Assume now, that we are given the probabilities of all the j-wise intersections of the events, for $j = 1, ..., k$. How well can we estimate the probability of the union $\cup A_i$? Linial and Nisan [10] followed by Kahn, Linial and Samorodnitsky [8] develop a method to approximate $Pr(\cup A_i)$ from the above data with an additive error of $exp\left(-\tilde{\Omega}\left(k^2/n\right)\right)$ and show this error to be optimal up to the logarithmic factor implicit in $\tilde{\Omega}$ notation.

The inclusion-exclusion formula is the Möbius inversion formula for the lattice of subsets of a finite set. One can also ask this question in other settings - in a recent paper Mavronicolas [12] deals with the question of "approximate Möbius inversion formula" for the lattice of subspaces of a finite dimensional vector space over a finite field.

One can try to figure out what "approximate Möbius inversion" might mean for a general poset. It is easy to generalize the question to the class of graded posets and to show, that the best possible approximation for such a poset P equals the distance (in the supremum norm) between a characteristic function of the zero element of P and a subspace spanned by characteristic functions of a certain family of upwards-monotone subsets of P. (In the special cases of the lattice of subsets and the lattice of subspaces this reduces to the results obtained in the above mentioned papers.)

For the lattice of subsets, we improve the additive error of the approximation to $O(\sqrt{n}) \cdot 2^{-n}\binom{n}{(n-k)/2}$, which is optimal up to a multiplicative factor of $O(\sqrt{n})$. A key role is played by three families of orthogonal polynomials: Discrete Chebyshev polynomials, Hahn polynomials and Krawchouk polynomials.

It turns out that this result has some simple but somewhat unexpected connections with coding and graph theory.

*Hebrew University and DIMACS Center, Rutgers University

1 Introduction

This paper returns once again to the question of approximated inclusion-exclusion, raised in [10] and addressed also in subsequent [8] and [12]. We refer to these papers for the computational motivation of the problem. Here we start with reviewing the approach of [10] to the solution of this question, continued in [8] and in [12]. Then we describe our results, and discuss the way they are related to the previous work.

Let us first describe the question: The inclusion-exclusion formula expresses the probability of the union of a family of sets in terms of the probabilities of their intersections:

$$Pr(\cup A_i) = \sum_i Pr(A_i) - \sum_{i<j} Pr(A_i \cap A_j) + ...$$

Assume now, that we are given the probabilities of all the j-wise intersections of the events, for $j = 1, ..., k$. How well can we estimate the probability of the union $\cup A_i$?

A precise formulation [10] is as follows:
Main question
Find

$$E(k,n) = \sup \left(Pr(\cup_{i=1}^n A_i) - Pr(\cup_{j=1}^n B_j) \right) \tag{1}$$

where the supremum ranges over all families of events $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ in any probability spaces, that satisfy:

$$Pr(\cap_{i \in S} A_i) = Pr(\cap_{j \in S} B_j)$$

for all $S \subset [n]$, such that $|S| \leq k$.

The approach suggested in [10] contains two main steps. In the first step the "target value" $E(k,n)$ is shown to be the optimum of a certain linear program. Passing to the dual produces a following (rather surprising) result:
Let δ_0 is a function on $0, 1, \ldots, n$ defined by $\delta_0(i) = \delta_{0,i}$. (Recall that *Kronecker's δ-function* $\delta_{i,j}$ is 1 if $i = j$ and is 0 otherwise.) We want to approximate δ_0 as close as possible by a polynomial of bounded degree. Namely, given $0 \leq k \leq n$, we ask: what is $e(k,n) = \inf_p \left(\max_{i=0,\ldots,n} |p(i) - \delta_0(i)| \right)$, where the infimum ranges over all polynomials p of degree $\leq k$. It turns out that $E(k,n) = 2 \cdot e(k,n)$, or in other words:

$$E(k,n) = 2 \cdot \|\delta_0 - \mathcal{P}_k\|_\infty, \tag{2}$$

where $\mathcal{P}_k$ is the space of polynomials of degree $\leq k$ on $0, 1, \ldots, n$.

In the second step [10] attacks the approximation question of finding $E(k,n)$ using techniques from the *theory of approximation of functions of the real variable.* It turns

out, that suitably scaled *Chebyshev polynomials* of degree k provide a good candidate for the approximation of δ_0, and this leads to the following bounds on $E(k,n)$:
1. For $k = o(\sqrt{n})$

$$E(k,n) = \Theta\left(1 - \frac{k^2}{n}\right).$$

2. For $k = \Theta\left(\sqrt{n}\right)$

$$E(k,n) = \Omega(1),$$

3. And for $k \gg \sqrt{n}$

$$E(k,n) \leq O\left(exp\left\{-\frac{2k}{\sqrt{n}}\right\}\right).$$

While this estimate of $E(k,n)$ turns to be optimal for $k \leq \sqrt{n}$, it's optimality for the other values of k was not at all clear. In fact [10] shows that, at least for a certain range of k, the estimate is not optimal, by computing $E(n-1,n) = \frac{1}{2^{n-1}}$ precisely.

The task of improving the estimates for $E(k,n)$ was continued in [8]. It was understood that the "problem" of Chebyshev polynomials is that they actually do too much - instead of providing a good approximation to zero in the integer points $1,\ldots,n$, they approximate zero on the *whole interval* $[1,n]$. A natural attempt to remedy this situation is to "tailor-make" the approximating polynomial q by constraining q to have a lot of integer zeroes in the "right" places on the interval $[1,n]$, and then, possibly, multiplying the obtained polynomial by a suitably scaled Chebyshev polynomial of a small degree - to "round things out". This approach leads [8] to an improved estimate on $E(k,n)$ for $k \gg \sqrt{n}$.

$$E(k,n) \leq exp\left\{-\Omega\left(\frac{k^2 \log n}{n}\right)\right\}.$$

This inequality turns out to be rather tight. Using linear duality once again [1] [8] provide also a lower estimate on $E(k,n)$:

$$E(k,n) \geq exp\left\{-O\left(\frac{k^2}{n \log n}\right)\right\}.$$

Using the convenient $\tilde{\Theta}$ notation we actually arrive to a "tight" estimate: $E(k,n) = exp\left\{-\tilde{\Theta}\left(\frac{k^2}{n}\right)\right\}$.

Now, let us change direction, and return once more to the inclusion-exclusion formula. It is well-known [19] that this formula is a special case of the *Möbius inversion formula* for partially ordered sets. (The underlying poset in our case is the lattice of all subsets of an n-set.) Therefore one can ask [10] whether an analog of the results described above exists for other posets. A step in this direction was taken by Mavronicolas [12], who,

[1]This is this rare case, in which passing to the dual twice doesn't lead back to the original question, since the underlying space had been meanwhile altered by considering only *symmetric events* [10], [8].

following [2] considers the *q-analog of the principle of inclusion-exclusion*, which means that the lattice of all subsets of an n-set is replaced by the lattice of all subspaces of an n-dimensional vector space over a finite field with q elements. Let us describe in some detail the framework addressed in [12]. Let V_q^n be an n-dimensional vector space over a finite field with q elements. Let X be a set such that with every vector $v \in V_q^n$ is associated a subset $\mathcal{A}_v \subseteq X$. Assume also that the association between elements of X and subspaces of V_q^n is regular in the following sense: for each $x \in X$ the set $V_x = \left\{v \in V_q^n \mid x \in \mathcal{A}_v\right\}$ is either empty or a subspace of V_q^n. For any subspace $T \sqsubseteq V_q^n$ let $S_T = |\left\{x \in X \mid T \sqsubseteq V_x\right\}|$. Then the q-analog of inclusion-exclusion principle is given by the following formula [19]:

$$|\cup_{v \in V_q^n} \mathcal{A}_v| = \sum_{T \sqsubseteq V_q^n} (-1)^{\dim(T)+1} q^{\binom{\dim(T)}{2}} S_T.$$

[12] deals with the following question: Suppose that we know S_T, for all subspaces T of V_q^n of dimension $j = 0, 1, \ldots, k$. How well can we estimate $|\cup_{v \in V_q^n} \mathcal{A}_v|$? More precisely, we can, without loss of generality [10] assume that X is a probability space and set

$$E^{(q)}(k,n) = sup\left(Pr(\cup_{v \in V_q^n} \mathcal{A}_v) - Pr(\cup_{v \in V_q^n} \mathcal{B}_v)\right), \tag{3}$$

where the supremum ranges over all families of events $\{\mathcal{A}_v\}_{v \in V_q^n}$ and $\{\mathcal{B}_v\}_{v \in V_q^n}$ in any probability spaces, that satisfy: $S_T^{\mathcal{A}} = S_T^{\mathcal{B}}$ for all $T \sqsubseteq V_q^n$, such that $\dim(T) \leq k$. (Here $S_T^{\mathcal{A}}$ and $S_T^{\mathcal{B}}$ have the obvious definition). What is $E^{(q)}(k,n)$?

The approach of [12] follows closely that of [10]. The value $E^{(q)}(k,n)$ is shown to be the optimum of a certain linear program. Passing to the dual produces a following result:

$$E^{(q)}(k,n) = 2 \cdot \inf_p \left(\max_{i=0,\ldots,n} |p(q^i) - \delta_0(q^i)| \right), \tag{4}$$

where the infimum ranges over all polynomials p of degree $\leq k$, and $\delta_0(q^i) = \delta_{0,i}$.

In other words:

$$E^{(q)}(k,n) = 2 \cdot ||\delta_0 - \mathcal{P}_k||_\infty, \tag{5}$$

where $\mathcal{P}_k$ is the space of polynomials of degree $\leq k$ on $1, q, \ldots, q^n$.

[12] then observes that the second step of [10] fails in this case, namely for any $1 \leq k \leq n$ a scaled Chebyshev polynomial of degree k provides only a weak (bounded below by a constant > 0) approximation of δ_0.

Now to the description of our results. There are two points we would like to make. The first is an easy generalization of the notion of "approximate Möbius inversion", namely the question how well Möbius inversion works given partial information, to a general class of graded posets. In our opinion the linear programming approach of [10] is easier and more natural to apply at this level of generality. It turns out that the best possible approximation for such a poset P equals the distance (in the supremum norm) between a characteristic function of the zero element of P and a subspace spanned by characteristic functions of a certain family of upwards-monotone subsets of P. This includes (2) and (5) as special cases.

Our second point is that one can essentially find this distance for the lattice of all subsets of an n-set and the lattice of all subspaces of an n-dimensional vector space over a finite field.

Let P be a *graded poset*[2] with a minimum $\hat{0}$, and let P_i denote the set of the elements of P of *rank* i, for $0 \le i \le n$, where n is the *rank* of P. For a real valued function $f : P \to \mathbb{R}$, let $F : P \to \mathbb{R}$ denote the *zeta transform* of f defined by $F(x) = \sum_{y \succeq x} f(y)$. Given F we can reconstruct f through the Möbius inversion formula: $f(x) = \sum_{y \succeq x} \mu(x, y)F(y)$, where $\mu : P \times P \to \mathbb{R}$ is the Möbius function of P. The question is, how well can we reconstruct f, given only a partial information on F, say the values of $F(x)$ for $x \in P_0 \cup P_1 \ldots \cup P_k$, for some $0 \le k \le n$.

Definition 1.1:

$$E_P(k, n) = \sup(f(0) - g(0)),$$

where the supremum ranges over all *nonnegative* functions $f, g : P \to \mathbb{R}$, such that: $\sum_x f(x) = \sum_y g(y) = 1$ and $F(x) = G(x)$ for all $x \in P_0 \cup P_1 \ldots \cup P_k$ where F, G are the zeta transforms of f and g correspondingly. ∎

It might not be immediately obvious that this definition indeed extends the definitions that we had for $E(k, n)$ (1) and for $E^{(q)}(k, n)$ (3), but this is actually just a reformulation in the language of functions rather than subsets and subspaces. We point that out in section 3.

Set $h = f - g$ and note, that since zeta transform is a linear transformation, determining $E_P(k, n)$ amounts to finding a norm of a linear functional $\phi(h) = h(0)$ on a subspace of all functions h such that the zeta transform of h vanishes on $P_0 \cup P_1 \ldots \cup P_k$. Passing to the dual one obtains the following characterization of $E_P(k, n)$:

Proposition 1.2:

$$E_P(k, n) = 2 \cdot \|\delta_0 - \mathcal{U}_k\|_\infty, \tag{6}$$

where $\delta_0 : P \to \mathbb{R}$ is the characteristic function of $\hat{0}$, namely $\delta_0(x) = \delta_{x,\hat{0}}$, and $\mathcal{U}_k$ is the space of functions from P to $\mathbb{R}$ spanned by the characteristic functions $1_{\{y \,:\, y \succeq x\}}$ for all $x \in P_0 \cup P_1 \ldots \cup P_k$.

[2] See next section for more information on posets and the Möbius inversion formula.

Remark 1.3: This claim is not very surprising. Indeed, by a basic theorem of functional analysis, a norm of a linear functional on a subspace M of a Banach space X equals to the distance in the dual space X^* between this functional and the subspace $M^\perp$ of all functionals which are identically zero on M - see section 2. ∎

Therefore, to find $E_P(k,n)$ we have to solve an approximation question in the space of real-valued functions on P endowed with the supremum norm. In full generality this seems to be a very hard question. More can be said if the poset P satisfies strong regularity conditions. We will call P *regular* if for any $0 \le i < j \le n$, for any $x_1, x_2 \in P_i$ and for any $y_1, y_2 \in P_j$ holds:

$$|\{z \in P_j \;:\; z \succeq x_1\}| = |\{z \in P_j \;:\; z \succeq x_2\}|,$$

and

$$|\{w \in P_i \;:\; w \preceq y_1\}| = |\{w \in P_i \;:\; w \preceq y_2\}|.$$

Namely, P is regular iff for any $0 \le i < j \le n$ the bi-partite graph on $P_i \cup P_j$ in which $x \in P_i$ and $y \in P_j$ are connected iff comparable is regular.

Proposition 1.4: *Let P be a regular graded poset of height n. For any $0 \le i, j \le n$ take $x \in P_j$ and define $f_j(i) = \frac{|\{z \in P_i \;:\; z \succeq x\}|}{|P_i|}$. Then*

$$E_P(k,n) = 2 \cdot \|\delta_0' - \mathcal{U}_k'\|_\infty,$$

where $\delta_0'(i) = \delta_{0,i}$, for $i = 0, \ldots, n$, and $\mathcal{U}_k'$ is the space of functions from $\{0, \ldots, n\}$ to $\mathbb{R}$ spanned by $f_0, \ldots, f_k$.

Clearly the lattice of all subsets of an n-set and the lattice of all subspaces of an n-dimensional vector space over a finite field are regular. [3]

Remark 1.5: If P is the lattice of all subsets of an n-set then $f_j(i) = \frac{\binom{n-j}{i-j}}{\binom{n}{i}}$, and if P is the lattice of all subspaces of an n-dimensional vector space over a finite field with q elements, then $f_j(i) = \frac{\left[{n-j \atop i-j}\right]_q}{\left[{n \atop i}\right]_q}$, where $\left[{m \atop k}\right]_q = \frac{(q^m-1)(q^{m-1}-1)\ldots(q^{m-k+1}-1)}{(q^k-1)(q^{k-1}-1)(q-1)}$ is the q-*Gaussian coefficient* [19].

Now, $\frac{\binom{n-j}{i-j}}{\binom{n}{i}}$ is a polynomial of degree j in i and $\frac{\left[{n-j \atop i-j}\right]_q}{\left[{n \atop i}\right]_q}$ is a a polynomial of degree j in q^i. Therefore, for the lattice of subsets, $\mathcal{U}_k'$ is the space of polynomials of degree $\le k$ on $0, 1, \ldots, n$. Similarly, for the lattice of subspaces, $\mathcal{U}_k'$ is the space of polynomials of degree $\le k$ on $1, q, \ldots, q^n$. This immediately implies the dual characterizations (2) of $E(k,n)$ and (5) of $E^{(q)}(k,n)$. ∎

[3] It is probably a good place to admit that we are not aware of any other *interesting* regular posets.

Let us now return to the lattices of subsets and subspaces. We would like to solve the polynomial approximation questions (2) and (5). It turns out that there is NO approximate inclusion-exclusion in the lattice of subspaces, and this can be established quite straightforwardly by the lower bound method developed in [8]. More precisely, we have the following proposition:

Proposition 1.6: *There exists an absolute constant c_0 (greater than $1/100$) such that for any prime power q and for any integer n*

$$E^{(q)}(n-1,n) > c_0.$$

Note that obviously for any q and n we have $E^{(q)}(n,n) = 0$, which is just to say once again that Möbius inversion works with full information, and therefore the statement of the proposition has the strongest possible form.

But the main goal of this paper is to determine how well the "original" approximate inclusion-exclusion works for the lattice of the subsets. It turns out that a very convenient way to deal with this question is using *orthogonal polynomials.* We will say more about orthogonal polynomials in section 2 but here let us just say that orthogonal polynomials associated with a measure on a set of points are intrinsically related to the l_2 norm defined by this measure. And the Euclidean norm enters through the following simple idea, which seems to work rather well in this setting: finding $E(k,n)$ amounts to solving an extremal question about polynomials of bounded degree, viewed as points in a normed vector space. The relevant norms are the l_1 norm and the l_∞ norm. Now, if we can, without giving away too much, work with an appropriate Euclidean norm instead, Parseval's identity will make our life much easier. In fact, it is possible to solve the corresponding extremal questions for the Euclidean norm, using appropriate families of orthogonal polynomials. This leads to good upper and lower bounds on $E(k,n)$.

With the help of this "philosophy" we prove the following fact:

Theorem 1.7: *For any integers n and $1 \le k \le n$:*

$$\frac{1}{2^{n-1}} \sum_{i=0}^{\frac{n-k}{2}} \binom{n}{i} \le E(k,n) \le O\left((n^2-k^2)^{\frac{1}{4}} \cdot \frac{\binom{n}{\frac{n-k}{2}}}{2^n} \right). \tag{7}$$

We present two proofs. The "dual" proof is a natural continuation of the approach of [10] and [8]. It uses the dual characterization (2) of $E(k,n)$ and constructs an almost optimal polynomial approximation to δ_0 with the help of a family of *Discrete Chebyshev polynomials*, attaining the upper bound in (7) [4]. This bound is shown to be rather tight,

[4]After this work was completed we learned, that Discrete Chebyshev polynomials were already applied, in a quite similar manner, in coding theory [5].

namely $E(k,n) \geq \Omega\left(\left(1-\frac{k^2}{n^2}\right)^{\frac{1}{4}} \cdot \frac{\binom{n}{\frac{n-k}{2}}}{2^n}\right)$. Note, that it is slightly weaker than the lower bound in (7).

The "primal" proof does not rely upon either theorem 2.2 or the duality theorem of linear programming. The main tool is the discrete Fourier analysis on $\mathbb{Z}_2^n$. Once again, $E(k,n)$ is shown to be the solution of an extremal problem (13) for polynomials on $0, 1, \ldots, n$.

We point out a connection, through this extremal problem, between $E(k,n)$ and the *Lövasz theta function* [11] of a certain family of graphs which occurs naturally in coding theory. These graphs $G_{n,k}$ are defined by two integer parameters n and $0 \leq k \leq n$. The vertex set of the graph $G_{n,k}$ is $\{0,1\}^n$ and two vertices x and y are connected iff they differ in at least $k+1$ places. It turns out [15] that $E(k,n) \approx \frac{\theta(G_{n,n-k})}{2^n}$, the $\approx$ sign hiding an $O(\sqrt{n})$ factor. (We speculate that these two values might actually be equal.)

The extremal problem (13) is essentially solved in [15] in order to get tight bounds on $\theta(G_{n,k})$. For completeness sake we present the easy solution here. The upper bound in (7) is obtained with the help of an orthogonal family of *Hahn polynomials*.
The lower bound in (7) follows from an obvious lower bound on $\theta(G_{n,n-k})$, which is the size of the maximal independent set in $G_{n,n-k}$.

2 Preliminaries

This section contains necessary definitions, terminology and references as well as some of the facts that are required later on.

Partially ordered sets [17]
Let P be a finite partially ordered set (poset in short). We say that P has a *minimum* $\hat{0}$ iff $\hat{0} \preceq x$ for all $x \in P$. A *chain* in P is a subset of elements of P any two of which are comparable.
A poset P with a minimum $\hat{0}$ is said to be *graded* if for every element $x \in P$, every maximal chain from $\hat{0}$ to x has the same length. This common length is defined to be the *rank* of x. The *rank of* P is the maximal rank of $x \in P$.
Consider matrices whose rows and columns are indexed by elements of P. One such matrix is ζ, *the zeta function of* P, defined by $\zeta(x,y) = 1$ if $x \preceq y$ and $\zeta(x,y) = 0$ otherwise. It turns out that ζ is always an invertible matrix. Let μ denote the inverse of ζ. This matrix is the *Möbius function* of P, and we have
The Möbius inversion formula: Let $f : P \to \mathbb{R}$ and let $F = \zeta \cdot f$ be the zeta transform of f, i.e. $F(x) = \sum_{y \succeq x} f(y)$. Then

$$f = \mu \cdot F, \qquad \text{namely} \qquad f(x) = \sum_{y \succeq x} \mu(x,y) F(y).$$

The Möbius function is known for some common partially ordered sets. We will need the following fact:
For the lattice of all subsets of an n-set, $\mu(x,y) = (-1)^{|y|-|x|}$ if $x \subseteq y$, and $\mu(x,y) = 0$ otherwise. Here $|x|$ denotes the size of the set x.

Orthogonal polynomials [18]

Let μ be a finite measure on an interval $[a,b]$. It defines a natural inner product on the space of real-valued functions on the interval, namely $< f, g >= \int_a^b f(x)g(x)d\mu(x)$. A sequence $p_0, p_1, \dots p_n \dots$ of polynomials are the *orthogonal polynomials associated with measure* μ if:

(1) For $i = 0, 1, \dots$, $\deg(p_i) = i$.
(2) For $i \neq j$, the polynomials p_i and p_j are orthogonal : $< p_i, p_j >= 0$.
One can choose different normalizations of polynomials $\{p_i\}$. We will require also:
(3) $< p_i, p_i >= 1$, for all $i = 0, 1, \dots$.

Given a measure, it seems, in general, quite hard to obtain an explicit construction of the associated orthogonal polynomials, or to decide whether these polynomials have some specific property. Fortunately, for some measures μ the corresponding orthogonal polynomials are sufficiently well-known. We are going to deal with two such measures, both of them discrete and supported on the integer points $0, 1, \dots, n$. Note, that in this case there are only $n+1$ orthogonal polynomials, since the space of real-valued functions on $0, 1, \dots, n$ is $n+1$ dimensional.

The first is the uniform measure, giving a unit weight to all the points $0, \dots, n$. The associated orthogonal polynomials are the *discrete Chebyshev polynomials* [18].
The second measure is the squared binomial measure on $0, 1, \dots, n$. This measure endows the point i with weight $\binom{n}{i}^2$. The associated orthogonal polynomials are a (special) family of *Hahn polynomials* [3, 13]. We are going to need only one property of the above-mentioned polynomials, namely their value at zero.

Lemma 2.1:
(1)Let p_i be the i'th discrete Chebyshev polynomial, $i = 0, 1 \dots, n$. Then [18], [13]

$$p_i(0) = \frac{\binom{n}{i}}{\sqrt{\binom{n+i+1}{2i+1}\binom{2i}{i}}}. \tag{8}$$

(2)Let q_i be the i'th Hahn polynomial, $i = 0, 1 \dots, n$. Then [13]

$$q_i(0) = \sqrt{\frac{\binom{2n}{i} - \binom{2n}{i-1}}{\binom{2n}{n}}}. \tag{9}$$

Harmonic Analysis on $\mathbb{Z}_2^n$ [3, 6]

$\mathbb{Z}_2^n$ is a finite Abelian group, therefore its characters $\{W_T\}_{T\in\mathbb{Z}_2^n}$ constitute a group (the *dual group* which is isomorphic to $\mathbb{Z}_2^n$.) The character W_T is a function from $\mathbb{Z}_2^n$ to $\{-1,1\}$, defined as: $W_T(S)=(-1)^{|T\cap S|}$. The characters $\{W_T\}_{T\in\mathbb{Z}_2^n}$ form an orthonormal basis in the space of real-valued functions on $\mathbb{Z}_2^n$, equipped with uniform probability distribution.

For $f:\mathbb{Z}_2^n\to\mathbb{R}$, define $\hat{f}:\mathbb{Z}_2^n\to\mathbb{R}$, as $\hat{f}(T)=\frac{1}{2^n}\sum_{S\in G} f(S)W_T(S)$. The function $\hat{f}$ is the *Fourier Transform* of f. See [6], for more on the Fourier Transform in $\mathbb{Z}_2^n$.

The Fourier Transform is a linear transformation on the space of real-valued functions on $\mathbb{Z}_2^n$. It commutes with another linear transformation π on this space, given by $(\pi f)(S)=\frac{1}{\binom{n}{|S|}}\sum_{|T|=|S|} f(T)$. π is the (symmetric) projection on the space of *symmetric functions* on $\mathbb{Z}_2^n$.

For $0\le k\le n$ define $K_k:\mathbb{Z}_2^n\to\mathbb{R}$ as $K_k(S)=\sum_{|T|=k}W_T(S)$. Clearly, $K_k(S)$ depends only on $|S|$, and can therefore be viewed as a function on integers $0\le s\le n$. With some abuse of notation we may, therefore, view K_k as $K_k(x)=\sum_{j=0}^{n}(-1)^j\binom{x}{j}\binom{n-x}{k-j}$, the k-th *Krawchouk polynomial* [9], a real polynomial of degree k.

A theorem from functional analysis [14]

Let X be a finite-dimensional real normed vector space with norm $\|\cdot\|_X$, and let the dual space X^* be the vector space of all linear functionals on X endowed with the *dual norm* $\|\cdot\|_{X^*}$, namely for $\phi\in X^*$ we define $\|\phi\|_{X^*}=\sup_{x\in X,\ \|x\|_X\le 1}\phi(x)$. For instance, let A be a finite set and X be the $|A|$-dimensional vector space of all real-valued functions on A with the l_1 norm: for $f\in X$ the norm is $\|f\|_X=\sum_{a\in A}|f(a)|$. Then X^* is the $|A|$-dimensional real vector space endowed with the *supremum* (or l_∞) norm: for $\phi\in X^*$ we have $\|\phi\|_{X^*}=\max_{a\in A}|\phi(a)|$. (The action of $\phi\in X^*$ on $x\in X$ is given by an inner product: $\phi(x)=<\phi,x>=\sum_{a\in A}\phi(a)x(a)$.)

Let M be a subspace of X. It is a normed vector space in it's own right, and we may consider the dual space M^*. Any linear functional ϕ on X is also a functional on M, and so may be considered an element of M^* (This transformation $X^*\to M^*$ is not one to one, M^* is actually a *quotient space* of X^*). We will require an answer to the following question: What is the norm of ϕ as an *element of* M^* ? The answer is given by following theorem:

Theorem 2.2: *In the notation above, let $M^\perp$ be the subspace of X^* given by all the functionals $\tau\in X^*$ such that $\tau_{|M}\equiv 0$. Then, for any $\phi\in X^*$:*

$$\|\phi\|_{M^*}=\|\phi-M^\perp\|_{X^*},$$

where the left hand side denotes the distance from ϕ to $M^\perp$ in the dual norm on X^*.

3 Approximate Möbius inversion on posets

We start with the proof that definition 1.1 of $E_P(k,n)$ indeed generalizes the notion of "approximate inclusion-exclusion", defined in (1) and in (3) for the lattice of subsets and the lattice of subspaces, to a general graded poset P.

Lemma 3.1: *1. Let P be the lattice of all subsets of an n-set, then $E_P(k,n) = E(k,n)$. 2. Let P be the lattice of all subspaces of an n-dimensional vector space over a finite field with q elements, then $E_P(k,n) = E^{(q)}(k,n)$.*

Proof: Let $A_1,\ldots,A_n$ and $B_1,\ldots,B_n$ be two families of events in probability spaces. We define two functions f and g on the lattice P of all subsets of an n-set, which will be convenient also to view as $\{0,1\}^n$, the set of all zero-one vectors of length n. For $x \in P$, set $f(x) = Pr\left(\bigcap_{i:x_i=1} A_i \cap \bigcap_{i:x_i=0} A_i^c\right)$. The function g is defined in the same way, using the family $B_1,\ldots,B_n$. Note that $|Pr(\cup_i A_i) - Pr(\cup_j B_j)| = |f(0) - g(0)|$.

The following claims are immediate:
1. The functions f and g are nonnegative and sum to 1.
2. The set families $A_1,\ldots,A_n$ and $B_1,\ldots,B_n$ satisfy $Pr\left(\cap_{i\in S} A_i\right) = Pr\left(\cap_{j\in S} B_j\right)$ for all $S \subset [n]$, such that $|S| \le k$, if and only if $F(x) = G(x)$ for all sets x of size $\le k$. Here F, G are zeta transforms of f and g.
Consequently, for any two families of sets satisfying the conditions of definition (1), there are two functions satisfying the conditions of definition 1.1. This implies $E_P(k,n) \ge E(k,n)$.

Conversely, take two functions f and g from $\{0,1\}^n$ to $\mathbb{R}$ satisfying the conditions of definition 1.1. Since f and g are nonnegative and sum to 1 they define two probability distributions on $\{0,1\}^n$. Consider two probability spaces $\Omega_1 = (\{0,1\}^n, f)$ and $\Omega_2 = (\{0,1\}^n, g)$. Let $A_1,\ldots,A_n \subseteq \Omega_1$ be defined by $A_i = \{x : x_i = 1\}$. Let Let $B_1,\ldots,B_n \subseteq \Omega_2$ be defined in the same way. It is immediate to check, that $|Pr(\cup_i A_i) - Pr(\cup_j B_j)| = |f(0) - g(0)|$ and the families $A_1,\ldots,A_n$ and $B_1,\ldots,B_n$ satisfy the conditions of definition (1).

This proves the first half of the lemma. The second half is proven similarly. ∎

Our main goal in this section is to prove the dual characterization of $E_P(k,n)$, namely proposition 1.2.

Remark 3.2: Observe, that although the following proof relies heavily on theorem 2.2, it is, essentially, just a reformulation of the linear programming proofs of [10] and [12]. In fact, these proofs simply reprove theorem 2.2 for finite-dimensional vector spaces with the l_1 norm. We prefer to use theorem 2.2, since it highlights the not quite obvious fact that $E_P(k,n)$ measures a distance between a point and a subspace in the vector space of real-values functions on P. (And besides, *it's already there.*) ∎

Proof: of proposition 1.2
Let f and g be two functions on P satisfying the conditions of definition 1.1, namely they are nonnegative, sum to 1 and $F(x) = G(x)$ for all x of rank at most k, where F and G are zeta transforms of f and g correspondingly. Set $h = f - g$, and consider the set $\mathcal{H} := \{h = f - g\}$ of all functions obtainable in this way. Clearly, $\mathcal{H}$ interests us, since $E_P(k, n)$ is just the maximum of a linear functional $\phi : h \to h(0)$ on $\mathcal{H}$. What is $\mathcal{H}$?

Lemma 3.3:

$$\mathcal{H} = \{h : \|h\|_1 \leq 2,\ H(x) = 0, \quad \textit{for all} \ \ x \in P_0 \cup P_1 \ldots \cup P_k\}\,, \tag{10}$$

where H is the zeta transform of h, and recall that P_i is the set of all elements of P of rank i.

Proof: Let $h \in \mathcal{H}, h = f - g$. Since $\|f\|_1 = \|g\|_1 = 1$, we have $\|h\|_1 \leq 2$ by the triangle inequality. Since $F(x) = G(x)$ for $x \in P_0 \cup P_1 \ldots \cup P_k$, and the zeta transform is linear, we have $H(x) = F(x) - G(x) = 0$ for x of rank $\leq k$.
Conversely, let h belong to the RHS of (10). Let $\|h\|_1 = 2a$. Set $\tilde{f} = h_+$, namely $\tilde{f}(x) = h(x)$ if $h(x) \geq 0$, and $\tilde{f}(x) = 0$ otherwise. Similarly, set $\tilde{g} = h_-$, which means $\tilde{g}(x) = -h(x)$ if $h(x) \leq 0$, and $\tilde{g}(x) = 0$ otherwise. These functions are, by definition, nonnegative and $h = \tilde{f} - \tilde{g}$. Therefore, taking the zeta transform: $\tilde{F}(x) = \tilde{G}(x)$ for $x \in P_0 \cup P_1 \ldots \cup P_k$. Note, that $\sum_x h(x) = H(\hat{0}) = 0$ implies $\sum_x \tilde{f}(x) = \sum_y \tilde{g}(y) = a$. If $a = 1$ we conclude by setting $f = \tilde{f}$ and $g = \tilde{g}$. If not, let $f(x) = \tilde{f}(x)$ for all x except $\hat{0}$, and take $f(\hat{0}) = \tilde{f}(\hat{0}) + (1 - a)$. The function g is defined similarly, and we are done. ∎

Let X denote the vector space of real-valued functions on P endowed with the l_1 norm, and let $M \sqsubset X$ be the subspace of all functions $h \in X$ with zeta transform $H = \zeta h$ vanishing on $P_0 \cup P_1 \ldots \cup P_k$. Then the lemma (see the definition of the dual norm of a linear functional in section 2) implies:

$$E_P(k, n) = 2 \cdot \|\phi\|_{M^*} = 2 \cdot \|\phi - M^\perp\|_{X^*} = 2 \cdot \|\phi - M^\perp\|_\infty.$$

The second equality follows from theorem 2.2. Therefore, to conclude the proof of the proposition, we only have to show $M^\perp = \text{Span}\left\{1_{\{y:y\succeq x\}} \ : \ x \in P_0 \cup P_1 \ldots \cup P_k\right\}$. Note, that since ζ is an invertible matrix. the image M_1 of M under ζ is just the subspace of all the functions vanishing on $P_0 \cup P_1 \ldots \cup P_k$, and therefore $M_1^\perp$ is easy to find:

$$M_1^\perp = \{f \in X^* \ : \ f(x) = 0, \quad \text{for all} \ \ x \in P_{k+1} \cup \ldots \cup P_n\} = \text{Span}_{x \in P_0 \cup P_1 \ldots \cup P_k} \{1_x\}\,.$$

Now, let $\tau \in M^\perp$. This means, that for any $h \in M$:
$0 =< \tau, h >=< \tau, \zeta^{-1} \cdot \zeta h >=< (\zeta^{-1})^t \tau, \zeta h >$. Consequently, $\tau \in M^\perp$ if and only if $(\zeta^{-1})^t \tau \in M_1^\perp$, namely $M^\perp = \zeta^t M_1^\perp$. In other words,

$$M^\perp = \text{Span}_{x \in P_0 \cup P_1 \ldots \cup P_k} \left\{\zeta^t 1_x\right\} = \text{Span}\left\{1_{\{y:y\succeq x\}} \ : \ x \in P_0 \cup P_1 \ldots \cup P_k\right\},$$

and we are done. ∎

3.1 Regular posets and symmetrization - proof of proposition 1.4

Let P be a graded poset of rank n. Proposition 1.2 tells us that to find $E_P(k,n)$ is the same as to find the distance of a function δ_0 to a subspace $\mathcal{U}_k$ of functions spanned by $\left\{1_{\{y:y\succeq x\}} \;:\; x \in P_0 \cup P_1 \ldots \cup P_k\right\}$. The next question is, what can we really do with this observation, in terms of computing $E_P(k,n)$. The structure of the space $\mathcal{U}_k$ seems, in general, to be rather complex. However, the function δ_0 is quite special. It obviously has a lot of nice properties, from which we exploit the following: δ_0 is constant on the level sets $P_0, \ldots, P_n$. Let us capture this property of δ_0 in a definition:

Definition 3.4: A function $f \;:\; P \to \mathbb{R}$ is a *symmetric function* if f is constant on the level sets $P_0, \ldots, P_n$. ∎

A space S of symmetric functions is an n-dimensional subspace of the $|P|$-dimensional space X of real-valued functions on P, and it comes with a natural projection $\pi \;:\; X \to S$, given by: $(\pi f)(x) = \frac{1}{|P_i|}\sum_{y\in P_i} f(y)$, where i is the rank of x. We need a following simple fact: π is a contraction in the l_∞ norm (in fact, in any l_p norm), namely for any $f, g \in X$ holds $\|\pi f - \pi g\|_\infty \leq \|f - g\|_\infty$. This leads to the following observation: Let $s \in S$ be a symmetric function, and let $M \sqsubseteq X$ be a subspace such that $\pi \;:\; M \to M$, i.e, the image of M under π is contained in M. Then: $\|s - M\|_\infty = \|s - \pi M\|_\infty$. This makes our life easier, narrowing the search for the point of M closest to s, to a small subspace πM. [5]

Accordingly, we would like to establish properties of poset P which will imply that the image of $\mathcal{U}_k$ under π is contained in $\mathcal{U}_k$. Since $\mathcal{U}_k = \text{Span}\left\{1_{\{y:y\succeq x\}} \;:\; x \in P_0 \cup P_1 \ldots \cup P_k\right\}$, the projection $\pi\mathcal{U}_k$ is spanned by $\left\{\pi 1_{\{y:y\succeq x\}} \;:\; x \in P_0 \cup P_1 \ldots \cup P_k\right\}$. Let P be a *regular* (see Introduction) poset. Then for any two elements $x, x' \in P$ of the same rank i holds $\pi 1_{\{y:y\succeq x\}} = \pi 1_{\{y:y\succeq x'\}}$, and therefore

$$\pi 1_{\{y:y\succeq x\}} = \frac{1}{|P_i|}\sum_{x'\in P_i} \pi 1_{\{y:y\succeq x'\}} = \pi\left(\frac{1}{|P_i|}\sum_{x'\in P_i} 1_{\{y:y\succeq x'\}}\right) = \frac{1}{|P_i|}\sum_{x'\in P_i} 1_{\{y:y\succeq x'\}},$$

since the last function is symmetric.

Therefore $\pi\mathcal{U}_k \sqsubseteq \mathcal{U}_k$ and it is spanned by $\tilde{f}_i = \frac{1}{|P_i|}\sum_{x'\in P_i} 1_{\{y:y\succeq x'\}}$, for $i = 0, \ldots, k$. We are interested in $\|\delta_0 - \mathcal{U}_k\|_\infty = \|\delta_0 - \pi\mathcal{U}_k\|_\infty = \|\delta_0 - \text{Span}\{\tilde{f}_0, \ldots, \tilde{f}_k\}\|_\infty$. Note, that instead of working in the space of symmetric functions on P, we may work in the space of real-valued functions on $0, 1, \ldots, n$, since there is a natural isometry between the two spaces: A function h on $\{0, 1 \ldots, n\}$ "lifts" to a symmetric function $\tilde{h}$, defined by $\tilde{h}(x) = h(rank(x))$. We conclude the proof of proposition 1.4 by observing that the symmetric function $\tilde{f}_j$ corresponds to a function $f_j(i) := \frac{|\{z\in P_i \;:\; z\succeq x\}|}{|P_i|}$ on $0, 1, \ldots, n$. (Here x is *any* element of P_j.) ∎

[5] Let us emphasize that this symmetric projection had already been used in [10] and [12]. We just apply it in a more general setting.

4 Tight estimates for approximate inclusion-exclusion using duality

In this section essentially tight estimates for $E_P(k,n)$ are given for the lattice of all subsets of an n-set, and for the lattice of all subspaces of an n-dimensional vector space over a finite field. We use the dual characterization (6) of $E_P(k,n)$. In the first short subsection we show that proposition 1.6 is a straightforward consequence of the lower bound method of [8]. In the second subsection we prove a statement which is only slightly weaker than theorem 1.7.

4.1 No q-analog for approximate inclusion-exclusion

We do not repeat the argument of [8] itself, and only state a lemma which is a slightly modified extraction from the "lower bound part" of theorem 2.1 of [8]

Lemma 4.1: *Assume that we can find distinct points $x_1,\ldots,x_{k+1}$ in $\{q,q^2,\ldots,q^n\}$ and an $\alpha>0$ such that*

$$\sum_{j=1}^{k+1}|\prod_{i\neq j}\frac{x_i}{x_i-x_j}|<\frac{1}{\alpha}.$$

Then $E^{(q)}(k,n)>\alpha$.

To prove proposition 1.6 we have no choice but to take $x_1=q$, $x_2=q^2$,...,$x_n=q^n$. So, we have to estimate $\sum_{j=1}^n|\prod_{i\neq j}\frac{q^i}{q^i-q^j}|$. Consider a summand $|\prod_{i\neq j}\frac{q^i}{q^i-q^j}|$. For $i<j$ we have $|\frac{q^i}{q^i-q^j}|=\frac{1}{q^{j-i}-1}\leq\frac{1}{q^{j-i-1}}$. Following this up, we get $|\prod_{i=1}^{j-1}\frac{q^i}{q^i-q^j}|\leq q^{-\frac{j(j-1)}{2}}$.

For $i>j$, we have:

$$|\prod_{i>j}\frac{q^i}{q^i-q^j}|=\prod_{i>j}\frac{q^{i-j}}{q^{i-j}-1}\leq\prod_{i>j}\left(1+\frac{2}{q^{i-j}}\right)\leq\prod_{i>j}e^{\frac{2}{q^{i-j}}}\leq e^{\frac{2}{q-1}}.$$

Therefore,

$$\sum_{j=1}^{n}|\prod_{i\neq j}\frac{q^i}{q^i-q^j}|\leq e^{\frac{2}{q-1}}\cdot\sum_j q^{-\frac{j(j-1)}{2}}<e^{\frac{4}{q-1}}.$$

Using the lemma, we conclude that for any n and q: $E(n-1,n)>e^{-\frac{4}{q-1}}\geq e^{-4}$, implying the claim of proposition 1.6 with $c_0=e^{-4}$.

4.2 A "dual" proof of theorem 1.7

The Setting: We work in the $n+1$-dimensional vector space X of real-valued functions on $0, 1 \ldots, n$. The function δ_0 is 1 at zero, and 0 everywhere else. The space $\mathcal{P}_k$ is the $k+1$-dimensional space of polynomials of degree $\leq k$ on $0, 1 \ldots, n$. We want to show:

$$\Omega\left(\left(1-\frac{k^2}{n^2}\right)^{\frac{1}{4}} \cdot \frac{\binom{n}{\frac{n-k}{2}}}{2^n}\right) \leq E(k,n) = \|\delta_0 - \mathcal{P}_k\|_\infty \leq O\left((n^2-k^2)^{\frac{1}{4}} \cdot \frac{\binom{n}{\frac{n-k}{2}}}{2^n}\right). \quad (11)$$

using the dual (2) characterization of $E(k,n)$.

Things become much easier, if we switch to the Euclidean norm.

Lemma 4.2:

$$\|\delta_0 - \mathcal{P}_k\|_2 = \Theta\left((n^2-k^2)^{\frac{1}{4}}\right) \cdot \frac{\binom{n}{\frac{n-k}{2}}}{2^n}$$

Proof: Finding the distance from a point δ_0 to a subspace $\mathcal{P}_k$ in a Euclidean space X is easy, given an orthonormal basis $v_0, \ldots, v_n$ of X such that $v_0, \ldots, v_k$ span $\mathcal{P}_k$. Indeed, expand δ_0 as a linear combination of $\{v_i\}$, namely $\delta_0 = \sum_{i=0}^n a_s v_s$, where $a_s = <\delta_0, v_s>$. For any point $p = \sum_{i=0}^k b_s v_s \in \mathcal{P}_k$ we have, by the Parseval identity, $\|\delta_0 - p\|_2^2 = \sum_{i=0}^k (a_s - b_s)^2 + \sum_{s=k+1}^n a_s^2$. Therefore, the point in $\mathcal{P}_k$ closest to δ_0 is its orthogonal projection $p^{(k)} = \sum_{i=0}^k a_s v_s$, and $\|\delta_0 - \mathcal{P}_k\|_2 = \sqrt{\sum_{s=k+1}^n a_s^2}$.

However, we do have such a basis for the inner product $< f, g >= \sum_{s=0}^n f(s)g(s)$! This basis is given by the *discrete Chebyshev polynomials* $p_1, \ldots, p_n$, see section 2. Plugging them in, we get:

$$\|\delta_0 - \mathcal{P}_k\|_2 = \sqrt{\sum_{s=k+1}^n <\delta_0, p_s>^2} = \sqrt{\sum_{i=k+1}^n p_i(0)^2} = \sqrt{\sum_{i=k+1}^n \frac{\binom{n}{s}^2}{\binom{n+s+1}{2s+1}\binom{2s}{s}}}.$$

The last equality follows from lemma 2.1.

The rest is a simple computation with binomial coefficients. Shuffling the factorials, one gets: $\frac{\binom{n}{s}^2}{\binom{n+s+1}{2s+1}\binom{2s}{s}} = \frac{1}{(n+1)\binom{2n+1}{n}} \binom{2n+1}{n+s+1} \cdot (2s+1)$. The following requires no explanation:

$$\sum_{s=k+1}^n \binom{2n+1}{n+s+1} \cdot (2s+1) = \sum_{s=k+1}^n \binom{2n+1}{n+s+1} \cdot (2n+2s+1) - 2n \sum_{s=k+1}^n \binom{2n+1}{n+s+1} =$$

$$(4n+2) \sum_{s=k+1}^n \binom{2n}{n+s} - 2n \sum_{s=k+1}^n \binom{2n+1}{n+s+1} =$$

$$2\sum_{s=k+1}^{n}\binom{2n}{n+s}+2n\sum_{s=k+1}^{n}\left(2\binom{2n}{n+s}-\binom{2n+1}{n+s+1}\right)=$$

$$2\sum_{s=k+1}^{n}\binom{2n}{n+s}+2n\sum_{s=k+1}^{n}\left(\binom{2n}{n+s}-\binom{2n}{n+s+1}\right)=\Theta\left(n\binom{2n}{n+k+1}\right).$$

The lemma now follows, when we recall [6] that for any n and $0 \leq t \leq n$ holds: $\binom{2n}{2t} = \Theta\left(\sqrt{\frac{t(n-t)}{n}}\right)\cdot\binom{n}{t}^2$. ∎

To conclude the proof, observe that for any n and k we have: $\frac{1}{\sqrt{n+1}}\|\delta_0 - \mathcal{P}_k\|_2 \leq \|\delta_0 - \mathcal{P}_k\|_\infty \leq \|\delta_0 - \mathcal{P}_k\|_2$.

∎

From the computational point of view, the important part of this result is the upper bound $\epsilon(k,n) := O\left(\sqrt{n}\cdot\frac{\binom{n}{\frac{n-k}{2}}}{2^n}\right)$ in (11). Given n events and the probabilities of their j-wise intersections, $j = 0, 1, \ldots, k$, we can now approximate the probability of the union efficiently, with an error not exceeding $\epsilon(k,n)$.

Corollary 4.3: *For any integers k, n there exist (explicitly given) constants $\alpha_0^{k,n}, \alpha_2^{k,n}, \ldots, \alpha_k^{k,n}$, such that for every collection of sets $A_1, \ldots, A_n$ in a probability space*

$$|Pr\left(\cup_{i=1}^n A_i\right) - \sum_{|T|\leq k}\alpha_{|T|}^{k,n} Pr\left(\cap_{i\in T} A_i\right)| \leq O\left(\sqrt{n}\cdot\frac{\binom{n}{\frac{n-k}{2}}}{2^n}\right). \tag{12}$$

Proof: The proof of lemma 4.2 gave us a polynomial $p = p^{(k)} \in \mathcal{P}_k$ such that $\|p - \delta_0\|_\infty \leq \epsilon(k,n)$.

We have two bases for the space $\mathcal{P}_k$. One of them is the usual monomial basis $1, x, \ldots, x^k$. The other (see proposition 1.4 and remark 1.5) is given by the functions $f_j(i) = \frac{\binom{n-j}{i-j}}{\binom{n}{i}}$ for $j = 0, \ldots, k$. Let $p = \sum_{s=0}^k a_s f_s$.

Lifting to the space S of symmetric functions on $\{0,1\}^n$ (see subsection 3.1) we get: $\tilde{p} = \sum_{s=0}^k a_s \tilde{f}_s$, where $\tilde{f}_i = \frac{1}{\binom{n}{i}}\sum_{x:|x|=i} 1_{\{y\succeq x\}}$, for $i = 0, \ldots, k$. (For a string $x \in \{0,1\}^n$, $|x|$ denotes the *Hamming norm* of x, namely the number of ones in x.) Also [7] $\|p - \delta_0\|_\infty = \|\tilde{p} - \delta_0\|_\infty \leq \epsilon(k,n)$.

Given sets $A_1, \ldots, A_n$, we define, as in lemma 3.1, a function $f : \{0,1\}^n \to \mathbb{R}$ by $f(x) = Pr\left(\bigcap_{i:x_i=1} A_i \cap \bigcap_{i:x_i=0} A_i^c\right)$. We have that $\|f\|_1 = \sum_x f(x) = 1$ and $f(0) = 1 - Pr\left(\cup_{i=1}^n A_i\right)$. Note also, that for any $x \in \{0,1\}^n$ holds $< f, 1_{\{y\succeq x\}} >= Pr\left(\cap_{i:x_i=1} A_i\right)$.

[6] This is an easy consequence of Stirling's formula for the factorial.

[7] Recall that (with much abuse of notation) the symbol δ_0 plays a multiple role in this paper, denoting in particular both a function on $0, 1, \ldots, n$ given by $\delta_0(i) = \delta_{0,i}$ and its lifting to S.

Consider the inner product $< f, \tilde{p} >$. On one hand, it is very close to $f(0)$, indeed:

$$| < f, \tilde{p} > - f(0)| = | < f, \tilde{p} > - < f, \delta_0 > | = | < f, \tilde{p} - \delta_0 > | \leq \|f\|_1 \cdot \|\tilde{p} - \delta_0\|_\infty \leq \epsilon(k, n).$$

On the other hand, it can be written as a linear combination of $Pr\left(\cap_{i \in T} A_i\right)$, for $|T| \leq k$:

$$< f, \tilde{p} > = < f, \sum_{s=0}^{k} a_s \tilde{f}_s > = \sum_{s=0}^{k} a_s < f, \tilde{f}_s > = \sum_{s=0}^{k} \frac{a_s}{\binom{n}{s}} \left(\sum_{x:|x|=s} < f, 1_{y \succeq x\}} > \right) =$$

$$\sum_{s=0}^{k} \frac{a_s}{\binom{n}{s}} \left(\sum_{T:|T|=s} Pr\left(\cap_{i \in T} A_i\right) \right).$$

Note, that $Pr\left(\cap_{i \in \emptyset} A_i\right) = 1$, and therefore we get (12) by setting $\alpha_0^{k,n} = 1 - a_0$, and $\alpha_i^{k,n} = -\frac{a_i}{\binom{n}{i}}$, for $i = 1, \ldots, k$.

The constants a_i for $i = 0, \ldots, k$ can be easily found from the expansion $p = \sum_{s=0}^{k} a_s f_s$, by solving a system of linear equations. The vectors $\{f_s\}$ are given, and p is a linear combination with known coefficients of the discrete Chebyshev polynomials $p_0, \ldots, p_k$, where [18] $p_s(i) = \sum_{j=0}^{s} (-1)^j \binom{s}{j} \binom{i}{j} \binom{n-i}{s-j}$. ∎

5 A "primal" proof of theorem 1.7

In this section we give a proof of theorem 1.7 which does not rely upon duality theory. Our main tools are the discrete Fourier analysis on $\mathbb{Z}_2^n$ and a family of orthogonal polynomials: the Hahn polynomials (see section 2). Elements of Z_2^n (or $\{0,1\}^n$) will be viewed as either vectors or as subsets of $\{1, \ldots, n\}$.

Applying lemma 3.3 to the definition 1.1 of $E(k, n)$, we obtain:

$$E(k, n) = \sup \left\{ h(0) \mid h : \{0,1\}^n \to \mathbb{R};\ \|h\|_1 \leq 2,\ (\zeta h)(x) = 0,\ for\ |x| \leq k \right\}.$$

Now we are going to "play" with this definition, modifying it until we obtain a convenient characterization (13), for which we will be able to provide tight bounds. First, let us translate the zeta transform properties of a function h to its Fourier transform properties.

Lemma 5.1: *For a function $h : \mathbb{Z}_2^n \to \mathbb{R}$, let ζh be the zeta transform of h and $\hat{h}$ its Fourier transform. Then $(\zeta h)(x) = 0$ for $|x| \leq k$ if and only if $\hat{h}(x) = 0$ for $|x| \leq k$.*

Proof: In the proof of proposition 1.2 we have shown that $(\zeta h)(x) = 0$ for $|x| \leq k$ iff h is orthogonal to Span $\left\{1_{\{y: y \succeq x\}}\ :\ |x| \leq k\right\}$. Therefore, we need only to show that

$\mathrm{Span}\left\{1_{\{y:y\succeq x\}}\ :\ |x|\le k\right\} = \mathrm{Span}\{W_x\ :\ |x|\le k\}$. (Recall, that $\{W_x\}$ are the characters of $\mathbb{Z}_2^n$, see section 2.)

Note, that both these spanning sets are linearly independent, and of the same cardinality. Therefore, it suffices to prove that if $|x|\le k$ then $f_x = 1_{\{y:y\succeq x\}}$ belongs to $\mathrm{Span}\{W_x\ :\ |x|\le k\}$.

Let us, for this purpose, compute the Fourier expansion of f_x. Let $|x| = a$, and w.l.o.g. assume that the last a bits of x are 1. Set $m = n-a$, and for $s\in\{0,1\}^n$ let s' be the restriction of s to the first m coordinates. Then:

$$< f_x, W_s >= \frac{1}{2^n}\sum_{y\succeq x}(-1)^{y\cap s} = \frac{1}{2^n}\sum_{y\succeq x}(-1)^{\sum_{i=1}^n y_i s_i} =$$

$$\frac{1}{2^n}\sum_{z\in\{0,1\}^m}(-1)^{\sum_{i=1}^m z_i s_i+\sum_{j=m+1}^n s_j} = \frac{(-1)^{\sum_{j=m+1}^n s_j}}{2^n}\cdot\sum_{z\in\{0,1\}^m}(-1)^{\sum_{i=1}^m z_i s'_i} =$$

$$\begin{cases}\frac{(-1)^{|s|}}{2^{|x|}} & s'=0\\ 0 & \text{otherwise}\end{cases}.$$

Clearly, $s'=0$ if and only if $s\preceq x$. Consequently $f_x = \sum_{s\preceq x}\frac{(-1)^{|s|}}{2^{|x|}}W_s \in \mathrm{Span}\{W_x\ :\ |x|\le k\}$, and we are done. ∎

In other words,

$$E(k,n) = \sup\left\{h(0)\ |\ \ \|h\|_1\le 2,\ \hat{h}(x)=0,\ for\ |x| = 0,1\ldots,k\right\}.$$

We would like, for a reason which will become clear in a moment, to move the zeroes of the Fourier transform of h from the lower levels of $\{0,1\}^n$ to the upper ones. For this purpose, observe that for any $f:\mathbb{Z}_2^n\to\mathbb{R}$, and for any $s\in\mathbb{Z}_2^n$ holds $\widehat{f\cdot W_s}(x) = \hat{f}(x+s)$. Let $s=\bar{1}$ be the vector of all ones. Multiplying every function h by $W_{\bar{1}}(x) = (-1)^{|x|}$, we get

$$E(k,n) = \sup\left\{h(0)\ |\ \ \|h\|_1\le 2,\ \hat{h}(x)=0,\ for\ |x| = n-k, n-k+1,\ldots,n\right\}.$$

Let π be the projection on the space of symmetric functions on $\{0,1\}^n$, defined by $(\pi f)(x) = \frac{1}{\binom{n}{i}}\sum_{|y|=|x|}f(y)$ (see subsection 3.1 and section 2). Since π is a contraction in the l_1 norm and commutes with Fourier transform, for any function h in our set, πh is also in the set and $h(0) = (\pi h)(0)$. That means, that without any loss we may consider only symmetric functions, and

$$E(k,n) = \sup\left\{h(0)\ |\ h \text{ symmetric}, \|h\|_1\le 2,\ \hat{h}(x)=0,\ for\ |x| = n-k, n-k+1,\ldots,n\right\}.$$

Our next step is to express $E(k,n)$ as a solution of an extremal problem for polynomials on $0,1,\ldots,n$. Recall that any function f on $0,1,\ldots,n$ lifts to a symmetric function $\tilde{f}$ on $\{0,1\}^n$, by setting $\tilde{f}(x) = f(|x|)$.

Let h be a symmetric function on $\{0,1\}^n$, and let $\hat{h}$ be its (symmetric) Fourier transform. Let $a = a_h$ be a function on $0, 1, \ldots, n$ lifting to h, and $b = b_h$ be a function on $0, 1, \ldots, n$ lifting to $\hat{h}$. Then we claim that $a = \sum_{s=0}^{n} b(s) K_s$, where $\{K_s\}_{s=0}^{n}$ are the Krawchouk polynomials (see section 2). Indeed, for $x \in \{0,1\}^n$ with $|x| = i$, we have $a(i) = h(x) = \sum_{y \in \{0,1\}^n} \hat{h}(y) W_y(x) = \sum_{s=0}^{n} b(s) \cdot \sum_{y:|y|=s} W_y(x) = \sum_{s=0}^{n} b(s) K_s(i)$.

Since K_s is a polynomial of degree s, it follows that symmetric functions h with $\hat{h}(x) = 0$ for $|x| = n-k, n-k+1, \ldots, n$ are exactly the functions for which a_h is a polynomial of degree $< n-k$. Plugging this in, we obtain a "primal" characterization of $E(n,k)$ in terms of polynomials of a single variable:

$$E(k,n) = \sup \left\{ f(0) \mid f \ : 0, 1 \ldots, n \to \mathbb{R}; \ \sum_{i=0}^{n} \binom{n}{i} |f(i)| \le 2, \ \text{degree of } f < n-k \right\}. \quad (13)$$

It turns out, that essentially tight bounds on the supremum in (13) were already obtained in [15] in the context of looking for tight bounds on the Lövasz θ function of a family of graphs $G_{n,k}$. (The graphs $G_{n,k}$are defined in the introduction.) The details regarding the computation of $\theta(G_{n,k})$ are provided in [15]. Here let us only say that, using one of the many properties of the θ-function, specifically the fact that for *subgraphs of association schemes* this function has a convenient representation as an optimum of a certain linear program [16], it is quite straightforward [15] to obtain:

$$\theta(G_{n,n-k}) = 2^n \cdot \sup \left\{ f(0) \mid f \ : 0, 1 \ldots, n \to \mathbb{R}; \ f \ge 0, \ \sum_{i=0}^{n} \binom{n}{i} f(i) \le 1, \ \text{degree of } f < n-k \right\}.$$

Therefore $E(k,n) \ge \frac{\theta(G_{n,n-k})}{2^{n-1}}$.

In fact, $E(k,n)$ is not much greater than $\frac{\theta(G_{n,n-k})}{2^{n-1}}$. More precisely [15]:

$$\frac{\theta(G_{n,n-k})}{2^{n-1}} \le E(k,n) \le \sqrt{n} \cdot \frac{\theta(G_{n,n-k})}{2^{n-1}}. \quad (14)$$

The approximate equality in (14) leads one to speculate that, in fact, a precise equality might hold here. So here is a

Question

Is it true that $E(k,n) = \frac{\theta(G_{n,k})}{2^{n-1}}$, and if yes, is there a straightforward combinatorial proof for this fact?

We continue the proof of theorem 1.7 by repeating the argument from [15]. In (13) we have to solve an extremal question about polynomials on $0, 1, \ldots, n$, and one of the conditions on the polynomials in question is given in terms of their *weighted* l_1 norm: $\|f\|_1 := \sum_{i=0}^{n} \binom{n}{i} |f(i)| \le 2$. It turns out, that the question becomes much simpler if we switch to the appropriate Euclidean norm (compare subsection 4.2).

Consider the following Euclidean norm: $\|f\|_2 = \sqrt{\sum_{i=0}^{n} \binom{n}{i}^2 f^2(i)}$. Clearly it is almost equivalent to the l_1 norm described above, namely $\|f\|_2 \leq \|f\|_1 \leq \sqrt{(n+1)}\|f\|_2$. Consequently, if we agree to an uncertainty of a $\sqrt{n}$ magnitude, it suffices to find

$$\tilde{E}(k,n) = \sup\left\{ f(0) \mid \sqrt{\sum_{i=0}^{n} \binom{n}{i}^2 f^2(i)} \leq 1, \text{ degree of } f < n-k \right\}.$$

Explicitly, the relation between these two values is given by: $\frac{1}{\sqrt{n+1}}\tilde{E}(k,n) \leq E(k,n) \leq \tilde{E}(k,n)$.

Lemma 5.2:

$$\tilde{E}(k,n) = \Theta\left((n^2-k^2)^{\frac{1}{4}}\right) \cdot \frac{\binom{n}{\frac{n-k}{2}}}{2^n}.$$

Proof: Let X be the $n+1$-dimensional vector space of real-valued functions on $0, 1 \ldots, n$, and let $\mathcal{P}_{n-k-1}$ be the $n-k$-dimensional subspace of polynomials of degree $< n-k$. The important ace up our sleeve, which justifies switching to the l_2 norm, is being in possession of an orthonormal (with respect to an inner product $< f, g >= \sum_{i=0}^{n} \binom{n}{i}^2 f(i)g(i)$) basis $v_0, \ldots, v_n$ of X such that $v_0, \ldots, v_{n-k-1}$ span $\mathcal{P}_{n-k-1}$. This basis is given by *Hahn polynomials* $\{q_s\}$, see section 2.

Expanding $f \in \mathcal{P}_{n-k-1}$ as a linear combination of $\{q_s\}$, namely $f = \sum_{s=0}^{n-k-1} a_s q_s$, and using the Parseval identity, we see that: $f(0) = \sum_{s=0}^{n-k-1} a_s q_s(0)$, and $\sum_{i=0}^{n} \binom{n}{i}^2 f^2(i) = \|f\|_2^2 = \sum_{s=0}^{n-k-1} a_s^2$.

Consider the space $V \cong \mathbb{R}^{n-k}$ of the coefficients $a_0, \ldots, a_{n-k-1}$ with the usual Euclidean norm $\|a\|_2 = \sqrt{\sum_{i=0}^{n-k-1} a_i^2}$. To find $\tilde{E}(k,n)$ we have to maximize, over all vectors $a = (a_0, \ldots, a_{n-k-1})$ of length 1, the inner product $< a, q >$, where $q := (q_0(0), \ldots, q_{n-k-1}(0))$. This maximum is clearly attained at $a = \frac{q}{\|q\|_2}$ implying $\tilde{E}(k,n) = \|q\|_2 = \sqrt{\sum_{s=0}^{n-k-1} q_s^2(0)}$. Recall, that we know $q_s(0) = \sqrt{\frac{\binom{2n}{s} - \binom{2n}{s-1}}{\binom{2n}{n}}}$, and therefore $\tilde{E}(k,n) = \sqrt{\frac{\binom{2n}{n-k-1}}{\binom{2n}{n}}}$.

We finish the proof of the lemma by once again (see previous section) using the fact that for any n and $0 \leq t \leq n$ holds: $\binom{2n}{2t} = \Theta\left(\sqrt{\frac{t(n-t)}{n}}\right) \cdot \binom{n}{t}^2$. ∎

Now we are in the same position as at the end of the dual proof presented in the previous section, namely we know that $\Omega\left(\left(1-\frac{k^2}{n^2}\right)^{\frac{1}{4}} \cdot \frac{\binom{n}{\frac{n-k}{2}}}{2^n}\right) \leq E(k,n) \leq O\left((n^2-k^2)^{\frac{1}{4}} \cdot \frac{\binom{n}{\frac{n-k}{2}}}{2^n}\right)$.

Here, however, we can do a bit better. To prove the lower bound in (7), recall that $E(k,n) \geq \frac{\theta(G_{n,n-k})}{2^{n-1}}$. We use the basic property of the θ-function of a graph, namely it being an upper bound on the size of the maximal independent set of this graph. Maximal independent sets in $G_{n,n-k}$ are well-known [7]. These are the Hamming balls of radius $\frac{n-k}{2}$, i.e. of size $\sum_{i=0}^{\frac{n-k}{2}} \binom{n}{i}$, and we are done. ∎

Let us compare the bounds of theorem 1.7 to these obtained in [10] and [8] (see Introduction). Note, that the upper and the lower bound in (7) differ by a factor of at most $O(\sqrt{n})$, and actually by only a constant factor when $k = O(\sqrt{n})$. Note also, that $\frac{\binom{n}{\frac{n-k}{2}}}{2^n} = exp\{-\Theta\left(\frac{k^2}{n}\right)\}$. These facts imply that our bounds here improve the bounds in [10] and [8] for $k = \Omega(\sqrt{n})$.

Interestingly enough, the tight answer for the remaining range of k, namely $k = o(\sqrt{n})$, which is obtained in [10], seems, so far, to be out of reach for the methods presented here. Namely, for these values of k, (7) gives only $1 - \Theta\left(\frac{k}{\sqrt{n}}\right) \leq E(k,n) \leq 1$, while we know the correct answer from [10]: $E(k,n) = 1 - \Theta\left(\frac{k^2}{n}\right)$.

In conclusion: to obtain the best known bounds on $E(k,n)$ for the whole range of k, one must use [10] for $k = o(\sqrt{n})$ and (7) for $k = \Omega(\sqrt{n})$.

6 The Boole problem and further questions

Attempts to approximate the probability of the union of a family of n sets, given the probabilities of the intersection for some of the subfamilies, have a long history. In fact in 1854, Boole [1] asked how well can we approximate the probability of the union, given the only the probabilities of the individual sets and of their pairwise intersections. This question, known [4] as *the Boole problem*, in our language translates simply to: "What is $E(2,n)$?"

In this sense, the question (1) of finding $E(k,n)$ for $k = 0,\ldots,n$ is a natural generalization of the Boole problem. And for this question we have essentially tight answers (7). However, one can generalize even further, and our understanding of the problem takes a dramatic plunge. For a family $\mathcal{F}$ of subsets of $\{0,1\}^n$, let

$$E(\mathcal{F},n) = \sup\left(Pr(\cup_{i=1}^n A_i) - Pr(\cup_{j=1}^n B_j)\right) \tag{15}$$

where the supremum ranges over all families of events $A_1,\ldots,A_n$ and $B_1,\ldots,B_n$ in any probability spaces, that satisfy:

$$Pr(\cap_{i\in S} A_i) = Pr(\cap_{j\in S} B_j)$$

for all $S \in \mathcal{F}$. <u>What is $E(\mathcal{F},n)$?</u>

This is, of course, a generalization of (1), since $E(k,n) = E(\{S \subseteq [n] \ : \ |S| \leq k\}, n)$. A straightforward modification of the proof of proposition 1.2 leads to a dual characterization of $E(\mathcal{F}, n)$:

$$E(\mathcal{F}, n) = 2 \cdot \|\delta_0 - \mathcal{U}_{\mathcal{F}}\|_{\infty}, \tag{16}$$

where $\delta_0 : \{0,1\}^n \to \mathbb{R}$ is the characteristic function of zero, and $\mathcal{U}_{\mathcal{F}}$ is the space of functions from $\{0,1\}^n$ to $\mathbb{R}$ spanned by the characteristic functions $1_{\{y \ : \ y \succeq x\}}$ for all $x \in \mathcal{F}$.

However, since the space $\mathcal{U}_{\mathcal{F}}$ is, in general, not invariant under the symmetric projection π, we do not have an equivalent of proposition 1.4. Consequently, the problem of finding $E(\mathcal{F}, n)$ seems much harder.

Question 1: Given a family $\mathcal{F} \subseteq \{0,1\}^n$, find good estimates on $E(\mathcal{F}, n)$.

The following question seems to be natural from the computational point of view:

Question 2: How large should the family $\mathcal{F}$ be, so that $E(\mathcal{F}, n) \to_{n \to \infty} 0$?

One can even entertain a hope to find such a family $\mathcal{F}$ of polynomial size, but this is shown to be impossible in the following lemma.

Lemma 6.1:

$$|\mathcal{F}| \geq \sum_{i=0}^{\lfloor \log\left(\frac{1}{E(\mathcal{F},n)}\right) \rfloor} \binom{n}{i}.$$

Proof: We use the dual characterization (16) of $E(\mathcal{F}, n)$. Let $f = \sum_{x \in \mathcal{F}} a(x) \cdot 1_{\{y \ : \ y \succeq x\}} \in \mathcal{U}_{\mathcal{F}}$ such that $\|f - \delta_0\|_{\infty} = E(\mathcal{F}, n)$. Extend a to a function $b : \{0,1\}^n \to \mathbb{R}$, by setting $b(x) = a(x)$ if $x \in \mathcal{F}$ and $b(x) = 0$ otherwise. Then, for any $z \in \{0,1\}^n$ holds $f(z) = \sum_{x \preceq z} b(x)$, namely f is the (transposed) zeta transform of b. We reconstruct b from f using the Möbius inversion formula for the boolean lattice $\{0,1\}^n$ (see section 2):

$$b(x) = \sum_{z \preceq x} \mu(z,x) f(z) = \sum_{z \preceq x} (-1)^{|x|-|z|} f(z) = (-1)^{|x|} f(0) + \sum_{z \preceq x, \ z \neq 0} (-1)^{|x|-|z|} f(z).$$

Since $\|f - \delta_0\|_{\infty} = E(\mathcal{F}, n)$, we have $|(-1)^{|x|} f(0) + \sum_{z \preceq x, \ z \neq 0} (-1)^{|x|-|z|} f(z)| \geq 1 - 2^{|x|} \cdot E(\mathcal{F}, n)$. Therefore, if $|x| < \log\left(\frac{1}{E(\mathcal{F},n)}\right)$ then $b(x) > 0$ implying $x \in \mathcal{F}$. The claim of the lemma follows. ∎

Our next two questions deal with unfinished business from previous sections: First, let us repeat the question from section 5:

Question 3: Is it true that $E(k,n) = \frac{\theta(G_{n,k})}{2^{n-1}}$, and if yes, is there a straightforward combinatorial proof for this fact?

Going back to the bounds (11) and (7), we still have a gap of order $\sqrt{n}$ between the upper and the lower bounds for $E(k,n)$. This gap, although small, is rather frustrating. The bounds (11) and (7) are obtained similarly: an extremal problem (2) (correspondingly (13)) is modified by replacing l_∞ (correspondingly l_1) norm with an appropriate Euclidean norm. The solution to the modified problem is shown to approximate the solution to the original problem by a factor of $O(\sqrt{n})$. The modified problem is then solved.

We point out, that the solution to the modified problem might actually give a better approximation than $O(\sqrt{n})$. Suppose, for instance, that the polynomial $p^{(k)}$ on which the distance $\|\delta_0 - \mathcal{P}_k\|_2$ (see subsection 4.2) is obtained, has the following property : $\|p^{(k)}\|_\infty = \Theta\left(\frac{1}{\sqrt{n}}\right)\|p^{(k)}\|_2$. Then it is not hard to see that $\|\delta_0 - \mathcal{P}_k\|_2 = \Theta(\|\delta_0 - \mathcal{P}_k\|_\infty)$ and $E(k,n) = \Theta\left((n^2-k^2)^{\frac{1}{4}} \cdot \frac{\binom{n}{\frac{n-k}{2}}}{2^n}\right)$. Note, that $p^{(k)}$ is a linear combination of Discrete Chebyshev polynomials with known coefficients.

Similarly, denote by $q^{(k)}$ the polynomial on which $\tilde{E}(k,n)$ is attained (see section 5). This polynomial is a linear combination of Hahn polynomials with known coefficients. If $\|q^{(k)}\|_1 = \Theta(\sqrt{n})\|q^{(k)}\|_2$, then $E(k,n) = \Theta\left(\left(1-\frac{k^2}{n^2}\right)^{\frac{1}{4}} \cdot \frac{\binom{n}{\frac{n-k}{2}}}{2^n}\right)$.

Question 4: Can a better analysis of either $p^{(k)}$ or $q^{(k)}$ (or both of them together(?)) lead to asymptotically tight bounds on $E(k,n)$?

7 Acknowledgement

I am grateful to Nathan Linial, Simon Litsyn and Michael Krivelevich for valuable conversations.

References

[1] G. Boole, **An investigation of the laws of thought on which are founded the mathematical theories of logic and probabilities**, Dover 1st printing, 1854.

[2] W. Chen and G.-C. Rota, q-Analogs of the inclusion-exclusion principle and permutations with restricted position, *Discrete Mathematics*, 104(1992), 7-22.

[3] Ph. Delsarte, *An algebraic approach to the association scheme of coding theory*, Philips Res. Rep., Suppl., vol. 10, 1973.

[4] M. M. Deza and M. Laurent, **Geometry of Cuts and Metrics**, Springer-Verlag, 1997.

[5] I. Honkala, T. Laihonen and S. Litsyn, On covering radius and Discrete Chebyshev polynomials, Applicable Algebra in Engineering, Communication and Computing , vol.8, 1997, pp.395–401.

[6] J. Kahn, G. Kalai and N. Linial, The influence of variables on boolean functions, *29th Symposium on the Foundations of Computer Science*, White Planes, 1988, 68 - 80.

[7] D. J. Kleitman, *On a combinatorial conjecture of Erdoš*, Journal of Comb. Theory 1, 1966, 209-214.

[8] J. Kahn, N. Linial and A. Samorodnitsky, Inclusion-exclusion: exact and approximate, *Combinatorica*, vol. 16(1996) .

[9] V. Levenshtein, *Krawchouk polynomials and universal bounds for codes and designs in Hamming spaces*, IEEE Trans. Inform. Theory, vol. IT-41, 1995, 1303-1321.

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

[11] L. Lovasz, *On the Shannon capacity of a graph*, IEEE Trans. Inform. Theory, vol. IT-25, 1979, 1-7.

[12] M. Mavronicolas, A q-analog of approximate inclusion-exclusion, *Advances in applied mathematics*, vol. 20(1998) 108-129.

[13] R. J. McEliece, E. R. Rodemich, H. Rumsey, Jr., and L. R. Welch, *New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities*, IEEE Trans. Inform. Theory, vol. IT-23, 1977, 157-166.

[14] W. Rudin, **Functional Analysis**, McGraw-Hill, 1973.

[15] A. Samorodnitsky, Extremal properties of solutions for Delsarte's linear program (1998), preprint.

[16] A. Schrijver, *A comparison of the Delsarte and Lovasz bounds*, IEEE Trans. Inform. Theory, vol. IT-25, 1979, 425-429.

[17] R. P. Stanley, **Enumerative Combinatorics**, vol. I, Cambridge University Press, 1997.

[18] G. Szego, **Orthogonal Polynomials**, American Mathematical Society, 1939.

[19] J. H. van Lint and R. M. Wilson, **A Course in Combinatorics**, Cambridge University Press, 1992.