ARE MOST BOOLEAN FUNCTIONS DETERMINED BY LOW FREQUENCIES?

ROMAN VERSHYNIN

Abstract. We ask whether most Boolean functions are determined by their low frequencies. We show a partial result: for almost every function $f : \{-1,1\}^p \to \{-1,1\}$ there exists a function $f' : \{-1,1\}^p \to (-1,1)$ that has the same frequencies as f up to dimension $(1/2 - o(1))p$.

1. Introduction

1.1. **Uniqueness of Boolean data.** When is a high-dimensional distribution determined by its low-dimensional marginals? To be specific, consider a random vector $X = (X_1, \ldots, X_p)$ that takes values in $\{0,1\}^p$. We may wonder if there exist a random vector $Y = (Y_1, \ldots, Y_p)$ that also takes values in $\{0,1\}^p$, whose all marginal distributions up to a given dimension $d < p$ are the same as those[1] of X, but whose distribution is different from that of X everywhere on the cube.[2] If this does happen, we call the distribution of X *non-unique* with respect to marginals up to dimension d.

Some distributions are very much unique. For example, it is easy to check the following:

Example 1.1. If $X_1 = \cdots = X_p$ almost surely, then the distribution of X is unique with respect to marginals up to dimension 2.

However, we will show that "most" distributions on the cube are not unique with respect to marginals up to almost half the dimension:

Theorem 1.2. *If p is sufficiently large, then for most of the subsets $S \subset \{0,1\}^p$ the uniform distribution on S is non-unique with respect to the marginals of dimension $0.49p$.*

As we will see from the proof, "most" means all but at most 2^{cp} subsets where $c \in (0,1)$ is an absolute constant, and the number 0.49 can be replaced by any constant smaller than 1/2.

Partially supported by NSF DMS-1954233, NSF DMS-2027299, U.S. Army 76649-CS, and NSF+Simons Research Collaborations on the Mathematical and Scientific Foundations of Deep Learning.

[1]By this we mean that for any subset of coordinates J or cardinality $|J| \leq d$, the distributions of the random vectors $(X_j)_{j \in J}$ and $(Y_j)_{j \in J}$ are the same.

[2]By this we mean that $\mathbb{P}\left\{X = \theta\right\} \neq \mathbb{P}\left\{Y = \theta\right\}$ for all $\theta \in \{0,1\}^p$.

We will prove Theorem 1.2 by showing that for most S, there exists a random vector $Y = Y(S) \in \{0,1\}^p$ that has the same marginals up to dimension $0.49p$ as random vector X uniformly distributed on S, and yet

$$0 < \mathbb{P}\left\{Y = \theta\right\} < \frac{1}{|S|} \quad \text{for all } \theta \in \{0,1\}^p.$$

This yields non-uniqueness, since $\mathbb{P}\left\{X = \theta\right\}$ takes values 0 and $1/|S|$ only.

1.2. **Uniqueness of Boolean functions.** Theorem 1.2 can be restated in a functional form. It is equivalent to saying that for most Boolean functions $f : \{0,1\}^p \to \{0,1\}$ there exists a function $f' : \{0,1\}^p \to (0,1)$ such that f and f' have the same marginals[3] up to dimension $0.49p$. (To see the connection, let $f(\theta)$ be the indicator function of S and set $f'(\theta) = |S|\,\mathbb{P}\left\{Y = \theta\right\}$.)

Furthermore, by translation we can replace 0 with -1 everywhere in the previous paragraph. This allows to put Theorem 1.2 in the context of Fourier analysis on the Boolean cube. Recall that *Rademacher functions* $r_j : \{-1,1\}^p \to \{-1,1\}$ are defined as

$$r_j(\theta) = \theta_j, \quad j = 1, \ldots, p.$$

Walsh functions $w_J : \{-1,1\}^p \to \{-1,1\}$ are indexed by subsets $J \subset [p]$ and are defined as

$$w_J = \prod_{j \in J} r_j, \tag{1.1}$$

with the convention $w_\emptyset = 1$. The canonical inner product on the Boolean cube is defined as

$$\langle f, g \rangle_{L^2} = \frac{1}{2^p} \sum_{\theta \in \{-1,1\}^p} f(\theta)\, g(\theta).$$

Walsh functions form an orthonormal basis of $L^2(\{-1,1\}^p)$.

In Section 3, we shall prove:

Theorem 1.3. *If p is sufficiently large, then for most of the functions $f : \{-1,1\}^p \to \{-1,1\}$ there exists a function $f' : \{-1,1\}^p \to (-1,1)$ that has the same frequencies[4] as f up to dimension $0.49p$.*

To see how Theorem 1.3 yields Theorem 1.2 in its functional form, note that the set of frequencies of f up to dimension d uniquely determines the set of marginals of f up to dimension d. For example, consider the two-dimensional marginal of f corresponding to setting the first coordinate to 1 and second to -1. We claim this marginal can be expressed in terms of

[3]This means that for any given set of coordinates $J \subset [p]$ with $|J| \le 0.49p$ and any given values $(\tau_j)_{j \in J}$, we have $\sum_\theta f(\theta) = \sum_\theta f'(\theta)$ where the summation is over all $\theta \in \{0,1\}^p$ whose values on the coordinates in J equal (τ_j).

[4]This means that for any set $J \subset [p]$ with $|J| \le 0.49p$, we have $\langle f, w_J \rangle = \langle f', w_J \rangle$.

the frequencies $\langle f, w_J\rangle$ up to dimension $|J| \le 2$. To do so, consider the set $\Theta = \{\theta \in \{-1,1\}^p :\ \theta_1 = 1,\, \theta_2 = -1\}$; then the marginal is

$$\begin{aligned}\frac{1}{2^p}\sum_{\theta\in\Theta} f(\theta) = \langle f, \mathbf{1}_\Theta\rangle = \Big\langle f, \Big(\frac{1+r_1}{2}\Big)\Big(\frac{1-r_2}{2}\Big)\Big\rangle \\ = \frac{1}{4}\Big(\langle f,1\rangle + \langle f,r_1\rangle - \langle f,r_2\rangle - \langle f,r_1r_2\rangle\Big) \\ = \frac{1}{4}\Big(\langle f,w_\emptyset\rangle + \langle f,w_{\{1\}}\rangle - \langle f,w_{\{2\}}\rangle - \langle f,w_{\{1,2\}}\rangle\Big).\end{aligned}$$

One can express other marginals in terms of frequencies similarly. This shows that Theorem 1.3 yields Theorem 1.2 indeed.

Question 1.4. *Can we replace the range $(-1,1)$ of f' by $\{-1,1\}$ in Theorem 1.3? In other words, is it true that most Boolean functions are not determined by frequencies up to almost half the dimension?*

Question 1.5. *Is half dimension an optimal threshold in Theorem 1.3 and/or Question 1.4?*

2. Background

Our proof of Theorem 1.3 is based on one combinatorial result about hyperplane arrangements and one probabilistic result – a version of Rudelson's sampling theorem.

2.1. **Hyperplane arrangements.**

A hyperplane arrangement is a collection of N hyperplanes in $\mathbb{R}^n$. Removing these hyperplanes from $\mathbb{R}^n$ leaves an open set, and the connected components of this set are called *regions*. Counting the regions of a given hyperplane arrangement is a well studied problem in enumerative combinatorics, see e.g. [1, 6, 7, 3].

For our purposes, two simple observations of Yu. Zuev [8] will be sufficient. Fix an arrangement of N hyperplanes in $\mathbb{R}^n$. The intersection of any subfamily of these original hyperplanes is called *an intersection subspace*. The dimension of an intersection subspace can range from zero (a single point) to n, since intersecting an empty set of hyperplanes yields the entire space $\mathbb{R}^n$.

Lemma 2.1 (Zuev [8])**.** *For any hyperplane arrangement, the number of regions is bounded below by the number of intersection subspaces.*

For example, an arrangement of N hyperplanes in general position in $\mathbb{R}^n$ produces exactly $\binom{N}{\le n}$ intersection subspaces, since intersecting any subset of such hyperplanes of cardinality at most n yields a different subspace. Zuev's Lemma 2.1 implies that such hyperplane arrangement has at least $\binom{N}{\le n}$ regions. This bound is in fact an identity, since any arrangement of N hyperplanes in $\mathbb{R}^n$ has at most $\binom{N}{\le n}$ regions [1]; see also [3, Proposition 2.4].

In general, it could be hard to count intersection subspaces directly. This task, however, can be facilitated by the following simple observation. For convenience, let us focus here on *central* hyperplane arrangements, those where all the hyperplanes pass through the origin. A central arrangement can be expressed as the form[5] $\{x_1^\perp, \ldots, x_N^\perp\}$ where $x_i \in \mathbb{R}^n$ is a vector orthogonal to the i-th hyperplane.

Definition 2.2 (Resilience)**.** *Fix a system of vectors $x_1, \ldots, x_N \in \mathbb{R}^n$. We call a subset $I \subset [N]$* resilient *if the linear span of $\{x_i\}_{i \in I}$ does not contain any vector from $\{x_i\}_{i \in I^c}$.*

Lemma 2.3 (Implicit in Zuev [8])**.** *For any central hyperplane arrangement $\{x_1^\perp, \ldots, x_N^\perp\}$, the number of intersection subspaces is bounded below by the number of resilient subsets.*

Proof. Any pair of distinct resilient subsets I and J satisfies

$$\operatorname{span}\{x_i\}_{i \in I} \neq \operatorname{span}\{x_j\}_{j \in J},$$

since for any $i_0 \in I \setminus J$, the definition of resilience yields $x_{i_0} \notin \operatorname{span}\{x_j\}_{j \in J}$. Hence the orthogonal complements of $\operatorname{span}\{x_i\}_{i \in I}$ and $\operatorname{span}\{x_j\}_{j \in J}$ are different, which we can write as

$$\bigcap_{i \in I} x_i^\perp \neq \bigcap_{j \in j} x_j^\perp.$$

Each side of this relation defines an intersection subspace. Thus, we obtained an injection from resilient subsets to intersection subspaces. The proof is complete. □

2.2. **Sampling.** In addition to hyperplane arrangements, our argument uses a version of Rudelson's sampling theorem [2]; see also [5, Theorem 5.6.1]. The version we need can be most conveniently deduced from *matrix Bernstein inequality*, which is due to J. Tropp [4]; see also [5, Theorem 5.4.1] for an exposition.

Theorem 2.4 (Matrix Bernstein's inequality)**.** *Let $Z_1, \ldots, Z_N$ be independent, mean zero, $k \times k$ symmetric random matrices, such that $\|Z_i\| \le M$ almost surely for all i. Then, for every $t \ge 0$, we have*

$$\mathbb{P}\Big\{\Big\|\sum_{i=1}^N Z_i\Big\| \ge t\Big\} \le 2k \exp\Big(-\frac{t^2/2}{\sigma^2 + Mt/3}\Big),$$

where $\sigma^2 = \Big\|\sum_{i=1}^N \mathbb{E}\, Z_i^2\Big\|$. Here $\|\cdot\|$ denotes the operator norm of a matrix.

We are ready to state and prove a version of Rudelson's sampling theorem.

[5]Throughout the paper, $x^\perp$ denotes the hyperplane orthogonal to the vector x in $\mathbb{R}^n$. More generally, for a subset $E \subset \mathbb{R}^n$, we denote by $E^\perp$ the set of vectors in $\mathbb{R}^n$ that are orthogonal to all vectors in E. In particular, if E is a linear subspace, $E^\perp$ is its orthogonal complement.

Theorem 2.5 (Sampling with independent selectors)**.** *Let $x_1, \dots, x_K \in \mathbb{R}^k$ be vectors satisfying*

$$\frac{1}{K}\sum_{i=1}^{K} x_i x_i^\mathsf{T} = \mathrm{Id}$$

and such that $\|x_i\|_2 \le 10\sqrt{k}$ for all i. Let $K \ge N \ge Ck\log k$ where C is a sufficiently large absolute constant. Consider independent Bernoulli random variables $\delta_1, \dots, \delta_K$ satisfying $\mathbb{E}\,\delta_i = N/K$. Then

$$0.99 \cdot \mathrm{Id} \preceq \frac{1}{N}\sum_{i=1}^{K} \delta_i x_i x_i^\mathsf{T} \preceq 1.01 \cdot \mathrm{Id}$$

with probability at least $1 - \frac{1}{4}k^{-10}$.

Proof. We are going to apply matrix Bernstein inequality for the random matrices

$$Z_i := (\delta_i - \delta)x_i x_i^\mathsf{T}, \quad \text{where } \delta := \mathbb{E}\,\delta_i = \frac{N}{K}.$$

By assumption, we have

$$\|Z_i\| = \left\|x_i x_i^\mathsf{T}\right\| = \|x_i\|_2^2 \le 100k =: M.$$

Furthermore,

$$0 \preceq \mathbb{E}\,Z_i^2 = \delta(1-\delta)\|x_i\|_2^2 \; x_i x_i^\mathsf{T} \preceq 100\delta k \cdot x_i x_i^\mathsf{T}.$$

Thus

$$\sigma^2 = \left\|\sum_{i=1}^{N} \mathbb{E}\,Z_i^2\right\| \le 100\delta k \left\|\sum_{i=1}^{K} x_i x_i^\mathsf{T}\right\| = 100\delta k K = 100kN.$$

Applying matrix Bernstein inequality (Theorem 2.4) and using the bounds on M and σ, we obtain

$$\begin{aligned}\mathbb{P}\left\{\left\|\frac{1}{N}\sum_{i=1}^{N} \delta_i x_i x_i^\mathsf{T} - \mathrm{Id}\right\| \ge 0.01\right\} &= \mathbb{P}\left\{\left\|\sum_{i=1}^{N} Z_i\right\| \ge 0.01N\right\} \\ &\le 2k \cdot \exp\Big(-\frac{cN}{k}\Big) \le \frac{1}{4}k^{-10}\end{aligned}$$

where the last inequality is guaranteed if $N \ge Ck\log k$ with a sufficiently large absolute constant C. This completes the proof. □

We will need a version of sampling theorem for a similar but not identical model of *sampling without replacement.*

Theorem 2.6 (Sampling without replacement)**.** *Let $x_1, \dots, x_K \in \mathbb{R}^k$ be vectors satisfying*

$$\frac{1}{K}\sum_{i=1}^{K} x_i x_i^\mathsf{T} = \mathrm{Id}$$

and such that $\|x_i\|_2 \le 10\sqrt{k}$ for all i. Let $N \ge Ck\log k$ where C is a sufficiently large absolute constant. Let I be a random subset of $[K]$ with cardinality $|I| = N$. Then

$$0.9 \cdot \mathrm{Id} \preceq \frac{1}{N} \sum_{i \in I} x_i x_i^\mathsf{T} \preceq 1.1 \cdot \mathrm{Id}$$

with probability at least $1 - k^{-10}$.

Proof. Apply Theorem 2.5 for $0.99N$ instead of N. It follows that a random set

$$I_0 \coloneqq \{i : \ \delta_i = 1\} \subset [K]$$

satisfies

$$\frac{1}{0.99N} \sum_{i \in I_0} x_i x_i^\mathsf{T} \succeq 0.99 \cdot \mathrm{Id} \tag{2.1}$$

with probability at least $1 - \frac{1}{4}k^{-10}$.

Since $\mathbb{E}\,\delta_i = 0.99N/K$, the expected cardinality of the set I_0 is $\mathbb{E}|I_0| = 0.99N$. Moreover, Chernoff inequality (see e.g. Exercise 2.3.5 in my book) implies that $|I_0|$ is concentrated around its expectation, and in particular

$$|I_0| \le N \tag{2.2}$$

with probability at least $1 - 2e^{-cN}$.

Let us create a random set I of cardinality exactly N from I_0 by the following rule. If $|I_0| < N$, add to I_0 exactly $N - |I_0|$ elements chosen from $[K] \setminus I_0$ at random and without replacement. If $|I_0| > N$, remove from I_0 exactly $|I_0| - N$ elements chosen at random and without replacement. Clearly, I obtained this way is a random subset of $[K]$ of cardinality $|I| = N$.

Suppose I_0 satisfies both (2.1) and (2.2); this occurs with probability at least $1 - \frac{1}{4}k^{-10} - 2e^{-cN} \ge 1 - \frac{1}{2}k^{-10}$. In this case, $I \supset I_0$ and so

$$\frac{1}{N} \sum_{i \in I} x_i x_i^\mathsf{T} \succeq \frac{1}{N} \sum_{i \in I_0} x_i x_i^\mathsf{T} \succeq 0.99^2 \cdot \mathrm{Id} \succeq 0.9 \cdot \mathrm{Id}.$$

A similar argument but for $1.01N$ instead of $0.99N$ yields

$$\frac{1}{N} \sum_{i \in I} x_i x_i^\mathsf{T} \preceq (1.01)^2 \cdot \mathrm{Id} \preceq 1.1 \cdot \mathrm{Id}$$

with probability at least $1 - \frac{1}{2}k^{-10}$. Taking the intersection of the two bounds completes the proof. □

3. Proof of Theorem 1.3

3.1. **Reduction to sign patterns.** Fix $d < p$. Let us call a function $f : \{-1,1\}^p \to \{-1,1\}$ *non-unique* if there exists a function $f' : \{-1,1\}^p \to (-1,1)$ that has the same frequencies as f' up to dimension d. Our goal is to show that most of the functions f are non-unique for $d = 0.49p$.

Consider the following linear subspace of real-valued functions on the Boolean cube:

$$\mathcal{H} := \big\{ h : \{-1,1\}^p \to \mathbb{R} : \ \langle h, w_J \rangle = 0 \quad \forall J \subset [p],\, |J| \le d \big\}. \tag{3.1}$$

Lemma 3.1. *f is non-unique if $f \equiv \operatorname{sign} h$ for some $h \in \mathcal{H}$.*

Proof. Suppose $f \equiv \operatorname{sign} h$ for some $h \in \mathcal{H}$. Let $\varepsilon > 0$ be small enough and set $f' := f - \varepsilon h$. By definition of $\mathcal{H}$, the functions f and f' have the same frequencies up to dimension d. Moreover, if $f(\theta) = 1$ then $h(\theta) > 0$ and hence $f'(\theta) < 1$. Similarly, if $f(\theta) = -1$ then $h(\theta) < 0$ and hence $f'(\theta) > -1$. Therefore, if ε is sufficiently small, $f'(\theta) \in (-1,1)$ for any θ. □

Lemma 3.1 shows that the number of non-unique functions f is bounded below by the number of functions of the form $\operatorname{sign} h$ where $h \in \mathcal{H}$, which we call *sign patterns* generated by $\mathcal{H}$. Hence, in order to show that most of the functions f are non-unique, it suffices to show that the subspace $\mathcal{H}$ generates a lot of sign patterns.

3.2. **From sign patterns to hyperplane arrangements.** To count different sign patterns generated by $\mathcal{H}$, we express this as a problem about hyperplane arrangements.

Let $e_\theta : \{-1,1\}^p \to \{0,1\}$ denote the point evaluation function at $\theta \in \{-1,1\}^p$. Consider two functions $h, h' \in \mathcal{H}$. We have $\operatorname{sign} h \not\equiv \operatorname{sign} h'$ if and only if there exists $\theta \in \{-1,1\}^p$ such that $\langle h, e_\theta \rangle$ and $\langle h', e_\theta \rangle$ have opposite signs. This happens if and only if h and h' are separated by at least one hyperplane $e_\theta^\perp \cap \mathcal{H}$ in $\mathcal{H}$. The latter is equivalent to h and h' lying in different regions of the hyperplane arrangement $\{e_\theta^\perp \cap \mathcal{H}\}_{\theta \in \{-1,1\}^p}$ in the subspace $\mathcal{H}$. Therefore, the number of different sign patterns generated by $\mathcal{H}$ (and thus also the number non-unique functions f) is bounded below by the number of regions of the hyperplane arrangement $\{e_\theta^\perp \cap \mathcal{H}\}_{\theta \in \{-1,1\}^p}$ in the subspace $\mathcal{H}$.

Denoting by $P_\mathcal{H}$ the orthogonal projection onto H, we see that the vector $P_\mathcal{H} e_\theta$ is orthogonal to the hyperplane $e_\theta^\perp \cap \mathcal{H}$ in $\mathcal{H}$. Now apply Lemmas 2.1 and 2.3 for our hyperplane arrangement in $\mathcal{H}$. We conclude that the number of regions (and thus also the number of sign patterns generated by $\mathcal{H}$, and thus also the number of non-unique functions f) is bounded below by the number of resilient subsets for the collection $\{P_\mathcal{H} e_\theta\}_{\theta \in \{-1,1\}^p}$.

3.3. **A necessary condition for non-resilience.** To complete the proof, we will show that most of the subsets are resilient. Let us see what happens when the complement Θ^c of some subset $\Theta \subset \{-1,1\}^p$ is *not* resilient. By definition, there exists $\theta_0 \in \Theta$ such that

$$P_\mathcal{H} e_{\theta_0} \in \operatorname{span}\{P_\mathcal{H} e_\theta\}_{\theta \in \Theta^c}.$$

This means that there exist real numbers $(a_\theta)_{\theta\in\Theta^c}$ such that

$$P_{\mathcal{H}} e_{\theta_0} = \sum_{\theta\in\Theta^c} a_\theta P_{\mathcal{H}} e_\theta.$$

This in turn means that

$$e_{\theta_0} - \sum_{\theta\in\Theta^c} a_\theta e_\theta \in H^\perp = \operatorname{span}\big\{ w_J : J \subset [p],\, |J| \le d \big\}.$$

Therefore, there exist real numbers $(b_J)_{J\subset[p],\,|J|\le d}$ such that

$$e_{\theta_0} - \sum_{\theta\in\Theta^c} a_\theta e_\theta = \sum_{J\subset[p],\,|J|\le d} b_J w_J.$$

If we evaluate this identity at any point $\theta \in \Theta \setminus \{\theta_0\}$, the left hand side of it vanishes, and we have

$$\sum_{J\subset[p],\,|J|\le d} b_J w_J(\theta) = 0.$$

Express this identity this as an orthogonality relation in $\mathbb{R}^{\binom{p}{\le d}}$, namely

$$\langle b, w(\theta)\rangle = 0$$

where

$$b = (b_J)_{J\subset[p],\,|J|\le d} \quad \text{and} \quad w(\theta) \coloneqq (w_J(\theta))_{J\subset[p],\,|J|\le d}. \tag{3.2}$$

Summarizing, we showed the following.

Lemma 3.2 (A necessary condition for non-resilience)**.** *If* Θ^c *is* not *resilient for some* $\Theta \subset \{-1,1\}^p$ *then there exists* $\theta_0 \in \Theta$ *and* $b \in \mathbb{R}^{\binom{p}{\le d}}$ *such that*

$$\langle w(\theta), b\rangle = 0 \quad \textit{for all } \theta \in \Theta \setminus \{\theta_0\}.$$

3.4. Applying the sampling theorem.

Lemma 3.3. *Assume that* $2^p \ge C\binom{p}{\le d}\log\binom{p}{\le d}$ *where* C *is a sufficiently large absolute constant. Let* Θ *be a random subset of* $\{-1,1\}^p$ *with cardinality* $|\Theta| = 0.1\cdot 2^p$*. Then, with probability at least* $1-p^{-10}$*, the following uniform lower bound holds:*

$$\frac{1}{0.1\cdot 2^p}\sum_{\theta\in\Theta} \langle w(\theta), b\rangle^2 \ge 0.9\|b\|_2^2 \quad \textit{for all } b \in \mathbb{R}^{\binom{p}{\le d}}. \tag{3.3}$$

Proof. Orthonormality of Walsh basis (1.1) can be expressed as

$$\frac{1}{2^p}\sum_{\theta\in\{-1,1\}^p} w_J(\theta)\, w_{J'}(\theta) = \delta_{J,J'} \quad \text{for any } J, J' \subset [p].$$

Using our notation (3.2), this becomes

$$\frac{1}{2^p}\sum_{\theta\in\{-1,1\}^p} w(\theta)\, w(\theta)^{\mathsf{T}} = \mathrm{Id}.$$

(The identity on the right side is in $\mathbb{R}^{\binom{p}{\le d}}$.) Moreover, since all coordinates of $w(\theta)$ are ± 1, we have

$$\left\|w(\theta)\right\|_2^2 = \binom{p}{\le d} \quad \text{for all } \theta. \tag{3.4}$$

Apply Rudelson's Sampling Theorem 2.6 for $k = \binom{p}{\le d}$ and $N = 0.1 \cdot 2^p$. We get

$$\frac{1}{0.1 \cdot 2^p} \sum_{\theta \in \Theta} w(\theta) w(\theta)^{\mathsf{T}} \succeq 0.9 \cdot \mathrm{Id}$$

with probability at least $1 - k^{-10} \ge 1 - p^{-10}$. Multiplying by b^{T} on the left and by b on the right, we obtain (3.3). □

3.5. **Completion of the proof.** If $1 \le d \le 0.49p$ then $2^p \ge C\binom{p}{\le d} \log \binom{p}{\le d}$, so the assumptions of Lemma 3.3 hold. Thus, the uniform lower bound (3.3) holds for most of the subsets $\Theta \subset \{-1,1\}^p$ with cardinality $0.1 \cdot 2^p$.

We claim that for any subset Θ satisfying (3.3), the complement Θ^c is resilient. This will be enough to complete the proof of Theorem 1.3. Indeed, it would follow that most of the subsets of $\{-1,1\}^p$ of cardinality $0.9 \cdot 2^p$ are resilient. Since a subset of a resilient subset is resilient, most of the subsets of $\{-1,1\}^p$ of cardinality bounded by $0.9 \cdot 2^p$ are resilient. Thus, most of the subsets of $\{-1,1\}^p$ are resilient, completing the proof.

Assume for contradiction that Θ^c is *not* resilient. Then the necessary condition (Lemma 3.2) implies the existence of $\theta_0 \in \Theta$ and $b \in \mathbb{R}^{\binom{p}{\le d}}$ such that

$$\sum_{\theta \in \Theta \setminus \{\theta_0\}} \langle w(\theta), b\rangle^2 = 0. \tag{3.5}$$

On the other hand,

$$\sum_{\theta \in \Theta \setminus \{\theta_0\}} \langle w(\theta), b\rangle^2 = \sum_{\theta \in \Theta} \langle w(\theta), b\rangle^2 - \langle w(\theta_0), b\rangle^2.$$

By Lemma 3.3, the first term on the right hand side is bounded below by $0.1 \cdot 2^p \cdot 0.9\|b\|_2^2$. The second term on the right hand side is bounded above by

$$\left\|w(\theta_0)\right\|_2^2 \cdot \|b\|_2^2 = \binom{p}{\le d} \|b\|_2^2$$

due to (3.4). Therefore we have

$$\sum_{\theta \in \Theta \setminus \{\theta_0\}} \langle w(\theta), b\rangle^2 > 0 \tag{3.6}$$

as long as

$$0.1 \cdot 2^p \cdot 0.9 > \binom{p}{\le d},$$

which is true if $d = 0.49p$ if p is sufficiently large. The contradiction of (3.5) and (3.6) completes the proof of Theorem 1.3. □

References

[1] Robert Creighton Buck. Partition of space. *The American Mathematical Monthly*, 50(9):541–544, 1943.

[2] Mark Rudelson. Random vectors in the isotropic position. *Journal of Functional Analysis*, 164(1):60–72, 1999.

[3] Richard P Stanley et al. An introduction to hyperplane arrangements. *Geometric combinatorics*, 13:389–496, 2004.

[4] Joel A Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics*, 12:389–434, 2012.

[5] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[6] RO Winder. Partitions of n-space by hyperplanes. *SIAM Journal on Applied Mathematics*, 14(4):811–818, 1966.

[7] Thomas Zaslavsky. *Facing up to Arrangements: Face-Count Formulas for Partitions of Space by Hyperplanes: Face-count Formulas for Partitions of Space by Hyperplanes*, volume 154. American Mathematical Soc., 1975.

[8] Yu A Zuev. Combinatorial-probability and geometric methods in threshold logic. *Diskretnaya Matematika*, 3(2):47–57, 1991.

Department of Mathematics, University of California, Irvine
Email address: rvershyn@uci.edu