EXPLICIT EXPONENTIAL LOWER BOUNDS FOR EXACT HYPERPLANE COVERS

Benjamin E. DIAMOND
Applied Cryptography
Coinbase
benediamond@gmail.com

Amir YEHUDAYOFF*
Department of Mathematics
Technion – IIT
amir.yehudayoff@gmail.com

Abstract
We describe an explicit and simple subset of the discrete hypercube which cannot be exactly covered by fewer than exponentially many hyperplanes. The proof exploits a connection to communication complexity, and relies heavily on Razborov’s lower bound for disjointness.

1 Introduction

The relationship between hyperplanes in Euclidean space and the discrete hypercube is fundamental and important. One basic problem entails understanding the number of hyperplanes which one must use to cover various subsets of the cube. A hyperplane is specified by a perpendicular vector \((a_1, \ldots, a_n) \in \mathbb{R}^n\) and an offset \(w \in \mathbb{R}\), defined specifically as the set \(H = \{x \in \mathbb{R}^n | \sum_{i=1}^{n} a_i \cdot x_i = w\}\). A collection of hyperplanes \(H_1, H_2, \ldots, H_N\) exactly covers a set \(S \subseteq \{0, 1\}^n\) if \(\bigcup_{i=1}^{N} H_i \cap \{0, 1\}^n = S\). The exact cover number \(ec(S)\) of a set \(S\) is the minimum cardinality \(N\) attained across all hyperplane configurations exactly covering \(S\).

The exact cover number \(ec(\{0, 1\}^n)\) of the full cube is 2. Seminal work of Alon and Füredi [AF93] shows that removing a single point makes exact covering much harder; the exact cover number of \(\{0, 1\}^n \setminus \{(0, \ldots, 0)\}\) is \(n\). The study of exact covers of arbitrary sets appears in recent work of Aaronson, Groenland, Grzesik, Johnston, and Kielak [AGG+21]. That work focuses on worst-case cardinalities of the form \(ec(n) := \max_{S \subseteq \{0, 1\}^n} ec(S)\); it proves that \(ec(n)\) is between \(\frac{2^n}{n^2}\) and \(2^n - \lceil \log n \rceil < 2 \cdot 2^n - \frac{n}{2}\). The work’s lower bound on \(ec(n)\) is not explicit, as it relies on a generic counting argument.

In this work, we analyze \(ec(S)\) for concrete subsets \(S \subseteq \{0, 1\}^n\). Beyond this problem’s intrinsic appeal, an additional strong source of motivation stems from forthcoming work by the first-listed author [Dia22], which links exact hyperplane covers to secure two-party computation. That work shows that exact covers yield protocols for secure computation by two malicious parties. The work’s protocols, moreover, are efficient when the relevant exact cover numbers are small. It is of interest, therefore, to determine which set families are and are not efficiently coverable.

Our main result exhibits a concrete set whose exact cover number is exponentially large.

Definition. We write \(D_n \subseteq \{0, 1\}^{\lceil n/2 \rceil} \times \{0, 1\}^{\lfloor n/2 \rfloor} \cong \{0, 1\}^n\) for the set
\[
D_n = \{(x, y) \in \{0, 1\}^{\lceil n/2 \rceil} \times \{0, 1\}^{\lfloor n/2 \rfloor} | \left( \bigvee_{i=1}^{\lceil n/2 \rceil} x_i \land y_i \right) = 0 \}.
\]
In other words, \(D_n\) consists exactly of those pairs \((x, y)\) for which \(x\) and \(y\)—interpreted as sets—are disjoint.

Theorem. \(ec(D_n) \geq 2^{\Omega(n)}\).

Though we focus throughout this work on hyperplanes defined over the real numbers, our results in fact carry through to any field, and even to “hyperplanes” defined over \(\mathbb{Z}\) (i.e., to rank-\(n-1\) submodules of \(\mathbb{Z}^n\)).

*Research funded by NSF–BSF grant number 2021674.
We briefly discuss a further interpretation of the theorem. Exact cover numbers can be understood as furnishing a sort of complexity measure on boolean functions. In computational complexity, functions are computed by devices, where each device has a cost; the complexity of a function is the minimum cost of a device computing it. In our setting, the “devices” are hyperplane covers, and the “cost” of a hyperplane cover is its cardinality. Specifically, to each boolean function $f : \{0,1\}^n \to \{0,1\}$, we may associate, for example, the complexity measure $ec(f) := \max \{ ec(f^{-1}(0)), ec(f^{-1}(1)) \}$.

Exact cover complexity relates in surprising ways to other, more standard, complexity measures, including, for example, that of constant-depth circuit complexity, a classically studied metric (see e.g. [Vol99 § 3.3]). We demonstrate this through the following example. A function $f$ is symmetric if it is invariant under arbitrary permutations of its inputs (parity and majority are symmetric functions, for example). Symmetric functions in general admit only superpolynomially sized constant-depth circuits (see e.g. Vollmer [Vol99 Cor. 3.32]).

On the other hand, if $f$ is symmetric, then $ec(f)$ is at most $n + 1$. This latter fact can be seen in the following way. For each symmetric $f$, $f^{-1}(1)$ is a union of sets of the form $S_j := \{ x \in \{0,1\}^n \mid \sum_{i=1}^n x_i = j \}$, for constants $j \in \{0, \ldots, n\}$. Each individual set $S_j$ can be exactly covered by a single hyperplane, so that $f^{-1}(1)$ can be exactly covered by $n + 1$ hyperplanes. A similar argument applies to $f^{-1}(0)$. We see that symmetric functions have small exact cover complexity, despite their large constant-depth circuit complexity.

In the opposite direction, our theorem shows that the exact cover complexity of a polynomially-sized, depth-two, monotone circuit can be exponential in $n$. This discrepancy demonstrates a strong separation between exact cover complexity and constant-depth circuit complexity.

We prove our lower bound on the exact cover number of $D_n$ by upper-bounding the sizes of certain sets of the form $H \cap D_n$, where $H \subseteq \mathbb{R}^n$ is a hyperplane. No useful such upper bound, of course, can possibly hold for all hyperplanes. Slightly abusing terminology, we say that a hyperplane $H \subseteq \mathbb{R}^n$ is contained in a set $S \subseteq \{0,1\}^n$, if $H \cap \{0,1\}^n \subseteq S$. If a hyperplane $H$ is contained in $S$, then, trivially, $|H \cap \{0,1\}^n| \leq |S|$ holds. The following lemma establishes an exponential improvement in the particular case of $D_n$:

**Lemma.** If a hyperplane $H \subseteq \mathbb{R}^n$ is contained in $D_n$, then $|H \cap \{0,1\}^n| \leq 2^{-\Omega(n)} \cdot |D_n|$.

The lemma implies the theorem by means of a covering argument, which we presently sketch. Each hyperplane $H$ which participates in an exact cover of $D_n$ must be contained in $D_n$. The lemma entails that each particular such $H$ may alone cover at most a proportion of $2^{-\Omega(n)}$ among $D_n$’s points. It follows that at least $2^{\Omega(n)}$ hyperplanes must be used in any configuration exactly covering $D_n$.

The lemma may be interpreted as a strong—though restricted—anti-concentration result. Classical anti-concentration results concern expressions of the form $\max_{w \in \mathbb{R}} \Pr[\sum_{i=1}^n a_i \cdot X_i = w]$, where $X$ is uniformly distributed in $\{0,1\}^n$. The Littlewood–Offord problem entails establishing anti-concentration when all of the coordinates of $(a_1, \ldots, a_n)$ are assumed to be nonzero [LO43]; the problem’s original motivation arose from the study of roots of random polynomials. Littlewood and Offord proved the preliminary upper bound of $O\left(\frac{n}{\sqrt{n}}\right)$. In a celebrated and sharp result, Erdős [Erd45] solved the Littlewood–Offord problem, proving the upper bound $2^{-n} \cdot \binom{n}{n/2} = \Theta\left(\frac{1}{\sqrt{n}}\right)$ using Sperner’s theorem on the sizes of antichains. Kleitman [Kle65], Frankl and Füredi [FF88], Griggs [Gri83], and others subsequently generalized the problem.

The lemma says that for each normal $a \in \mathbb{R}^n$, we have $\max_{w \in \mathbb{R}} \Pr[\sum_{i=1}^n a_i \cdot X_i = w] \leq 2^{-\Omega(n)} \cdot \Pr[X \in D_n]$, where the maximum is taken not over all constants $w \in \mathbb{R}$, but rather over only those for which the hyperplane $\{ x \in \mathbb{R}^n \mid \sum_{i=1}^n a_i \cdot x_i = w \}$ is contained in $D_n$. The lemma holds for all $a$, and guarantees exponentially strong anti-concentration; on the other hand, the bound holds only for certain $w$. The fact that the bound holds only for some among the values $w$ makes it difficult to use known techniques to prove anti-concentration (like extremal combinatorics, or Fourier analysis).

Our main high-level contribution is a bridge between this sort of restricted anti-concentration and two-party communication complexity (see the textbook [RY20] and references within). Our proof follows the ideas of Razborov’s [Raz92] famous lower bound for the distributional two-party communication complexity of disjointness. This bridge is built by the means of a certain decomposition. For each hyperplane $H = \{ (x, y) \in \{0,1\}^{n/2} \times \{0,1\}^{n/2} \mid \sum_{i=1}^{n/2} a_i \cdot x_i + \sum_{i=1}^{n/2} b_i \cdot y_i = w \}$, we have:

$$H \cap D_n = \bigcup_{k \in \mathbb{R}} (A_k \times B_k) \cap D_n,$$

(1)

where $A_k := \{ x \in \{0,1\}^{n/2} \mid \sum_{i=1}^{n/2} a_i \cdot x_i = k \}$ and $B_k := \{ y \in \{0,1\}^{n/2} \mid \sum_{i=1}^{n/2} b_i \cdot y_i = w - k \}$. 

2
In the language of communication complexity, for each $k$, the set $A_k \times B_k$ is a rectangle (i.e., a product set in $\{0,1\}^{n/2} \times \{0,1\}^{n/2}$); the assumption $H \subseteq D_n$ entails precisely that each rectangle $A_k \times B_k$ is monochromatic (i.e., it satisfies $A_k \times B_k \subseteq D_n$). In a nutshell, we see that if $\text{ec}(D_n)$ were small, then $D_n$ would resemble the on-set of a function with low communication complexity. The full picture is in fact subtler, as we explain below.

The crux of Razborov’s lower bound asserts that, in the setting of disjointness, all 1-monochromatic rectangles are small. To assess their sizes, Razborov carefully constructs a probability measure $\rho$ on the cube, and proves that each 1-monochromatic rectangle has an exponentially small mass—say, $2^{-c \cdot n}$—under this measure. In our language, Razborov proves that if a rectangle $A \times B$ is contained in $D_n$, then $\rho(A \times B) \leq 2^{-c \cdot n}$. We must surmount a few barriers in order to apply Razborov’s ideas in our context.

The first among these barriers concerns the possible number of rectangles used (i.e., the number of $\rho$-upper-bound the total probability mass they represent. The key observation which overcomes this barrier is that our rectangles have a very specific structure; indeed, the sets $A_\rho \in \{0,1\}$ are themselves pairwise disjoint in $\{0,1\}^{n/2}$, and likewise for the sets $B_\rho \in \{0,1\}^{n/2}$. This observation, together with a more careful analysis, allows us to overcome the first barrier. Interestingly, we exploit the structure of $H$ as a hyperplane during our proof only in our use of this simple property of the set families $\{A_\rho\}_{\rho \in \mathbb{R}}$ and $\{B_\rho\}_{\rho \in \mathbb{R}}$.

A second barrier that we must overcome stems from the fact that the distribution on $D_n$ we consider is uniform; Razborov’s argument exploits the carefully constructed distribution $\rho$. This difference introduces several technical difficulties. Roughly speaking, we use measure concentration to reduce our problem to a setting closer to Razborov’s; we then analyze a “perturbed” variant of his distribution (see Claim 2.1 below).

Our main lemma above conceals an implicit small linear constant within its expression $\Omega(n)$, which is moreover ineffective throughout our proof. We suspect that the following precise variant of our main result holds:

**Conjecture.** If a hyperplane $H \subseteq \mathbb{R}^n$ is contained in $D_n$, then $|H \cap \{0,1\}^n| \leq 2^{n/2}$.

That is, we suspect the precise variant of our above lemma whereby $|H \cap \{0,1\}^n| \leq 2^{(1-\log_2 3) \cdot n/2} \cdot |D_n|$. This conjecture is sharp, in that there exist hyperplanes $H$ contained in $D_n$ for which $|H \cap \{0,1\}^n| = 2^{n/2}$. For example, we may take as $H$ the hyperplane $\{(x, y) \in \{0,1\}^{n/2} \times \{0,1\}^{n/2} \mid \sum_{i=1}^{n/2} a_i \cdot x_i + \sum_{i=1}^{n/2} b_i \cdot y_i = w\}$. The intersection of this $H$ with $\{0,1\}^{n/2} \times \{0,1\}^{n/2}$ is exactly the set $\{(0, \ldots, 0) \times \{0,1\}^{n/2} \text{ consisting of pairs } (x, y) \text{ for which } x \text{ is empty} \}$.

This set is obviously contained in $D_n$, and consists of exactly $2^{n/2}$ points.

## 2 Proving the Lemma

We fix even $n$ and a hyperplane

$$H = \{(x, y) \in \{0,1\}^{n/2} \times \{0,1\}^{n/2} \mid \sum_{i=1}^{n/2} a_i \cdot x_i + \sum_{i=1}^{n/2} b_i \cdot y_i = w\} \subseteq \mathbb{R}^n,$$

which we moreover assume is contained in $D := D_n$. We write $\mu$ for the uniform distribution on $D \subseteq \{0,1\}^n$. Throughout, we frequently interpret elements $x$ and $y$ of $\{0,1\}^{n/2}$ as subsets of $\{1, \ldots, n/2\}$, and use corresponding notation. For example, we write $|x|$ and $|y|$ for the cardinalities—that is, the Hamming weights—of $x$ and $y$. We partition $D$ along the sizes of its two constituent sets, in the following way. For integers $\ell_x$ and $\ell_y$ in $\{0, \ldots, \frac{n}{2}\}$, we set:

$$D_{\ell_x, \ell_y} := \{(x, y) \in \{0,1\}^{n/2} \times \{0,1\}^{n/2} \mid |x| = \ell_x \wedge |y| = \ell_y\}.$$

We first argue that all but an exponentially vanishing proportion of the mass of $\mu$ is concentrated within those $D_{\ell_x, \ell_y}$ for which both $\ell_x$ and $\ell_y$ are simultaneously near $\frac{n}{4}$.

**Claim 2.1.** For each constant $\delta > 0$, we have $\mu\left(\bigcup_{n/4 - \delta \leq \ell_x, \ell_y \leq n/4 + \delta} D_{\ell_x, \ell_y}\right) > 1 - 2^{-\Omega(n)}$. 

3
Proof. A pair \((X, Y)\) distributed according to \(\mu\) may be sampled using \(\frac{n}{2}\) i.i.d. pairs \((X_1, Y_1), \ldots, (X_{n/2}, Y_{n/2})\), each uniform in \(\{(0, 0), (0, 1), (1, 0)\}\). The expected values of \(\sum_{i=1}^{n/2} X_i\) and of \(\sum_{i=1}^{n/2} Y_i\) are exactly \(\frac{n}{6}\). The union bound, together with a standard application of Chernoff’s bound, completes the proof. 

We write \(\mu_{\ell_x, \ell_y}\) for the distribution \(\mu\) conditioned on \(D_{\ell_x, \ell_y}\). In light of Claim 2.1, it suffices to prove that \(\mu_{\ell_x, \ell_y}(H) \leq 2^{-\Omega(n)}\) holds whenever \(\ell_x\) and \(\ell_y\) simultaneously reside in \([n \cdot (\frac{1}{6} - \delta), n \cdot (\frac{1}{6} + \delta)]\), for some appropriate fixed \(\delta > 0\). Throughout the remainder of the proof, we fix \(\delta := \frac{1}{500}\), as well as arbitrary integers \(\ell_x\) and \(\ell_y\) in \([\frac{1}{150} \cdot n, \frac{51}{150} \cdot n]\). To bound \(\mu_{\ell_x, \ell_y}(H)\), we use the decomposition (1). For fixed \(k \in \mathbb{R}\), we write:

\[
A_k := \left\{ x \in \{0, 1\}^{n/2} \left| \sum_{i=1}^{n/2} a_i \cdot x_i = k \right. \right\} \quad \text{and} \quad B_k := \left\{ y \in \{0, 1\}^{n/2} \left| \sum_{i=1}^{n/2} b_i \cdot y_i = w - k \right. \right\}.
\]

A key step in our proof involves sampling from \(\mu_{\ell_x, \ell_y}\) in a more informative way.

Remark. We denote random variables by capital letters, and by lowercase letters the values they attain. In what follows, we impose the further assumption whereby \(\frac{n}{2}\) is odd, so that \(m := \frac{n - 2}{4}\) is an integer. This restriction is not necessary, but simplifies notation below.

We write \(T := (Z_x, Z_y, \{I\})\) for a uniformly random partition of \(\{1, \ldots, \frac{n}{2}\}\) into subsets sized exactly \(m\), \(m\), and 1, respectively. We write \(X\) for a uniformly random subset of \(Z_x \cup \{I\}\) of cardinality exactly \(\ell_x\) and \(Y\) for a uniformly random subset of \(Z_y \cup \{I\}\) of cardinality exactly \(\ell_y\) \((X, Y)\) are chosen independently). We moreover write \(X_0\) and \(Y_0\) for \(X\) and \(Y\) conditioned on \(I \not\subseteq X\) and \(I \not\subseteq Y\), respectively. Finally, we write \(X_1\) and \(Y_1\) for \(X\) and \(Y\) conditioned on \(I \subseteq X\) and \(I \subseteq Y\), respectively.

We intuitively characterize the role which this sampling procedure plays in our proof. The core of our argument shows that for each arbitrary rectangle \(A \times B \subseteq \{0, 1\}^{n/2} \times \{0, 1\}^{n/2}\) (including, crucially, the rectangles \(A_k \times B_k\) constructed above), the probability that \((X_0, Y_0) \in A \times B\) cannot significantly exceed the probability that \((X_1, Y_1) \in A \times B\). This implies in particular that \(H \cap D_n\) is small for any \(H \subseteq \mathbb{R}^n\) contained in \(D_n\), since \((X_1, Y_1)\) never falls within \(D_n\), whereas \((X_0, Y_0)\) always does.

2.1 Good partitions

In this subsection, we introduce and analyze the notion of “good” partitions, and describe their basic properties.

Definition 2.2. For each \(k \in \mathbb{R}\), we define sets of “good” partitions in the following way:

\[
G_x^k := \left\{ t = (z_x, z_y, \{i\}) \left| \Pr[X_1 \in A_k \mid T = t] \geq \frac{1}{10} \cdot \Pr[X_0 \in A_k \mid T = t] - 2^{-\varepsilon \cdot n} \right. \right\}
\]

and

\[
G_y^k := \left\{ t = (z_x, z_y, \{i\}) \left| \Pr[Y_1 \in B_k \mid T = t] \geq \frac{1}{10} \cdot \Pr[Y_0 \in B_k \mid T = t] - 2^{-\varepsilon \cdot n} \right. \right\},
\]

where \(\varepsilon := \frac{1}{500}\).

Roughly speaking, a partition \(t = (z_x, z_y, \{i\})\) is “good” with respect to \(k \in \mathbb{R}\) if, conditioned upon it, the distributions \((X_0, Y_0)\) and \((X_1, Y_1)\) do not intersect the sets \(A_k\) and \(B_k\) excessively differently. The main result of this subsection is the following proposition, which states that most partitions are “good”. We write \(\chi_x^k(t)\) and \(\chi_y^k(t)\) for the indicator functions of the events \(t \in G_x^k\) and \(t \in G_y^k\), respectively.

Proposition 2.3. For each \(k \in \mathbb{R}\),

\[
\mathbb{E}_T \left[ \Pr[X_0 \in A_k \mid T] \cdot \Pr[Y_0 \in B_k \mid T] \cdot \chi_x^k(T) \cdot \chi_y^k(T) \right] \geq \frac{1}{10} \cdot \mathbb{E}_T \left[ \Pr[X_0 \in A_k \mid T] \cdot \Pr[Y_0 \in B_k \mid T] \right].
\]

Before proving the proposition, we establish a few preliminary claims. The first records a fact pertaining to the structure of this probability distribution upon conditioning.
Claim 2.4. For each $k \in \mathbb{R}$, as the partition $t = (z_x, z_y, \{i\})$ varies, the numbers $\Pr[X \in A_k \mid T = t]$ and $\Pr[Y_0 \in B_k \mid T = t]$ depend only on $z_y$ and the numbers $\Pr[Y \in B_k \mid T = t]$ and $\Pr[X_0 \in A_k \mid T = t]$ depend only on $z_x$.

The following claim compares the probability that a random $\ell_x$-element subset of $z_x \subseteq \{1, \ldots, \frac{n}{2}\}$ resides in $A_k \subseteq \{0, 1\}^{n/2}$ with the probability that a random $\ell_x$-element subset of $z_x \cup \{i\}$ does. It shows that the latter exceeds the former by much more than threefold. It exploits the fact that the probability that a random $\ell_x$-element subset of $z_x \cup \{i\}$ does not contain $i$ is close to $\frac{1}{3}$.

Claim 2.5. For each fixed scalar $k \in \mathbb{R}$ and partition $t = (z_x, z_y, \{i\})$, we have

$$\Pr[X_0 \in A_k \mid T = t] \leq \frac{25}{8} \cdot \Pr[X \in A_k \mid T = t]$$

and

$$\Pr[Y_0 \in B_k \mid T = t] \leq \frac{25}{8} \cdot \Pr[Y \in B_k \mid T = t].$$

Proof. We prove the first conclusion; the second is similar. By the definitions of the distributions $X$ and $X_0$,

$$\Pr[X \in A_k \mid T = t] = \Pr[i \notin X \mid T = t] \cdot \Pr[X_0 \in A_k \mid T = t] + \Pr[i \in X \mid T = t] \cdot \Pr[X_0 \in A_k \mid T = t] \geq \Pr[i \notin X \mid T = t] \cdot \Pr[X_0 \in A_k \mid T = t].$$

The proportion of $\ell_x$-element subsets of $z_x \cup \{i\}$ which do not contain $i$ is

$$\Pr[i \notin X \mid T = t] = \binom{n}{\ell_x} = \frac{m+1-\ell_x}{m+1} \geq 1 - \frac{n}{m+2},$$

$$\frac{17}{25} \geq \frac{8}{25}.$$

The following claim shows that, for each fixed $z_y \subseteq \{1, \ldots, \frac{n}{2}\}$, as the index $i \in \{1, \ldots, \frac{n}{2}\} - z_y$ varies, most among the resulting partitions $(z_x, z_y, \{i\})$ are “good”; a symmetrical statement holds for each fixed $z_x \subseteq \{1, \ldots, \frac{n}{2}\}$ as the index $i \in \{1, \ldots, \frac{n}{2}\} - z_x$ varies.

Claim 2.6. For each fixed scalar $k \in \mathbb{R}$, and arbitrary fixed subsets $z_x$ and $z_y$ of $\{1, \ldots, \frac{n}{2}\}$, we have:

$$\Pr[T \notin G^k_x \mid Z_y = z_y] < \frac{1}{7}$$

and

$$\Pr[T \notin G^k_y \mid Z_x = z_x] < \frac{1}{7}.$$

Proof. We prove only the first inequality, as the second is similar. We first handle the case in which $\Pr[X \in A_k \mid Z_y = z_y] < 2^{-c_n}$. In light of Claim 2.4, we note that $\Pr[X \in A_k \mid Z_y = z_y] = \Pr[X \in A_k \mid T = t]$ holds for each particular $t$ partitioned from the distribution $(Z_x, z_y, \{I\})$. Using Claim 2.5, we see that if any particular such $t$ moreover satisfied $t \notin G^k_x$, then we would have:

$$\Pr[X_1 \in A_k \mid T = t] < \frac{1}{30} \cdot \Pr[X_0 \in A_k \mid T = t] - 2^{-c_n} \leq \frac{1}{30} \cdot \Pr[X \in A_k \mid T = t] - 2^{-c_n} < 0,$$

a contradiction, so that $\Pr[T \notin G^k_x \mid Z_y = z_y] = 0$, and the claim is proved.

We thus assume that $\Pr[X \in A_k \mid Z_y = z_y] \geq 2^{-c_n}$. We write $\overline{z_y}$ for the complement of $z_y$ in $\{1, \ldots, \frac{n}{2}\}$, and abbreviate $\widehat{A}_k := A_k \cap (\overline{z_y})$ for the set of $\ell_x$-element subsets of $\overline{z_y}$ which reside in $A_k$. We note that:

$$\Pr[X \in A_k \mid Z_y = z_y] = \frac{|\widehat{A}_k|}{\binom{m+1}{\ell_x}},$$

so that $|\widehat{A}_k| \geq \binom{m+1}{\ell_x} \cdot \Pr[X \in A_k \mid Z_y = z_y] \geq \binom{m+1}{\ell_x} \cdot 2^{-c_n}$. We moreover record the lower bound

$$\log_2 \left( \frac{m+1}{\ell_x} \right) \geq \log_2 \left( \frac{m+1}{\left\lfloor \frac{17}{25} \cdot n \right\rfloor} \right) \geq \log_2 \left( \frac{m+1}{\left\lfloor \frac{17}{25} \cdot (m+1) \right\rfloor} \right) \geq (0.9 - o(1)) \cdot (m+1);$$

the last inequality is a standard consequence of Stirling’s approximation, together with the binary entropy inequality $H \left( \frac{17}{25} \right) > 0.9$. These facts together imply that $\log_2 \left( |\widehat{A}_k| \right) \geq (0.9 - o(1) - 4 \cdot \varepsilon) \cdot (m+1)$. 

5
We write $\hat{X}$ for a uniformly random element of $\hat{A}_k$. We fix a partition $t = (z_x, z_y, \{i\})$, and abbreviate $\hat{A}_{k,1}$ for the set of elements of $\hat{A}_k$ which contain $i$; we now have that:

$$\Pr[X \in A_k \mid T = t] \cdot \Pr[i \in \hat{X}] = \frac{|\hat{A}_k|}{(m+1)} \cdot \frac{|\hat{A}_{k,1}|}{|A_k|} = \frac{|\hat{A}_{k,1}|}{(\ell_x - 1)} = \Pr[X_1 \in A_k \mid T = t] \cdot \frac{\ell_x}{m + 1},$$

so that $\Pr[X \in A_k \mid T = t] \cdot \Pr[i \in \hat{X}] \leq \Pr[X_1 \in A_k \mid T = t] \cdot \frac{17}{25}$. By this inequality and Claim 2.5, we see that if \textit{moreover} $t \notin G_x^k$ holds, then we have:

$$\Pr[X \in A_k \mid T = t] \cdot \Pr[i \in \hat{X}] < \frac{1}{105} \cdot \Pr[X_0 \in A_k \mid T = t] \leq \frac{1}{30} \cdot \Pr[X \in A_k \mid T = t],$$

so that $\Pr[i \in \hat{X}] < \frac{1}{30}$. Equivalently, if the partition $t = (z_x, z_y, \{i\})$ is not “good”, then the component of the joint distribution $\hat{X}$ corresponding to the element $i \in \pi_y$ has success probability less than $\frac{1}{30}$.

We write $\hat{X}_t^k$ for the indicator function of the event $i_j \in \hat{X}$, where $\{i_1, \ldots, i_{m+1}\}$ are the elements of $\pi_y$, so that $\hat{X} = (\hat{X}_1^1, \ldots, \hat{X}_{m+1}^1)$. We observe that if the claim were false—and, in particular, $\Pr[i \in \hat{X}] < \frac{1}{30}$ held for a proportion consisting of at least $\frac{1}{4}$ among the $m + 1$ elements $i \in \pi_y$—then the binary entropy of $\hat{X}$ would satisfy:

$$(0.9 - o(1) - 4 \cdot \varepsilon) \cdot (m + 1) \leq H(\hat{X}) \quad \text{(by } H(\hat{X}) = \log_2 \left(|\hat{A}_k|\right) \text{ and the above)}$$

$$\leq \sum_{j=1}^{m+1} H(\hat{X}_j) \quad \text{(by the sub-additivity of entropy)}$$

$$< (\frac{4}{7} + \frac{1}{7} \cdot H(\frac{1}{30})) \cdot (m + 1) \quad \text{(by the assumption that the claim is false)}$$

$$\leq 0.89 \cdot (m + 1).$$

This contradiction completes the proof of the claim.

We are now ready to prove the main proposition.

\textit{Proof of Proposition 2.3}. Because $1 - \chi^k_x(t) \cdot \chi^k_y(t) \leq 1 - \chi^k_x(t) + 1 - \chi^k_y(t)$ holds for each $t$, and by linearity of expectation and symmetry, it suffices to prove that

$$E_T \left[ \Pr[X_0 \in A_k \mid T] \cdot \Pr[Y_0 \in B_k \mid T] \cdot (1 - \chi^k_x(T)) \right] \leq \frac{9}{25} \cdot E_T \left[ \Pr[X_0 \in A_k \mid T] \cdot \Pr[Y_0 \in B_k \mid T] \right].$$

To prove this, it in turn suffices to show that, for each fixed $m$-element subset $z_y \subseteq \{1, \ldots, \frac{1}{2}\}$, it holds that:

$$E_T \left[ \Pr[X_0 \in A_k \mid T] \cdot \Pr[Y_0 \in B_k \mid T] \cdot (1 - \chi^k_x(T)) \mid Z_y = z_y \right] \leq \frac{9}{25} \cdot E_T \left[ \Pr[X_0 \in A_k \mid T] \cdot \Pr[Y_0 \in B_k \mid T] \mid Z_y = z_y \right].$$

We prove this latter claim in the following way:

$$E_T \left[ \Pr[X_0 \in A_k \mid T] \cdot \Pr[Y_0 \in B_k \mid T] \cdot (1 - \chi^k_x(T)) \mid Z_y = z_y \right] = \Pr[Y_0 \in B_k \mid Z_y = z_y] \cdot E_T \left[ \Pr[X_0 \in A_k \mid T] \cdot (1 - \chi^k_x(T)) \mid Z_y = z_y \right] \quad \text{(by Claim 2.4)}$$

$$\leq \frac{25}{27} \cdot \Pr[Y_0 \in B_k \mid Z_y = z_y] \cdot E_T \left[ \Pr[X \in A_k \mid T] \cdot (1 - \chi^k_x(T)) \mid Z_y = z_y \right] \quad \text{(by Claim 2.5)}$$

$$= \frac{25}{27} \cdot \Pr[Y_0 \in B_k \mid Z_y = z_y] \cdot \Pr[X \in A_k \mid Z_y = z_y] \cdot E_T \left[ 1 - \chi^k_x(T) \mid Z_y = z_y \right] \quad \text{(by Claim 2.4)}$$

$$\leq \frac{25}{27} \cdot \frac{1}{7} \cdot \Pr[Y_0 \in B_k \mid Z_y = z_y] \cdot \Pr[X \in A_k \mid Z_y = z_y] \quad \text{(by Claim 2.6)}$$

$$= \frac{25}{27} \cdot \Pr[Y_0 \in B_k \mid Z_y = z_y] \cdot \Pr[X \in A_k \mid T] \mid Z_y = z_y].$$

The final equality above amounts to the following calculation:

$$\Pr[X \in A_k \mid Z_y = z_y] = \frac{|A_k \cap (\frac{7}{\ell_x})|}{m+1} \cdot \frac{|A_k \cap (\frac{7}{\ell_x})| \cdot (m + 1 - \ell_x)}{m+1} = \frac{1}{m+1} \cdot \sum_{i \in \pi_y} |A_k \cap (\frac{7}{\ell_x} - (i))|,$$

where the rightmost expression is precisely $E_T [\Pr[X_0 \in A_k \mid T] \mid Z_y = z_y]$, by definition, and the final equality stems from a double-counting argument; indeed, $\sum_{i \in \pi_y} |A_k \cap (\frac{7}{\ell_x} - (i))|$ counts each distinct element of $A_k \cap (\frac{7}{\ell_x})$ exactly $m + 1 - \ell_x$ times. An additional application of Claim 2.4 completes the proof. □


2.2 Completing the argument

We record the following claim:

Claim 2.7. $\sum_{k \in \mathbb{R}} \mathbb{E}_T [\Pr [X_0 \in A_k \mid T]] \leq 1$ and $\sum_{k \in \mathbb{R}} \mathbb{E}_T [\Pr [Y_0 \in B_k \mid T]] \leq 1$.

Proof. We prove the first inequality. Because the sets $A_k$—as $k$ ranges throughout $\mathbb{R}$—are pairwise disjoint, $\sum_{k \in \mathbb{R}} \mathbb{E}_T [\Pr [X_0 \in A_k \mid T]] = \sum_{k \in \mathbb{R}} \Pr [X_0 \in A_k] \leq \Pr [X_0 \in \bigcup_{k \in \mathbb{R}} A_k] \leq 1$.

This completes the proof.

We are now in a position to prove the lemma. Invoking finally the hypothesis whereby $H$ is contained in $D_n$, we have that:

$0 = \Pr [(X_1, Y_1) \in D_n] \geq \Pr [(X_1, Y_1) \in H]$.

Applying now the decomposition (1), we have:

$0 = \Pr [(X_1, Y_1) \in H] \geq \sum_{k \in \mathbb{R}} \mathbb{E}_T [\Pr [X_1 \in A_k \mid T] \cdot \Pr [Y_1 \in B_k \mid T] \cdot \chi^k_x(T) \cdot \chi^k_y(T)]$.

Invoking Definition 2.2, we see further that:

$0 \geq \sum_{k \in \mathbb{R}} \mathbb{E}_T \left[ \left( \frac{1}{\sqrt{n}} \cdot \Pr [X_0 \in A_k \mid T] - 2^{-\varepsilon n} \right) \cdot \left( \frac{1}{\sqrt{n}} \cdot \Pr [Y_0 \in B_k \mid T] - 2^{-\varepsilon n} \right) \cdot \chi^k_x(T) \cdot \chi^k_y(T) \right]$. 

Applying Claim 2.7, we have that:

$0 \geq \frac{1}{10 \cdot \sqrt{n}} \cdot \sum_{k \in \mathbb{R}} \mathbb{E}_T [\Pr [X_0 \in A_k \mid T] \cdot \Pr [Y_0 \in B_k \mid T] \cdot \chi^k_x(T) \cdot \chi^k_y(T)] - 2^{-\Omega(n)}$.

Using Proposition 2.3 we conclude that:

$0 \geq \frac{1}{10 \cdot \sqrt{n}} \cdot \sum_{k \in \mathbb{R}} \mathbb{E}_T [\Pr [X_0 \in A_k \mid T] \cdot \Pr [Y_0 \in B_k \mid T]] - 2^{-\Omega(n)}$

$= \frac{1}{10 \cdot \sqrt{n}} \cdot \mathbb{E}_T \left[ \sum_{k \in \mathbb{R}} \Pr [X_0 \in A_k \mid T] \cdot \Pr [Y_0 \in B_k \mid T] \right] - 2^{-\Omega(n)}$

$= \frac{1}{10 \cdot \sqrt{n}} \cdot \mu_{x, \ell_x}(H) - 2^{-\Omega(n)}$.

In light of Claim 2.1, this calculation completes the proof of the lemma.

References


