COMBINATORICA
Akadémiai Kiadó – Springer-Verlag

COMBINATORICA **10** (4) (1990) 349–365

APPROXIMATE INCLUSION-EXCLUSION

NATHAN LINIAL and NOAM NISAN*

Received May 28, 1989

The Inclusion-Exclusion formula expresses the size of a union of a family of sets in terms of the sizes of intersections of all subfamilies. This paper considers approximating the size of the union when intersection sizes are known for only some of the subfamilies, or when these quantities are given to within some error, or both.

In particular, we consider the case when all k-wise intersections are given for every $k \leq K$. It turns out that the answer changes in a significant way around $K = \sqrt{n}$: if $K \leq O(\sqrt{n})$ then any approximation may err by a factor of $\Theta(n/K^2)$, while if $K \geq \Omega(\sqrt{n})$ it is shown how to approximate the size of the union to within a multiplicative factor of $1 \pm e^{-\Omega(K/\sqrt{n})}$.

When the sizes of all intersections are only given approximately, good bounds are derived on how well the size of the union may be approximated. Several applications for Boolean function are mentioned in conclusion.

1. Introduction

Are all the terms in the inclusion-exclusion formula

$$|A_1 \cup A_2 \cup \ldots \cup A_n| =$$
$$\sum_i |A_i| - \sum_{i<j} |A_i \cap A_j| + \sum_{i<j<k} |A_i \cap A_j \cap A_k| - \ldots + (-1)^n |A_1 \cap \ldots \cap A_n|$$

really necessary? The obvious answer is positive. In the absence of even a single term the size of the union is not uniquely specified. But can the size of the union be *approximated* well, given only some of the terms? Also, if terms are given to within some error, can the size of the union approximated? The present article answers questions of this general character.

Our interest in these problem arose from some computational considerations: Many computational problems may be viewed as asking for the size of a union of a collection of sets. On some instances it turns out that while computing the size of the union is rather difficult, computing the sizes of members in the family, or even of arbitrary intersections thereof is easy. In these cases, the inclusion-exclusion formula may be used to find the size of the union.

Perhaps the most obvious example is the problem of computing the number of satisfying assignments to a DNF formula (a problem known to be #P-complete [10]). This problem can be stated as that of computing the size of the union of the sets of assignments that satisfy the various clauses of the DNF formula. The number of

AMS subject classification (1980): 05 A 20

*Partially supported by NSF 865727-CCR and ARO DALL03-86-K-017. Part of this work was done in U.C. Berkeley, supported by NSF CCR-8411954.

assignments satisfying an intersection of clauses is either zero or 2^m where m is the number of variables which appear in none of these clauses. So inclusion-exclusion may be applied to derive the size of the union. This procedure takes time which is exponential in the number of *clauses*, and seems to be the best algorithm known for this problem when the number of clauses is less than the total number of variables that occur in the formula. In fact, in these cases, this method even seems to be the quickest way known to check whether *every* assignment satisfies the DNF formula, i.e. to check if the complement, a CNF formula, is satisfiable.

A somewhat more subtle example is Ryser's formula [9] for computing the permanent (also a #P-complete problem [10]). Ryser essentially reduces the problem of computing the permanent to a problem of computing the size of a union of sets, where the sizes of intersections of all subfamilies can be easily computed. The inclusion-exclusion formula is then used to compute the size of the union. This is the quickest method known to compute the permanent, as it requires $2^{n+o(n)}$ operations to compute the permanent of an n by n matrix, instead of the trivial $n!$.

The obvious drawback of using the inclusion-exclusion formula is the fact that it has an exponential number of terms, and that, as mentioned, all terms are necessary, i.e. if the size of the intersection of any subcollection is missing, then the size of the union cannot be computed. This prompted our interest in *approximate* versions of the inclusion-exclusion formula.

We start, in Section 2, with the following version of this problem: Let A_1, A_2, ..., A_n be a collection of sets. Suppose that $\left|\bigcap_{i\in S} A_i\right|$ is given for every subset $S \subset [n]$ of cardinality $|S| \leq k$. How well can $\left|\bigcup A_i\right|$ be approximated based only on this information? Equivalently, let $A_1, A_2, \ldots, A_n$ and $B_1, B_2, \ldots, B_n$ be two collections of sets such that for every $S \subset [n]$ of cardinality $|S| \leq k$ there holds $\left|\bigcap_{i\in S} A_i\right| = \left|\bigcap_{i\in S} B_i\right|$. How different can $\left|\bigcup A_i\right|$ and $\left|\bigcup B_i\right|$ be?

A naive approach to the problem would be to truncate the exclusion-inclusion formula up to the k-terms. This approach is easily seen to fail completely e.g. when all sets are identical.

We give a nearly complete answer to this question, essentially showing that for $k < O(\sqrt{n})$ no good approximation is possible, while for larger k a good approximation is possible. The essence of our main result may be formulated as:

Theorem 1. *Let k and n be integers and let $A_1, A_2, \ldots, A_n$ and $B_1, B_2, \ldots, B_n$ be collections of sets where not all B_i are empty and where:*

$$\left|\bigcap_{i\in S} A_i\right| = \left|\bigcap_{i\in S} B_i\right|$$

for every subset $S \subset [n]$ such that $|S| \leq k$, then

1. *For $k \geq \Omega(\sqrt{n})$*

$$\frac{\left|\bigcup_{i=1}^n A_i\right|}{\left|\bigcup_{i=1}^n B_i\right|} = 1 + O\left(e^{-2k/\sqrt{n}}\right)$$

2. *For $k \leq O(\sqrt{n})$*

$$\frac{\left|\bigcup_{i=1}^n A_i\right|}{\left|\bigcup_{i=1}^n B_i\right|} = O\left(\frac{n}{k^2}\right)$$

This bound is tight up to a constant factor for $k \leq \sqrt{n}$.

The main observation used in the proof of this theorem is that the problem may be reduced, via linear programming, to questions in approximation theory and in particular to the theory of Chebyshev polynomials.

Earlier references for related work can be found in Prekopa [7]. In particular, the theory of linear programming seems to have first been employed by Kwerel [5], but the roots of this idea go back at least as far as Bonferroni [3]. More recently this question has been studied by Alon and Hastad [1], who, using different techniques than ours, could show than when $k \leq O(\log \log n)$, the size of the union cannot be approximated well, and solve the case $k = n - 1$ completely.

Theorem 2, appearing in Section 2, shows how to effectively derive a good approximation for the size of the union. The approximation is given by a linear form, and is essentially as good as Theorem 1 implies is possible.

Section 3 considers how well the size of a union can be approximated when the sizes of all of the k-wise intersections are only given *approximately*. An almost complete answer to this question is given.

In Sections 4 and 5 we present further computational motivation for our problems: the study of Boolean functions. We present several applications of our results to questions related to the computational complexity of Boolean functions. These questions and similar ones from circuit complexity were in fact the point of departure for the present research.

In Section 4 the following problem is considered: Let $f_1, \ldots, f_n$, be n Boolean functions, none of which can even approximate a target function g, and moreover, the conjunction of any subset of these functions cannot approximate g. Is it possible that the *disjunction* of these functions approximates the target function g? Rather tight bounds are established on the extent to which the original functions must not approximate g as to insure a similar conclusion for the disjunction.

Section 5 contains some comments on constant depth circuits. We conjecture that any distribution which is t-wise independent, for large enough t, "looks random" to any small constant depth circuit. The methods of this paper yield only a weak result of this form for depth 2 circuits.

2. Approximation Using Initial Terms

2.1. Main result

Let $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ and $\mathcal{B} = (B_1, B_2, \ldots, B_n)$ be two collections of sets. Assume that for any subset $S \subset [n]$ of cardinality $|S| \leq k$ it is true that $\left|\bigcap_{i \in S} A_i\right| = \left|\bigcap_{i \in S} B_i\right|$. How different can $\left|\bigcup A_i\right|$ and $\left|\bigcup B_i\right|$ be?

This problem is clearly scalable, i.e. multiplying each size by a constant will change every answer by the same constant. It is therefore without loss of generality that we restrict our attention to events in a probability space.

Definition 1.

$$E(k, n) = \sup\left(Pr\left[\bigcup_{i=1}^{n} A_i\right] - Pr\left[\bigcup_{i=1}^{n} B_i\right]\right)$$

where the supremum ranges over all collections of events, in all probability spaces, $A_1, A_2, \ldots, A_n$ and $B_1, B_2, \ldots, B_n$ that satisfy

$$Pr\Big[\bigcap_{i\in S} A_i\Big] = Pr\Big[\bigcap_{i\in S} B_i\Big]$$

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

Our aim in this section is to derive bounds for $E(k,n)$. The first fact to notice is that there is no loss of generality in assuming symmetry. A j-atom of $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ is defined to be the intersection of j members in $\mathcal{A}$ and the complements of the other $n-j$ members. A collection of events $\mathcal{A}$ is called *symmetric* if for every $1 \leq j \leq n$, all j-atoms of $\mathcal{A}$ are of the same probability.

Lemma 1. *The supremum, $E(k,n)$, remains unchanged even when $\mathcal{A}$ and $\mathcal{B}$ are restricted to be symmetric.*

Proof. Given non-symmetric $\mathcal{A}$ and $\mathcal{B}$, we construct symmetric collections $\mathcal{A}'$ and $\mathcal{B}'$ with the same difference in the probability of their union. The probability of each j-atom in $\mathcal{A}'$ is defined to be the average of the probabilities of all j-atoms in $\mathcal{A}$, and similarly for $\mathcal{B}'$. ∎

From now on $\mathcal{A} = (A_1, A_2, \ldots, A_n)$ and $\mathcal{B} = (B_1, B_2, \ldots, B_n)$ are always assumed to be symmetric. Here is some notation: For $1 \leq j \leq n$ let a_j (resp. b_j) be the probability of the union of all $j-$atoms in $\mathcal{A}$ (resp $\mathcal{B}$.) i.e.

$$a_j = \binom{n}{j} Pr\Big[\bigcap_{i\in S} A_i \cap \bigcap_{i\notin S} A_i{}^c\Big]$$

and

$$b_j = \binom{n}{j} Pr\Big[\bigcap_{i\in S} B_i \cap \bigcap_{i\notin S} B_i{}^c\Big]$$

where S is any set of cardinality j. For $1 \leq j \leq k$ let r_j be the sum of the probabilities of all j-intersections, i.e.,

$$r_j = \binom{n}{j} Pr\Big[\bigcap_{i\in S} A_i\Big] = \binom{n}{j} Pr\Big[\bigcap_{i\in S} B_i\Big]$$

where S is any set of cardinality j. For $1 \leq j \leq k$ define the linear form

$$E_j(x_1, \ldots, x_n) = \sum_{i=j}^{n} \binom{i}{j} x_i.$$

The next lemma indicates the role of linear forms E_j:

Lemma 2. *For every symmetric collection of events $A_1, A_2, \ldots, A_n$, and for every $1 \leq j \leq k$:*

$$r_j = E_j(\overrightarrow{a})$$

Proof. Consider any $i-$atom, say the one corresponding to the set S of size i. For any $j \leq i$, the weight of this atom is counted once towards r_j for any subset of size j of S, namely, $\binom{i}{j}$ times. ∎

Observe first that $E(k,n)$ can be expressed as the value of a certain linear program:

Lemma 3. *$E(k,n)$ is the optimum of the following linear program:*

$$Maximize \sum_{i=1}^{n} x_n$$

Subject to the constraints:

(1) $$For\ 1 \leq j \leq k\ :\ E_j(\overrightarrow{x}) = 0$$

(2) $$For\ any\ S \subseteq [n]\ :\ -1 \leq \sum_{i \in S} x_i \leq 1$$

Proof. Let $A_1, A_2, \ldots, A_n$ and $B_1, B_2, \ldots, B_n$ be (symmetric) collections of events. Let a_i and b_i be defined as above and let $x_i = a_i - b_i$. The previous lemma immediately implies that the x_i's satisfy constraints of type 1. Constraints of type 2 are satisfied, as the a_i's are probabilities of disjoint events, and so are the b_i's. Note that $\sum_{i=1}^{n} x_i$ is exactly the difference in the probabilities of the union of the A_i's and of the B_i's. Thus the optimum of the linear program is at least $E(k,n)$.

On the other hand, let $x_1, \ldots, x_n$ be reals satisfying constraints 1 and 2. Define a_i to be x_i if $x_i > 0$, and 0 otherwise, and define b_i to be $-x_i$ if $x_i < 0$ and 0 otherwise. Consider collections of events $\mathcal{A}$ and $\mathcal{B}$ as follows: each j-atom of $\mathcal{A}$ has probability $a_j \Big/ \binom{n}{j}$. Such a collection exists, because the a_i's are all non-negative, and sum to at most 1. $\mathcal{B}$ is defined similarly with the b_i's. Note that the difference in the probabilities of the union of the A_i's and of the union of the B_i's is exactly $\sum_i x_i$. Also the sizes of the j-intersections of the A_i's and of the B_i's are equal for all $1 \leq j \leq k$ as they are both given by $E_j(\overrightarrow{a}) = E_j(\overrightarrow{b})$ (by constraint 2). ∎

By passing to the dual some useful insight may be gained:

Lemma 4. *E(k,n) is given by the optimum of the following problem:*

$$Minimize \max_{i\ integer,\ 1 \leq i \leq n} (1 - f_i)$$

over all linear forms $f = \sum_{i=1}^{n} f_i x_i$ that are linear combinations of the linear forms E_j for $1 \leq j \leq k$, and satisfy $f_i \leq 1$ for every integer $1 \leq i \leq n$.

Proof. In the dual optimization problem linear combinations of the equations (1) and inequalities (2) are considered which yield the vector $x_1 + x_2 + \ldots + x_n$. Equations of type (1) may appear with an arbitrary coefficient, since they add nothing to the cost of the dual, and only inequalities of type (2) contribute to it. Consider an optimal combination of inequalities of both types and concentrate on the contribution of type

(1) equations. Let c_i be the coefficient of x_i in this restricted combination, and let $c_+ = \max(0, \max_i c_i - 1)$ and $c_- = \max(0, 1 - \min_i c_i)$.

Our first claim is that the cost of the dual is at least $c_+ + c_-$. If $c_- > 0$, consider the index i where c_- is attained. In the full combination, the coefficient of x_i must be 1, which must come from the right hand side of inequalities of type (2), adding ε to the cost per each ε in the coefficient of x_i. These terms must supply the missing c_- in the coefficient of x_i. A similar argument applies to c_+ where the left side of (2) is used.

Secondly, observe that a cost of $c_+ + c_-$ can indeed be achieved. Consider the set S_1 of all i such that $c_i < 1$, and let ε_1 be $\min_{i \in S_1}(1 - c_i)$. Let S_2 be the set of all i such that $c_i < 1 - \varepsilon_1$, and let ε_2 be $\min_{i \in S_2}(1 - c_i)$, and so on. Combine now ε_1 time the r.h.s. of of the type (2) inequality corresponding to S_1, ε_2 times the inequality corresponding to S_2, etc. to correct all the coefficients smaller than 1 to be 1, for a total cost of c_-. A similar fix works for c_+.

So far it was shown that $E(k, n)$ is the minimum of $c_- + c_+$ over linear combinations of equations of type (1). Our next claim is that in the optimum of the dual, $c_+ = 0$. Consider any combination with $c_+ > 0$, with a cost of $c_+ + c_-$. Divide all the coefficients in the combination by $(1 + c_+)$. This yields a combination where all the coefficients are between $(1 - c_-)/(1 + c_+)$ and 1. The cost associated with the new combination does not exceed $1 - (1 - c_-)/(1 + c_+)$, less than the original $c_- + c_+$. ∎

The main observation underlying the proof is presented in the next lemma, where the problem is stated in terms of approximations by polynomials.

Lemma 5.

$$E(k, n) = \inf_q \left(\max_{m=1,\ldots,n} (1 - q(m)) \right)$$

where the infimum ranges over all polynomials q of degree at most k that have zero constant term and satisfy $q(m) \leq 1$ for all integer $1 \leq m \leq n$.

Proof. Consider the linear forms E_j as functions on $1, \ldots, n$, assigning to each i the coefficient of x_i. Viewed this way E_j is the function $\binom{x}{j}$, a polynomial of degree j. Thus the linear span of $E_1, \ldots, E_k$ consists of all $k-$th degree polynomials with a zero constant term.

The present lemma is now seen to be nothing but a restatement of the previous one. ∎

It will be easier to estimate $E(k, n)$ in terms of a related quantity:

Definition 2.

$$D(k, n) = \inf_q \left(\max_{m=1,\ldots,n} |q(m) - 1| \right)$$

where the infimum ranges over all polynomials q of degree at most k that have zero constant term.

Lemma 6.

$$E(k, n) = \frac{2D(k, n)}{1 + D(k, n)}$$

Proof. Let q be a polynomial achieving $D(k,n)$, and consider $p = q/(1+D(k,n))$. It is clear that for every integer $1 \leq m \leq n$, $[1-D(k,n)]/[1+D(k,n)] \leq p(m) \leq 1$. This implies that $E(k,n) \leq [2D(k,n)]/[1+D(k,n)]$.

Conversely, if p is a polynomial achieving $E(k,n)$, then define $q=2p/(2-E(k,n))$, and the other side follows similarly. ∎

Consider an optimization problem similar to the one posed in lemma 5 but where the variable m is any real between 1 and n, rather than an integer in that range. This continuous version is fairly close to standard questions from analysis on approximating functions throughout an interval under L_∞ (max) norm. A prototype of questions like this asks for a polynomial $P(x)$ of a given degree, with a leading coefficient 1 which minimizes $max|P(x)|$ where x ranges over $-1 \leq x \leq 1$. This specific problem is solved by Chebyshev polynomials which, not too surprisingly, play an important role in the present article as well. The interested reader may find a detailed analysis of Chebyshev polynomials in many texts on approximation theory ([4], [8]). Here are some of their properies which will be required for the present discussion.

1. The k'th Chebyshev polynomial $T_k(x)$ is a polynomial of degree k and is given by:
$$T_k(x) = \frac{(x+\sqrt{x^2-1})^k + (x-\sqrt{x^2-1})^k}{2}.$$
2. For every $-1 \leq x \leq 1$: $|T_k(x)| \leq 1$
3. There are exactly $k+1$ different points $-1 \leq x \leq 1$ for which $|T_k(x)| = 1$. The sign of $T_k(x)$ alternates between any two consecutive ones.
4. The derivative of T_k satisfies $T_k'(x) \leq k^2$ for every $-1 \leq x \leq 1$.

A word of intuition may be useful at this point. First of all, using a linear transformation, the interval in question is changed to be $[-1,1]$ rather than $[1,n]$. It turns out that for a given k and a large n, the discrete problem where test points are integers, is sufficiently close to the continuous one, which is optimized by Chebyshev polynomials. Actually, Chebyshev polynomials come very close to optimizing the discrete problem as well. This happens because for a large n the set of points at which the polynomial is tested is sufficiently dense to make the problem almost identical with the continuous problem, where all real points are examined. This heuristic argument carries through as long as the distance between test points $(1/n)$ is sufficiently smaller than the distance between any two consecutive zeros of the Chebyshev polynomial. It is known that the least distance between roots of T_k is $\Theta(1/k^2)$ which explain the transition that occurs around $k = \sqrt{n}$.

Lemma 7.
$$\frac{1-\frac{k^2}{n-1}}{\left|T_k\left(\frac{-(n+1)}{n-1}\right)\right|} \leq D(k,n) \leq \frac{1}{\left|T_k\left(\frac{-(n+1)}{n-1}\right)\right|}$$

Proof. Consider a polynomial $q_{k,n}$ which results from a linear transformation applied to the k'th Chebyshev polynomial.
$$q_{k,n}(x) = 1 - \frac{T_k\left(\frac{2x-(n+1)}{n-1}\right)}{T_k\left(\frac{-(n+1)}{n-1}\right)}$$

The upper bound for $D(k,n)$ is implied by noting that q has the following properties:

- It is a polynomial of degree k with a zero constant term (it can be easily verified that $q_{k,n}(0) = 0$).
- For any $1 \leq x \leq n$:

$$|q_{k,n}(x) - 1| \leq \frac{1}{\left|T_k\left(\frac{-(n+1)}{n-1}\right)\right|}$$

This is so since for all such x, $[2x-(n+1)]/[n-1]$ is between -1 and 1, and thus $|T_k([2x-(n+1)]/[n-1])| \leq 1$.

Turn back now to the lower bound for $D(k,n)$. Assume to the contrary that a polynomial $p(x)$ of degree k with zero constant term satisfies

$$|p(x) - 1| < \frac{1 - \frac{k^2}{n-1}}{\left|T_k\left(\frac{-(n+1)}{n-1}\right)\right|}$$

for all integers $1 \leq x \leq n$. The properties of Chebyshev polynomials mentioned above imply the following for $q_{k,n}$:

- There are exactly $k+1$ points $1 \leq x \leq n$ such that

$$|q_{k,n}(x) - 1| = \frac{1}{\left|T_k\left(\frac{-(n+1)}{n-1}\right)\right|}$$

and the sign of $q_{k,n}(x) - 1$ alternates between each two consecutive points.

- The derivative of $q_{k,n}$ satisfies

$$\left|q'_{k,n}(x)\right| \leq \frac{k^2}{2(n-1)\left|T_k\left(\frac{-(n+1)}{n-1}\right)\right|}$$

for all $1 \leq x \leq n$.

Let us examine the $k+1$ extrema of $q_{k,n}$, and consider the integer points nearest to them, which we call $z_1, \ldots, z_{k+1}$. Each of these points is at most $1/2$ away from an extremum, and by the bound on the derivative $q'_{k,n}$ for all $i = 1, \ldots, k+1$:

$$|q_{k,n}(z_i) - 1| \geq \frac{1 - \frac{k^2}{n-1}}{\left|T_k\left(\frac{-(n+1)}{n-1}\right)\right|}$$

and, moreover, $q_{k,n}(z_i) - 1$ changes sign between any two consecutive z_i's. Now consider the polynomial $p(x) - q_{k,n}(x)$. It changes sign between any two consecutive z_i's, so it must have at least k roots between 1 and n. But it is a polynomial of degree at most k, which vanishes at 0 as well, a contradiction. ∎

This finishes the derivation of our bounds for $E(k,n)$. The next theorem summarizes our results:

Theorem 1. *Let k and n be integers and let $A_1, A_2, \ldots, A_n$ and $B_1, B_2, \ldots, B_n$ be collections of sets that satisfy:*

$$\left|\bigcap_{i \in S} A_i\right| = \left|\bigcap_{i \in S} B_i\right|$$

for every subset $S \subset [n]$ *with* $|S| \leq k$, *then*

$$\frac{\left|\bigcup_{i=1}^{n} A_i\right|}{\left|\bigcup_{i=1}^{n} B_i\right|} \leq \left(\frac{\lambda^k + 1}{\lambda^k - 1}\right)^2$$

where $\lambda = (\sqrt{n}+1)/(\sqrt{n}-1)$. *In particular*

1. *For* $k \geq \Omega(\sqrt{n})$

$$\frac{\left|\bigcup_{i=1}^{n} A_i\right|}{\left|\bigcup_{i=1}^{n} B_i\right|} \leq 1 + O(e^{-\frac{2k}{\sqrt{n}}})$$

2. *and for* $k \leq O(\sqrt{n})$

$$\frac{\left|\bigcup_{i=1}^{n} A_i\right|}{\left|\bigcup_{i=1}^{n} B_i\right|} \leq O\left(\frac{n}{k^2}\right)$$

Moreover, the inequality in this range is optimal as there exist collections of sets satisfying the requirements of the theorem and yet:

$$\frac{\left|\bigcup_{i=1}^{n} A_i\right|}{\left|\bigcup_{i=1}^{n} B_i\right|} = \Omega\left(\frac{n}{k^2}\right)$$

Proof. As pointed out already, there is no loss of generality in assuming the sets to be events in a probability space with cardinalities replaced by probabilities. The ratio between the probabilities of the two unions is seen now to be at most $1/(1-E(k,n)) = (1+D(k,n))/(1-D(k,n))$. But $D(k,n) \leq |T_k(-(n+1)/(n-1))|^{-1} = 2/(\lambda^k + \lambda^{-k})$, so the upper bound on the ratio follows. The optimality in the range $k \leq O(\sqrt{n})$ follows from the lower bounds on $D(k,n)$. The asymptotic results follow now from standard estimates. ∎

While this theorem gives nearly optimal bounds for $k \leq O(\sqrt{n})$, there is evidence that for larger k a better bound can be proved. A special case which we, as well as [1], managed to solve is $k = n-1$. In that case another family of classical orthogonal polynomials, viz., Krawchouk polynomials replace the Chebyshev polynomials yielding an approximation to within a factor of $1 + 2^{-\Omega(n)}$ rather than the $1 + 2^{-\Omega(\sqrt{n})}$ implied by this theorem. A derivation of this bound appears in next subsection. We do not know at present what the best approximation is in the range $n-1 > k > \omega(\sqrt{n})$.

Theorem 1 only estimates the quality of approximations obtainable from the sizes of all intersections of up to k sets. The next theorem indicates how to effectively compute an approximation which attains this bound. As the reader probably expects the result is obtained from an appropriate linear combination.

Theorem 2. *For any integers* k, n *there exist (explicitly given) constants* $(\alpha_1^{k,n}, \alpha_2^{k,n}, \ldots, \alpha_k^{k,n})$ *such that for every collection of sets* $A_1, A_2, \ldots, A_n$, *the quantity*

$$\sum_{|S| \leq k} \alpha_{|S|}^{k,n} \left|\bigcap_{i \in S} A_i\right|$$

differs from $\left|\bigcup_{i=1}^{n} A_i\right|$ *by at most a factor of:*

1. $1 + O\left(e^{-2k/\sqrt{n}}\right)$ *if* $k \geq \Omega(\sqrt{n})$.
2. $O(n/(k^2))$ *if* $k \leq O(\sqrt{n})$.

Proof. The statement of the lemma follows on observing that the linear program in lemma 3 may be slightly modified to yield an approximation for the size of the union. The modification is to replace every equation of type (1) $E_j(\vec{x}) = 0$ with $E_j(\vec{x})$ ="size of intersection", and replace all inequalities of type (2) by the inequalities $x_j \geq 0$ for all $1 \leq i \leq n$. Now the same linear combination that achieves the bound $E(k,n)$ for the program in lemma 3, can be used to obtain a good approximation for the size of the union.

Thus the real numbers $\alpha_1^{k,n}, \alpha_2^{k,n}, \ldots, \alpha_k^{k,n}$ are defined to be the coefficients of the linearly transformed Chebyshev polynomials expressed in terms of the polynomials $\binom{x}{1}, \binom{x}{2}, \ldots, \binom{x}{k}$, that is

$$q_{k,n} = \sum_{i=1}^{k} \alpha_i^{k,n} \binom{x}{i}.$$

The vector $\vec{\alpha} = (\alpha_1^{k,n}, \alpha_2^{k,n}, \ldots, \alpha_k^{k,n})$ can be calculated by as follows. Consider the above polynomial identity for $x = 1, \ldots, k$. There results a system of linear equations in the α's:

$$\vec{\alpha} M = \vec{t}$$

where M is the matrix whose (i,j) entry is $\binom{j}{i}$ and the $j-$th element in $\vec{t}$ is $q_{k,n}(j)$. Now it is easily verified that the (i,j) entry of the inverse M^{-1} is $(-1)^{i+j}\binom{j}{i}$, so one can calculate $\vec{\alpha} = \vec{t} M^{-1}$. ∎

2.2. The case $k = n - 1$

As mentioned previously, we can improve on Theorem 1 for the case $k = n - 1$.

Theorem 3.

$$D(n-1,n) = \frac{1}{2^n - 1}$$

Proof. By Lemmas 5 and 6, $D(n-1,n)$ is the optimum of the following linear program:

$$\min u$$

under the constraint that for all integers $1 \leq t \leq n$ there holds

$$1 - u \leq \sum_{j=1}^{n-1} a_j t^j \leq 1 + u.$$

From LP duality it follows that the optimum for u is obtained by linear combinations of these inequalities. In such a combination all terms involving the a_j have to cancel out. So if $\vec{z}$ is the vector yielding the optimal bound, then necessarily $\vec{z}M = 0$ where M is the matrix of the LP whose (i,j) entry is i^j for $1 \le i \le n$ and $1 \le j \le n-1$. Now any $n-1$ rows of M form a van der Monde matrix so the rank of M is $n-1$ and consequently the space of such $\vec{z}$ is one-dimensional. It is easily verified that this space is spanned by the vector whose $i-$th entry is $(-1)^i\binom{n}{i}$. In other words, the combination in question is that where for t odd (even) the left (right) side of the $t-$th inequality is multiplied by $\binom{n}{t}$. The resulting inequality is easily seen to be

$$u \ge \frac{1}{2^n - 1}.$$

Since the null space of M is one-dimensional this is the optimal bound for u. ∎

When translated back to the language of polynomials the optimal solution is seen to be closely related to Krawtchuk polynomials:

$$P_l(x;m) = \sum_0^l (-1)^j \binom{x}{j}\binom{m-x}{l-j}.$$

The exact relationship is

$$q(x) = 1 - (1+\varepsilon)\binom{n-x-1}{n-1} + \varepsilon P_{n-1}(x; n-1)$$

where $\varepsilon = 1/(2^n - 1)$. To check this, note that $P_{n-1}(r; n-1) = (-1)^r$ for $r = 0, 1, \ldots, n-1$ which readily implies that $q(0) = 0$ and $q(r) = 1 + (-1)^r\varepsilon$ for $r = 1, \ldots, n-1$. The only fact to verify is that the same holds also at $r = n$, but $P_{n-1}(n, n-1) = (-1)^{n-1}(2^n - 1) = (-1)^{n-1}/\varepsilon$ and the validity for $r = n$ follows, too.

The fact that this problem is optimized at two ranges by (simple variants of) Chebyshev and Krawchuk polynomials is very suggestive, of course. Is it not the case that in other ranges the optimum is obtained through other classical families of orthogonal polynomials, perhaps other instances of Racach or Hahn polynomials? Unfortunately, we do not know the answer to this question for other values of k at the time of writing.

3. Using Approximate Information

This section estimates the size of the union when intersections sizes are given only to within an error.

Let $A_1, A_2, \ldots, A_n$ be events in a probability space. Say that for any subfamily, the probability of the intersection is given to within an additive error of ε. How good an estimate for the probability of the union can be derived? Again, the problem is restated in terms of two collections of events:

Definition 3. $B(n,\varepsilon)$ is defined to be

$$\sup\left|Pr\Big[\bigcup_{i=1}^{n} A_i\Big] - Pr\Big[\bigcup_{i=1}^{n} B_i\Big]\right|$$

where $A_1, A_2, \ldots, A_n$ and $B_1, B_2, \ldots, B_n$ are any two collections of events satisfying:

$$\left|Pr\Big[\bigcap_{i\in S} A_i\Big] - Pr\Big[\bigcap_{i\in S} B_i\Big]\right| \leq \varepsilon$$

for every subset $S \subset [n]$.

The results of the last section imply good bounds on $B(n,\varepsilon)$:

Theorem 4. *For $\varepsilon = \varepsilon(n) > 0$ there holds:*

1. *If $\log(1/\varepsilon) = \Omega(\sqrt{n}\log n)$ then $B(n,\varepsilon) \leq \varepsilon^{\Omega(1/(\sqrt{n}\log n))} = o(1)$.*
2. *If $\log(1/\varepsilon) = o(\sqrt{n})$ then $B(n,\varepsilon) = 1 - o(1)$.*

Proof. To prove part (1), apply Theorem 2 for good estimates on $Pr\Big[\bigcup A_i\Big]$ and $Pr\Big[\bigcup B_i\Big]$, using $k = O(\log\frac{1}{\varepsilon}/\log n)$. Theorem 2 then yields estimates of a error no bigger than $e^{-\Omega(k/\sqrt{n})} = \varepsilon^{\Omega(1/(\sqrt{n}\log n))}$. These approximations must be very close to each other, for their difference is given by:

$$\sum_{S\subset[n],\,|S|\leq k} \alpha_{|S|}^{k,n}\Big(Pr\Big[\bigcap_{i\in S} A_i\Big] - Pr\Big[\bigcap_{i\in S} B_i\Big]\Big).$$

There are clearly at most n^k terms in this expression, none exceeding $\alpha_i^{k,n}\varepsilon$. The considerations of Theorem 2 imply that that $\left|\alpha_i^{k,n}\right|$ is bounded by, say, n^k. Thus, the whole difference does not exceed $n^{2k}\varepsilon = \varepsilon^{\Omega(1/(\sqrt{n}\log n))}$ and part (1) of the theorem follows.

Part (2) follows from the fact that $B(n,2^{-t}) \geq E(n/2,t)$ together with the part of Theorem 1 stating that when $t = o(\sqrt{n})$, $E(n/2,t) = 1 - o(1)$. Let there be given two collections $A_1', \ldots, A_{n/2}'$ and $B_1', \ldots, B_{n/2}'$ where

$$Pr\Big[\bigcap_{i\in S} A_i'\Big] = Pr\Big[\bigcap_{i\in S} B_i'\Big]$$

if $|S| \leq t$. We show how to construct families $A_1, A_2, \ldots, A_n$ and $B_1, B_2, \ldots, B_n$, so that

$$Pr\Big[\bigcup_{i=1}^{n} A_i\Big] = Pr\Big[\bigcup_{i=1}^{n/2} A_i'\Big]\ ,\ Pr\Big[\bigcup_{i=1}^{n} B_i\Big] = Pr\Big[\bigcup_{i=1}^{n/2} B_i'\Big]$$

for every $S \subset [n]$, and

$$\left|Pr\Big[\bigcap_{i\in S} A_i\Big] - Pr\Big[\bigcap_{i\in S} B_i\Big]\right| \leq 2^{-t}.$$

To construct $A_1, A_2, \ldots, A_n$ and $B_1, B_2, \ldots, B_n$, first augment each collection with $n/2$ empty sets, then apply lemma 1 to "symmetrize" them thus obtaining symmetric collections $A_1, A_2, \ldots, A_n$ and $B_1, B_2, \ldots, B_n$. In other words, if S has size m:

$$Pr\Big[\bigcap_{i\in S} A_i\Big] = \frac{\sum_{S'\subset[n/2],\,|S'|=m} Pr\Big[\bigcap_{i\in S'} A'_i\Big]}{\binom{n}{m}}$$

and similarly for the B_i's.

That

$$Pr\Big[\bigcup_{i=1}^{n} A_i\Big] = Pr\Big[\bigcup_{i=1}^{n/2} A'_i\Big]$$

holds is obvious. As for

$$\left|Pr\Big[\bigcap_{i\in S} A_i\Big] - Pr\Big[\bigcap_{i\in S} B_i\Big]\right| \le 2^{-t}$$

for every $S \subset [n]$, note that if $|S| \le t$ the difference is zero. For $|S| = m > t$,

$$Pr\Big[\bigcap_{i\in S} A_i\Big] \le \binom{n/2}{m} \Big/ \binom{n}{m} \le 2^{-t}$$

and the claim follows.

This completes the proof of the lemma, because for any solution of the $E(n/2, t)$ problem, a solution for the $B(n, 2^{-t})$ problem is presented, with the same value for the target function. ∎

Let us note that we have actually showed that an estimate of size of the union may be effectively computed by the same formula given by Theorem 2. Also note that this formula does not require the sizes of intersections of a large number of sets.

4. Disjunctions of Boolean Functions

This section addresses the approximation of Boolean functions by other Boolean functions.

Definition 4. Let f and g be Boolean functions. The *advantage* of f on g is

$$Adv(f, g) = |Pr[f(x) = g(x)] - Pr[f(x) \neq g(x)]|$$

where the probability distribution is uniform over all binary $n-$vectors x.

The advantage of f on g is a measure of how well f (or its negation) approximate g. The following question comes up naturally in the study of circuit complexity of Boolean functions: Given are Boolean functions $f_1, \ldots, f_n$, and a target function g. Suppose that the conjunction of any subset of $f_1, \ldots, f_n$ has advantage of less than ε on g. How large can the advantage of the disjunction of these functions be on g?

Lemma 8. *Let $g, f_1, f_2, \ldots, f_n$ be Boolean functions such that*

$$\text{For all } S \subset [n] : \; Adv\Big(\bigwedge_{i \in S} f_i, g\Big) \leq \varepsilon$$

then $Adv\Big(\bigvee_{i=1}^{n} f_i, g\Big) \leq B(n, \varepsilon)$. This bound is optimal in that there exist Boolean functions satisfying the conditions of the lemma such that $Adv\Big(\bigvee_i f_i, g\Big)$ is arbitrarily close to $B(n, \varepsilon)$.

Proof. Define

$$A_i = \{x : g(x) = f_i(x)\}$$
$$B_i = \{x : g(x) \neq f_i(x)\}$$

Our assumptions on $g, f_1, \ldots, f_n$ translate to:

$$\left| Pr\Big[\bigcap_{i \in S} A_i\Big] - Pr\Big[\bigcap_{i \in S} B_i\Big] \right| \leq \varepsilon$$

for every $S \subset [n]$, and the advantage is given by:

$$Adv\Big(\bigvee_i f_i, g\Big) = \left| Pr\Big[\bigcup_{i=1}^{n} A_i\Big] - Pr\Big[\bigcup_{i=1}^{n} B_i\Big] \right|.$$

It is thus clear that the advantage is bounded by $B(n, \varepsilon)$.

Conversely, let $A_1, A_2, \ldots, A_n$ and $B_1, B_2, \ldots, B_n$ be collections of events that achieve the value of $B(n, \varepsilon)$. They can be viewed, w.l.o.g. as subsets of $\{0,1\}^l$ for some l, because for a sufficiently large l the original distributions can be approximated arbitrarily well. Consider the following Boolean functions on $\{0,1\} \times \{0,1\}^l$.

$$g(x_0, x_1, \ldots, x_l) = x_0$$

$$f_i(x_0, x_1, \ldots, x_l) = \begin{cases} 1 & \text{if } x_0 = 0 \text{ and } \langle x_1, \ldots, x_l \rangle \in A_i \\ 1 & \text{if } x_0 = 1 \text{ and } \langle x_1, \ldots, x_l \rangle \in B_i \\ -1 & \text{otherwise} \end{cases}$$

Direct translation of the properties of $A_1, A_2, \ldots, A_n$ and $B_1, B_2, \ldots, B_n$ implies that any conjunction of f_i has advantage of at most ε on g, yet their disjunction has an advantage of $B(n, \varepsilon)$ on g. ∎

Combining with Theorem 4 we obtain:

Theorem 5. *Let $f_1, f_2, \ldots, f_n$ and g be Boolean functions that satisfy*

$$\text{For all } S \subset [n] : \; Adv\Big(\bigwedge_{i \in S} f_i, g\Big) \leq 2^{-t}$$

where $t \geq \Omega(\sqrt{n} \log n)$. Then

$$Adv\Big(\bigvee_{i=1}^{n} f_i, g\Big) \leq 2^{-\Omega(t/(\sqrt{n} \log n))}$$

Moreover, this is optimal in the sense that for any $t = t(n) = o(\sqrt{n})$ there exist Boolean functions $f_1, f_2, \ldots, f_n$ and g that satisfy

$$\text{For all } S \subset [n] \; : \; Adv\Big(\bigwedge_{i \in S} f_i, g\Big) \leq 2^{-t}$$

and yet

$$Adv\Big(\bigvee_{i=1}^{n} f_i, g\Big) \geq 1 - o(1) \; .$$

5. t-wise Independence and Constant Depth Circuits

Some of the original questions motivating this research were ones regarding "what looks random to constant depth circuits?" Although the present results do not provide a strong answer to this problem, they do enable us to make some nontrivial remarks. Here is a statement of the problem and our results.

Consider balanced probability distributions on $x_1, \ldots, x_n$, i.e. distributions that satisfy for every i, $Pr[x_i = 0] = Pr[x_i = 1] = 1/2$. Such a distribution is called t-wise independent if the induced distribution on any t variables $x_{i_1}, \ldots, x_{i_t}$ is uniform.

Definition 5. Let $f(x_1, \ldots, x_m)$ be a Boolean function. Say that f is fooled by t-wise independence if:

$$|Pr[f(x_1, \ldots, x_m) = 1] - Pr[f(y_1, \ldots, y_m) = 1]| \; \leq \; 0.1$$

whenever $x_1, \ldots, x_m$ are chosen independently uniformly at random and $y_1, \ldots, y_m$ are chosen at random according to any balanced distribution which is t-wise independent.

Conjecture 1. *Any function which is computed by a Boolean circuit of depth d and size s is fooled by $(\log s)^{d-1}$-wise independence.*

(The circuits in question have unbounded fanin "and" and "or" gates, all negations are applied to variables). This conjecture essentially states that a bit generator producing distributions which are polylog-wise independent is a pseudorandom generator for AC^0 tests (polynomial size, constant depth Boolean circuits). This would generalize the two known pseudorandom generators for AC^0 [2] [6], both of which indeed are at least polylog-wise independent.

We are only able to show:

Theorem 6. *Any function which is computed by a Boolean circuit of depth 2 and size s is fooled by $\Omega(\sqrt{s} \log s)$-wise independence.*

Proof. Assume w.l.o.g. the circuit to be a DNF-formula with terms $f_1, \ldots f_s$. Consider a distribution μ on $x_1, \ldots, x_m$ which is t-wise independent. Define A_i to be the event that $f_i(x_1, \ldots, x_m)$ is true where x is chosen uniformly at random, and define B_i as the event that $f_i(y_1, \ldots, y_m)$ is true where y is chosen according to distribution μ. We now claim that for any subset $S \subset \{1 \ldots m\}$:

$$\left| Pr\Big[\bigcap_{i \in S} A_i\Big] - Pr\Big[\bigcap_{i \in S} B_i\Big] \right| \; \leq \; 2^{-t}$$

There are two cases: if the conjunction of the f_i's has less that t literals then the two probabilities are exactly equal, as μ is t-wise independent. Otherwise, both probabilities are bounded from above by 2^{-t}. The statement of the theorem now follow directly from Theorem 4. ∎

6. Open Problems

1. Is there an efficient *adaptive* algorithm to approximate the size of a union of events given an oracle for sizes of intersections?
2. The inclusion-exclusion formula is the Möbius inversion formula for the full Boolean lattice. Are there results similar to the present ones for other lattices?
3. Theorem 1 shows the existence of two collections of sets where small subcollections have intersections of the same size, whose unions differ much by size. The proof does not supply an example. What do such families look like? Is it possible to achieve stronger results for some interesting special cases of collections of sets?
4. Theorem 1 is nearly optimal for $k \leq \sqrt{n}$. Can the bounds for larger k be improved? As mentioned, we know the best result for $k = n - 1$ and it is better than what is supplied by the theorem.
5. A proof or counterexample to Conjecture 1 would be interesting.

Acknowledgments. We are grateful to John Tomlin who ran for us his computer programs for linear programming to solve some small cases of the optimization problem in lemma 3 which we could not solve by hand. These solutions helped us gain insight for the general case. The second author would like to thank Dick Karp for many helpful discussions.

References

[1] N. ALON, and J. HASTAD: Private communication. 1988.

[2] M. AJTAI, and A. WIGDERSON: Deterministic simulation of probabilistic constant depth circuits, In 26^{th} *Annual Symposium on Foundations of Computer Science, Portland, Oregon,* (1985), 11–19.

[3] C. E. BONFERRONI: Teoria statistica delle classi e calcolo dell probabilita, In *volume in onore di Ricardo Dalla Volta,* Universita di Firenze, (1937), 1–62.

[4] E. W. CHENEY.: *Approximation Theory,* McGraw-Hill Book Co., **1966**.

[5] S. M. KWEREL: Most stringent bounds on aggregated probabilities of partially specified dependent probability systems, *J. Am. Stat. Assoc.* **70**, (1975), 472–479.

[6] N. NISAN, and A. WIGDERSON: Hardness vs. randomness. In 29^{th} *Annual Symposium on Foundations of Computer Science, White Plains, New York,* October (1988), 2–12.

[7] A. PRÉKOPA: Boole-Bonferroni inequalities and linear programming, *Operation Research* **36**, (1988), 145–162.

[8] THEODORE J. RIVLIN: *An Introduction to the Approximation of Functions,* Blaisdell Publishing Company, **1969**.

[9] H. J. RYSER: *Combinatorial mathematics*, The Mathematical Association of America, **1963**.

[10] L.G. VALIANT: The complexity of computing the permanent, *Theor. Comp. Sci.*, **8**, (1979), 189–201.

Nathan Linial

IBM-Almaden, Stanford University and Hebrew University.

Noam Nisan

Laboratory for Computer science, MIT, 545 Tech. sq., Cambridge, MA 02139. U.S.A.