Inclusion-exclusion: exact and approximate

Jeff Kahn * Nathan Linial † Alex Samorodnitsky ‡

Abstract

It is often required to find the probability of the union of given n events $A_1, \ldots, A_n$. The answer is provided, of course, by the inclusion-exclusion formula: $\Pr(\cup A_i) = \sum_i \Pr(A_i) - \sum_{i<j} \Pr(A_i \cap A_j) \pm \ldots$. Unfortunately, this formula has exponentially many terms, and only rarely does one manage to carry out the exact calculation. From a computational point of view, finding the probability of the union is an intractable, #P-hard problem, even in very restricted cases. This state of affairs makes it reasonable to seek approximate solutions that are computationally feasible. Attempts to find such approximate solutions have a long history starting already with Boole [1]. A recent step in this direction was taken by Linial and Nisan [4] who sought approximations to the probability of the union, given the probabilities of all j-wise intersections of the events for $j = 1, \ldots k$. They developed a method to approximate $\Pr(\cup A_i)$, from the above data with an additive error of $\exp(-O(k/\sqrt{n}))$. In the present article we develop an expression that can be computed in polynomial time, that, given the sums $\sum_{|S|=j} \Pr(\cap_{i\in S} A_i)$ for $j = 1, \ldots k$, approximates $\Pr(\cup A_i)$ with an additive error of $\exp(-\tilde{\Omega}(k^2/n))$. This error is optimal, up to the logarithmic factor implicit in the $\tilde{\Omega}$ notation.

The problem of enumerating satisfying assignments of a boolean formula in DNF form $F = \vee_1^m C_i$ is an instance of the general problem that had been extensively studied [7]. Here A_i is the set of assignments that satisfy C_i, and $\Pr(\cap_{i\in S} A_i) = a_S/2^n$ where $\wedge_{i\in S} C_i$ is satisfied by a_S assignments. Judging from the general results, it is hard to expect a decent approximation of F's number of satisfying assignments, without knowledge of the numbers a_S for, say, all cardinalities $1 \le |S| \le \sqrt{m}$. Quite surprisingly, already the numbers a_S over $|S| \le \log(n+1)$ *uniquely determine the number of satisfying assignments for* F.

We point out a connection between our work and the edge-reconstruction conjecture. Finally we discuss other special instances of the problem, e.g., computing permanents of $0,1$ matrices, evaluating chromatic polynomials of graphs and for families of events whose VC dimension is bounded.

*Rutgers University. Work supported in part by a grant of the Binational Israel-US Science Foundation.

†Hebrew University. Work supported in part by a grant of the Binational Israel-US Science Foundation and by the Israel Science Foundation.

‡Hebrew University

1 Introduction

Let $A_1, \ldots, A_n$ be events in a probability space. The inclusion-exclusion formula expresses the probability of their union:

$$\Pr(\cup A_i) = \sum_i \Pr(A_i) - \sum_{i<j} \Pr(A_i \cap A_j) + \sum_{i<j<k} \Pr(A_i \cap A_j \cap A_k) \mp \ldots =$$

$$= \sum_{[n]\supseteq S \neq \emptyset} (-1)^{|S|-1} \Pr(\cap_{i\in S} A_i).$$

If any of the $2^n - 1$ terms is unknown, it is not possible, in general, to exactly evaluate $\Pr(\cup A_i)$. Many investigators (e.g. [2] and the references therein) had asked how well $\Pr(\cup A_i)$ can be *approximated* given only partial information on the numbers $\Pr(\cap_{i\in S} A_i)$.

Linial and Nisan [4] showed that if $\Pr(\cap_{i\in S} A_i)$ is known whenever $|S| \leq k$, then $\Pr(\cup A_i)$ may be be approximated as follows: If $k = O(\sqrt{n})$, then it is possible to estimate the probability of the union to within an additive error of $O(1 - k^2/n)$. Moreover, for $k = O(\sqrt{n})$, this is also essentially optimal. However, for larger k, [4] fails to provide a full answer. A method of approximation which is developed in that paper offers an approximation to within $e^{-\Omega(k/\sqrt{n})}$ of the truth. This degree of approximation has been shown (ibid.) to be suboptimal. This problem is resolved in the present article: Given the numbers $\sum_{|S|=j} \Pr(\cap_{i\in S} A_i)$ for all $j \leq k$, we can approximate the probability of the union to within an additive error of $e^{-\tilde{\Omega}(k^2/n)}$. Moreover, the approximation can be computed in polynomial time. The result is tight in the sense that given the numbers $\Pr(\cap_{i\in S} A_i)$ for all $|S| \leq k$, it is in general impossible to approximate $\Pr(\cup A_i)$ to within an additive error smaller than $e^{-\tilde{O}(k^2/n)}$ (Regardless of the computational complexity involved in the approximation).

The problem of enumerating the satisfying assignments of a DNF formula is an instance of the general problem. Already this special case is known to be complete for the class #P . Much attention has been given to efficient algorithms for *approximating* this number, see [7] and the references therein. To put the DNF problem in the general context of our problem, let the probability space be $\{0,1\}^n$ under uniform distribution. Associated with every clause in the DNF formula is the event that this clause is satisfied. Each such event is, in fact, a subcube of the n-dimensional cube. For this problem something quite remarkable happens: While any decent approximation for the general problem requires information on $\Omega(\sqrt{m})$-wise intersections, here the answer is *uniquely determined* from the probabilities of $\leq (\log n + 1)$-wise intersections.

Our methods offer also some new insight into the edge–reconstruction problem from graph theory. In particular Müller's [8] Theorem can be reproved and put in a more general context that may possibly lead to further progress on this conjecture.

We then point out that calculating $0, 1$ permanents may also be viewed as an instance of the general problem, and similarly the problem of computing chromatic polynomials of graphs. Some comments are made on the possibility of getting estimates for these cases that are better than those achievable in the general case. Finally we derive a theorem similar to the one for DNF formulas, in case the VC dimension of our family of events is bounded.

2 Near-tight approximations for general inclusion-exclusion

The main result of this section is:

Theorem 2.1: *Let $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ be collections of events in some probability space where:*

$$\Pr\left(\bigcap_{i\in S} A_i\right) = \Pr\left(\bigcap_{i\in S} B_i\right)$$

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

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

Moreover, there are coefficients $\lambda_j = \lambda_{j,k,n}$ such that

$$\left|\sum_{j=1}^{k} \lambda_j \sum_{|S|=j} \Pr\left(\bigcap_{i\in S} A_i\right) - \Pr\left(\bigcup_{i=1}^{n} A_i\right)\right| \leq e^{-\Omega(\frac{k^2}{n\log n})}$$

and these coefficients λ_j can be found in time polynomial in n.

On the other hand, families $A_1, \ldots, A_n$ and $B_1, \ldots, B_n$ exist for which $\Pr\left(\bigcap_{i\in S} A_i\right) = \Pr\left(\bigcap_{i\in S} B_i\right)$ whenever $|S| \leq k$ and

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

Remark 2.2: We follow the notation in [4] and denote by $E(k, n)$ the maximum difference between $\Pr(\cup A_i)$ and $\Pr(\cup B_i)$, where there are n events A_i and n events B_i such that the intersection of any $\leq k$ of the A_i's has the same probability of the corresponding event with B_i's. The theorem states that $E(k, n) = \exp(-\tilde{\Theta}(\frac{k^2}{n}))$. Our belief is that the logarithmic terms hidden in the "soft Θ" notation are redundant and the truth is $E(k, n) = \exp(-\Theta(\frac{k^2}{n}))$. Moreover, we think that the present methods are powerful enough to establish this statement and no essential new ideas will be required. ∎

Combining two lemmas from [4] (pp. 354, 357) we get the following:

Lemma 2.3: *$E(k,n) \leq \delta$ iff there is a real polynomial q of degree at most k whose constant term is zero, so that $|q(m)-1| \leq \frac{\delta}{2-\delta}$ holds for every integer $m = 1,\ldots,n$. Moreover, if $q(x) = \sum_{j=1}^{k} \lambda_j \binom{x}{j}$, then*

$$\left| \sum_{j=1}^{k} \lambda_j \sum_{|S|=j} \Pr\left(\bigcap_{i \in S} A_i \right) - \Pr\left(\bigcup_{i=1}^{n} A_i \right) \right| \leq \delta.$$

Proof: (Of Theorem 2.1) We follow the approach suggested by Lemma 2.3 and explicitly construct polynomials $q(x)$ that satisfy the lemma with $\delta = \exp(-\Omega(\frac{k^2}{n \log n}))$. The coefficients of this polynomial, expressed as a linear combination of $\binom{x}{j}$ over $j = 1,\ldots,k$ will satisfy the claim made in the Theorem. By a simple change of variable, we need to construct a real polynomial t of degree k which satisfies $t(n) = 1$ and $\Delta \geq \max_{m=0,\ldots,n-1} |t(m)|$, where $\Delta = \exp(-\Omega(\frac{k^2}{n \log n}))$.

To begin, we choose for t to vanish at integer points near the ends of the interval $[0,\ldots,n-1]$. That is, we let

$$s(x) = \prod_{0}^{a-1} (x-i) \cdot \prod_{n-b}^{n-1} (x-j)$$

and $t(x) := s(x)/s(n)$. The integers a and b depend on n and their sum is k. The maximum of $|t(m)|$ over $m = 0,\ldots,n-1$ is the maximum of $|\frac{s(m)}{s(n)}|$ over the same set of m's. A direct calculation with the polynomial s yields:

$$|\frac{s(m)}{s(n)}| = \frac{\binom{m}{a}\binom{n-m-1}{b}}{\binom{n}{a}} \leq \frac{\binom{n}{a+b}}{\binom{n}{a}} \leq 2^{n(H(\frac{a+b}{n}) - H(\frac{a}{n}) + o(1))}. \tag{1}$$

If $k \geq 3n/5$ we select $a = n/2$. It is easily verified that for every choice of m in our range, the right hand side in Equation 1 is exponentially small in n, as needed.

For $k < 3n/5$ a more complicated construction is called for. We still guarantee that t vanishes on integral points near the ends of the interval $[0,\ldots,n-1]$. Around the center of this interval, we control the growth of t using a (linearly transformed) Tchebyshef polynomial. Let

$$s(x) = \prod_{0}^{a-1} (x-i) \cdot \prod_{n-b}^{n-1} (x-j) \cdot \tau_r(x)$$

where τ_r is a linearly transformed Tchebyshef polynomial:

$$\tau_r(x) = T_r(\frac{x-a}{n-b-a}).$$

Here $T_r(x)$ is the r-th Tchebyshef polynomial, and $a+b+r = k$. We also let $\alpha = a/n$ and $\beta = b/n$. For $n-b \geq m \geq a$ the same calculation carried out for $k \geq 3n/5$ can be repeated. Since a Tchebyshef polynomial varies between -1 and 1 when the argument is a real in $[-1,1]$, we conclude that

$$|\frac{s(m)}{s(n)}| \leq 2^{n(H(\alpha+\beta)-H(\alpha))} \ / \ T_r(\frac{n-a}{n-a-b}). \tag{2}$$

The Tchebyshef polynomial can be written as $T_r(x) = \frac{1}{2}\left((x+\sqrt{x^2-1})^r + (x-\sqrt{x^2-1})^r\right)$, which is convenient for estimations. Using this expression, it not hard to show that:

$$T_r(\frac{n-a}{n-a-b}) = T_r(\frac{1-\alpha}{1-\alpha-\beta}) \geq (1+\sqrt{2\beta})^r/2.$$

To get an upper bound on $|\frac{s(m)}{s(n)}|$, we select $\alpha = \Theta(\frac{k}{n\log n})$ and $\beta = \alpha^2$. Together with the lower bound on T_r, we obtain, after some calculations, an upper bound in Equation 2:

$$\forall m = 1,\ldots,n \quad |\frac{s(m)}{s(n)}| \leq \exp(-\Omega(\frac{k^2}{n\log n}))$$

as claimed.

We now turn to a lower bound on $E(k,n)$. By a slight modification of Lemma 2.3, this amounts to showing that there is no polynomial t of degree k with $t(0)=1$ and with $|t(m)| < \epsilon$ for every $m = 1,\ldots,n$ where $\epsilon = \exp(-\Omega(\frac{k^2\log n}{n}))$. Letting $t(x) = 1 + \sum_1^k a_i x^i$ we need to show that the following system of linear inequalities (in the a_i) is inconsistent:

$$\forall m = 1,\ldots,n \qquad -\epsilon < 1 + \sum_{i=1}^{k} a_i m^i < \epsilon.$$

Inconsistency will be established by linearly combining $k+1$ of these inequalities. Our plan is to find $x_1,\ldots,x_{k+1}$ (integers) which are the indices for inequalities participating in this combination, the x_j-th inequality being weighed by w_j ($j = 1,\ldots,k+1$). (In fact, LP duality says we have no choice here, and this is *the* way to derive a contradiction). A contradiction is obtained if following conditions hold:

$$\forall i = 1,\ldots,k \quad \sum_{j=1}^{k+1} w_j x_j^i = 0$$

which means that all nonconstant terms get eliminated, and

$$\sum_1^{k+1} w_j > \epsilon \sum_1^{k+1} |w_j|$$

which means that the combination of constant terms is a contradiction. It is convenient to normalize with $\sum_1^{k+1} w_j = 1$, thus transforming the latter condition to:

$$\frac{1}{\epsilon} > \sum_1^{k+1} |w_j|.$$

Observe that the w_j satisfy a linear system of equations, and can, therefore be expressed in terms of the x_j by Cramer's rule. The matrix of this linear system of equations is a Vandermonde, so the expressions for w_j are convenient:

$$w_j = \pm \frac{\prod_{i\neq j} x_i}{\prod_{i\neq j}(x_i - x_j)}.$$

Our goal is, then, to find integers $x_1, \ldots, x_{k+1}$ for which

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

Our choice for the x_i is as follows: Let $R := \lfloor \frac{k^2}{n} \rfloor$. For $i = 1, \ldots, R$ we let $x_i = i$. For $R < i \leq k+1$ we let $x_i = \lfloor \frac{i^2}{R} \rfloor$.

Proposition 2.4: *With the above choice of the x_i, for every i,*

$$\prod_{j\neq i} \frac{x_j}{|x_i - x_j|} < \exp(O(R \log n)).$$

Remark 2.5: For future reference, let Y be the left expression in this inequality. ∎

Proof: Calculate the numerator first. Whether x_i is missing from this product is inconsequential for the type of estimate we are seeking.

$$\prod x_j = R! \prod_{j=R+1}^{k+1} \lfloor \frac{j^2}{R} \rfloor \leq R! \prod \frac{j^2}{R} = \frac{(k+1)!^2}{R! R^{k-R+1}}.$$

Now let us turn to the denominator. Here we have to distinguish between the cases where i is among the first R indices, or is bigger.

If $1 \leq i \leq R$ and $x_i = i$, then

$$\prod_{1\leq j\neq i\leq R} |i - x_j| \cdot \prod_{j=R+1}^{k+1} (x_j - i) \geq (i-1)!(R-i)! \prod_{j=R+1}^{k+1} (\frac{j^2}{R} - i - 1) \geq$$

$$(i-1)!(R-i)! \prod_{R+2}^{k+1} \frac{(j-R-1)(j+R-1)}{R} = (i-1)!(R-i)! R^{R-k}(k-R)!(k+R)!/(2R)!$$

Dividing out, an upper bound for Y is obtained:

$$Y \leq \frac{(2R)!(k+1)!^2}{R!(i-1)!(R-i)!(k-R)!(k+R)!} \leq n^{O(R)}.$$

Now consider the case where $i > R$ and $x_i = \lfloor \frac{i^2}{R} \rfloor$. In this case,

$$\prod_{j=1}^{R}(x_i - j) \prod_{R+1 \le j \ne i \le k+1} |x_i - x_j| \ge \frac{\lfloor \frac{i^2}{R} \rfloor !}{\lfloor \frac{i^2}{R} - R \rfloor !} \cdot \prod_{R+1 \le j \ne i, i+1 \le k+1} \frac{|j^2 - i^2 - i|}{R} \ge$$

$$\left(\frac{i^2 - R^2}{R} \right)^R \cdot R^{R-k+1} \cdot \prod_{R+1 \le j \ne i, i+1 \le k+1} |(j-i-1)(j+i)| \ge$$

$$\frac{(i^2 - R^2)^R}{R^{k-1}} \cdot \frac{(i-R-2)!(k-i)!(k+i)!}{(R+i)!}$$

Dividing the numerator by the denominator yields:

$$Y \le \frac{(R+i)! R^R}{(i-R-2)!(i^2-R^2)^R} \cdot \frac{(k+1)!^2}{(k-i)!(k+i)!} \le k^4 \frac{(R+i)^{2R} R^R}{(i^2-R^2)^R} = k^4 R^R \left(\frac{i+R}{i-R} \right)^R \le n^{O(R)},$$

as claimed.

3 On enumerating satisfying assignments for DNF

In this section we show:

Theorem 3.1: *Let F be a DNF formula in n boolean variables with clauses $C_1, \ldots, C_m$. For $S \subseteq [m]$, let a_S be the number of satisfying assignments for $\wedge_{i \in S} C_i$. The numbers a_S where S ranges over all subsets of no more than $\log n + 1$ members of $[m]$, already uniquely determine the number of satisfying assignments for F.*

Remark 3.2: Observe that the number a_S is always either zero or 2^{n-k}, where k is the number of distinct variables which appear in all C_i over $i \in S$. In particular, these numbers are very easy to evaluate. ∎

Proof: We start with the following lemma

Lemma 3.3: *Let A_i and B_i be two families of events ($i = 1, \ldots, n$). For $S \subseteq [n]$, let $a_S := \Pr(\bigcap_{i \in S} A_i)$. Also let*

$$\alpha_S := \Pr(\bigcap_{i \in S} A_i \cap \bigcap_{j \notin S} A_j^c).$$

Define b_S and β_S analogously. Suppose that for every subset $S \neq [n]$, $a_S = b_S$. Then there is a real ϵ of absolute value at most $(\frac{1}{2})^{n-1}$, such that for every $S \subseteq [n]$ there holds $\alpha_S = \beta_S + (-1)^{|S|}\epsilon$.

Proof: Although this statement implicitly appears in [4] (the case $k = n-1$), we cannot resist presenting the following short and simple proof. Rather than think of the two families of events A_i and B_i, we represent the situation through α and β which are viewed as real functions on all subsets of $[n]$, or, what is the same, on the n-dimensional cube. Also, let $\gamma_S := \alpha_S - \beta_S$, and $c_S := a_S - b_S$. By inclusion-exclusion:

$$\gamma_S = \sum_{T \supseteq S} (-1)^{|T \setminus S|} c_T \quad \text{and} \quad c_S = \sum_{T \supseteq S} \gamma_T,$$

for every S. In the present case, $c_S = 0$ for every $S \neq [n]$ and the conclusion follows. To see that $|\epsilon| \leq (\frac{1}{2})^{n-1}$, observe that since γ is the difference between two probability distributions, obviously $\sum_S \gamma_S^+ \leq 1$, but $\sum_S \gamma_S^+ = 2^{n-1}|\epsilon|$ which concludes the proof. ∎

We can turn now to a proof of Theorem 3.1: Let $F = \vee_1^m C_i$ and $F' = \vee_1^m C_i'$ be two DNF formulae on variables $x_1, \ldots, x_n$. The integers a_S and a_S' are defined as before, and we assume that if $|S| \leq \log n + 1$, then $a_S = a_S'$. If this last equality holds for *all* S, then obviously F and F' have an equal number of satisfying assignments. Observe that if $a_S = 0$, then there are two clauses C_i and C_j with $i, j \in S$ whose conjunction is a contradiction i.e., $a_{\{i,j\}} = 0$. If $T \subseteq [n]$ satisfies $a_T = 0$, then by this observation there is a two-element subset P of T for which $a_P = 0$, hence also $a_P' = 0$ and so $a_T' = 0$.

We want to consider a minimal set $S \subseteq [n]$ for which $a_S \neq a_S'$. By assumption $|S| \geq \log n + 2$. Also, the previous remarks allow us to assume that $a_S, a_S' \neq 0$. Therefore, we are allowed to assume that there are no negated literals in the clauses C_i and C_i' over $i \in S$. Having made this assumption, let Q_i and Q_i' be the set of variables in C_i (resp. C_i'). It follows that for any $T \subseteq S$,

$$a_T = 2^{n - |\cup_{i \in T} Q_i|}$$

and similarly in the primed case. It follows now that we are in the following situation with respect to the families Q_i and Q_i': For every $T \subset S$, other than $T = S$, $|\bigcup_T Q_i| = |\bigcup_T Q_i'|$, while for $T = S$ the two sides disagree. From the inclusion-exclusion formula we can conclude that also $|\bigcap_T Q_i| = |\bigcap_T Q_i'|$ for every T which is a proper subset of S, but not for $T = S$. We now invoke Lemma 3.3, and conclude that

$$-(\frac{1}{2})^{|S|-1} \leq \frac{(|\bigcap_S Q_i| - |\bigcap_S Q_i'|)}{n} \leq (\frac{1}{2})^{|S|-1}.$$

This is obtained by placing a uniform distribution on $[n]$ and viewing Q_i and Q_i' as events. Since we are assuming that the middle term does not vanish, its absolute value is at least $\frac{1}{n}$ and we conclude that $|S| \leq \log n + 1$, a contradiction which completes the proof. ∎

Remark 3.4: While this result is satisfying in terms of the intersection sizes that are being considered, at this writing this statement is only existential. We do not know any effective way of actually reading the number of satisfying assignments from the integers a_S as above. ∎

For an application of this result in Learning Theory, see [3].

4 Inclusion-exclusion and the edge-reconstruction problem

The *deck* associated with a graph $G = (V, E)$ is the list of unlabeled graphs $\{G \backslash e | e \in E\}$. The well-known edge-reconstruction conjecture states that every graph with four edges or more is uniquely determined by its deck. The most successful approach to this problem, initiated by Lovász [5] and improved by Müller [8] proceeds as follows: Let G and H be two graphs with the same deck. There is no loss in assuming that $V(G) = V(H) = [n]$. Let $X = X_n$ be the probability space of all permutations on $[n]$ under uniform distribution. For $S \subseteq E(G)$, let A_S be the event

$$A_S = \{\pi \in X | E(\pi(G)) \setminus E(G) \supseteq S\}.$$

Likewise,

$$B_S = \{\pi \in X | E(\pi(G)) \setminus E(H) \supseteq S\}.$$

(Here $E(\pi(G))$ is the edge-set of the vertex-permuted graph.) Since G, H have the same deck, it can be shown that $\Pr(A_S) = \Pr(B_S)$ for every proper subset $S \subset E(G)$. Consider now two families of events $\{A_e | e \in E(G)\}$ and $\{B_e | e \in E(G)\}$. If $\Pr(A_S) = \Pr(B_S)$ also for $S = E(G)$, then by inclusion-exclusion, corresponding atoms in these two families are equiprobable. In particular, since the identity permutation maps G to itself, there is also some $\pi \in X$ for which $E(\pi(G)) = E(H)$, namely G and H are isomorphic.

As the reader can see, we are exactly in the situation covered by Lemma 3.3, and we may conclude:

Proposition 4.1: *Let G, H be a pair of graphs with n vertices and m edges for which the edge-reconstruction conjecture fails. Let α_S (resp. β_S) be the probability (in $X = X_n$) that $E(\pi(G)) \setminus E(G) = S$ (resp. $E(\pi(G)) \setminus E(H) = S$). Then there is a real ϵ of absolute value no bigger than $(\frac{1}{2})^{m-1}$ such that $\alpha_S - \beta_S = (-1)^{|S|}\epsilon$ for every $S \subseteq E(G)$.*

Thus, a counterexample to the conjecture must satisfy a large number of additional constraints. Müller's Theorem, that $m \leq \log_2(n!) + 1$ follows immediately: For nonisomorphic G and H, necessarily $\epsilon \neq 0$, so $(\frac{1}{2})^{m-1} \geq |\epsilon| \geq \frac{1}{n!}$.

5 Other instances of inclusion-exclusion and their approximations

In this section we consider two families of enumeration problems which may be approached via inclusion-exclusion.

Remark 5.1: Ryser's formula for computing the permanent of an $n \times n$ matrix A (see [9]) is based on a slight extension of the inclusion-exclusion principle:

$$Per(A) = \sum_{S \subseteq [n]} (-1)^{n-|S|} \prod_{i=1}^{n} \left(\sum_{j \in S} a_{ij} \right).$$

We concentrate on matrices of zeros and ones, where Ryser's formula is a special case of the usual inclusion-exclusion principle. Under some mild assumptions the permanent of A is uniquely determined by the terms which correspond to sets S of cardinality $\geq n - \log n$ in the above formula. We do not go into the exact statement of these mild assumptions and only mention that (i) This statement is not true unconditionally - take A to be the identity matrix, B the zero matrix. (ii) An example where our (unspecified) assumption do apply is that where A has an all-ones column. ∎

Remark 5.2: Other instances of the inclusion-exclusion principle may yield results which exceed the general case. We point out that *chromatic polynomials* of graphs may be viewed as a class of such examples. (For the basic theory of chromatic polynomials, see Lovász [6] Chapter 9). Briefly, let $G = (V, E)$ be a graph and let λ be a positive integer. Denote by $P_G(\lambda)$ the number of proper coloring of G with λ colors. $P_G(\cdot)$ is called the chromatic polynomial of G. As we presently show, it is indeed a polynomial of degree n, the number of vertices in G. To this end, consider the collection Ψ of all λ^n mappings $f : V \to \{1, \ldots, \lambda\}$. For every edge $u = [x, y] \in E$, let A_u be the family of mappings $f \in \Psi$ that fail to be a proper coloring on the edge u, i.e. those mapping satisfying $f(x) = f(y)$. It is easily seen that for every edge u the cardinality of A_u is λ^{n-1}. Moreover, if $U \subseteq E$ is a set of edges, then $|\cap_{u \in U} A_u|$ is λ^k for some integer $k < n$: Simply shrink the edges $u \in U$, if the resulting graph has k vertices, then $|\cap_{u \in U} A_u| = \lambda^k$. A mapping $f : V \to \{1, \ldots, \lambda\}$ is a proper coloring of G iff it is in none of the sets $A_u (u \in E)$, whence the number of proper λ-colorings of G is $\lambda^n - |\cup_{u \in E} A_u|$. An application of the inclusion-exclusion formula yields an expression for the number of proper colorings as a signed sum of powers of λ, i.e., a polynomial with integer coefficients. Some other known properties of chromatic polynomials can be deduced from this perspective. We raise the problem of approximating the number of proper λ-colorings of graphs using our approach to approximate inclusion-exclusion. ∎

6 Families with a bounded VC dimension

Definition 6.1: We recall the definition of $d_{VC}(\mathcal{F})$, the Vapnik-Chervonenkis (VC) dimension of a family of sets $\mathcal{F} \subset 2^{\mathrm{X}}$ (see [10]). A subset $S \subset X$, is said to be *shattered* by $\mathcal{F}$ if for every subset $T \subset S$, there is some $F \in \mathcal{F}$ with $T = S \cap F$. Define $d_{VC}(\mathcal{F})$ as the largest cardinality of a set shattered by $\mathcal{F}$. ∎

Theorem 6.2: *Let $\mathcal{A} = \{\mathrm{A_i}\}_{\mathrm{i \in I}}$ be a family of measurable subsets of a measure space (X, μ) with $d_{VC}(\mathcal{A}) = \mathrm{d}$. As before, let $a_S := \mu(\bigcap_{i \in S} A_i)$ for $S \subseteq I$. The numbers a_S where $|S| \leq 2^{d+1}$ determine a_T on all (finite) subsets T of I.*

Proof: Let $\{A_i\}, \{B_j\}$ be two families of sets with VC dimension at most d, such that $a_S = b_S$ for all S, $|S| \leq 2^{d+1}$. Consider a minimal finite set $N \subseteq I$ for which $a_N \neq b_N$. For convenience assume that $N = [n]$, i.e., $a_S = b_S$ for all $S \subset [n]$ and $a_{[n]} \neq b_{[n]}$. We will prove that either $d_{VC}\{A_i\}$ or $d_{VC}\{B_j\}$ is greater than d and thus reach a contradiction.

Also, define as before $\alpha_S := \mu(\bigcap_{i \in S} A_i \cap \bigcap_{j \in [n] \setminus S} A_j^c)$. By Lemma 3.3 $\alpha_S - \beta_S = (-1)^{|S|}\epsilon$. Assume that $\epsilon > 0$, so $\alpha_T > 0$ for even $|T|$ whence the sets $\bigcap_{i \in T} A_i \cap \bigcap_{j \notin T} A_j^c$ are not empty when $|T|$ is even.

View $U = [2^{(n-1)}]$ as the family of all subsets of $[n]$ having an even cardinality. For $i \in [n]$ let $L_i \subseteq U$, be the family of all $T \subseteq [n]$ of even cardinality with $i \in T$.

The family $\{L_i\}_{i \in [n]}$ has VC dimension $k = [\log n]$. Indeed, $\mathcal{T} = \mathrm{T_1..T_k}$ is shattered by $\{L_i\}$ iff for every $S \subseteq [k]$ $\cap_{i \in S} T_i$ is not a subset of $\cup_{j \notin S} T_j$. To construct such $\mathcal{T}$ take $T_i = \{x \in [n] \mid$ the i-th digit in binary representation of x is $\epsilon_i\}$. (If n is odd or a multiple of four than ϵ_i can always be chosen so that $|T_i|$ is even. Otherwise, note that the sets $\bigcap_{i \in S} B_i \cap \bigcap_{j \notin S} B_j^c$ are not empty for odd $|S|$, and carry out the whole argument with $\{B_j\}$.)

Let $T_1 ... T_k$ be the k coordinates shattered by $\{L_i\}$. Chose, for $1 \leq r \leq k$, a point x_r in $\bigcap_{i \in T_r} A_i \cap \bigcap_{j \notin T_r} A_j^c$. Put $X = \{x_r\}_{r \in [k]}$. We will show that X is shattered by A_i and so $d_{VC}\{A_i\} \geq k > d$. Take $S \subseteq [k]$. Let $i \in [n]$ be such that T_j is in L_i iff j is in S. Then $X \cap A_i = \{x_r\}_{r \in S}$, completing the proof.

∎

Example 6.3: The VC dimension of any family of compact triangles in the plane is easily shown to be ≤ 7. Therefore for any such $\mathcal{F}$ and a measure μ on the plane we may conclude as follows: The measures of all up to 256-wise intersections: $A_T = \cap_{i \in T} A_i, |T| \leq 256$ uniquely determine the measure of the intersection of any finite subfamily of $\mathcal{F}$. ∎

7 Acknowledgment

We are grateful to Yuri Rabinovich for stimulating discussions.

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] J. Galambos, T. Xu, A new method for generating Bonferroni-type inequalities by iteration, Math. Proc. Cambridge Philos. Soc. 107 (1990), no. 3, 601-607.

[3] M. Kearns, Y. Mansour, D. Ron, R. Rubinfeld, R. E. Schapire and L. Sellie, On the learnability of discrete distributions, STOC 26 (1994), 273-282.

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

[5] L. Lovász, A note on the line reconstruction problem, J. Comb. Theory [B], 13 (1972), 309-310.

[6] L. Lovász, **Combinatorial Problems and Exercises**, North Holland, 1979.

[7] M. Luby and B. Velickovic, On deterministic approximation of DNF, STOC 23 (1991), 430-438.

[8] W. Müller, The edge reconstruction hypothesis is true for graphs with more than $n \log_2 n$ edges, J. Comb. Theory [B], 22 (1977), 281-283.

[9] H. J. Ryser, **Combinatorial Mathematics**, The Mathematical Association of America, 1963.

[10] V. N. Vapnik and A. Ya. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, Theoret. Probl. and Its Appl. 16, 2 (1971), 264-280.