Optimal Space Lower Bounds for all Frequency Moments

David Woodruff *
MIT
dpwood@mit.edu

Abstract

We prove that any one-pass streaming algorithm which (ϵ, δ)-approximates the kth frequency moment F_k, for any real $k \neq 1$ and any $\epsilon = \Omega\left(\frac{1}{\sqrt{m}}\right)$, must use $\Omega\left(\frac{1}{\epsilon^2}\right)$ bits of space, where m is the size of the universe. This is optimal in terms of ϵ, resolves the open questions of Bar-Yossef *et al* in [3, 4], and extends the $\Omega\left(\frac{1}{\epsilon^2}\right)$ lower bound for F_0 in [11] to much smaller ϵ by applying novel techniques. Along the way we lower bound the one-way communication complexity of approximating the Hamming distance and the number of bipartite graphs with minimum/maximum degree constraints.

1 Introduction

Computing statistics on massive data sets is increasingly important these days. Advances in communication and storage technology enable large bodies of raw data to be generated daily, and consequently, there is a rising demand to process this data efficiently. Since it is impractical for an algorithm to store even a small fraction of the data stream, its performance is typically measured by the amount of space it uses. In many scenarios, such as internet routing, once a stream element is examined it is lost forever unless explicitly saved by the processing algorithm. This, along with the sheer size of the data, makes multiple passes over the data infeasible. In this paper we restrict our attention to one-pass streaming algorithms and we investigate their space complexity.

Let $\mathbf{a} = a_1, \ldots, a_q$ be a stream of q elements drawn from a universe of size m, which we denote by $[m] = \{1, \ldots, m\}$, and let f_i denote the number of occurrences of the ith universe element in $\mathbf{a}$. For any real k, the kth frequency moment F_k is defined by:

$$F_k = \sum_{i=1}^{m} f_i^k.$$

Interpreting $0^0 = 0$, we see that F_0 is the number of distinct elements in $\mathbf{a}$, F_1 is the stream size q, and F_2 is the *repeat rate*, also known as *Gini's index of homogeneity* [10]. Efficient algorithms for computing F_0 are important to the database community since query optimizers can use them for finding the number of unique values of an attribute without having to perform an expensive sort on the values. Efficient algorithms for F_2 are useful for determining the output size of self-joins in databases and for computing the *surprise index* of a data sequence [10]. Higher frequency moments are used to determine data *skewness* which is important in parallel database applications [8].

An algorithm A (ϵ, δ)-approximates F_k if A outputs a number $\tilde{F}_k$ such that $\Pr[|\tilde{F}_k - F_k| > \epsilon F_k] < \delta$. Since there is an $\Omega(m)$ space lower bound [1] for any deterministic algorithm computing F_k exactly or even approximating F_k within a multiplicative factor of $(1 \pm \epsilon)$, considerable effort has been invested into randomized approximation algorithms for the problem. In [1, 3, 7, 9] various algorithms are given to (ϵ, δ)-approximate F_0 with the best known algorithm (in terms of space complexity) given in [3] achieving space $O\left(\frac{1}{\epsilon^2} \log\log m + \log m \log \frac{1}{\epsilon}\right)$ [1]. Alon *et al* [1] present the best algorithm for (ϵ, δ)-approximating F_2 which achieves space $O\left(\frac{1}{\epsilon^2}(\log m + \log q)\right)$, and the best algorithm for (ϵ, δ)-approximating F_k which achieves space $O\left(\frac{(\log m + \log q)}{\epsilon^2} m^{1-\frac{1}{k}}\right)$ for any integer constant $k \geq 1$.

This paper is concerned with space lower bounds for the problem - we show that for any $\epsilon = \Omega\left(\frac{1}{\sqrt{m}}\right)$, any one-pass streaming algorithm which (ϵ, δ)-approximates F_k, for any real $k \neq 1$ [2], must use $\Omega\left(\frac{1}{\epsilon^2}\right)$ bits of space. Prior to our work the only known space lower bounds in terms of the approximation error ϵ were for F_0. For F_0 an $\Omega(\log m)$ space lower bound was established in [1], an $\Omega\left(\frac{1}{\epsilon}\right)$ lower bound in [4], and an $\Omega\left(\frac{1}{\epsilon^2}\right)$ lower bound for $\epsilon = \Omega\left(m^{\frac{-1}{9+c}}\right)$ for any $c > 0$ in [11]. Note that one cannot hope for the $\Omega\left(\frac{1}{\epsilon^2}\right)$ lower bound to

*Supported by a DoD NDSEG fellowship.

[1] In this paper we take the error probability δ to be a constant, i.e., a value independent of m.

[2] Note that F_1 can be computed trivially and exactly in space $O(\log q)$.

hold for $\epsilon = o\left(\frac{1}{\sqrt{m}}\right)$ since there is an $O(m)$ algorithm computing F_0 exactly and an $O(m \log q)$ computing F_k exactly for any $k \notin \{0, 1\}$.

As in previous papers [1, 4, 5, 6, 11], to show space lower bounds we lower bound the one-way communication complexity of a boolean function f and reduce the computation of f to that of F_k. More precisely, there are two parties Alice and Bob holding inputs x and y respectively who wish to compute $f(x, y)$ with error probability at most δ. Suppose that Alice and Bob can associate x, y with data streams $\mathbf{a}_x, \mathbf{a}_y$. Let A be an algorithm which (ϵ, δ)-approximates F_k. Then Alice can compute $A(\mathbf{a}_x)$ and transmit the state S of A to Bob. Bob can feed S into his copy of A and continue the computation to obtain $\tilde{F}_k(\mathbf{a}_x \circ \mathbf{a}_y)$. If $\tilde{F}_k(\mathbf{a}_x \circ \mathbf{a}_y)$ can determine $f(x, y)$ with probability at least $1 - \delta$, then the space used by A must be at least the one-way communication complexity of f. The cleverness is in choosing f and bounding its one way complexity.

Let $\Delta(\cdot, \cdot)$ denote Hamming distance and set $t = \Theta\left(\frac{1}{\epsilon^2}\right)$. We consider the following function f suggested in [11]. Alice and Bob are given $x, y \in \{0, 1\}^t$ with the promise that either $\Delta(x, y) \leq \frac{t}{2} - \sqrt{t}$, in which case $f(x, y) = 0$, or $\Delta(x, y) > \frac{t}{2}$, in which case $f(x, y) = 1$. The authors of [11] were not able to lower bound the one-way complexity of f directly, and instead considered a related function g with rational inputs $x, y \in [0, 1]^t$. They used a low distortion embedding to reduce a bound on g's complexity to a bound on F_0's space complexity. This indirect approach led to an additional assumption on ϵ, namely, that their bound held only for $\epsilon = \Omega\left(m^{\frac{-1}{9+c}}\right)$ for any $c > 0$. We instead lower bound the one-way complexity of f *directly* using novel techniques, and hence our $\Omega\left(\frac{1}{\epsilon^2}\right)$ bound holds for all $\epsilon = \Omega\left(\frac{1}{\sqrt{m}}\right)$ and all $k \neq 1$, which is optimal. To lower bound f's one-way complexity, we use shatter coefficients [6] which generalize the VC-dimension [12, 14]. The tricky part is proving our main theorem, which essentially computes the largest shatter coefficient of f. We use the probabilistic method in an elaborate way and a correlation inequality due to Kleitman [2].

Our main theorem establishes some additional results. Consider the problem: Alice and Bob have inputs x, y respectively and wish to (ϵ, δ)-approximate $\Delta(x, y)$. Such a protocol necessarily computes $f(x, y)$ with error probability at most δ. Hence, we obtain the first (in terms of ϵ) lower bound on the one-way communication complexity of (ϵ, δ)-approximating the Hamming distance.

Finally, in the proof of our main theorem it is shown that the number of m by n binary matrices M with majority one in each column and majority one in each row is at least $2^{mn-zm-n}$ for a constant $z < 1$. Here $m = \omega(1)$ for $n \to \infty$. Using the natural association between bipartite graphs on n by m vertices with binary m by n matrices, we obtain a nontrivial lower bound on the number of bipartite graphs on n by m vertices where each left vertex has degree at most (resp. at least) $\frac{m}{2}$ and each right vertex has degree at most (resp. at least) $\frac{n}{2}$. Our presentation is much simpler than that in [13], although our result is only a lower bound. As far as we are aware, this is the first nontrivial lower bound for the class of bipartite graphs [3].

2 Preliminaries

We adopt some of the definitions/notation given in [4, 11]. For $x, y \in \{0, 1\}^n$, let $x \oplus y$ denote vector addition over $GF(2)$, $\overline{x}$ complementation, $\Delta(x, y)$ Hamming distance, and $\mathbb{Z}$ the integers. The characteristic vector of a stream $\mathbf{a}$ is the length-m bit vector with ith bit set to 1 iff $f_i > 0$.

2.1 One-Way Communication Complexity

Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be a boolean function. In this paper we consider two parties, Alice and Bob, receiving x and y respectively, who wish to compute $f(x, y)$. In our protocols Alice computes some function $A(x)$ of x and sends the result to Bob. Bob then attempts to compute $f(x, y)$ from $A(x)$ and y. Note that only one message is sent, and it must be from Alice to Bob.

Definition 2.1. *For each randomized protocol Π as described above for computing f, the **communication cost** of Π is the expected length of the longest message sent from Alice to Bob over all inputs. The δ-**error randomized communication complexity** of f, $R_\delta(f)$, is the communication cost of the optimal protocol computing f with error probability δ (that is, $\Pr[\Pi(x, y) \neq f(x, y)] \leq \delta$).*

For deterministic protocols with input distribution μ, define $D_{\mu,\delta}(f)$, the **δ-error μ-distributional communication complexity** of f, to be the communication cost of an optimal such protocol. Using the Yao Minimax Principle, $R_\delta(f)$ is bounded from below by $D_{\mu,\delta}$ for any μ [15].

2.2 VC dimension and Shatter Coefficients

Let $\mathcal{F} = \{f : \mathcal{X} \to \{0, 1\}\}$ be a family of Boolean functions on a domain $\mathcal{X}$. Each $f \in \mathcal{F}$ can be viewed as a $|\mathcal{X}|$-bit string $f_1 \ldots f_{|\mathcal{X}|}$.

[3] The presentation in [13] was a characterization for general graphs.

Definition 2.2. *For a subset $\mathcal{S} \subseteq \mathcal{X}$, the* ***shatter coefficient*** *$SC(f_S)$ of $\mathcal{S}$ is given by $|\{f|_\mathcal{S}\}_{f\in\mathcal{F}}|$, the number of distinct bit strings obtained by restricting $\mathcal{F}$ to $\mathcal{S}$. The l-th shatter coefficient $SC(\mathcal{F},l)$ of $\mathcal{F}$ is the largest number of different bit patterns one can obtain by considering all possible $f|_S$, where S ranges over all subsets of size l. If the shatter coefficient of $\mathcal{S}$ is $2^{|\mathcal{S}|}$, then $\mathcal{S}$ is* ***shattered*** *by $\mathcal{F}$. The* ***VC dimension*** *of $\mathcal{F}$, $VCD(\mathcal{F})$, is the size of the largest subset $\mathcal{S} \subseteq \mathcal{X}$ shattered by $\mathcal{F}$.*

The following theorem [6] lower bounds the one-way complexity of f in terms of information theory.

Theorem 2.1. *For every function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, every $l \geq VCD(f_\mathcal{X})$, and every $\delta > 0$, there exists a distribution μ on $\mathcal{X} \times \mathcal{Y}$ such that:*

$$D_{\mu,\delta}(f) \geq \log(SC(f_\mathcal{X}, l)) - l \cdot H_2(\delta).$$

2.3 Properties of the Binomial Distribution We need some properties of the binomial distribution in the proof of our main theorem. The following lemmas follow easily from Stirling's formula. Let n be odd and let X be the sum of n independent unbiased Bernoulli random variables $X_1, \ldots, X_n$.

Lemma 2.1. *For any constant $c > 0$, and for sufficiently large n,*

$$\Pr[X > \frac{n}{2} + c\sqrt{n}] > \frac{1}{2} - c\sqrt{\frac{2}{\pi}}$$

Lemma 2.2.

$$\forall i \ \Pr[X_i = 1 \mid X > \frac{n}{2}] = \frac{1}{2} + \sqrt{\frac{2}{\pi n}}\,(1 + o(1))$$

2.4 A Theorem of Kleitman We also need the following theorem due to Kleitman [2]. We say a set family $\mathcal{A}$ of a finite set N is *monotone increasing* if whenever $S \in \mathcal{A}$ and $S \subseteq T \subseteq N$, then $T \in \mathcal{A}$. If $\mathcal{A}$ and $\mathcal{B}$ are monotone increasing, then their intersection $\{S \mid S \in \mathcal{A} \text{ and } S \in \mathcal{B}\}$ is monotone increasing.

Theorem 2.2. (Kleitman) *Let N be a set of size n. Consider the symmetric probability space whose elements are the members of the power set of N, that is, for any $A \subseteq N$, $\Pr[A] = 2^{-n}$. Let $\mathcal{A}$ and $\mathcal{B}$ be two monotone increasing families of subsets of N. Then,*

$$\Pr[\mathcal{A} \cap \mathcal{B}] \geq \Pr[\mathcal{A}] \cdot \Pr[\mathcal{B}]$$

3 Applications of the Main Theorem

The main theorem intuitively says that there is a set $S \subseteq \{0,1\}^n$ of n elements such that for many subsets T of S, one can find a word $y_T \in \{0,1\}^n$ that *separates* T from its complement $S - T$. By y_T separating T from $S - T$, we mean that y_T is *closer* to every element of T than to any element of $S - T$. We measure closeness in terms of Hamming distance. For one of our applications we also need to ensure that y_T is not *too close* to any element of T. We give the formal theorem statement now and defer its proof to section 4:

Theorem 3.1. (Main) *There exist constants $c, c' > 0$ such that for sufficiently large n there is a set $S \subseteq \{0,1\}^n$ of size n such that for $2^{\Omega(n)}$ subsets T of S, there exists a $y = y_T \in \{0,1\}^n$ such that for all $t \in T$, $c'n \leq \Delta(y,t) \leq \frac{n}{2} - c\sqrt{n}$, and for all $t \in S - T$, $\Delta(y,t) > \frac{n}{2}$.*

We say that a set $T \subseteq S$ is *good* if there is a $y_T \in \{0,1\}^n$ which separates T from its complement. More precisely, T is *good* if for all $t \in T$, $c'n \leq \Delta(y,t) \leq \frac{n}{2} - c\sqrt{n}$, and for all $t \in S - T$, $\Delta(y,t) > \frac{n}{2}$.

3.1 One-way Communication Complexity of Approximating the Hamming Distance Let $\epsilon = \Omega\left(\frac{1}{\sqrt{m}}\right)$ and $t = \Theta\left(\frac{1}{\epsilon^2}\right)$, where we assume t is a power of 2 without loss of generality (wlog). Let S, c be as in the main theorem, applied with $n = t$, and define $\mathcal{Y} = \{y_T \mid T \subseteq S \text{ and } T \text{ is good}\}$, using the notation above. We assume ϵ is small enough so that t is sufficiently large to apply the main theorem with $n = t$. Setting ϵ to be less than a small constant suffices. Define the promise problem:

$$L = \{(y,s) \in \mathcal{Y} \times S \text{ s.t. } \Delta(y,s) \leq \frac{t}{2} - c\sqrt{t} \text{ or } \Delta(y,s) > \frac{t}{2}\}$$

Define $f : \mathcal{Y} \times S \to \{0,1\}$ as $f(y,s) = 1$ if $\Delta(y,s) > \frac{t}{2}$ and $f(y,s) = 0$ if $\Delta(y,s) \leq \frac{t}{2} - c\sqrt{t}$, and define the function family $\mathcal{F} = \{f_y \mid y \in \mathcal{Y}\}$ where $f_y : S \to \{0,1\}$ is defined by $f_y(s) = f(y,s)$.

Consider the (ϵ,δ)-Hamming Distance Approximation Problem ((ϵ,δ)-HDAP): Alice, Bob have $x, y \in \{0,1\}^m$ respectively, and wish to output $\tilde{\Delta}(x,y)$ with $\Pr[|\tilde{\Delta}(x,y) - \Delta(x,y)| > \epsilon\Delta(x,y)] < \delta$. The claim is that provided $t \leq m$, the randomized one-way communication complexity $R_\delta(f)$ of deciding L is a lower bound on the one-way communication complexity of the (ϵ,δ)-HDAP. Indeed, a special case of the (ϵ,δ)-HDAP is when Alice is given a random element x of $\mathcal{Y}$, padded with $m - t$ zeros, and Bob a random element y of S, padded with $m - t$ zeros. Then with probability at least $1 - \delta$, if $\Delta(x,y) \leq \frac{t}{2} - c\sqrt{t}$, $\tilde{\Delta}(x,y) \leq (1+\epsilon)\left(\frac{t}{2} - c\sqrt{t}\right)$, and if $\Delta(x,y) > \frac{t}{2}$, then $\tilde{\Delta}(x,y) \geq (1-\epsilon)\frac{t}{2}$. For appropriately small $\epsilon = \Theta\left(\frac{1}{\sqrt{t}}\right)$, these two cases can be distinguished. Hence, the output $\tilde{\Delta}(x, y)$ can decide L with probability $1-\delta$.

We now show $R_\delta(f) = \Omega(t)$, and hence that the one-way complexity of the (ϵ, δ)-HDAP is $\Omega\left(\frac{1}{\epsilon^2}\right)$.

THEOREM 3.2. *The $\frac{t}{4}$th shatter coefficient of $\mathcal{F}$ is $2^{\Omega(t)}$.*

Proof. The claim is that there are $2^{\Omega(t)}$ distinct bitstrings in the truth table of $\mathcal{F}$. Indeed, for every $y \in \mathcal{Y}$, there exists a good subset $T \subseteq S$ such that $y = y_T$. For $s \in T$, $f(y, s) = 0$ and for $s \in S - T$, $f(y, s) = 1$. Viewing f_y as a bitstring (see section 2), it follows that $f_y \neq f_{y'}$ for $y \neq y'$ since if $T' \subseteq S$ is such that $y' = y_{T'}$, T' and T differ in at least one element. Hence there are $|\mathcal{Y}| = 2^{\Omega(t)}$ distinct bitstrings, so the shatter coefficient is $2^{\Omega(t)}$. ■

COROLLARY 3.1. *The randomized one-way communication complexity $R_\delta(f)$ is $\Omega(t) = \Omega\left(\frac{1}{\epsilon^2}\right)$.*

Proof. Follows immediately from theorem 2.1. ■

3.2 Space Complexity of Approximating the Frequency Moments From the previous section, we know that for $\epsilon = \Omega\left(m^{-\frac{1}{2}}\right)$, the one-way communication complexity of deciding L with error probability at most δ is $\Omega\left(\frac{1}{\epsilon^2}\right)$. We now give a protocol for any $\epsilon = \Omega\left(m^{-\frac{1}{2}}\right)$ which decides L with probability at least $1-\delta$ with communication cost equal to the space of any (ϵ, δ) F_k-approximation algorithm for any $k \neq 1$. It follows that for any $k \neq 1$ and any $\epsilon = \Omega\left(m^{-\frac{1}{2}}\right)$, any (ϵ, δ) F_k-approximation algorithm must use $\Omega\left(\frac{1}{\epsilon^2}\right)$ space. In particular, for all smaller ϵ, any such algorithm must use $\Omega(m)$ space. For $k = 0$ this is optimal since one can keep a length-m bit vector to compute F_0 exactly. For $k \notin \{0, 1\}$ this is optimal up to a factor of $\log q$ since one can keep a length-m vector with ith entry set to f_i.

Let $t = \Theta\left(\frac{1}{\epsilon^2}\right)$ as before. Alice and Bob are given random $y \in \mathcal{Y}$ and $s \in \mathcal{S}$, respectively, and wish to determine $f(y, s)$. The protocol is as follows: Alice chooses a stream $\mathbf{a}_y$ with characteristic vector $y \circ 0^{m-t}$. Let M be an (ϵ, δ) F_k-approximation algorithm for some constant $k \neq 1$. Alice runs M on $\mathbf{a}_y$. When M terminates, she transmits the state S of M to Bob along with $wt(y)$. Bob chooses a stream $\mathbf{a}_s$ with characteristic vector $s \circ 0^{m-t}$ and feeds both S and $\mathbf{a}_s$ into his copy of M. Let $\tilde{F_k}$ be the output of M. The claim is that $\tilde{F_k}$ along with $wt(y)$ and $wt(s)$ can be used to determine $f(y, s)$ (and hence decide L) with probability at least $1-\delta$. We first decompose F_k:

$$F_k(\mathbf{a}_y \circ \mathbf{a}_s) = \sum_{i \in [m]} f_i^k = 2^k wt(y \wedge s) + 1^k \Delta(y, s)$$

$$= 2^{k-1}(wt(y) + wt(s) - \Delta(y, s)) + \Delta(y, s)$$

$$= 2^{k-1}(wt(y) + wt(s)) + (1 - 2^{k-1})\Delta(y, s)$$

and hence for $k \neq 1$,

$$\Delta(y, s) = \frac{2^{k-1}}{2^{k-1}-1}(wt(y) + wt(s)) - \frac{F_k(\mathbf{a}_y \circ \mathbf{a}_s)}{2^{k-1}-1} \qquad (3.1)$$

We want a $(1 \pm \epsilon')$ approximation to F_k to result in a $(1 \pm \epsilon)$ approximation to $\Delta(y, s)$ for some $\epsilon' = \Theta(\epsilon)$. Specifically, if $k < 1$ we want:

$$(1-\epsilon)\Delta(y, s) \leq$$

$$\frac{2^{k-1}}{2^{k-1}-1}(wt(y) + wt(s)) - (1-\epsilon')\frac{F_k(\mathbf{a}_y \circ \mathbf{a}_s)}{2^{k-1}-1}$$

and

$$\frac{2^{k-1}}{2^{k-1}-1}(wt(y) + wt(s)) - (1+\epsilon')\frac{F_k(\mathbf{a}_y \circ \mathbf{a}_s)}{2^{k-1}-1} \leq$$

$$(1+\epsilon)\Delta(y, s),$$

whereas for $k > 1$ we want:

$$(1-\epsilon)\Delta(y, s) \leq$$

$$\frac{2^{k-1}}{2^{k-1}-1}(wt(y) + wt(s)) - (1+\epsilon')\frac{F_k(\mathbf{a}_y \circ \mathbf{a}_s)}{2^{k-1}-1}$$

and

$$\frac{2^{k-1}}{2^{k-1}-1}(wt(y) + wt(s)) - (1-\epsilon')\frac{F_k(\mathbf{a}_y \circ \mathbf{a}_s)}{2^{k-1}-1} \leq$$

$$(1+\epsilon)\Delta(y, s).$$

After some algebraic manipulation, we see that these properties hold iff:

$$\epsilon' \quad \leq \quad \frac{\epsilon|2^{k-1}-1|\Delta(y, s)}{F_k(\mathbf{a}_y \circ \mathbf{a}_s)}.$$

Now, $F_k(\mathbf{a}_y \circ \mathbf{a}_s) = O(t)$. Hence, for any $k \neq 1$ we will have $\epsilon' = \Theta(\epsilon)$ if there exists a positive constant p so that for all pairs of inputs y, s, $\Delta(y, s) > pt$. For $n = t$ in the main theorem, we see that this condition is satisfied for $p = c'$.

We conclude that Alice and Bob can choose $\epsilon' = \Theta(\epsilon)$ such that Bob can use his knowledge of $wt(y)$, $wt(s)$, and an (ϵ', δ) approximation to F_k (i.e., $\tilde{F_k}$), to compute $\frac{2^{k-1}}{2^{k-1}-1}(wt(y) + wt(s)) - \frac{\tilde{F_k}(\mathbf{a}_y \circ \mathbf{a}_s)}{2^{k-1}-1}$, which is a $(1 \pm \epsilon)$-approximation to $\Delta(y, s)$. Hence, as in the analysis of the (ϵ, δ)-HDAP, Bob can decide L

with probability at least $1-\delta$. One may worry that the $\log t = O(\log m)$ bits used to transmit $wt(y)$ will dominate the space of the F_k-approximation algorithm for large ϵ. Fortunately, there is also an $\Omega(\log m)$ space lower bound [1] for approximating F_k for any $k \neq 1$ [4], so if indeed $\log m = \omega\left(\frac{1}{\epsilon^2}\right)$, the $\Omega\left(\frac{1}{\epsilon^2}\right)$ lower bound is absorbed in the $\Omega(\log m)$ lower bound. From the reduction we see that the F_k-approximation algorithm must use $\Omega\left(\frac{1}{\epsilon^2}\right)$ space.

3.3 Lower Bound for Bipartite Graphs with Given Maximum/Minimum Degree There is a bijective correspondence between m by n binary matrices M and bipartite graphs G on $m+n$ vertices, where $M_{ij} = 1$ iff there is an edge from the ith left vertex to the jth right vertex in G. From corollary 4.1 (see the end of section 4) we see that the number of bipartite graphs on $m+n$ vertices where each left vertex has degree at least $\frac{n}{2}$ and each right vertex has degree at least $\frac{m}{2}$, is at least $2^{mn-zm-n}$ for a constant $z < 1$. Interchanging the role of 1s and 0s, it follows that the number of bipartite graphs with each left vertex having degree at *most* $\frac{n}{2}$ and each right vertex having degree at most $\frac{m}{2}$, is at least $2^{mn-zm-n}$.

Note that a trivial lower bound on the number of such graphs can be obtained from theorem 2.2. Indeed, if $\mathcal{C}$ is the event that each column of M is majority 1 and $\mathcal{R}$ the event that each row is majority 1, $\mathcal{C}$ and $\mathcal{R}$ represent monotone families of subsets of $[mn]$, so by theorem 2.2, $\Pr[\mathcal{R} \cap \mathcal{C}] \geq 2^{-m} \cdot 2^{-n} = 2^{-m-n}$, and hence the number of such M is at least $2^{mn} \cdot 2^{-m-n} = 2^{(mn-m-n)}$. Since $z < 1$ in our bound, our bound is strictly stronger.

4 Proof of the Main Theorem

We use the probabilistic method to prove our main theorem, repeated here for convenience:

THEOREM 4.1. *There exist constants $c, c' > 0$ such that for sufficiently large n there is a set $S \subseteq \{0,1\}^n$ of size n such that for $2^{\Omega(n)}$ subsets T of S, there exists a $y = y_T \in \{0,1\}^n$ such that for all $t \in T$, $c'n \leq \Delta(y,t) \leq \frac{n}{2} - c\sqrt{n}$, and for all $t \in S - T$, $\Delta(y,t) > \frac{n}{2}$.*

Proof. Let $c, c' > 0$ be constants to be determined. We assume $n \equiv 1 \bmod 4$ in what follows, so that n and $\lceil \frac{n}{2} \rceil$ are odd. Choose n elements $r_1, \ldots, r_n$ uniformly at random from $\{0,1\}^n$ with replacement, and put $S = \{r_1, \ldots, r_n\}$. Note that S may be a multiset; we correct this later. Set $m = \lceil \frac{n}{2} \rceil$ and let T be an arbitrary subset of S of size m. We omit ceilings if not essential.

For notational convenience put $T = \{r_1, \ldots, r_m\}$. Let $y = y_T$ be the majority codeword of T, that is, $y_j = \text{majority}(r_{1j}, \ldots, r_{mj})$ for all $1 \leq j \leq m$. The map $f_y(x) = \overline{x \oplus y}$ preserves Hamming distances, so wlog, assume $y = 1^n$.

We say that T is *good* if for all $t \in T$, $c'n \leq \Delta(y,t) \leq \frac{n}{2} - c\sqrt{n}$, and for all $t \in S - T$, $\Delta(y,t) > \frac{n}{2}$. We show the probability that T is good is greater than 2^{-zn} for a constant $z < 1$. It follows that the expected number of good subsets of S of size m is $\binom{n}{m} 2^{-zn} = 2^{H_2(\frac{1}{2})n + o(1)n - zn} = 2^{\Omega(n)}$. Hence, there exists an S with $2^{\Omega(n)}$ good subsets. It remains to lower bound the probability that T is good.

The probability that T is good is just the product:

$$\Pr[\ \forall t \in S - T,\ \Delta(y,t) > \frac{n}{2}\ ] \cdot$$

$$\Pr[\ \forall t \in T,\ c'n \leq \Delta(y,t) \leq \frac{n}{2} - c\sqrt{n}\ ],$$

since these events are independent. Since y is independent of $S - T$,

$$(4.2) \quad \Pr[\ \forall t \in S - T,\ \Delta(y,t) > \frac{n}{2}\ ] \quad = \quad 2^{m-n}.$$

We find $\Pr[\ \forall t \in T,\ \Delta(y,t) \leq \frac{n}{2} - c\sqrt{n}\ ]$. We force $\Delta(y,t) \geq c'n$ later. Let M be the binary $m \times n$ matrix whose ith row is r_i. Let $m = m_1 + m_2$ for m_1, m_2 positive integers to be determined. Let $\mathcal{R}_1$ be the event that M has at least $\frac{n}{2} + c\sqrt{n}$ ones in each of its first m_1 rows, $\mathcal{R}_2$ the event that M has at least $\frac{n}{2} + c\sqrt{n}$ ones in each of its remaining m_2 rows, and $\mathcal{C}$ the event that M has at least $\frac{m}{2}$ ones in each column. Then,

$$\Pr[\ \forall t \in T,\ \Delta(y,t) \leq \frac{n}{2} - c\sqrt{n}\ ] = \Pr[\ \mathcal{R}_1 \cap \mathcal{R}_2 \mid \mathcal{C}\ ]$$

$$= \frac{\Pr[\ \mathcal{R}_1 \cap \mathcal{R}_2 \cap \mathcal{C}\ ]}{\Pr[\ \mathcal{C}\ ]}$$

M can be viewed as the characteristic vector of a subset of $[mn] = \{0, \ldots, mn-1\}$. Under this correspondence, each of $\mathcal{R}_1, \mathcal{R}_2$, and $\mathcal{C}$ represent monotone families of subsets of $[mn]$. Applying Theorem 2.2,

$$\begin{aligned} \Pr[\ \mathcal{R}_1 \cap \mathcal{R}_2 \cap \mathcal{C}\ ] &\geq \Pr[\ \mathcal{R}_1 \cap \mathcal{C}\ ] \Pr[\ \mathcal{R}_2\ ] \\ &= \Pr[\ \mathcal{R}_1 \mid \mathcal{C}\ ] \Pr[\ \mathcal{C}\ ] \Pr[\ \mathcal{R}_2\ ] \end{aligned}$$

[4] In [1] the authors only explicitly state the $\Omega(\log m)$ lower bound for $k \in \{0, 2\}$, but their argument in propositions 3.7 and 4.1 is easily seen to hold for any fixed $k \neq 1$ (even nonintegral) for sufficiently small, but constant ϵ.

and hence,

$$\text{(4.3)} \qquad \Pr[\ \forall t \in T,\ \Delta(y,t) \le \frac{n}{2} - c\sqrt{n}\ ]$$

$$\ge \Pr[\ \mathcal{R}_1 \mid \mathcal{C}\ ] \Pr[\ \mathcal{R}_2\ ]$$

Computing $\Pr[\ \mathcal{R}_2\ ]$ is easy since M's entries are independent in this case. There are m_2 independent rows, and each row is a sum of n independent unbiased Bernoulli variables. By lemma 2.1,

$$\text{(4.4)} \qquad \Pr[\ \mathcal{R}_2\ ] \quad > \quad \left(\frac{1}{2} - c\sqrt{\frac{2}{\pi}}\right)^{m_2}$$

To compute $\Pr[\mathcal{R}_1|\mathcal{C}]$, let Y be the number of ones in M. We compute

$$\Pr[\mathcal{R}_1|\mathcal{C}] = \sum_s \Pr[\mathcal{R}_1|Y = s, \mathcal{C}] \cdot \Pr[Y = s \mid \mathcal{C}].$$

The following insight simplifies this calculation:

Lemma 4.1. $\Pr[Y = \frac{nm}{2} + n\sqrt{\frac{2m}{\pi}}(1 + o(1)) \ \mid \mathcal{C}\ ] = 1 - o(1)$.

Proof. Let Y_i be the number of ones in column i, for $1 \le i \le n$. From lemma 2.2, $\mathbf{E}[Y_i|\mathcal{C}] = \frac{m}{2} + \sqrt{\frac{2m}{\pi}}(1 + o(1))$. Hence, $\mathbf{E}[Y|\mathcal{C}] = \frac{nm}{2} + n\sqrt{\frac{2m}{\pi}}(1 + o(1))$. Since the columns are i.i.d., $\mathbf{Var}[Y|\mathcal{C}] = n\mathbf{Var}[Y_i|\mathcal{C}] \le \frac{nm}{4}$. Chebyshev's inequality establishes the lemma:

$$\Pr[|Y|\mathcal{C} - \mathbf{E}[Y|\mathcal{C}]| > \omega(n)] \le \frac{\mathbf{Var}[Y|\mathcal{C}]}{\omega(n^2)} = o(1). \qquad \blacksquare$$

Put $s = \frac{nm}{2} + n\sqrt{\frac{2m}{\pi}}(1 + o(1))$. It follows that:

$$\text{(4.5)} \quad \Pr[\mathcal{R}_1|C] \quad \ge \quad (1 - o(1)) \Pr[\mathcal{R}_1|Y = s, \mathcal{C}].$$

Technically speaking, s represents a set of values, all of which are of the form $\frac{nm}{2} + n\sqrt{\frac{2m}{\pi}}(1 + o(1))$. We abuse notation and say $Y = s$, when in fact Y assumes a value in this set.

Define X_{ij} to be the (i,j)th entry of M, *conditioned* on events $Y = s$ and $\mathcal{C}$, and define $X_i = \sum_j X_{ij}$. Now put $c = \frac{2r}{\sqrt{\pi}}$ and $d = \frac{2(2-r)}{\sqrt{\pi}}$ for a constant $0 < r < 1$ to be determined, and let $\mathcal{E}_i$ be the event:

$$\frac{n}{2} + c\sqrt{n} < X_i < \frac{n}{2} + d\sqrt{n}.$$

for $1 \le i \le m_1$. Clearly,

$$\text{(4.6)} \Pr[\mathcal{R}_1 \mid \mathcal{C},\ Y = s] > \prod_{i=1}^{m_1} \Pr[\mathcal{E}_i \mid \cap_{l=1}^{i-1} \mathcal{E}_l].$$

The idea is to bound $\mathbf{E}[X_i | \cap_{l=1}^{i-1} \mathcal{E}_l]$ and to show $\mathbf{Var}[X_i | \cap_{l=1}^{i-1} \mathcal{E}_l]$ is small so that we can use Chebyshev's inequality on each multiplicand in the RHS of 4.6.

We first bound $\mathbf{E}[X_i | \cap_{l=1}^{i-1} \mathcal{E}_l]$. Given $\cap_{l=1}^{i-1} \mathcal{E}_l$, we know that $\sum_{l=1}^{i-1} X_i$ is at least $(i-1)\left(\frac{n}{2} + c\sqrt{n}\right)$ and at most $(i-1)\left(\frac{n}{2} + d\sqrt{n}\right)$. To ensure that $\mathbf{E}[X_i | \cap_{l=1}^{i-1} \mathcal{E}_l]$ doesn't vary much with i, we restrict m_1 from being too large by setting $m_1 = vm$ for a constant $0 < v < 1$ to be determined. Since there are s ones in M, and $\mathbf{E}[X_{j_1} \mid \cap_{l=1}^{i-1} \mathcal{E}_l] = \mathbf{E}[X_{j_2} \mid \cap_{l=1}^{i-1} \mathcal{E}_l]$ for all $j_1, j_2 \ge i$,

$$\frac{s - (i-1)\left(\frac{n}{2} + d\sqrt{n}\right)}{m - (i-1)}$$

$$= \frac{\frac{mn}{2} + 2m\sqrt{\frac{n}{\pi}} + o\left(n^{\frac{3}{2}}\right) - (i-1)\left(\frac{n}{2} + d\sqrt{n}\right)}{m - (i-1)}$$

$$= \frac{n}{2} + \sqrt{n}\left(\frac{2}{\sqrt{\pi}} - \frac{(i-1)\left(d - \frac{2}{\sqrt{\pi}}\right)}{m - i + 1}\right) + o\left(n^{\frac{1}{2}}\right)$$

$$\le \ \mathbf{E}[X_i | \cap_{l=1}^{i-1} \mathcal{E}_l].$$

From a similar calculation,

$$\mathbf{E}[X_i | \cap_{l=1}^{i-1} \mathcal{E}_l] \quad \le$$

$$\frac{n}{2} + \sqrt{n}\left(\frac{2}{\sqrt{\pi}} + \frac{(i-1)\left(\frac{2}{\sqrt{\pi}} - c\right)}{m - i + 1}\right) + o\left(n^{\frac{1}{2}}\right)$$

Setting $i = m_1 + 1$ in the above, we obtain bounds independent of i which hold for all $1 \le i \le m_1$,

$$\frac{n}{2} + \sqrt{n}\left(\frac{2}{\sqrt{\pi}} - \frac{v\left(d - \frac{2}{\sqrt{\pi}}\right)}{1 - v}\right) + o\left(n^{\frac{1}{2}}\right)$$

$$\le \mathbf{E}[X_i | \cap_{l=1}^{i-1} \mathcal{E}_l] \le$$

$$\frac{n}{2} + \sqrt{n}\left(\frac{2}{\sqrt{\pi}} + \frac{v\left(\frac{2}{\sqrt{\pi}} - c\right)}{1 - v}\right) + o\left(n^{\frac{1}{2}}\right)$$

Define k_i to be

$$\min$$

$$\left(\mathbf{E}[X_i | \cap_{l=1}^{i-1} \mathcal{E}_l] - \frac{n}{2} - c\sqrt{n},\ \frac{n}{2} + d\sqrt{n} - \mathbf{E}[X_i | \cap_{l=1}^{i-1} \mathcal{E}_l]\right)$$

and note that k_i measures how far $X_i | \cap_{l=1}^{i-1} \mathcal{E}_l$ has to deviate from its expectation for $\overline{\mathcal{E}_i} \mid \cap_{l=1}^{i-1} \mathcal{E}_l$ to occur. We will use k_i in Chebyshev's inequality below. Simplifying k_i using our bounds, after some algebra we obtain:

$$k_i \quad = \quad \sqrt{n}\left(1-\frac{v}{1-v}\right)\left(\frac{2-2r}{\sqrt{\pi}}\right)+o\left(n^{\frac{1}{2}}\right),$$

using the definitions of c and d, which were defined to be symmetric around $\frac{2}{\sqrt{\pi}}$. Note that for sufficiently large n, k_i is positive provided $v < \frac{1}{2}$, which we hereby enforce.

We show that $\mathbf{Var}[X_i | \cap_{l=1}^{i} \mathcal{E}_l]$ is small by showing the entries in the ith row are negatively correlated:

LEMMA 4.2. *For any $2 \leq i \leq m_1$ and any $1 \leq j < k \leq n$,*

$$\frac{\boldsymbol{Cov}[X_{ij}, X_{ik} \mid \cap_{l=1}^{i-1} \mathcal{E}_l\ ]}{\Pr[X_{ik}=1 \mid \cap_{l=1}^{i-1} \mathcal{E}_l\ ]} =$$

$$\Pr[X_{ij}=1 \mid X_{ik}=1,\ \cap_{l=1}^{i-1}\mathcal{E}_l] - \Pr[X_{ij}=1 \mid \cap_{l=1}^{i-1}\mathcal{E}_l] < 0$$

Proof. Interpreting $\binom{n}{x}=0$ for $x<0$, we have:

$$\Pr[X_{ij}=1 \mid \cap_{l=1}^{i-1} \mathcal{E}_l] =$$

$$\sum_{t=0}^{n} \Pr[X_{ij}=1 \mid X_i=t,\ \cap_{l=1}^{i-1}\mathcal{E}_l] \cdot \Pr[X_i=t,\ \cap_{l=1}^{i-1}\mathcal{E}_l] =$$

$$\sum_{t=1}^{n} \frac{\binom{n-1}{t-1}}{\binom{n}{t}} \Pr[X_i=t,\ \cap_{l=1}^{i-1}\mathcal{E}_l]\ >$$

$$\sum_{t=1}^{n} \frac{\binom{n-2}{t-2}}{\binom{n-1}{t-1}} \Pr[X_i=t,\ \cap_{l=1}^{i-1}\mathcal{E}_l] =$$

$$\sum_{t=0}^{n}(\Pr[X_{ij}=1 \mid X_{ik}=1,\ X_i=t,\ \cap_{l=1}^{i-1}\mathcal{E}_l] \cdot$$

$$\Pr[X_i=t,\ \cap_{l=1}^{i-1}\mathcal{E}_l])$$

$$= \Pr[X_{ij}=1 \mid X_{ik}=1,\ \cap_{l=1}^{i-1}\mathcal{E}_l],$$

where we used the fact that conditioned on $X_i = t$, every t-combination in the ith row is equally likely by symmetry. ■

It follows that for all i,

$$\mathbf{Var}[X_i \mid \cap_{l=1}^{i-1} \mathcal{E}_l] =$$

$$\sum_{j=1}^{n} \mathbf{Var}[X_{ij} \mid \cap_{l=1}^{i-1} \mathcal{E}_l] + \sum_{j \neq k} \mathbf{Cov}[X_{ij},\ X_{ik} \mid \cap_{l=1}^{i-1} \mathcal{E}_l]$$

$$< \sum_{j=1}^{n} \mathbf{Var}[X_{ij} \mid \cap_{l=1}^{i-1} \mathcal{E}_l] \leq \frac{n}{4}.$$

We now apply Chebyshev's inequality to each row:

$$\Pr[\mathcal{E}_i \mid \cap_{l=1}^{i-1} \mathcal{E}_l] =$$

$$\Pr[\frac{n}{2}+c\sqrt{n} < X_i < \frac{n}{2}+d\sqrt{n} \mid \cap_{l=1}^{i-1} \mathcal{E}_l] \geq$$

$$1-\Pr[\ |X_i - \mathbf{E}[X_i | \cap_{l=1}^{i-1} \mathcal{E}_l]\ | > k_i\ ] \ \geq$$

$$1-\frac{\mathbf{Var}[X_i \mid \cap_{l=1}^{i-1} \mathcal{E}_l]}{{k_i}^2} \geq$$

$$1-\frac{n}{4{k_i}^2} \ = \ 1-\frac{\pi}{4\left(\frac{1-2v}{1-v}\right)^2(2-2r)^2-o(1)}.$$

To simplify this expression, we choose $v = \left(\frac{\sqrt{2}-1}{2\sqrt{2}-1}\right) < \frac{1}{2}$. The above inequality becomes

$$\text{(4.7)} \quad \Pr[\mathcal{E}_i \mid \cap_{l=1}^{i-1} \mathcal{E}_l] \quad \geq \quad 1-\frac{\pi}{8(1-r)^2-o(1)}$$

From equations 4.3, 4.4, 4.5, 4.6, and 4.7, we conclude:

$$\text{(4.8)} \qquad \Pr[\ \forall t \in T,\ \Delta(y,t) \leq \frac{n}{2}-c\sqrt{n}\ ] >$$

$$\left(1-\frac{\pi}{8(1-r)^2-o(1)}\right)^{m_1}(1-o(1))\left(\frac{1}{2}-c\sqrt{\frac{2}{\pi}}\right)^{m_2}$$

We say that T is *almost good* if for all $t \in T$, $\Delta(y,t) \leq \frac{n}{2}-c\sqrt{n}$, and for all $t \in S-T$, $\Delta(y,t) > \frac{n}{2}$. Note that these two events are independent and that T is good if and only if T is almost good and for all $t \in T, \Delta(y,t) \geq c'n$. Combining equations 4.2 and 4.8, we have:

$$\Pr[T \text{ is almost good }] =$$

$$\Pr[\ \forall t \in S-T,\ \Delta(y,t) > \frac{n}{2}\ ] \cdot$$

$$\Pr[\ \forall t \in T, \Delta(y,t) \leq \frac{n}{2}-c\sqrt{n}\ ]$$

$$> 2^{m-n}\left(\frac{1}{2}-c\sqrt{\frac{2}{\pi}}\right)^{m_2} \cdot$$

$$\left(1-\frac{\pi}{8\left(1-r\right)^2-o(1)}\right)^{m_1}(1-o(1))$$

$$= 2^{m-n}\left(\frac{1}{2}-\frac{2r\sqrt{2}}{\pi}\right)^{(1-v)m} \cdot$$

$$\left(1-\frac{\pi}{8\left(1-r\right)^2-o(1)}\right)^{vm}(1-o(1))$$

Taking logarithms base 2 and dividing by n we obtain:

$$\text{(4.9)} \qquad \frac{\log(\Pr[T \text{ is almost good }])}{n} = -\frac{1}{2}$$

$$+\frac{(1-v)}{2}\log_2\left(\frac{1}{2}-\frac{2r\sqrt{2}}{\pi}\right)$$

$$+\frac{v}{2}\log_2\left(1-\frac{\pi}{8\,(1-r)^2-o(1)}\right)$$

$$+\ \log_2\left(1-o(1)\right)$$

Observe that the RHS of equation 4.9 is continuous in r for $0 \le r < 1$ and for $r = 0$ is just:

$$(4.10)\ -1+\frac{v}{2}\left(1+\log_2\left(1-\frac{\pi}{8-o(1)}\right)\right)+$$

$$\log_2(1-o(1)).$$

Let $n_1 \in \mathbb{Z}$ be such that for all $n > n_1$, (4.10) is less than $l = -1 + \frac{v}{2}\left(1+\log_2\left(1-\frac{\pi}{9}\right)\right) > -1$. Let n be larger than n_1 and large enough to satisfy all previous steps where n needed to be sufficiently large. Then (4.10) is larger than a *constant* larger than -1. Since equation 4.9 is continuous in r, there exists a constant $r > 0$ so that for sufficiently large n, the RHS of equation 4.9 is larger than a constant larger than -1. Hence for sufficiently large n, there exists a constant $z < 1$ so that $\Pr[T$ is almost good $] > 2^{-zn}$.

We compute $\Pr[\ \forall t \in T,\ \Delta(y,t) \ge c'n]$. Fix $t \in T$. From lemma 2.2, there is a constant $u > 0$ with:

$$\begin{aligned}
&\Pr[\Delta(y,t) \le c'n] \\
\le\ & \sum_{i=(1-c')n}^{n}\binom{n}{i}\left(\frac{1}{2}+\frac{u}{\sqrt{n}}\right)^{i}\left(\frac{1}{2}-\frac{u}{\sqrt{n}}\right)^{n-i} \\
\le\ & \binom{n}{(1-c')n}\left(\frac{1}{2}+\frac{u}{\sqrt{n}}\right)^{n}c'n \\
\le\ & 2^{H_2(1-c')n+O(\log n)-\alpha n}
\end{aligned}$$

for any constant $\alpha < 1$ and sufficiently large n. Hence,

$$\Pr[\exists t \in T \text{ such that } \Delta(y,t) \le c'n]$$

$$\le n2^{H_2(1-c')n+O(\log n)-\alpha n} \le 2^{H_2(1-c')n-\alpha' n},$$

for any $\alpha' < \alpha$ and large enough n. By the union bound,

$$\Pr[\ T \text{ is good }] =$$

$$\Pr[\ \forall t \in T,\ c'n \le \Delta(y,t) \le \frac{n}{2}-c\sqrt{n}\ ] \ge$$

$$\Pr[\ T \text{ is almost good }]-$$

$$\Pr[\exists t \in T \text{ such that } \Delta(y_T,t) \le c'n] \ge$$

$$2^{-zn}-2^{H_2(1-c')n-\alpha' n}$$

We choose c', α, α' so that $\alpha' - H_2(1-c') > z$ by choosing c' close to 0 and α close to 1. Hence, $\Pr[\ T$ is good $] > 2^{-z'n}$ for any $z' > z$ and large enough n. Since $z < 1$, we can choose $z' < 1$, as needed.

The only loose end to tie up is that S may be a multiset. But for any $i \ne j$, $\Pr[r_i = r_j] = 2^{-n}$, so:

$$\Pr[\exists i \ne j \text{ such that } r_i = r_j] \le \binom{n}{2}2^{-n} = 2^{-n+O(\log n)},$$

and hence for any specific T,

$$(4.11)\ \ \Pr[T \text{ is not good or } S \text{ is a multiest }] <$$

$$1-2^{-z'n}+2^{-n+O(\log n)},$$

so that for sufficently large n and for any $1 > z'' > z'$,

$$\Pr[T \text{ is good } |S \text{ is not a multiset }] \ge$$

$$\Pr[T \text{ is good and } S \text{ is not a multiset }] > 2^{-z''n}$$

Thus, the expected number of good subsets of S, given that S is not a multiset, is $2^{\Omega(n)}$, as before. This completes the proof. ■

Corollary 4.1. *The number of m by n binary matrices M with more ones than zeros in each column and more ones than zeros in each row is at least $2^{mn-zm-n}$ for a constant $z < 1$.*

Proof. Using the notation of the proof, the probability that a (uniformly) random m by n binary matrix M has majority 1 in each row, given that it has majority 1 in each column, is $\Pr[\ \mathcal{R}_1 \mid \mathcal{C}\ ] \cdot \Pr[\mathcal{R}_2]$ with $r = 0$ (and hence $c = 0$). Note that the proof holds for any superconstant value of m, even though we only needed $m = \lceil \frac{n}{2} \rceil$ before. As $n \to \infty$, $\Pr[\ \mathcal{R}_1 \mid \mathcal{C}\ ] \cdot \Pr[\mathcal{R}_2]$ approaches $\left(\frac{1}{2}\right)^{(1-v)m}\left(1-\frac{\pi}{8}\right)^{vm}$ (see equations 4.4, 4.5, 4.6, 4.7), which is $2^{-z'm}$ for a constant $z' < 1$. Hence for large enough n, we can get rid of the $o(1)$ terms (see the RHS of equation 4.8) and have $\Pr[\mathcal{R}_1|\mathcal{C}]\cdot\Pr[\mathcal{R}_2] \ge 2^{-zm}$ for a constant z with $z' < z < 1$. Thus, the probability that M has majority 1 in each row *and* majority 1 in each column is at least $2^{-zm} \cdot 2^{-n} = 2^{-zm-n}$. Since there are 2^{mn} total binary matrices, the number of such M is at least $2^{mn-zm-n}$. ■

5 Acknowledgment

The author thanks Piotr Indyk for helpful discussion and checking the proof of the main theorem.

References

[1] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. In *Proceedings of the 28th Annual ACM Symposum on the Theory of Computing*, p. 20-29, 1996.

[2] N. Alon and J. Spencer. *The Probabilistic Method*, Wiley Interscience, New York, 1992, 86—87.

[3] Z. Bar Yossef, T.S. Jayram, R. Kumar, D. Sivakumar, and Luca Trevisan. Counting distinct elements in a data stream. *RANDOM 2002, 6th. International Workshop on Randomization and Approximation Techniques in Computer Science*, p. 1-10, 2002.

[4] Z. Bar Yossef. The complexity of massive data set computations. Ph.D. Thesis, U.C. Berkeley, 2002.

[5] Z. Bar Yossef, T.S. Jayram, R. Kumar, and D. Sivakumar. Information statistics approach to data stream and communication complexity. *Foundations of Computer Science*, p.209-218, 2002.

[6] Z. Bar Yossef, T.S. Jayram, R. Kumar, and D. Sivakumar. Information Theory Methods in Communication Complexity. *17th IEEE Annual Conference on Computational Complexity*, p.93-102, 2002.

[7] G. Cormode, M. Datar, P. Indyk, and M. Muthukrishnan. Comparing Data Streams Using Hamming Norms. *28th International Conference on Very Large Databases (VLDB)*, 2002.

[8] D.J. DeWitt, J.F. Naughton, D.A. Schneider, and S. Seshadri. Practical Skew Handling in Parallel Joins. *Proc. of the 18th Int'l Conf. Very Large Data Bases*, p. 27, 1992.

[9] P. Flajolet and G.N. Martin. Probabilistic Counting Algorithms for Data Base Applications. *Journal of Computer and System Sciences*, 18(2) 143-154, 1979.

[10] I.J. Good. Surprise Indexes and P-values. *Journal of Statistical Computation and Simulation* 32, p 90-92, 1989.

[11] P. Indyk and D. Woodruff. Tight Lower Bounds for the Distinct Elements Problem. To appear: *Foundations of Computer Science*, 2003. Available: http://web.mit.edu/dpwood/www

[12] I. Kremer, N. Nisan, and D. Ron. On randomized one-round communication complexity. *Computational Complexity*, 8(1):21-49, 1999.

[13] B.D. McKay, I.M. Wanless, and N.C. Wormald. Asymptotic enumeration of graphs with a given upper bound on the maximum degree, *Combinatorics, Probability and Computing 11* p. 373-392, 2002.

[14] V.N. Vapnik and A.Y. Chervonenkis. On the uniform converges of relative frequences of events to their probabilities. *Theory of Probability and its Applications*, XVI(2):264-280, 1971.

[15] A. C-C. Yao. Lower bounds by probabilistic arguments. In *Proceedings of the 24th Annual IEEE Symposium on Foundations of Computer Science*, p. 420-428, 1983.