THE DISTANCE APPROACH TO APPROXIMATE COMBINATORIAL COUNTING

Alexander Barvinok and Alex Samorodnitsky

January 2001

Abstract. We develop general methods to obtain fast (polynomial time) estimates of the cardinality of a combinatorially defined set via solving some randomly generated optimization problems on the set. Examples include enumeration of perfect matchings in a graph, linearly independent subsets of a set of vectors and colored spanning subgraphs of a graph. Geometrically, we estimate the cardinality of a subset of the Boolean cube via the average distance from a point in the cube to the subset with respect to some distance function. We derive asymptotically sharp cardinality bounds in the case of the Hamming distance and show that for small subsets a suitably defined "randomized" Hamming distance allows one to get tighter estimates of the cardinality.

1. Introduction

A general problem of combinatorial counting can be stated as follows: given a family $\mathcal{F} \subset 2^X$ of subsets of the ground set X, compute or estimate the cardinality $|\mathcal{F}|$ of the family. We would like to do the computation efficiently, *in polynomial time*. Of course, one should clarify what "given" means, especially since in most interesting cases $|\mathcal{F}|$ is exponentially large in the cardinality $|X|$ of the ground set. We assume that the family $\mathcal{F}$ is defined by its *Optimization Oracle*:

(1.1) Optimization Oracle defining a family $\mathcal{F} \subset 2^X$

Input: A set of integer weights $\gamma_x : x \in X$.

Output: The number $\displaystyle\min_{Y \in \mathcal{F}} \sum_{x \in Y} \gamma_x$.

Key words and phrases. combinatorial counting, perfect matchings, Hamming distance, polynomial time algorithms, isoperimetric inequalities, Boolean cube, concentration of measure.

The research of the first author was partially supported by NSF Grant DMS 9734138. The research of the second author was partially supported by a State of New Jersey grant.

That is, for any given integer weighting $\{\gamma_x\}$ on the set X, we should be able to produce the minimum weight of a subset $Y \in \mathcal{F}$. The following example was our main motivation.

(1.2) Example: Perfect matchings in a graph. Let $G = (V, E)$ be a graph with the set V of vertices and set E of edges. We assume that G has no loops (edges whose endpoints coincide). A set $Y \subset E$ of edges is called a *matching* in G if every vertex of G is incident to at most one edge from Y. A matching Y is called *perfect* if every vertex of G is incident to precisely one edge from Y. Let $\mathcal{F} \subset 2^E$ be the set of all *perfect matchings* in G. The problem of computing or estimating $|\mathcal{F}|$ efficiently is one of the hardest and most intriguing problems of combinatorial counting, see, for example, [Lovász and Plummer 86], [Jerrum and Sinclair 89], [Jerrum 95], [Jerrum and Sinclair 97] and [Jerrum, *et al.* 00]. It is known that the problem of exact counting of perfect matchings is hard. It belongs to the class of # P-hard problems, see Chapter 18 of [Papadimitriou 94] for discussion of computational complexity in enumeration problems. Recently, Jerrum, Sinclair and Vigoda [Jerrum *et al.* 00] found a way to estimate $|\mathcal{F}|$ within any prescribed relative error $\epsilon > 0$ in time polynomial in n and ϵ^{-1} when G is a bipartite graph (in fact, they solved a more general problem of approximating the permanent of a non-negative matrix). However, for general graphs G the problem of efficient approximation of $|\mathcal{F}|$ remains open.

We observe that Optimization Oracle 1.1 can be efficiently constructed. Indeed, if we assign integer weights γ_e: $e \in E$ to the edges of the graph, the minimum weight of a perfect matching can be computed in $O(|V|^3)$ time, see, for example, Section 11.3 of [Papadimitriou and Steiglitz 98].

The following example shows that sometimes optimization is nearly trivial but counting is still hard.

(1.3) Example: Linearly independent subsets. Let $\mathbb{F}$ be a field and let $X \subset \mathbb{F}^d$ be a finite set of vectors. Given a number $k \leq d$, let $\mathcal{F} \subset 2^X$ be the set of all linearly independent k-subsets of X. Optimization Oracle 1.1 for $\mathcal{F}$ is supplied by the following *greedy algorithm*: given integer weights $\{\gamma_x\}$ on X, we construct a linearly independent k-subset $Y \subset X$ of the minimum weight by successively choosing vectors $x_1, x_2, \ldots, x_k$ of the minimum possible weight such that each set $\{x_1\}, \{x_1, x_2\}, \ldots, \{x_1, \ldots, x_k\}$ is linearly independent. When $\mathbb{F} = \mathbb{Q}$, special cases of the counting problem for $\mathcal{F}$ include counting *forests* (acyclic subgraphs) with k edges in a given graph and counting *spanning subgraphs* (connected subgraphs containing all the vertices of the graph) with k edges in a given connected graph. The last problem has interesting relations to percolation and network reliability. These counting problems appear to be difficult, except in some special cases, see [Jerrum and Sinclair 97] for discussion, generalization to counting in *matroids* and some interesting conjectures and Section 12.4 of [Papadimitriou and Steiglitz 98] for optimization via greedy algorithms.

Finally, we give an example of a more complicated structure, where no existing approaches to efficient counting seem to work.

(1.4) Example: Colored spanning subgraphs. Let $G = (V, E)$ be a connected graph with the set V of vertices and set E of edges. Suppose that the edges of G are colored in r colors (that is, the set E is represented as a union $E = E_1 \cup \ldots \cup E_r$ of r non-empty disjoint subsets). Let $\mathcal{F} \subset 2^E$ consist of all subsets of edges such that the underlying graph is a spanning subgraph of G containing precisely one edge of each color. Optimization Oracle 1.1 for $\mathcal{F}$ can be efficiently constructed using the matroid intersection algorithm (see, for example, Section 7.5 of [Grötschel *et al.* 93]). The problem of counting or estimating $|\mathcal{F}|$ appears to be rather difficult.

Next, we would like to discuss what "polynomial time" means. Typically, we are dealing with an infinite family of counting problems $(\mathcal{F}_i, X_i) : i \in I$ (for example, X_i may range over the sets of edges of all finite graphs G_i and $\mathcal{F}_i \subset 2^{X_i}$ may be the set of all perfect matchings in G_i). We would like to construct an algorithm which works for every particular instance $(\mathcal{F}_i, X_i)$. If there exists a univariate polynomial *poly* such that the running time of the algorithm on the instance $(\mathcal{F}_i, X_i)$ is bounded by $poly(|X_i|)$, we say that the algorithm is polynomial time. Our algorithms are *randomized*, that is, the algorithms rely on some coin tossing on the way, so the outcome is a random variable, which, with some high probability (say, 0.9) satisfies the desired properties. The probability of success can be made arbitrarily close to 1 by running the algorithm several times and taking a version of the majority vote (cf. Theorem 3.6 and Remark that follows). For a general reference in the area of computational complexity and algorithms, see [Papadimitriou 94].

The most general approach to combinatorial counting has been via Monte Carlo method. The key component of the method is the ability to sample a random point from the (almost) uniform distribution on $\mathcal{F}$. Often, to achieve this, a Markov chain on the set $\mathcal{F}$ is generated, so that it converges rapidly to the uniform distribution on $\mathcal{F}$ (see [Jerrum and Sinclair 97] for a survey). Spectacular successes of this approach are finding a polynomial time randomized algorithm to count matchings of all sizes in a given graph [Jerrum and Sinclair 89] and to count perfect matchings in a given bipartite graph [Jerrum *et al.* 00], both within any prescribed relative error. When the Markov chain approach works, it produces incomparably better results than the method of this paper. However, for many important counting problems, some of which are mentioned above, it is either not clear how to generate a rapidly mixing Markov chain (as in Example 1.4) or, when there is a natural candidate (as in Example 1.3, see [Jerrum and Sinclair 97]), it seems to be extremely hard to prove that the chain is indeed converging rapidly enough to the steady state. In contrast, our approach produces very crude bounds, but it is totally insensitive to the fine structure of $\mathcal{F}$, so it is ready to handle a broad class of problems. A slight modification of the problem (for example, given small integer weights on the edges of the graph in Example 1.2, estimate the number of perfect matchings of the prescribed total weight) may lead to drastic changes in the construction of the underlying Markov chain but has almost no effect in our approach.

(1.5) The distance approach. The main idea of our approach is as follows. Given a family $\mathcal{F}$, we identify it with a subset F of a metric space (Ω, d), such that for any given point $x \in \Omega$ the distance $d(x, F) = \min_{y \in F} d(x, y)$ can be quickly computed using Optimization Oracle 1.1 for $\mathcal{F}$. Then we estimate the cardinality $|F|$ from the distance $d(x, F)$ for a typical $x \in \Omega$. Intuitively, if $|F|$ is small, we expect the distance $d(x, F)$ from a random point $x \in \Omega$ to be large and vice versa. In this paper, Ω is the Boolean cube $\{0,1\}^n$ and d is either the Hamming distance or its modification, although as we discuss in Section 5, some other possibilities may be of interest. Thus our approach can be considered as a refinement of the classical Monte Carlo method: we do not only register how often a randomly sampled point $x \in \Omega$ lands in the target set F, but also take into account the distance $d(x, F)$. This allows us to get non-trivial bounds even when $|F|$ is exponentially small with respect to $|\Omega|$ so that x typically misses F.

The paper is organized as follows.

In Section 2, we introduce a "geometric cousin" of Optimization Oracle 1.1. Distance Oracle 2.2 describes a subset F of the Boolean cube $\{0,1\}^n$ by computing a suitably defined distance d from a given point in the cube to the set. We show how to construct embeddings $\phi : \mathcal{F} \longrightarrow \{0,1\}^n$, so that the Distance Oracle for the image $F = \phi(\mathcal{F})$ is derived from the Optimization Oracle for $\mathcal{F}$. We show that in some important cases (for example, when $\mathcal{F}$ is the set of Example 1.2 of perfect matchings in a graph or the set of colored spanning subgraphs of Example 1.4), we can "squeeze" $\mathcal{F}$ into a substantially smaller cube than we would have expected for a general family $\mathcal{F}$.

In Section 3, we describe the bounds obtained by choosing d to be the Hamming distance in the cube. The bounds are sharp, meaning that we can't possibly estimate (in polynomial time) the cardinality of a subset $F \subset \{0,1\}^n$ better if the only information available is the Hamming distance from any given point $a \in \{0,1\}^n$ to the set F.

In Section 4, we describe how to get better bounds for small sets by using a suitably defined "randomized Hamming distance", which ignores a (random) part of the information contained in the standard Hamming distance. The isoperimetric problems arising here seem to be interesting in their own right. The proofs are not complicated but somewhat lengthy and therefore postponed till Section 6.

In Section 5, we discuss the types of estimates which can be obtained for particular counting problems (such as in Examples 1.2–1.4), possible ramifications of our approach and its relations with the Monte Carlo method.

2. Distance Oracle and Cubical Embeddings

The idea of our method is to represent $\mathcal{F}$ geometrically as a subset F of the Boolean cube and then derive estimates of $|\mathcal{F}|$ using the average distance from a point in the cube to F.

(2.1) Definitions. Let $C_n = \{0,1\}^n$ be the Boolean cube and let dist be the Hamming distance in C_n, that is

$$\operatorname{dist}(a,b) = \sum_{i:\alpha_i \neq \beta_i} 1 \qquad \text{for} \quad a = (\alpha_1, \ldots, \alpha_n), \ b = (\beta_1, \ldots, \beta_n) \in C_n.$$

More generally, let us fix n functions $d_i : \{0,1\} \times \{0,1\} \longrightarrow \mathbb{Z}$, $i = 1, \ldots, n$, which we interpret as *penalties*. We assume that $d_i \geq 0$ and that $d(0,0) = d(1,1) = 0$. Finally, let

$$d(a,b) = \sum_{i=1}^{n} d_i(\alpha_i, \beta_i), \quad \text{where} \quad a = (\alpha_1, \ldots, \alpha_n) \quad \text{and} \quad b = (\beta_1, \ldots, \beta_n)$$

be the L^1 distance function determined by the penalties $\{d_i\}$. If $d_i(\alpha, \beta) = 1$ whenever $\alpha \neq \beta$ then $d(a,b) = \operatorname{dist}(a,b)$.

For a subset $B \subset C_n$ and a point $a \in C_n$, let

$$d(a,B) = \min_{b \in B} d(a,b)$$

be the distance from a to B. In particular, let

$$\operatorname{dist}(a,B) = \min_{b \in B} \operatorname{dist}(a,b)$$

be the Hamming distance from a point a to the subset B.

We will be working with the following "geometric cousin" of Optimization Oracle 1.1.

(2.2) Distance Oracle defining a set $F \subset C_n$

Input: A point $a \in C_n$ and penalties $d_i : \{0,1\} \times \{0,1\} \longrightarrow \mathbb{Z}$, $i = 1, \ldots, n$.

Output: The number $d(a, F)$.

There is an obvious way to associate with a family $\mathcal{F} \subset 2^X$ a subset $F \subset C_n$, $n = |X|$ of the Boolean cube.

(2.3) Straightforward embedding. Let us identify the ground set X with the set $\{1, \ldots, n\}$, $n = |X|$. Let $\mathcal{F}$ be a family of subsets of $\{1, \ldots, n\}$ given by its Optimization Oracle. For a subset $Y \in \mathcal{F}$ let us define the indicator $y \in C_n$, $y = (\eta_1, \ldots, \eta_n)$ by

$$\eta_i = \begin{cases} 1 & \text{if } i \in Y \\ 0 & \text{if } \ i \notin Y. \end{cases}$$

Let $F = \bigl\{ y \in C_n : Y \in \mathcal{F} \bigr\}$ be the set of all indicators of subsets $Y \in \mathcal{F}$.

Let us construct the Distance Oracle for the set $F \subset C_n$. Given a point $a = (\alpha_1, \ldots, \alpha_n) \in C_n$ and penalties d_i, $i = 1, \ldots, n$, let us define weights γ_i by $\gamma_i = d_i(\alpha_i, 1) - d_i(\alpha_i, 0)$. Then for a set $Y \subset \{1, \ldots, n\}$ and its indicator $y = (\eta_1, \ldots, \eta_n) \in C_n$, we have

$$\sum_{i \in Y} \gamma_i = \sum_{i \in Y} \big(d_i(\alpha_i, 1) - d_i(\alpha_i, 0)\big) = \sum_{i=1}^{n} d_i(\alpha_i, \eta_i) - \sum_{i=1}^{n} d_i(\alpha_i, 0) = d(a, y) - d(a, 0).$$

Hence, given the output

$$\lambda = \min_{Y \in \mathcal{F}} \sum_{i \in Y} \gamma_i$$

of Oracle 1.1 for the family $\mathcal{F}$, we can easily compute the output

$$d(a, F) = \lambda + d(a, 0)$$

of Oracle 2.2 for the set F. Thus, given an Optimization Oracle 1.1 for a family $\mathcal{F} \subset 2^X$, we can efficiently construct a Distance Oracle 2.2 for a set $F \subset C_n$, $n = |X|$, such that $|F| = |\mathcal{F}|$.

To be able to estimate the cardinality $|\mathcal{F}|$ with a better precision, we would like to embed $\mathcal{F}$ into a smaller Boolean cube. Sometimes this is indeed possible.

(2.4) Economical embedding. Suppose that the ground set X can be represented as a union $X = X_1 \cup \ldots \cup X_k$ of (not necessarily disjoint) parts X_i, so that $|Y \cap X_i| = 1$ for every subset $Y \in \mathcal{F}$ and every X_i. In other words, every member of $\mathcal{F}$ is a transversal of the cover of X by $X_1, \ldots, X_k$. Let

$$m_i = \lceil \log_2 |X_i| \rceil \quad \text{and} \quad m = \sum_{i=1}^{k} m_i.$$

We construct an embedding $\mathcal{F} \longrightarrow C_m$ as follows.

First, we index the elements of X_i by distinct binary strings of length m_i, that is, we choose an embedding $\phi_i : X_i \longrightarrow C_{m_i}$. Thus for any $x \in X_i$ the point $\phi_i(x)$ is a binary string of length m_i and $\phi_i(x) \neq \phi_i(y)$ provided $x \neq y$.

Let us identify

$$C_m = C_{m_1} \times \ldots \times C_{m_k}.$$

For a subset $Y \in \mathcal{F}$, let us define $y \in C_m$ as

$$y = (y_1, \ldots, y_k), \quad \text{where} \quad y_i = \phi_i(Y \cap X_i) \in C_{m_i}.$$

Note that y is well-defined, since every intersection $Y \cap X_i$ consists of a single point. Let $F = \big\{ y \in C_m : Y \in \mathcal{F} \big\}$. Clearly, $|F| = |\mathcal{F}|$.

Given an Optimization Oracle 1.1 for $\mathcal{F}$, let us construct a Distance Oracle 2.2 for F. The input of Oracle 2.2 consists of a point $a \in C_m$ (binary string of length m) and penalty functions $\{d_i : i = 1, \ldots m\}$. We view a as

$$a = (a_1, \ldots , a_k), \quad \text{where} \quad a_i \in C_{m_i}.$$

The penalties d_i, $i = 1, \ldots , m$ give rise to the L^1 distance function d on binary strings, cf. Definition 2.1. For a point $x \in X$, let us define its weight γ_x by

$$\gamma_x = \sum_{i:\ x \in X_i} d\big(a_i, \phi_i(x)\big). \tag{2.4.1}$$

Let $Y \in \mathcal{F}$ be a set and let $y \in C_m$ be the point representing Y. We observe that

$$\sum_{x \in Y} \gamma_x = \sum_{x \in Y} \sum_{i:\ x \in X_i} d\big(a_i, \phi_i(x)\big) = \sum_{i=1}^{k} d(a_i, y_i) = d(a, y).$$

Hence, the outputs of Oracles 1.1 and 2.2 coincide:

$$\min_{Y \in \mathcal{F}} \sum_{x \in Y} \gamma_x = \min_{y \in Y} d(a, y).$$

Thus, given an Optimization Oracle 1.1 for a family $\mathcal{F} \subset 2^X$, we can efficiently construct a Distance Oracle 2.2 for a set $F \subset C_m$, such that $|F| = |\mathcal{F}|$. More precisely, given a point $a \in C_m$ and penalties $\{d_i\}$, we compute weights $\{\gamma_x\}$ on X by (2.4.1) in $O(k|X| \ln |X|)$ time and then apply Optimization Oracle 1.1 to find the minimum weight λ of a subset $Y \in \mathcal{F}$ in this weighting. The distance $d(a, F)$ is equal to λ.

(2.5) Example: Embedding perfect matchings and colored spanning subgraphs. Let $\mathcal{F}$ be the family of all perfect matchings in a graph $G = (V, E)$, see Example 1.2. The straightforward embedding (2.3) identifies $\mathcal{F}$ with a subset F of the Boolean cube $\{0, 1\}^{|E|}$ and provides us with Distance Oracle 2.2 for F. We will be better off using the economical embedding (2.4). Indeed, for a vertex $v \in V$ of G, let E_v be the set of edges of G incident with v. Then $E = \cup_{v \in V} E_v$ and every perfect matching has exactly one edge in every set E_v. Hence the embedding (2.4) identifies $\mathcal{F}$ with a subset F of the Boolean cube $\{0, 1\}^m$, where

$$m = \sum_{v \in V} \lceil \log_2 |E_v| \rceil$$

and provides us with Distance Oracle 2.2 for F. Given a point $a \in C_m$, by (2.4.1) we compute weights γ_e on the edges E in $O(|E| \ln |E|)$ time (since every edge $e \in E$ belongs to exactly two sets E_v) and then find the minimum weight λ of a perfect matching in G in $O(|V|^3)$ time. The distance $d(a, F)$ from a to F is equal to λ.

Typically, if the graph has $|V| = n$ vertices and $\Omega(n^2)$ edges, the dimension of the straightforward embedding will be $\Omega(n^2)$, whereas the dimension of the economical embedding will be $O(n \ln n)$.

Similarly, if $\mathcal{F}$ is the set of properly colored spanning subgraphs of Example 1.4, the partition $E = E_1 \cup \ldots \cup E_r$ gives rise to the economical embedding of $\mathcal{F}$ into the Boolean cube $\{0,1\}^m$ of dimension $m = \sum_{i=1}^r \lceil \log_2 |E_i| \rceil$, as opposed to the dimension $|E| = \sum_{i=1}^r |E_i|$ of the straightforward embedding.

3. Estimating Cardinality from the Hamming Distance

In this section, we obtain estimates of the cardinality of a subset $F \subset C_n$ if we choose $d_i(0,1) = d_i(1,0) = 1$, $i = 1, \ldots, n$ in Distance Oracle 2.2. In other words, we estimate $|F|$, provided we can compute the Hamming distance $\operatorname{dist}(x, F)$ to F from any given point $x \in C_n$, cf. Definitions 2.1. Our main tool is the *average* Hamming distance from a point to the set.

(3.1) Definition. Let $A \subset C_n$ be a subset of the Boolean cube. Let

$$\Delta(A) = \frac{1}{2^n} \sum_{x \in C_n} \operatorname{dist}(x, A)$$

be the average Hamming distance from a point in the cube to the set A.

Obviously, $\Delta(A) \leq \Delta(B)$ if $B \subset A$.

(3.2) Example: Set consisting of a single point. Suppose that the set A is a point. Without loss of generality we assume that $A = \{(0, \ldots, 0)\}$. Then, for $x = (\xi_1, \ldots, \xi_n)$ we have $\operatorname{dist}(x, A) = \operatorname{dist}(x, 0) = \xi_1 + \ldots + \xi_n$ and

$$\Delta(A) = \frac{1}{2^n} \sum_{x \in C_n} \operatorname{dist}(x, A) = \frac{1}{2^n} \sum_{x \in C_n} (\xi_1 + \ldots + \xi_n) = \frac{n}{2}.$$

It follows then that $\Delta(A) \leq n/2$ for any non-empty $A \subset C_n$ and that $\Delta(A) = n/2$ if and only if A consists of a single point.

Our first objective is to present a probabilistic algorithm that computes $\Delta(A)$ approximately by averaging $\operatorname{dist}(x, A)$ for a number of randomly chosen $x \in C_n$.

(3.3) Algorithm for computing $\Delta(A)$

Input: A set $A \subset C_n$ defined by its Distance Oracle 2.2 and a number $\epsilon > 0$.

Output: A number α approximating $\Delta(A)$ within error ϵ.

Algorithm: Let $k = \lceil 3n/2\epsilon^2 \rceil$. Sample k points $x_1, \ldots, x_k \in C_n$ independently at random from the uniform distribution in the cube C_n. Apply Distance Oracle 2.2 to find $\operatorname{dist}(x_i, A)$, $i = 1, \ldots, k$. Compute $\alpha = \dfrac{1}{k} \displaystyle\sum_{i=1}^{k} \operatorname{dist}(x_i, A)$. Output α.

To prove that Algorithm 3.3 indeed approximates $\Delta(A)$ with the desired accuracy, we need a couple of technical results. The first lemma supplies us with important *concentration inequalities* for the Boolean cube.

(3.4) Lemma. *Let $\xi_1, \ldots, \xi_N$ be independent random variables taking values in $\{0,1\}$. Let $f : C_N \longrightarrow \mathbb{R}$ be a function such that $|f(x) - f(y)| \leq 1$ whenever $\operatorname{dist}(x,y) \leq 1$ and let η be the random variable $f(\xi_1, \ldots, \xi_N)$. Then for any $\delta > 0$*

$$\mathbf{P}\ \Big\{\eta : |\eta - \mathbf{E}\ (\eta)| \geq \delta\Big\} \leq 2\exp\Big\{\frac{-2\delta^2}{N}\Big\}.$$

Proof. This is a special case of Lemma 1.2 of [McDiarmid 89]. □

The next lemma provides a useful "scaling" trick.

(3.5) Lemma. *Let us fix positive integers k and n and let $N = kn$. Let us identify $C_N = C_n \times \ldots \times C_n = (C_n)^k$. Thus a point $x \in C_N$ is identified with a k-tuple $x = (x_1, \ldots, x_k)$, where $x_i \in C_n$ for $i = 1, \ldots, k$.*

For a subset $A \subset C_n$, let $B = A \times \ldots \times A = A^k \subset C_N$. Then

$$\operatorname{dist}(x, B) = \sum_{i=1}^k \operatorname{dist}(x_i, A) \quad \textit{for any} \quad x = (x_1, \ldots, x_k) \in C_N$$

and

$$\Delta(B) = k\Delta(A).$$

Proof. Clearly,

$$\operatorname{dist}(x, y) = \sum_{i=1}^k \operatorname{dist}(x_i, y_i) \quad \text{for all} \quad x, y \in C_N,$$

hence the first identity follows. Next,

$$\begin{aligned}\Delta(B) &= \frac{1}{2^N} \sum_{x \in C_N} \operatorname{dist}(x, B) = \frac{1}{2^N} \sum_{x_1, \ldots, x_k \in C_n} \sum_{i=1}^k \operatorname{dist}(x_i, A) \\ &= \frac{k2^{n(k-1)}}{2^{nk}} \sum_{x \in C_n} \operatorname{dist}(x, A) = \frac{k}{2^n} \sum_{x \in C_n} \operatorname{dist}(x, A) = k\Delta(A).\end{aligned}$$

□

Now we can prove correctness of Algorithm 3.3.

(3.6) Theorem. *With probability at least 0.9, the output α of Algorithm 3.3 satisfies the inequality $|\Delta(A) - \alpha| \leq \epsilon$.*

Proof. Let $N = nk$ and let us identify $C_N = (C_n)^k$ as in Lemma 3.5. Let $B = A^k \subset C_N$. Let $f : C_N \longrightarrow \mathbb{R}$ be defined by $f(x) = \operatorname{dist}(x, B)$. Interpreting $x = (\xi_1, \ldots, \xi_N)$ as a vector of N independent random variables ξ_i uniformly distributed on $\{0,1\}$, applying Lemma 3.4 with $\delta = k\epsilon$ and observing that $\mathbf{E}\ (f) = \Delta(B)$, we conclude that

$$\mathbf{P}\ \Big\{x : |\operatorname{dist}(x,B) - \Delta(B)| \geq k\epsilon\Big\} \leq 2\exp\Big\{-\frac{2(\epsilon k)^2}{N}\Big\} = 2\exp\Big\{-\frac{2\epsilon^2 k}{n}\Big\} \leq 0.1.$$

Since by Lemma 3.5

$$\Delta(B) = k\Delta(A) \quad \text{and} \quad \frac{1}{k}\sum_{i=1}^{k} \operatorname{dist}(x_i, A) = \frac{1}{k}\operatorname{dist}(x,B)$$

for $x = (x_1, \ldots, x_k)$, we conclude that

$$\mathbf{P}\ \Big\{x_1, \ldots, x_k : \Big|\frac{1}{k}\sum_{i=1}^{k}\operatorname{dist}(x_i, A) - \Delta(A)\Big| \geq \epsilon\Big\} =$$
$$\mathbf{P}\ \Big\{x : |\operatorname{dist}(x,B) - \Delta(B)| \geq k\epsilon\Big\} \leq 0.1,$$

and the proof follows. □

Remark. Hence to evaluate $\Delta(A)$ within error ϵ we have to average $O(n\epsilon^{-2})$ values $\operatorname{dist}(x_i, A)$. By doing that, we allow probability 0.1 of failure. As usual, to attain a lower probability $\delta > 0$ of failure, one should run Algorithm 3.3 $O(\ln \delta^{-1})$ times and then select the median of the computed α's (cf. [Jerrum *et al.* 86]). For all applications, choosing $\epsilon = 1$ will suffice and in many cases $\epsilon = \sqrt{n}$ will do (cf. Section 5.1). Hence, often we will have to apply Oracle 2.2 only a constant number of times.

We would like to relate the value of $\Delta(A)$ to the cardinality $|A|$.

(3.7) Definition. Entropy Function. For $0 \leq x \leq 1/2$ let

$$H(x) = x\log_2\frac{1}{x} + (1-x)\log_2\frac{1}{1-x}.$$

We agree that $H(0) = 0$. Thus H is an increasing concave function on the interval $[0, 1/2]$.

We use the following estimate (see, for example, Theorem 1.4.5 of [van Lint 99])

(3.7.1) $$\sum_{k=0}^{r}\binom{n}{k} \leq 2^{nH(r/n)} \qquad \text{for} \quad r \leq n/2.$$

Also, we remark that around $x = +0$ we have

(3.7.2) $$H(x) = x\log_2\frac{1}{x} + O(x) \quad \text{and} \quad H\Big(\frac{1}{2} - x\Big) = 1 - \frac{2}{\ln 2}x^2 + O(x^3)$$

We will use the classical isoperimetric inequality for the Boolean cube (see, for example, [Leader 91]).

(3.8) Harper's Theorem. *Let $A \subset C_n$ be a set such that*

$$|A| \geq \sum_{k=0}^{r} \binom{n}{k}$$

for some integer r. Then, for any non-negative integer t

$$\Big|\big\{x \in C_n : \operatorname{dist}(x, A) \leq t\big\}\Big| \geq \sum_{k=0}^{r+t} \binom{n}{k}.$$

We are going to obtain an estimate of the cardinality of a set $A \subset C_n$ in terms of the average Hamming distance $\Delta(A)$ from a point $x \in C_n$ to A. It is convenient to express the estimate in terms of a related quantity

$$\rho = \rho(A) = \frac{1}{2} - \frac{\Delta(A)}{n}.$$

As follows from Example 3.2, for every non-empty set $A \subset C_n$ we have $0 \leq \rho(A) \leq 1/2$. We observe that $\rho(A) = 0$ if and only if A consists of a single point and that $\rho(A) = 1/2$ if and only if A is the whole cube C_n.

(3.9) Theorem. *Let $A \subset C_n$ be a non-empty set. Let*

$$\rho = \frac{1}{2} - \frac{\Delta(A)}{n}.$$

Then

$$1 - H\Big(\frac{1}{2} - \rho\Big) \leq \frac{\log_2 |A|}{n} \leq H(\rho).$$

Before we proceed with a formal proof, we would like to highlight some ideas.

(3.10) The idea of the proof. Extremal sets. Let $A \subset C_n$ be a set. Concentration inequalities (Lemma 3.4) imply that the average distance $\Delta(A)$ is approximately equal to the distance $\operatorname{dist}(x, A)$ from a "typical" point $x \in C_n$ to A (see also Sections 6.2 and 7.9 of [Milman and Schechtman 86]). For a given positive integer t, let us consider the t-neighborhood $A_t = \big\{x \in C_n : \operatorname{dist}(x, A) \leq t\big\}$ of A. We expect that $\Delta(A) \approx t_1$, where t_1 is the smallest value of t such that A_t covers "almost all" cube C_n. The neighborhood A_t grows the slowest when A is a ball in the Hamming metric, that is when $A = \big\{x : \operatorname{dist}(x, x_0) \leq r\big\}$ for some $x_0 \in C_n$ and some $r > 0$, as follows from Harper's Theorem 3.8, cf. also [Leader 91]. Hence the upper bound for $n^{-1} \log_2 |A|$ in Theorem 3.9 is attained (up to an $O(n^{-1/2})$ error term) when A is a ball. The neighborhood A_t grows the fastest when the points of A are spread around in C_n. In any case, the size $|A_t|$ does not exceed the sum of sizes of the balls of radius t centered at the points of A. Thus the lower bound for $n^{-1} \log_2 |A|$ in Theorem 3.9 is obtained from this "packing" type argument. One can show that if the points of A are chosen at random in C_n, then with high probability the lower bound is indeed attained asymptotically. More precisely, let us fix a number $0 < \beta < 1$ and let A be the set of $\lfloor 2^{\beta n} \rfloor$ points chosen at random from C_n. Then with the probability that tends to 1 as n grows to infinity, $\beta = 1 - H\left(\frac{1}{2} - \rho\right) + O(n^{-1/2})$. The proof is straightforward, but technical and therefore omitted.

Finally, we note that using average distance $\Delta(A)$ and the scaling trick (Lemma 3.5) allows us to get rid of $O(n^{-1/2})$ error terms in the proof.

Proof of Theorem 3.9. Let us choose a positive even integer m, let $N = mn$ and let us identify $C_N = (C_n)^m$, as in Lemma 3.5. Let $B = A^m \subset C_N$. Let us fix the uniform probability measure $\mathbf{P}$ on C_N.

Let $\alpha = \log_2 |A|/n$, so $|A| = 2^{\alpha n}$ and $|B| = 2^{\alpha N}$. Let $0 \le \gamma \le 1/2$ be a number such that $H(\gamma) = \alpha$ and let $r = \lfloor N\gamma \rfloor$. Then by (3.7.1)

$$|B| = 2^{N \cdot H(\gamma)} \ge \sum_{k=0}^{r} \binom{N}{k}.$$

Then Theorem 3.8 implies that

$$\left| \left\{ x \in C_N : \operatorname{dist}(x, B) \le N/2 - r \right\} \right| \ge \sum_{k=0}^{N/2} \binom{N}{k} = 2^{N-1}.$$

Therefore,

$$\mathbf{P} \left\{ x \in C_N : \operatorname{dist}(x, B) \le \frac{N}{2} - r \right\} \ge \frac{1}{2}.$$

We have that $x = (x_1, \ldots, x_m)$ for some $x_i \in C_n$ and that $\operatorname{dist}(x, B) = \operatorname{dist}(x_1, A) + \ldots + \operatorname{dist}(x_m, A)$ (see Lemma 3.5). Therefore,

$$\mathbf{P} \left\{ (x_1, \ldots, x_m) : \frac{1}{m} \sum_{i=1}^{m} \operatorname{dist}(x_i, A) \le \frac{N}{2m} - \frac{r}{m} \right\} \ge \frac{1}{2}. \tag{1}$$

Now we observe that

$$\frac{N}{2m} - \frac{r}{m} \longrightarrow \frac{n}{2} - n\gamma \quad \text{as} \quad m \longrightarrow +\infty. \tag{2}$$

Furthermore, by the Law of Large Numbers,

$$\frac{1}{m} \sum_{i=1}^{m} \operatorname{dist}(x_i, A) \longrightarrow \Delta(A) \quad \text{in probability} \qquad \text{as} \quad m \longrightarrow +\infty. \tag{3}$$

Hence the assumption that $\Delta(A) > n/2 - n\gamma$ would contradict (1)–(3). Thus we must have $\Delta(A) \le n/2 - n\gamma$, which implies that $\gamma \le \rho(A)$. Hence $\alpha = H(\gamma) \le H(\rho)$ and the upper bound is proven.

Let us prove the lower bound. We observe that for every point $b \in C_N$ and any $N/2 \geq s \geq 0$

$$\left|\{x \in C_N : \operatorname{dist}(x, b) \leq s\}\right| = \sum_{k=0}^{s} \binom{N}{k} \leq 2^{N \cdot H(s/N)}.$$

Therefore,

$$\left|\{x \in C_N : \operatorname{dist}(x, B) \leq s\}\right| \leq |B| 2^{N \cdot H(s/N)} = 2^{N \cdot (H(s/N)+\alpha)}.$$

Hence

$$\mathbf{P} \left\{x \in C_N : \operatorname{dist}(x, B) \leq s\right\} \leq 2^{N \cdot (H(s/N)+\alpha-1)}.$$

Therefore,

$$\mathbf{P} \left\{(x_1, \dots, x_m) : \frac{1}{m} \sum_{i=1}^{m} \operatorname{dist}(x_i, A) \leq s/m\right\} \leq 2^{N \cdot (H(s/N)+\alpha-1)}. \tag{4}$$

If $\Delta(A) = n/2$ then A is a point and the lower bound in Theorem 3.9 is satisfied. Otherwise, let us fix an $\epsilon > 0$ such that $(1+\epsilon)\Delta(A)/n < 1/2$ and let $s = \lceil m(1+\epsilon)\Delta(A) \rceil$. We have

$$s/m \longrightarrow (1+\epsilon)\Delta(A) \quad \text{and} \quad s/N \longrightarrow (1+\epsilon)\Delta(A)/n \quad \text{as} \quad m \longrightarrow +\infty. \tag{5}$$

Hence the assumption that $H\big((1+\epsilon)\Delta(A)/n\big)+\alpha-1 < 0$ would contradict (3)–(5). Therefore, $H\big((1+\epsilon)\Delta(A)/n\big)+\alpha-1 \geq 0$ for any $\epsilon > 0$ and $H\big(\Delta(A)/n\big)+\alpha-1 \geq 0$. Since $\Delta(A)/n = \frac{1}{2} - \rho$, the proof follows. □

For applications, the most interesting case is when $n^{-1} \log_2 |A|$ is small, that is $\rho \approx 0$.

(3.11) Corollary. *There exist positive constants c_1 and c_2 such that for any non-empty set $A \subset C_n$ and for $\rho = \dfrac{1}{2} - \dfrac{\Delta(A)}{n}$ we have*

$$c_1 \cdot \rho^2 \leq \frac{\ln |A|}{n} \leq c_2 \cdot \rho \ln \frac{1}{\rho}.$$

In particular, for any $c_1 < 2$ and any $c_2 > 1$, the inequality holds in a sufficiently small neighborhood of $\rho = 0$.

Proof. Follows from Theorem 3.9 by (3.7.2). □

(3.12) Discussion. Figure 1 depicts the feasible region for $n^{-1}\log_2|A|$ as described by Theorem 3.9.

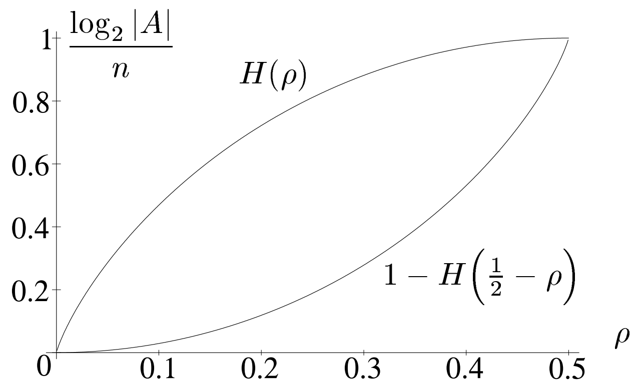

Figure 1

Thus possible values of $n^{-1}\log_2|A|$ with the given value of ρ form a vertical interval between the two curves. As we discussed in Section 3.10, asymptotically both bounds are sharp. Remarkably, the bounds converge at $\rho = 0$ and $\rho = 0.5$. On the other hand, the difference is the greatest when $\rho = 1/4$. Thus, if the average Hamming distance from a point $x \in C_n$ to a set $A \subset C_n$ is $n/4$, the set A can contain as many as $2^{0.811n}$ points and as few as $2^{0.189n}$ points. We note that if A is a face (subcube) of the Boolean cube then the corresponding point lies on the straight line $n^{-1}\log_2|A| = 2\rho(A)$.

Corollary 3.11 (with somewhat weaker constants and stated in different terms) together with the observation that the distance $\text{dist}(x, A)$ for a randomly chosen point $x \in C_n$ allows one to estimate ρ up to an $O(n^{-1/2})$ error constitute the main result of the earlier paper [Barvinok 97]. Consequently, the main conclusion of [Barvinok 97] is equivalent to stating that the Hamming distance to A from a random point x in the Boolean cube allows one to decide whether $|A|$ is exponentially large in n. In this paper, we make improvements in several direction. The most important one is that the optimization functionals of [Barvinok 97] are recognized as *distances*, which allows us to establish connections with isoperimetric inequalities. In particular, Theorem 3.9 provides us with sharp bounds valid for all $0 \le \rho \le 1/2$. Also, using Algorithm 3.3 and Theorem 3.6 we get rid of the $O(n^{-1/2})$ error term. This allows us to obtain meaningful cardinality estimates for sets with small values of ρ. Curiously, we can even distinguish in polynomial time between a set consisting of a single point ($\rho = 0$) and a set having more than one point (one can show that $\rho = \Omega(1/n)$ in that case), although apparently we can't distinguish between sets consisting of 2 and 3 points respectively. Finally, construction 2.4 of "economical embedding" allows us to obtain tighter bounds for a wide class of problems by lowering the dimension of the ambient Boolean cube.

In Section 4, we show how using "randomized Hamming distance" allows one to get better estimates for sets A with $n^{-1} \log_2 |A|$ small, which is the case in most applications. Essentially, the randomized Hamming distance will allow us to "sandwich" such a set between a random set and a face (subcube), as opposed to a random set and a Hamming ball in the case of the standard Hamming metric, cf. Section 3.10.

4. Randomized Hamming Distance

Let us fix a number $0 < p \leq 1$ and let $q = 1 - p$. In this section, we construct a quantity $\Delta(A, p)$, which measures the cardinality of "small" subsets $A \subset C_n$ of the Boolean cube in a somewhat more precise way than the average Hamming distance $\Delta(A)$ discussed in Section 3. In fact, $\Delta(A, 1) = \Delta(A)$, so $\Delta(A)$ is a particular case of $\Delta(A, p)$.

(4.1) Definitions. Let Λ_n be a copy of the Boolean cube $\{0, 1\}^n$. We make Λ_n a probability space by letting

$$\mathbf{P}\ \{l\} = p^{|l|} q^{n-|l|}, \quad \text{where} \quad |l| = \lambda_1 + \ldots + \lambda_n \quad \text{for} \quad l = (\lambda_1, \ldots, \lambda_n).$$

Hence a vector $l = (\lambda_1, \ldots, \lambda_n)$ from Λ_n is interpreted as a realization of n independent random variables λ_i such that $\mathbf{P}\ \{\lambda_i = 1\} = p$ and $\mathbf{P}\ \{\lambda_i = 0\} = q$.

For $x, y \in C_n$ and an $l \in \Lambda_n$, where $x = (\xi_1, \ldots, \xi_n)$, $y = (\eta_1, \ldots, \eta_n)$ and $l = (\lambda_1, \ldots, \lambda_n)$, let

$$d_l(x, y) = \sum_{i: \xi_i \neq \eta_i} \lambda_i.$$

In other words, we count disagreement in the i-th coordinate of x and y if and only if the value of λ_i is 1. Thus if $l = (1, \ldots, 1)$, we have $d_l(x, y) = \operatorname{dist}(x, y)$, the usual Hamming distance.

For $l \in \Lambda_n$ and a set $A \subset C_n$, let

$$d_l(x, A) = \min_{y \in A} d_l(x, y).$$

Finally, let

$$\Delta(A, p) = \sum_{l \in \Lambda_n} \sum_{x \in C_n} d_l(x, A) \frac{p^{|l|} q^{n-|l|}}{2^n}.$$

In other words, $\Delta(A, p)$ is the expected value of $d_l(x, A)$, where $x = (\xi_1, \ldots, \xi_n)$ and $l = (\lambda_1, \ldots, \lambda_n)$ are vectors of independent random variables such that $\mathbf{P}\ \{\lambda_i = 1\} = p$, $\mathbf{P}\ \{\lambda_i = 0\} = q$ and $\mathbf{P}\ \{\xi_i = 0\} = \mathbf{P}\ \{\xi_i = 1\} = 1/2$. Obviously, $\Delta(A, p) \leq \Delta(B, p)$ if $B \subset A$.

It follows that for a fixed non-empty $A \subset C_n$, the value $\Delta(A, p)$ is a polynomial in p of degree at most n.

(4.2) Example: Set consisting of a single point. Suppose that the set A consists of a single point. Without loss of generality we assume that $A = \{(0, \dots, 0)\}$. Then for $x = (\xi_1, \dots, \xi_n)$ and $l = (\lambda_1, \dots, \lambda_n)$,

$$d_l(x, A) = \sum_{i=1}^{n} \lambda_i \xi_i.$$

Interpreting λ_i and ξ_i, $i = 1, \dots, n$ as independent random variables such that $\mathbf{P}\ \{\xi_i = 1\} = \mathbf{P}\ \{\xi_i = 0\} = 1/2$ and $\mathbf{P}\ \{\lambda_i = 1\} = p$, $\mathbf{P}\ \{\lambda_i = 0\} = q$, we get

$$\Delta(A, p) = \mathbf{E}\ \sum_{i=1}^{n} \lambda_i \xi_i = \sum_{i=1}^{n} (\mathbf{E}\ \lambda_i)(\mathbf{E}\ \xi_i) = \frac{np}{2}.$$

It follows then that for any non-empty set $A \subset C_n$ we have $\Delta(A, p) \leq np/2$ and that $\Delta(A, p) = np/2$ if and only if A consists of a single point (we agreed that $p > 0$).

As was the case with $\Delta(A)$, the functional $\Delta(A, p)$ can be easily computed by averaging. For a set $A \subset C_n$ defined by its Distance Oracle 2.2 and any $l = (\lambda_1, \dots, \lambda_n)$ the value of $d_l(x, A)$ is computed by choosing the penalties $d_i(0, 1) = d_i(1, 0) = 1$ when $\lambda_i = 1$ and $d_i = 0$ when $\lambda_i = 0$.

(4.3) Algorithm for Computing $\Delta(A, p)$

Input: A set $A \subset C_n$ given by its Distance Oracle 2.2, a number $1 \geq p > 0$ and an $\epsilon > 0$.

Output: A number α approximating $\Delta(A, p)$ within error ϵ.

Algorithm: Let $k = \lceil 3n/\epsilon^2 \rceil$. Sample k points $x_1, \dots, x_k \in C_n$ independently at random from the uniform distribution in C_n and k points $l_1, \dots, l_k \in \Lambda_n$ independently at random from the distribution in Λ_n. Apply Distance Oracle 2.2 to compute $d_{l_i}(x_i, A)$, $i = 1, \dots, k$. Compute $\alpha = \frac{1}{k} \sum_{i=1}^{k} \operatorname{dist}_{l_i}(x_i, A)$. Output α.

(4.4) Theorem. *With probability at least 0.9, the output α of Algorithm 4.3 satisfies the inequality $|\Delta(A, p) - \alpha| \leq \epsilon$.*

We postpone the proof till Section 6.

We are going to obtain estimates of the cardinality $|A|$ of a set $A \subset C_n$ in terms of the quantity $\Delta(A, p)$. As in Section 3, it is convenient to work with a related quantity

$$\rho = \rho(A, p) = \frac{p}{2} - \frac{\Delta(A, p)}{n}.$$

From Definitions 4.1, for any non-empty $A \subset C_n$, the function $\rho(A, p)$ is a polynomial in p of degree at most n. As follows from Example 4.2, $0 \leq \rho \leq p/2$ for any non-empty set $A \subset C_n$. Our estimate will be useful for "small" sets A where $n^{-1} \ln |A|$ is close to 0.

(4.5) Theorem. *Let $A \subset C_n$ be a non-empty set. Let*

$$\rho = \frac{p}{2} - \frac{\Delta(A,p)}{n}.$$

Then

$$\frac{\rho^2}{p} \leq \frac{\ln|A|}{n}. \tag{4.5.1}$$

Suppose that $\rho \leq 1/4$ and that

$$p \geq \frac{\ln 2 + \ln(1-2\rho)}{\ln(1-2\rho) - \ln(2\rho)}. \tag{4.5.2}$$

Then

$$\frac{\ln|A|}{n} \leq 2\rho \ln\frac{1}{2\rho} + (1-2\rho)\ln\frac{1}{1-2\rho}. \tag{4.5.3}$$

We obtain the following counterpart of Corollary 3.11.

(4.6) Corollary. *Let us choose any $c_3 < 1/(\ln 2) \approx 1.44$ and any $c_4 > 2$. Then there exists a $\delta > 0$ such that for any non-empty $A \subset C_n$ with $n^{-1}\ln|A| \leq \delta$ there exists a $0 < p \leq 1$ such that for $\rho = \dfrac{p}{2} - \dfrac{\Delta(A,p)}{n}$ one has*

$$c_3 \cdot \rho^2 \ln\frac{1}{\rho} \leq \frac{\ln|A|}{n} \leq c_4 \cdot \rho \ln\frac{1}{\rho}.$$

Proof. By (4.5.1), we have $\rho \leq \sqrt{n^{-1}\ln|A|} \leq \sqrt{\delta}$, so $\rho(A,p)$ is small if δ is small, no matter what p is. We observe that for small positive ρ the right hand side of (4.5.2) is of the order $(\ln 2)\ln^{-1}(1/\rho)$ and the right hand side of (4.5.3) is of the order $2\rho\ln(1/\rho)$.

Given $c_3 < (\ln 2)^{-1}$ and $c_4 > 2$, let us choose $1/16 > \delta > 0$ in such a way that the right hand side of (4.5.2) does not exceed $(c_3)^{-1}\ln^{-1}(1/\rho)$ and the right hand side of (4.5.3) does not exceed $c_4\rho\ln(1/\rho)$ for all $0 < \rho < \sqrt{\delta} < 1/4$.

We recall that $|A| = 1$ if and only if $\rho = 0$, in which case the bounds of Corollary 4.6 are satisfied by default. If $|A| > 1$, one can see (cf. also Section 4.7 below) that $\rho(A,p) = \Omega(p/n)$. Given a set $A \subset C_n$, $|A| > 1$, let us choose the smallest $p \geq 0$ that satisfies the inequality (4.5.2). Then $0 < p < 1$ since the right hand side of (4.5.2) is $\Omega\bigl(\ln^{-1}(n/p)\bigr)$ (and therefore goes to 0 slower than p), and smaller than 1 for $0 < \rho < 1/4$. Since $\rho(A,p)$ depends continuously on p, we must have equality in (4.5.2) (otherwise, we could have taken a smaller p). Thus $p \leq (c_3)^{-1}\ln^{-1}(1/\rho)$ and the proof follows by (4.5.1)–(4.5.3). □

(4.7) Extremal sets. Let us fix a $0 < p \leq 1$ and an $\epsilon > 0$. Then there exists an $\alpha = \alpha(p, \epsilon) > 0$ with the following property: if $A \subset C_n$ is a set of $\lfloor 2^{\alpha n} \rfloor$ points randomly chosen from the Boolean cube, then with the probability that tends to 1 as n grows to infinity, $n^{-1} \ln |A| < (2+\epsilon)\rho^2/p$. Hence for any $p > 0$ the bound (4.5.1) is tight up to a constant factor for sufficiently small random sets. The proof is rather technical and therefore omitted.

One can show that the bound (4.5.3) is asymptotically tight on small *faces* of the cube C_n. More precisely, let us fix a $\delta > 0$ (to be adjusted later), let $m = \lfloor \delta n \rfloor$ and let $A \subset C_n$ be an m-dimensional face of the Boolean cube:

$$A = \Big\{ (\xi_1, \ldots, \xi_n) : \xi_i = 0 \quad \text{for} \quad i = m+1, \ldots, n \Big\}.$$

Thus $|A| = 2^m$. Moreover, a computation similar to that of Example 4.2 shows that $\rho(A, p) = pm/2n$. Hence we have

$$\frac{\ln |A|}{n} = \frac{2 \ln 2}{p} \rho(A, p).$$

We observe that $\rho(A, p) \leq \delta/2$. Hence for any small $\epsilon > 0$ one can find $\delta = \delta(\epsilon) > 0$ such that there exists p satisfying (4.5.2) and such that $p < (1+\epsilon)(\ln 2) \ln^{-1}(1/\rho)$. For such a p, we have

$$\frac{\ln |A|}{n} \geq \frac{2}{1+\epsilon} \rho \ln \frac{1}{\rho},$$

so the bound (4.5.3) is indeed asymptotically tight for small sets.

Apparently, the sets A having the largest cardinality among all sets with the given value of $\rho(A, p)$ evolve from the balls in the Hamming metric for $p = 1$ (see Section 3.10) to faces at $p \longrightarrow 0$. Since faces are packed somewhat less tightly than balls, we gain in Corollary 4.6 as compared to Corollary 3.11.

The proof of Theorem 4.5 is postponed till Section 6.

(4.8) Discussion. Corollary 4.6 implies that for small sets A by "tuning up" p we can get an additional logarithmic factor which brings the lower bound for $n^{-1} \ln |A|$ a little closer to the upper bound compared to the bound of Corollary 3.11. Any p which is only slightly bigger than the bound (4.5.2) will do. Suppose, for example, that $A \subset C_n$ is a set such that $n^{-1} \ln |A| \sim n^{-\alpha}$ for some $0 < \alpha < 1$. By applying Algorithm 3.3 to approximate $\Delta(A) = \Delta(A, 1)$ and Theorem 3.9 to interpret the results, the worst lower bound we can get for $n^{-1} \ln |A|$ is $\sim n^{-2\alpha}/\ln^2 n$ (this happens when A is a ball in the Hamming metric, but we think it is a "random set", see Section 3.10) and the worst upper bound we can get is $\sim n^{-\alpha/2} \ln n$ (this happens when A is a "random set" but we think that it is a ball). Now, by (4.5.1) it follows that $\rho(A, p) = O(n^{-\alpha/2})$ for any p. Then we can choose some $p = O(\ln^{-1} n)$ that satisfies (4.5.2). Applying Algorithm 4.3 to approximate $\Delta(A, p)$ and Theorem 4.5 to interpret the results, for $n^{-1} \ln |A|$ we would obtain a lower bound of the form

$\sim n^{-2\alpha}/\ln n$ at worst (this happens when A is a face but we think it is a random set) and an upper bound of the form $\sim n^{-\alpha/2}\sqrt{\ln n}$ at worst (this happens when A is a random set but we think it is a face).

To find a particular suitable p for a small set A, we note that if $n^{-1}\ln|A| \leq \delta$ for some $\delta < 1/4$, the value of p obtained by substituting $\rho = \sqrt{\delta}$ into the right hand side of (4.5.2) would satisfy the inequality since the right hand side is an increasing function of ρ and since $\rho(A,p) \leq \sqrt{\delta}$ for any p by (4.5.1). It is interesting that an improvement in the cardinality estimate can be achieved by simply ignoring a (random) part of the information contained in the standard Hamming distance.

5. Corollaries, Remarks and Possible Ramifications

(5.1) Testing emerging exponential growth. As an illustration, we show that our approach allows one to test in polynomial time whether the number of perfect matchings (Example 1.2) is exponentially large in the number of vertices of the graph (in the sense defined below). Let $G = (V, E)$ be a graph with $|V| = n$ vertices and let $\mathcal{F}$ be the set of all perfect matchings in G. Economical embedding (2.4) allows us to identify $\mathcal{F}$ with a subset F of the Boolean cube $\{0,1\}^m$ with $m \leq n(\log_2 n + 1)$ and to construct efficiently Distance Oracle 2.2 for F. Suppose that $|\mathcal{F}| \leq \exp\{n^\alpha\}$ for some $\alpha < 1$. Corollary 3.11 implies that $\rho(F) = O\big(n^{\alpha/2}m^{-1/2}\big)$. Thus using Algorithm 3.3 to compute $\Delta(F)$ within error $\epsilon = \sqrt{n}$ and then Corollary 3.11 to interpret the result, we will be able to conclude that $|\mathcal{F}| = O(\exp\{n^\beta\})$ for any $\beta > (1+\alpha)/2$. Similarly, if $|\mathcal{F}| \geq \exp\{n^\alpha\}$ for some $\alpha > 0$, our method would allow us to conclude for any $\beta < 2\alpha - 1$ that $|\mathcal{F}| = \Omega\big(\exp\{n^\beta\}\big)$. The estimate is, of course, void for $\alpha < 1/2$ but it improves as α approaches 1. For example, if $|\mathcal{F}|$ has the order of $\exp\{n^{0.95}\}$, our method would allow us to conclude that $|\mathcal{F}|$ is greater than $\exp\{n^{0.89}\}$ and smaller than $\exp\{n^{0.98}\}$. Thus we can tell the order $|\mathcal{F}| \approx \exp\{n^{0.99}\}$ from the order $|\mathcal{F}| \approx \exp\{n\}$ and to distinguish them we have to solve the minimum weight matching problem a constant number of times. Bounds of this type for perfect matchings in general graphs are new.

Similarly, we can test whether the number of colored spanning subgraphs is exponentially large in the number r of colors (see Example 1.4).

Implementing our approach, Ryckman [Ryckman 00] wrote an experimental C++ code to estimate the number of perfect matchings in a given bipartite graph (or, equivalently, to estimate the permanent of a 0-1 matrix). In the case of a bipartite graph, Optimization Oracle 1.1 is especially easy to construct. In this case, the corresponding problem, known as the *Assignment Problem*, is not only "theoretically easy", but in practice large instances are routinely solved as particular cases of the minimum cost network flow problem, see for example Section 11.2 of [Papadimitriou and Steiglitz 98]. Theoretically, the algorithm can not compete in precision with the recent polynomial time approximation scheme of [Jerrum *et al.* 00]. However, in practice the code appears to be working extremely fast on fairly large graphs (it was tested on graphs with up to 256 vertices) and produces estimates which, although crude, are often non-trivial. Generally speaking, we think that our approach can be useful for problems of large size where some fairly crude estimates of the cardinality are needed and where the underlying optimization problem is especially easy (as in Example 1.3).

(5.2) Connections to Monte Carlo methods. The main idea of our approach can be described as follows: given a (finite) ambient space Ω and a set $A \subset \Omega$, we estimate the cardinality $|A|$ by choosing a certain distance function d in Ω and estimating the average distance

$$\Delta(A) = \frac{1}{|\Omega|} \sum_{x \in \Omega} d(x, A), \quad \text{where} \quad d(x, A) = \min_{y \in A} d(x, y)$$

from $x \in \Omega$ to A. We get the classical Monte Carlo method if the distance function d is the simplest possible:

$$d(x, y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } \ x = y. \end{cases}$$

In this case, $\Delta(A) = 1 - |A|/|\Omega|$, so there is a direct relation between $\Delta(A)$ and $|A|$. It is well understood that the main difficulty with the Monte Carlo method is that if $|A|$ is "exponentially small" compared to $|\Omega|$ then to get a non-trivial bound for $|A|$, we have to compute $\Delta(A)$ with exponentially high precision. In this paper, we showed that in many interesting cases one can choose a different distance function d, so that the distance $d(x, A)$ from a point $x \in \Omega$ to A is efficiently computable and to get a meaningful estimate of $|A|$ even for exponentially small sets A, one needs to compute $\Delta(A)$ with a polynomial precision. Hence our approach can be considered as a natural extension of the Monte Carlo method.

In this paper, we choose Ω to be the Boolean cube endowed either with the standard Hamming distance (Section 3) or with its randomized version (Section 4). In many cases, other embeddings might be of interest. In Example 1.3, all subsets $Y \in \mathcal{F}$ have the same cardinality k. Straightforward embedding 2.3 identifies $\mathcal{F}$ with a subset F of the section

$$\Omega = \Big\{(\xi_1, \ldots, \xi_n) \in \{0,1\}^n : \sum_{i=1}^{n} \xi_i = k\Big\}$$

of the Boolean cube $\{0,1\}^n$. Tighter estimates for $|F|$ can be obtained by considering F to be a subset of Ω and not of the whole cube $\{0,1\}^n$.

In Example 1.4, the set $\mathcal{F}$ of properly colored spanning subgraphs can be naturally identified with a subset F of the direct product $\Omega = E_1 \times \ldots \times E_r$. Let d_i be a distance function on E_i for $i = 1, \ldots, r$ and let d be the corresponding L^1 distance on Ω. One can show that Distance Oracle 2.2 can be efficiently constructed for any choice of d_i. How should we choose d_i to get the best possible estimates for the cardinality $|F|$? Note, that looking for such d_i, we are trying to satisfy two competing requirements: neighborhoods of "small" sets $F \subset \Omega$ should be as small as possible while neighborhoods of "large" sets $F \subset \Omega$ should be as large as possible. Perhaps one should use a whole family of distance functions d_i and combine the resulting estimates. The general isoperimetric inequality of [Alon *et al.* 98] can be very useful for investigating that. Note, that economical embedding 2.4 results in choosing d_i so that E_i becomes isometric to a subset of the Boolean cube $\{0,1\}^{m_i}$ with $m_i = \lceil \log_2 |E_i| \rceil$.

(5.3) Weighted counting. Let $\mathcal{F} \subset 2^X$ be a family of subsets of the ground set $X = \{1, \ldots, n\}$ and let $\mu(i) = p_i/q_i > 0$ be a rational weight of $i \in X$, where $p_i, q_i \in \mathbb{N}$. Let us define

$$\mu(Y) = \prod_{i \in Y} \mu(i) \quad \text{for} \quad Y \in \mathcal{F} \qquad \text{and} \qquad \mu(\mathcal{F}) = \sum_{Y \in \mathcal{F}} \mu(Y).$$

We may be interested to estimate $\mu(\mathcal{F})$. There are several ways to extend our methods to problems of this type, here we sketch one. For every $i \in X$, let $m_i = \lceil \log_2(p_i + q_i) \rceil$. Let us choose subsets $A_i \subset C_{m_i}$ and $B_i \subset C_{m_i}$ such that $|A_i| = p_i$, $|B_i| = q_i$ and $A_i \cap B_i = \emptyset$. Let $m = m_1 + \ldots + m_n$ and let us identify

$$C_m = C_{m_1} \times \ldots \times C_{m_n}.$$

For $Y \subset \mathcal{F}$ let $Z_Y \subset C_m$ be the direct product of n factors, the i-th factor being A_i if $i \in Y$ and B_i if $i \notin Y$. Finally, let $F \subset C_m$ be the union of all Z_Y for $Y \in \mathcal{F}$. We see that $\mu(\mathcal{F}) = (q_1 \cdots q_n)^{-1} |F|$. Moreover, one can define subsets A_i and B_i in such a way that Optimization Oracle 1.1 for $\mathcal{F}$ gives rise to Distance Oracle 2.2 for F. This construction corresponds to the straightforward embedding (2.3). In some cases, there is a way to come up with an economical embedding in the spirit of (2.4).

6. Proofs of Theorems 4.4 and 4.5

(6.1) Definition. We recall that C_N is the Boolean cube $\{0,1\}^N$ endowed with the uniform probability measure and that Λ_N is the Boolean cube $\{0,1\}^N$ endowed with the probability measure of Definition 4.1. Let $\Omega_N = C_N \times \Lambda_N$. We consider the product measure on Ω_N, so

$$\mathbf{P}\ \{(x,l)\} = p^{|l|} q^{N-|l|} 2^{-N}, \quad \text{where} \quad |l| = \lambda_1 + \ldots + \lambda_N \quad \text{for} \quad l = (\lambda_1, \ldots, \lambda_N).$$

Hence a point $(x,l) \in \Omega_N$ is interpreted as a vector of $2N$ independent random variables $(\xi_1, \ldots, \xi_N; \lambda_1, \ldots, \lambda_N)$, where $\mathbf{P}\ \{\xi_i = 0\} = \mathbf{P}\ \{\xi_i = 1\} = 1/2$, $\mathbf{P}\ \{\lambda_i = 1\} = p$ and $\mathbf{P}\ \{\lambda_i = 0\} = q$. We observe that

$$\Delta(A,p) = \mathbf{E}\ d_l(x,A). \tag{6.1.1}$$

(6.2) Lemma. *Let $A \subset C_N$ be a set. Then for every $\delta \geq 0$*

$$\mathbf{P}\ \Big\{(x,l) \in \Omega_N : |d_l(x,A) - \Delta(A,p)| \geq \delta\Big\} \leq 2e^{-\delta^2/N}.$$

Proof. Since $d_l(x,A)$ is a function of $2N$ independent random variables, the proof follows by Lemma 3.4. □

Next, we need an analogue of the scaling trick (3.5).

(6.3) Lemma. *Let us fix positive integers k and n and let $N = kn$. Let us identify $C_N = (C_n)^k$, $\Lambda_N = (\Lambda_n)^k$ and $\Omega_N = (\Omega_n)^k$. Thus a point $(x,l) \in \Omega_N$ is identified with $x = (x_1, \ldots, x_k; l_1, \ldots, l_k)$, where $x_i \in C_n$ and $l_i \in \Lambda_n$.*

For a subset $A \subset C_n$, let $B = A^k \subset C_N$. Then

$$d_l(x,B) = \sum_{i=1}^k d_{l_i}(x_i, A) \quad \text{and} \quad \Delta(B,p) = k\Delta(A,p).$$

Proof. Clearly,

$$d_l(x,y) = \sum_{i=1}^k d_{l_i}(x_i, y_i) \quad \text{for all} \quad x,y \in C_N$$

and the first identity follows. Now, by (6.1.1)

$$\Delta(B,p) = \mathbf{E}\ d_l(x,B) = \sum_{i=1}^k \mathbf{E}\ d_{l_i}(x_i, A) = k\Delta(A,p).$$

□

Now we are ready to prove Theorem 4.4.

Proof of Theorem 4.4. Let $N = nk$ and let us identify $C_N = (C_n)^k$, $\Lambda_N = (\Lambda_n)^k$ and $\Omega_N = (\Omega_n)^k$. Let $B = A^k \subset C_N$ as in Lemma 6.3. Applying Lemma 6.2, we get

$$\mathbf{P}\ \Big\{(x,l) \in \Omega_N : |d_l(x,B) - \Delta(B,p)| \geq \delta\Big\} \leq 2e^{-\delta^2/N}$$

for any $\delta \geq 0$. Using Lemma 6.3, we conclude:

$$\mathbf{P}\ \left\{(x,l) \in \Omega_N : \left|\frac{1}{k}\sum_{i=1}^k d_{l_i}(x_i, A) - \Delta(A,p)\right| \geq \delta/k\right\} \leq 2e^{-\delta^2/N}.$$

Let us choose $\delta = \epsilon k$. Hence

$$\mathbf{P}\ \left\{(x,l) \in \Omega_N : \left|\frac{1}{k}\sum_{i=1}^k d_{l_i}(x_i, A) - \Delta(A,p)\right| \geq \epsilon\right\} \leq 2e^{-\epsilon^2 k/n}.$$

Since $k \geq 3n/\epsilon^2$, the proof follows. □

Next, we need a (crude) version of inequality (3.7.1).

(6.4) Lemma. *Let $\epsilon \geq 0$, let $r(\epsilon) = pN(1-\epsilon)/2$. Let $y \in C_N$ be a point. Then*

$$\mathbf{P}\ \Big\{(x,l) \in \Omega_N : d_l(x,y) \leq r(\epsilon)\Big\} \leq e^{-\epsilon^2 pN/4}.$$

Proof. Without loss of generality we may assume that $y = 0$. Then

$$\mathbf{P}\ \Big\{(x,l) \in \Omega_N : d_l(x,0) \leq r(\epsilon)\Big\} = \mathbf{P}\ \Big\{(x,l) \in \Omega_N : \sum_{i=1}^{N} \xi_i \lambda_i \leq r(\epsilon)\Big\},$$

where $x = (\xi_1, \ldots, \xi_N)$ and $l = (\lambda_1, \ldots, \lambda_N)$. Let $\zeta_i = \xi_i \lambda_i$. Then ζ_i, $i = 1, \ldots, N$ are independent random variables such that $\mathbf{P}\ \{\zeta_i = 1\} = p/2$ and $\mathbf{P}\ \{\zeta_i = 0\} = 1 - p/2$. Hence

$$\mathbf{P}\ \Big\{(x,l) \in \Omega_N : d_l(x,y) \leq r(\epsilon)\Big\} = \mathbf{P}\ \Big\{\zeta_1 + \ldots + \zeta_N \leq r(\epsilon)\Big\} \leq e^{-\epsilon^2 pN/4}$$

by a corollary of Hoeffding's inequality (see Corollary 5.6 of [McDiarmid 89]). $\square$

Now we are ready to prove the first part of Theorem 4.5.

Proof of inequality (4.5.1). Let us choose a positive integer m, let $N = mn$, let $C_N = (C_n)^m$, and let $\Lambda_N = (\Lambda_n)^m$. Let $B = A^m \subset C_N$ as in Lemma 6.3.

Let us choose an $\alpha > 0$. Applying Lemma 6.4, we obtain

$$\mathbf{P}\ \Big\{(x,l) \in \Omega_N : d_l(x,B) \leq pN(1-\sqrt{\alpha})/2\Big\} \leq |B| e^{-\alpha pN/4} = (|A| e^{-\alpha pn/4})^m.$$

Therefore, by Lemma 6.3

$$\mathbf{P}\ \Big\{(x,l) \in \Omega_N : \frac{1}{m}\sum_{i=1}^{m} d_{l_i}(x_i, A) \leq pn(1-\sqrt{\alpha})/2\Big\} \leq (|A| e^{-\alpha pn/4})^m.$$

The right hand side of the inequality tends to 0 provided $\alpha > 4\ln|A|/pn$. Since by the Law of Large Numbers

$$\frac{1}{m}\sum_{i=1}^{m} d_{l_i}(x_i, A) \longrightarrow \Delta(A,p) \quad \text{in probability} \quad \text{as} \quad m \longrightarrow +\infty,$$

we must have

$$\Delta(A,p) \geq pn(1-\sqrt{\alpha})/2 \quad \text{for any} \quad \alpha > 4\ln|A|/pn.$$

Hence

$$\Delta(A,p) \geq pn(1-\sqrt{\alpha})/2 \quad \text{for} \quad \alpha = 4\ln|A|/pn,$$

which is equivalent to (4.5.1). $\square$

In Section 3, we used the sharp isoperimetric inequality (Theorem 3.8) for the Hamming distance in C_n to get a sharp upper bound for $n^{-1}\log_2|A|$. Unfortunately, we don't know of a similar result for the randomized Hamming distance. To prove (4.5.2)–(4.5.3), we proceed by induction on n in a way resembling that of [Talagrand 95] (see also Remark 6.9).

We start with a simple technical result.

(6.5) Lemma. *For any $0 \leq \epsilon \leq 1$, any $\gamma \geq 0$ and any $0 < p \leq 1$ and $q = 1 - p$ we have*

$$\min\Bigl\{\frac{p\gamma}{2} + \ln\frac{1}{1+\epsilon}, \quad p\ln\frac{1}{1-\epsilon} + q\ln\frac{1}{1+\epsilon}\Bigr\} \leq \max\Bigl\{0, \quad \ln(1+e^{\gamma/2}) - \frac{q\gamma}{2} - \ln 2\Bigr\}.$$

Proof. Fixing p, q and γ, let

$$f(\epsilon) = \frac{p\gamma}{2} + \ln\frac{1}{1+\epsilon} \quad \text{and} \quad g(\epsilon) = p\ln\frac{1}{1-\epsilon} + q\ln\frac{1}{1+\epsilon}.$$

Then $f(0) \geq 0$ and $f(\epsilon)$ is decreasing whereas $g(\epsilon)$ behaves as follows: $g(0) = 0$ and if $p \geq q$ then $g(\epsilon)$ is increasing and if $p < q$ then $g(\epsilon)$ is decreasing for $0 < \epsilon < q - p$ and increasing for $q - p < \epsilon < 1$. Furthermore, $f(\epsilon_0) = g(\epsilon_0)$ at the single point $\epsilon_0 = (e^{\gamma/2} - 1)/(1 + e^{\gamma/2})$, where $f(\epsilon_0) = g(\epsilon_0) = \ln(1 + e^{\gamma/2}) - q\gamma/2 - \ln 2$. The proof now follows. □

(6.6) Definition. Let μ_n (or simply μ) denote the uniform probability measure in C_n. Hence $\mu(A) = |A|/2^n$.

The induction is based on the following lemma.

(6.7) Lemma. *Let $A \subset C_{n+1}$ be a set. Let*

$$A_0 = \bigl\{x \in C_n : (x, 0) \in A\bigr\} \quad \textit{and} \quad A_1 = \bigl\{x \in C_n : (x, 1) \in A\bigr\}.$$

For $l \in \Lambda_n$ let $(l, 0) \in \Lambda_{n+1}$ denote l appended by $\lambda_{n+1} = 0$ and let $(l, 1) \in \Lambda_{n+1}$ denote l appended by $\lambda_{n+1} = 1$. Let

$$\Delta_0(A, p) = \mathbf{E}\ d_{(l,0)}(x, A) \quad \textit{and} \quad \Delta_1(A, p) = \mathbf{E}\ d_{(l,1)}(x, A),$$

where the expectation is taken with respect to a random $(x, l) \in C_{n+1} \times \Lambda_n$. Then

$$\frac{\mu_n(A_0) + \mu_n(A_1)}{2} = \mu_{n+1}(A); \tag{6.7.1}$$

$$\Delta(A, p) = q\Delta_0(A, p) + p\Delta_1(A, p); \tag{6.7.2}$$

$$\Delta_0(A, p) \leq \Delta(A_i, p) \quad \textit{for} \quad i = 0, 1; \tag{6.7.3}$$

$$\Delta_1(A, p) \leq \Delta(A_i, p) + \frac{1}{2} \quad \textit{for} \quad i = 0, 1; \tag{6.7.4}$$

$$\Delta_1(A, p) \leq \frac{\Delta(A_0, p) + \Delta(A_1, p)}{2}. \tag{6.7.5}$$

Proof. The proof is straightforward, cf. also proof of Lemma 2.1.2 of [Talagrand 95] and proof of Lemma 2.5 of [Barvinok 97]. □

Now we use induction to get a preliminary bound.

(6.8) Lemma. *Suppose that for some* $\gamma \geq 0$, $0 < p \leq 1$ *and* $q = 1 - p$,

$$\ln\bigl(1 + e^{\gamma/2}\bigr) - \frac{q\gamma}{2} - \ln 2 \geq 0.$$

Then for any non-empty set $A \subset C_n$ *we have*

$$\gamma\Delta(A, p) + \ln \mu(A) \leq n\Bigl(\ln\bigl(1 + e^{\gamma/2}\bigr) - \frac{q\gamma}{2} - \ln 2\Bigr).$$

Proof. We proceed by induction on n. If $n = 1$ then the two cases are possible:

A consists of a single point, $\mu(A) = 1/2$ and $\Delta(A, p) = p/2$ (see Example 4.2);

$A = \{0, 1\}$, $\mu(A) = 1$ and $\Delta(A, p) = 0$.

In either case, the inequality holds.

Suppose that the inequality holds for non-empty subsets of C_n. Let us prove that it holds for non-empty $A \subset C_{n+1}$. Let us define $A_0, A_1 \subset C_n$ as in Lemma 6.7. From (6.7.1) it follows that either

$$\mu_n(A_0) = (1 - \epsilon)\mu_{n+1}(A) \quad \text{and} \quad \mu_n(A_1) = (1 + \epsilon)\mu_{n+1}(A)$$

or

$$\mu_n(A_1) = (1 - \epsilon)\mu_{n+1}(A) \quad \text{and} \quad \mu_n(A_0) = (1 + \epsilon)\mu_{n+1}(A)$$

for some $0 \leq \epsilon \leq 1$.

Let B be the one of the sets A_0, A_1 that has a bigger measure μ_n (either of the two if $\mu_n(A_0) = \mu_n(A_1)$) and let D be the one of the sets A_0, A_1 that has a bigger value of $\Delta(\cdot, p)$ (either of the two if $\Delta(A_0, p) = \Delta(A_1, p)$). Then

$$\mu_n(B) \geq (1 + \epsilon)\mu_{n+1}(A) \quad \text{and} \quad \mu_n(D) \geq (1 - \epsilon)\mu_{n+1}(A).$$

Furthermore, by (6.7.3)

$$\Delta_0(A, p) \leq \Delta(B, p) \quad \text{and} \quad \Delta_0(A, p) \leq \Delta(D, p)$$

whereas by (6.7.3) and (6.7.5)

$$\Delta_1(A, p) \leq \Delta(B, p) + \frac{1}{2} \quad \text{and} \quad \Delta_1(A, p) \leq \Delta(D, p).$$

Hence we get

$$\gamma\Delta_0(A, p) + \ln \mu_{n+1}(A) \leq \gamma\Delta(B, p) + \ln \mu_n(B) + \ln \frac{1}{1 + \epsilon}$$

and

$$\gamma\Delta_1(A,p)+\ln\mu_{n+1}(A)\leq$$
$$\min\Big\{\gamma\Delta(B,p)+\ln\mu_n(B)+\ln\frac{1}{1+\epsilon}+\frac{\gamma}{2},\quad \gamma\Delta(D,p)+\ln\mu_n(D)+\ln\frac{1}{1-\epsilon}\Big\}.$$

Clearly, B is non-empty. Assume first, that D is non-empty as well. Applying the induction hypothesis to B and D, we conclude that

$$\gamma\Delta_0(A,p)+\ln\mu_{n+1}(A)\leq n\Big(\ln\big(1+e^{\gamma/2}\big)-\frac{q\gamma}{2}-\ln 2\Big)+\ln\frac{1}{1+\epsilon}$$

and

$$\gamma\Delta_1(A,p)+\ln\mu_{n+1}(A)\leq n\Big(\ln\big(1+e^{\gamma/2}\big)-\frac{q\gamma}{2}-\ln 2\Big)+\min\Big\{\ln\frac{1}{1+\epsilon}+\frac{\gamma}{2},\quad \ln\frac{1}{1-\epsilon}\Big\}.$$

Adding the first inequality multiplied by q and the second inequality multiplied by p and using (6.7.2), we get

$$\gamma\Delta(A,p)+\ln\mu_{n+1}(A)\leq$$
$$n\Big(\ln\big(1+e^{\gamma/2}\big)-\frac{q\gamma}{2}-\ln 2\Big)+\min\Big\{\frac{p\gamma}{2}+\ln\frac{1}{1+\epsilon},\quad p\ln\frac{1}{1-\epsilon}+q\ln\frac{1}{1+\epsilon}\Big\}.$$

The desired inequality follows by Lemma 6.4.

If D is empty then $\mu_n(B)=2\mu_{n+1}(A)$ and we obtain

$$\gamma\Delta_0(A,p)+\ln\mu_{n+1}(A)\leq\gamma\Delta(B,p)+\ln\mu_n(B)-\ln 2$$

and

$$\gamma\Delta_1(A,p)+\ln\mu_{n+1}(A)\leq\gamma\Delta(B,p)+\ln\mu_n(B)-\ln 2+\frac{\gamma}{2}$$

Adding the first inequality multiplied by q to the second inequality multiplied by p and using (6.7.2) and the induction hypothesis, we get:

$$\begin{aligned}\gamma\Delta(A,p)+\ln\mu_{n+1}(A)&\leq\gamma\Delta(B,p)+\ln\mu_n(B)-\ln 2+\frac{p\gamma}{2}\\&\leq n\Big(\ln\big(1+e^{\gamma/2}\big)-\frac{q\gamma}{2}-\ln 2\Big)+\Big(\frac{\gamma}{2}-\frac{q\gamma}{2}-\ln 2\Big)\\&\leq(n+1)\Big(\ln\big(1+e^{\gamma/2}\big)-\frac{q\gamma}{2}-\ln 2\Big),\end{aligned}$$

which completes the proof. □

Now we are ready to complete the proof of Theorem 4.5.

Proof of (4.5.2)–(4.5.3). By Lemma 6.8,

$$\frac{\ln|A|}{n}=\frac{\ln\mu_n(A)}{n}+\ln 2\leq\ln(1+e^{\gamma/2})-\frac{q\gamma}{2}-\frac{\gamma\Delta(A,p)}{n}=\ln(1+e^{\gamma/2})-\frac{\gamma}{2}+\gamma\rho$$

provided

$$\ln\bigl(1+e^{\gamma/2}\bigr)-\frac{q\gamma}{2}-\ln 2\geq 0.$$

We optimize the inequality on $\gamma\geq 0$. Let

$$\gamma=2\ln\Bigl(\frac{1}{2\rho}-1\Bigr).$$

Since we assumed that $\rho\leq 1/4$, we have $\gamma\geq 0$. Furthermore,

$$\begin{aligned}\ln\bigl(1+e^{\gamma/2}\bigr)-\frac{q\gamma}{2}-\ln 2&=\ln\frac{1}{2\rho}-q\ln\Bigl(\frac{1}{2\rho}-1\Bigr)-\ln 2\\ &=-\ln(1-2\rho)+p\bigl(\ln(1-2\rho)-\ln(2\rho)\bigr)-\ln 2\geq 0,\end{aligned}$$

because of (4.5.2). Therefore,

$$\frac{\ln|A|}{n}\leq\ln\frac{1}{2\rho}-\ln\frac{1-2\rho}{2\rho}+2\rho\ln\frac{1-2\rho}{2\rho}=2\rho\ln\frac{1}{2\rho}+(1-2\rho)\ln\frac{1}{1-2\rho}$$

and (4.5.3) follows. □

(6.9) Remark. Our proof of (4.5.2)–(4.5.3) can be considered as an "additive" version of Talagrand's method [Talagrand 95]. Indeed, Talagrand's approach very roughly can be can be stated as follows. Let Ω be a space with the distance function d and probability measure μ. To prove an isoperimetric inequality for $A\subset\Omega$, we first find a uniform bound for the expression $\mu^{\alpha}(A)\cdot\mathbf{E}\ \exp\{\tau d(x,A)\}$ and then adjust parameters $\alpha>0$ and $\tau>0$. This way tight inequalities are obtained in [Talagrand 95] for sets A of large measure, most often with $\mu(A)\geq 1/2$. We are mostly interested in sets of a small measure. One can check that for "small sets" A the inequalities of [Talagrand 95] are very far from sharp, which is, of course, should not be perceived as a "fault" of the method, since the method was designed for totally different problems. We find a uniform bound for the expression $\ln\mu(A)+\tau\mathbf{E}\ d(x,A)$, which looks like Talagrand' functional with "exp" removed. Our method seems to produce reasonably good bounds for small sets A but it fails miserably for large A, with $\mu(A)=1/2$, say. As should have been expected, the case of "middle-sized" sets is the most complicated.

Acknowledgment

The authors are grateful to M. Gromov, B. Sudakov and G. Kalai for many helpful discussions and to the anonymous referee for suggestions.

References

[Alon *et al.* 98] N. Alon, R. Boppana and J. Spencer, An asymptotic isoperimetric inequality, *Geom. Funct. Anal.* **8** (1998), 411–436.

[Barvinok 97] A. Barvinok, Approximate counting via random optimization, *Random Structures & Algorithms*, **11**(1997), no. 2, 187–198.

[Grötschel *et al.* 93] M. Grötschel, L. Lovász and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, second ed., Algorithms and Combinatorics, **2**, Springer-Verlag, Berlin, 1993.

[Jerrum 95] M. Jerrum, The computational complexity of counting, *Proceedings of the International Congress of Mathematicians*, Vol. 1,2 (Zürich, 1994), 1407–1416, Birkhäuser, Basel, 1995.

[Jerrum and Sinclair 89] M. Jerrum and A. Sinclair, Approximating the permanent, *SIAM J. Comput.*, **18** (1989), no. 6, 1149–1178.

[Jerrum and Sinclair 97] M. Jerrum and A. Sinclair, The Markov chain Monte Carlo Method: an approach to approximate counting and integration, in: *Approximation Algorithms for NP-hard Problems*, ed. D.S. Hochbaum, PWS, Boston, 1997, 483–520.

[Jerrum *et al.* 00] M. Jerrum, A. Sinclair and E. Vigoda, A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries, *Electronic Colloquium on Computational Complexity*, Report TR00-079, ftp://ftp.eccc.uni-trier.de/pub/eccc/reports/2000/TR00-079/index.html

[Jerrum *et al.* 86] M. Jerrum, L.G. Valiant and V.V. Vazirani, Random generation of combinatorial structures from a uniform distribution, *Theoret. Comput. Sci.*, **43** (1986), no. 2-3, 169–188.

[Leader 91] I. Leader, Discrete isoperimetric inequalities, in: *Probabilistic Combinatorics and its Applications (San Francisco, CA, 1991)*, 57–80, Proc. Sympos. Appl. Math., **44**, Amer. Math. Soc., Providence, RI, 1991.

[van Lint 99] J.H. van Lint, *Introduction to Coding Theory*, Third edition. Graduate Texts in Mathematics, **86**, Springer-Verlag, Berlin, 1999.

[Lovász and Plummer 86] L. Lovász and M.D. Plummer, *Matching Theory*, North-Holland Mathematics Studies, **121**, Annals of Discrete Mathematics, **29**, North-Holland Publishing Co., Amsterdam-New York; Akadémiai Kiadó (Publishing House of the Hungarian Academy of Sciences), Budapest, 1986.

[McDiarmid 89] C. McDiarmid, On the method of bounded differences, in: *Surveys in Combinatorics*, 1989 (Norwich, 1989), 148–188, London Math. Soc. Lecture Note Ser., **141**, Cambridge Univ. Press, Cambridge, 1989.

[Milman and Schechtman 86] V.D. Milman and G. Schechtman, *Asymptotic Theory of Finite-Dimensional Normed Spaces. With an Appendix by M. Gromov*, Lecture Notes in Mathematics, **1200**, Springer-Verlag, Berlin-New York, 1986.

[Papadimitriou 94] C.H. Papadimitriou, *Computational Complexity*, Addison-Wesley, Reading, Mass., 1994.

[Papadimitriou and Steiglitz 98] C.H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Dover, NY, 1998.

[Ryckman 00] E.M. Ryckman, Code for permanent approximations, *experimental C++ code*, available at http://www.math.lsa.umich.edu/~barvinok/papers.html

[Talagrand 95] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, *Inst. Hautes Études Sci. Publ. Math.* No. 81, (1995), 73–205.

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109-1109
E-mail address: barvinok@math.lsa.umich.edu

Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540
E-mail address: asamor@ias.edu