On the weight distribution of random binary linear codes

Nati Linial * Jonathan Mosheiff †

Abstract

We investigate the weight distribution of random binary linear codes. For $0 < \lambda < 1$ and $n \to \infty$ pick uniformly at random λn vectors in $\mathbb{F}_2^n$ and let $C \leq \mathbb{F}_2^n$ be the orthogonal complement of their span. Given $0 < \gamma < 1/2$ with $0 < \lambda < h(\gamma)$ let X be the random variable that counts the number of words in C of Hamming weight γn. In this paper we determine the asymptotics of the moments of X of all orders $o(\frac{n}{\log n})$.

1 Introduction

Random linear codes play a major role in the theory of error correcting codes, and are also important in other areas such as information theory, theoretical computer science and cryptography [8, 12, 2, 1]. Nevertheless, not much seems to be known about their properties. As already demonstrated in Shannon's foundational paper [13], random linear codes occupy a particularly prominent position in coding theory. This is arguably the simplest construction to achieve channel capacity in the binary symmetric channel, as well as the Gilbert-Varshamov bound for minimal distance. The present paper is motivated by the contrast between the importance of random codes and the lack of our understanding. Our main aim is to improve our comprehension of the weight distribution of random binary linear codes.

The two most basic parameters of a code $C \subseteq \mathbb{F}_2^n$ are its *rate* $R = \frac{\log_2 |C|}{n}$ and its *relative distance* $\delta = \frac{\min\{\|x-y\| \mid x,y \in C\ x \neq y\}}{n}$, where $\|\cdot\|$ is the Hamming norm. Clearly, the rate of a $\mathcal{D}$-dimensional linear code $C \subseteq \mathbb{F}_2^n$ is $\frac{\mathcal{D}}{n}$, and its relative distance is $\frac{\min\{\|w\| \mid w \in C\ w \neq 0\}}{n}$.

*Department of Computer Science, Hebrew University, Jerusalem 9190401. e-mail: nati@cs.huji.ac.il. Supported by ERC grant 339096 "High-dimensional combinatorics".

†Department of Computer Science, Hebrew University, Jerusalem 9190401. e-mail: yonatanm@cs.huji.ac.il. Supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities.

It is a major challenge to understand the trade-off between rate and distance for linear as well as general codes. Concretely, given $0 < \delta < \frac{1}{2}$, we wish to know the value of $\limsup R(C)$ where the the $\limsup$ is taken over all binary codes of relative distance at least δ. The Gilbert-Varshamov (GV) lower bound (e.g., [7], lec. 2) states that $R \geq 1 - h(\delta)$ is achievable, where h is the binary entropy function. Despite many attempts, this bound has not been improved, nor shown to be tight, through over 60 years of intense investigations. The best known upper bound, from 1977 (MRRW), is due to McEliece, Rodemich, Rumsey and Welch [9]. An alternative proof of this bound, using harmonic analysis on $\mathbb{F}_2^n$, was given in 2007 by Navon and Samorodnitsky [10]. Note that this is an upper bound on *all* codes.

Curiously, neither bound, GV and MRRW, exhibits any distinction between linear and nonlinear codes. Of course, the realm of nonlinear codes is much richer than the linear one, but whether or not nonlinear codes perform better than linear ones remains a mystery. One would thus expect that both the lower bounds and the upper bounds for nonlinear codes be higher than for linear codes, but neither one is the case at present. Since both GV and MRRW are several decades old, it is of interest to find any key ways in which linear and nonlinear codes differ. As this paper shows, the weight distribution of random linear codes is very different from that of random nonlinear codes.

This paper concerns the weight distribution of random linear codes. Concretely, fix two rational numbers $0 < \gamma < \frac{1}{2}$ and $0 < \lambda < h(\gamma)$, and let $n \in \mathbb{N}$ be such that λn is an integer and γn is an even integer[1]. Let $C = C_{n,\lambda}$ be a random subspace of $\mathbb{F}_2^n$ that is defined via $C := \{x \in \mathbb{F}_2^n | Kx = 0\}$ where K is a uniformly random $\lambda n \times n$ binary matrix. Clearly $\dim C \geq (1-\lambda)n$, and with very high probability equality holds. Denote $L = L_{n,\gamma} = \{x \in \mathbb{F}_2^n \mid \|x\| = \gamma n\}$. We investigate the distribution of the random variable $X = X_{n,\gamma,\lambda} = |C \cap L|$ for fixed γ and λ when $n \to \infty$. Clearly $\mathbb{E}(X) = N^{-\lambda}\binom{n}{\gamma n} = N^{h(\gamma)-\lambda+o(1)}$, where $N = 2^n$. This follows since every $x \in L_{n,\gamma}$ belongs to a random $C_{n,\lambda}$ with probability $N^{-\lambda}$. Also, $\lim_{n\to\infty} \mathbb{E}(X) = \infty$, since, by assumption $\lambda < h(\gamma)$.

It is instructive to compare what happens if rather than a random linear code C, we consider a uniformly random subset $C' \subset \mathbb{F}_2^n$, where every vector in $\mathbb{F}_2^n$ independently belongs to C' with probability $N^{-\lambda}$. In analogy, we define $X' = |C' \cap L|$, and the distribution of X' is clearly approximately normal. It would not be unreasonable to guess that X behaves similarly, and in particular that its limit distribution, as $n \to \infty$ is normal. However, as we show, the code's linear structure has a rather strong effect. Indeed X does not converge to a normal random variable, and moreover, only a few of its central moments are bounded.

[1]For other ranges of the problem - See our Discussion.

1.1 Rough outline of how we compute the moments

We seek to approximate the central k-th moments of X for all $k \le o(\frac{n}{\log n})$. In Section 2 we reduce this question to an enumeration problem that we describe next. We say that a linear subspace $U \le \mathbb{F}_2^k$ is *robust* if every system of linear equations that defines it involves all k coordinates. Given a subspace $U \le \mathbb{F}_2^k$, let T_U be the set of all $k \times n$ binary matrices where every column is a vector in U and every row has weight γn and let $|T_U|$ denote the cardinality of this set. We show that

$$\mathbb{E}\left((X - \mathbb{E}(X))^k\right) = \Theta\left(\sum_{\mathcal{D}=0}^{k-1} N^{-\lambda\mathcal{D}} \sum_{\substack{V \le \mathbb{F}_2^k \\ \dim(V)=\mathcal{D} \\ V \text{ robust}}} |T_V|\right), \tag{1}$$

The main challenge is to estimate the internal sum, but understanding the interaction with the outer sum is nontrivial either. The reason that we can resolve this problem is that the main contributors to the internal sum are fairly easy to describe. As it turns out, this yields a satisfactory answer even though we provide a rather crude upper bound on all the other terms.

A key player in this story is the space of even-weight vectors $V = \mathbb{V}^k \le \mathbb{F}_2^k$. In Section 3 we solve this enumeration problem for this space, and show that $|T_{\mathbb{V}^k}| \approx N^{F(k,\gamma)}$ up to a factor that is polynomial in n and exponential in k. Here $F(k,\gamma)$ is the entropy of a certain entropy maximizing probability distribution on $\mathbb{V}^k$. In our proof, we generate a $k \times n$ matrix A with i.i.d. columns sampled from this distribution, and compute the probability that $A \in T_{\mathbb{V}^k}$. The function F has the explicit description

$$F(k,\gamma) = \min_{1>x>0} \log_2\left((1+x)^k + (1-x)^k\right) - k\gamma \log_2 x - 1$$

and its asymptotic behavior for large k is:

$$F(k,\gamma) = kh(\gamma) - 1 + O((1-2\gamma)^k).$$

In Section 4 we use the result of Section 3 to bound $|T_U|$ for a general robust $U \le \mathbb{F}_2^k$. Consider a robust space $U \le \mathbb{F}_2^k$ of the form $\bigoplus_{i=1}^c \mathbb{V}^{m_i}$, where $\sum m_i = k$. Clearly, $|T_U| = \prod_{i=1}^c |T_{\mathbb{V}^{m_i}}| \approx N^{\sum_{i=1}^c F(m_i,\gamma)}$. Hence, finding a space of this form of given dimension that maximizes $|T_U|$ translates into a question about the dependence of $F(m,\gamma)$ on m. We show (Lemma 25) that this function is convex, so that the optimum is attained at $m_1 = k - 2c + 2$ and $m_2 = m_3 = \ldots = m_c = 2$.

We show that if $U \le \mathbb{F}_2^k$ is robust and not a product of *Even* spaces, then there is some V of this form and of the same dimension with $|T_V| \ge |T_U|$. We reduce the proof of this claim (Equation (23)) to the analysis of $m \times n$ matrices where

every row weighs γn, the first δn columns have odd weight and the last $(1-\delta)n$ ones are even. A key step in the proof (Lemma 24) shows that the number of such matrices decreases with δ.

Finally, in Section 5, the results of the previous sections are put together to find the dominating terms of Equation (1), yielding the moments of X. For even k, we show that the dominating terms are those corresponding to either $\mathcal{D} = \frac{k}{2}$ or $\mathcal{D} = k-1$, and respectively, to the subspaces $\bigoplus_{i=1}^{k/2} \mathbb{V}^2$ or $\mathbb{V}^k$. More precisely, there exists some $k_0(\gamma, \delta)$ such that the former dominates when $k \leq k_0$ and the latter when $k > k_0$. The behavior of odd order moments is similar, although slightly more complicated to state.

Theorems 2 and 3 in Section 5, deal with even and odd order moments, respectively. Theorem 1 gives the central moments of the normalized variable $\frac{X}{\sqrt{\mathrm{Var}(X)}}$.

Theorem 1. *Fix $\gamma < \frac{1}{2}$ and $0 < \lambda < h(\gamma)$, let $X = X_{n,\gamma,\lambda}$, and let*

$$k_0 = \min\left\{ m \mid F(m,\gamma) - (m-1)\lambda > \frac{m}{2}(h(\gamma) - \lambda) \right\}.$$

Then, for $2 \leq k \leq o(\frac{n}{\log n})$,

$$\frac{\mathbb{E}((X - \mathbb{E}(X))^k)}{\mathrm{Var}(X)^{\frac{k}{2}}} = \begin{cases} o(1) & \text{if } k \text{ is odd and } < k_0 \\ (1+o(1)) \cdot k!! & \text{if } k \text{ is even and } < k_0 \\ N^{F(k,\gamma) - \frac{k}{2}h(\gamma) - (\frac{k}{2}-1)\lambda - \frac{k \log n}{4n} + O(\frac{k}{n})} & \text{if } k \geq k_0 \end{cases}$$

We call the reader's attention to the following interesting point on which we elaborate below. For given γ and λ there is a bounded number of moments for which our distribution behaves as if it were normal, but from that index k_0, the code's linear structure starts to dominate the picture and the normalized moments become unbounded as $n \to \infty$. (See Figure 1).

1.2 Preliminaries

General

- Unless stated otherwise, all logarithms here are to base 2.
- Our default is that an asymptotic statement refers to $n \to \infty$, while the parameters γ and λ take fixed arbitrary values within their respective domains. Other parameters such as k may or may not depend on n.
- We denote a binomial distribution with n trials of probability p by $B(n,p)$.

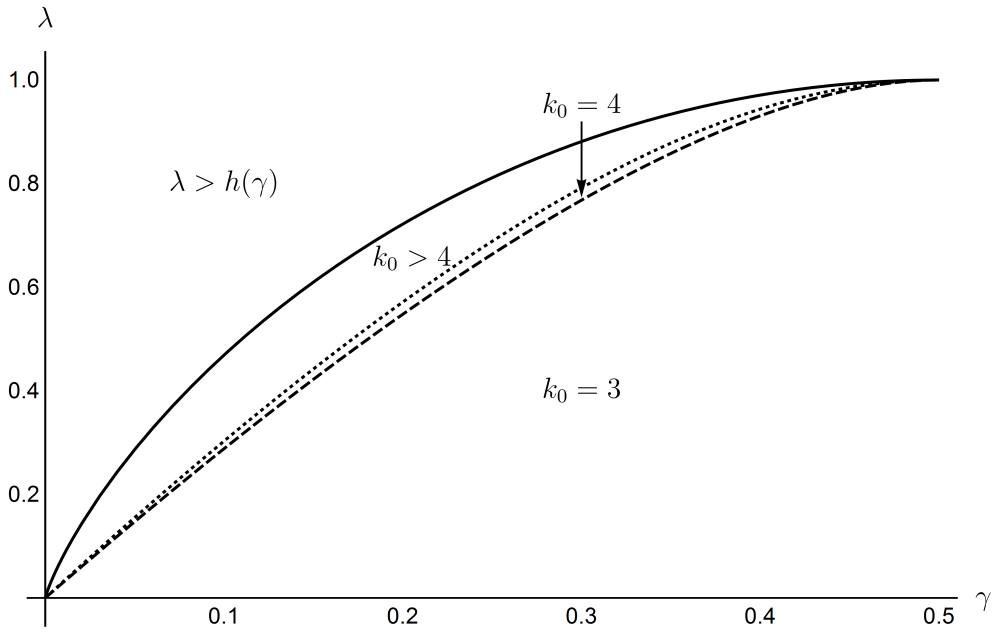

Figure 1: Illustration for Theorem 1. For $k < k_0 = k_0(\gamma, \lambda)$ the k-th moment of X is that of a normal distribution. The relevant range $\lambda < h(\gamma)$ is below the solid line. Note that $k_0 = 3$ for much of the parameters range.

Entropy

- We use the standard notation $h(t) = -t \cdot \log t - (1-t) \cdot \log(1-t)$. Entropy and conditional entropy are always binary.

Linear algebra

- $U \leq V$ means that U is a linear subspace of the vector space V. The *weight*, $\|u\|$ of a vector $u \in \mathbb{F}_2^n$ is the number of its 1 coordinates. Accordingly we call u even or odd. Likewise, the weight $\|A\|$ of a binary matrix A, is the number of its 1 entries.

- The sets of even and odd vectors in $\mathbb{F}_2^n$ are denoted by $\mathbb{V}^n$ and $\mathbb{D}^n$.

- The i-th row of a matrix A is denoted by A_i. If $I \subseteq [k]$ then A_I is the sub-matrix consisting of the rows $\{A_i \mid i \in I\}$. Also v_I is the restriction of the vector v to the coordinates in I.

- For a subspace $U \leq \mathbb{F}_2^k$ and $I \subset [k]$ we denote by U_I the projection of U to the coordinates in I, i.e., $U_I = \{u_I \mid u \in U\}$, and we use the shorthand $\mathcal{D}_I(U) = \dim U_I$, and $\mathcal{D}(U) = \dim U$.

2 From moments to enumeration

To recap: $C = C_{n,\lambda}$ is a random linear subspace of $\mathbb{F}_2^n$, and $L = L_{n,\gamma}$ is the γn-th layer of $\mathbb{F}_2^n$. We fix $0 < \gamma < 1$, $0 < \lambda < h(\gamma)$, so that λn is an integer and γn is an even integer, and we start to investigate the moments of $X = |C \cap L|$, as $n \to \infty$.

The probability that C contains a given subset of $\mathbb{F}_2^n$ depends only on its linear dimension:

Proposition 1. *If $Y \subseteq \mathbb{F}_2^n$ has dimension $\dim(Y) = \mathcal{D}$, then $\Pr(Y \subseteq C) = N^{-\lambda\mathcal{D}}$.*

Proof. As mentioned, we think of C as the kernel of a uniform random $\lambda n \times n$ binary matrix K, so $Y \subseteq C$ iff every row of K is orthogonal to Y. The probability of this event is $2^{-\mathcal{D}}$ for a given row, and $2^{-\lambda n\mathcal{D}} = N^{-\lambda\mathcal{D}}$ for all rows together. □

2.1 Interpreting the central moments of X

We turn to express X and its moments in terms of indicator random variables.

Definition 2. *For a vector $u \in \mathbb{F}_2^n$, let Y_u be the indicator for the event that $u \in C$. For a binary $k \times n$ matrix A we let Y_A be the indicator random variable for the event that every row of A is in C.*

Proposition 1 plainly yields the first two central moments of X.

$$\mathbb{E}(X) = \sum_{u \in L} \mathbb{E}(Y_u) = |L| N^{-\lambda} = \binom{n}{\gamma n} N^{-\lambda} = N^{h(\gamma) - \lambda - \frac{\log n}{2n} + O(\frac{1}{n})}.$$

Proposition 1 also implies that $\mathrm{Cov}(Y_u, Y_v) = 0$ for every $u \neq v \in L$. Hence,

$$\mathrm{Var}(X) = \sum_{u \in L} \mathrm{Var}(Y_u) = \binom{n}{\gamma n} N^{-\lambda}(1 - N^{-\lambda}) = N^{h(\gamma) - \lambda - \frac{\log n}{2n} + O(\frac{1}{n})}.$$

In words, the first two moments of X are not affected by the linearity of C.

We now turn to higher order moments. Specifically we wish to compute the k-th central moment of X for any $2 < k \leq o(\frac{n}{\log n})$.

Definition 3. *We denote by $W_k = W_{k,\gamma}$ the set of binary $k \times n$ matrices in which every row has weight γn.*

Definition 4. *For a subspace $U \leq \mathbb{F}_2^k$ we denote*

$$T_{U,n,\gamma} = T_U = \{A \in W_k \mid \mathrm{Im}\, A \subseteq U\}$$

and

$$\overline{T}_{U,n,\gamma} = \overline{T}_U = \{A \in W_k \mid \mathrm{Im}\, A = U\}.$$

Let us expand the k-th central moment.

$$\mathbb{E}\left((X-\mathbb{E}(X))^k\right) = \mathbb{E}\left(\left(\sum_{u\in L} Y_u - \sum_{u\in L}\mathbb{E}(Y_u)\right)^k\right)$$
$$= \sum_{u_1,\ldots,u_k\in L}\sum_{I\subseteq[k]} \mathbb{E}\left(\prod_{i\in I} Y_{u_i}\right)\prod_{j\in[k]\setminus I}\left(-\mathbb{E}\left(Y_{u_j}\right)\right). \qquad (2)$$

If A is the matrix with rows $u_1,\ldots,u_k$, then by Proposition 1 this equals

$$\sum_{A\in W_k}\sum_{I\subseteq[k]} (-1)^{k-|I|}\cdot N^{-\lambda\cdot(\operatorname{rank} A_I + k - |I|)}.$$

We group the matrices $A \in W_k$ with the same image U and rewrite the above as

$$\sum_{U\leq\mathbb{F}_2^k} |\overline{T}_U| \sum_{I\subseteq[k]} (-1)^{k-|I|}\cdot N^{-\lambda\cdot(\mathcal{D}_I(U)+k-|I|)},$$

which we restate as

$$\mathbb{E}\left((X-\mathbb{E}(X))^k\right) = \sum_{U\leq\mathbb{F}_2^k} |\overline{T}_U| R_U, \qquad (3)$$

where for any $U \leq \mathbb{F}_2^k$

$$R_U = \sum_{I\subseteq[k]} (-1)^{k-|I|}\cdot N^{-\lambda\cdot(\mathcal{D}_I(U)+k-|I|)}. \qquad (4)$$

We proceed as follows:

1. We recall the notion of a *robust* linear subspace of $\mathbb{F}_2^k$, and bound R_U separately for robust and non-robust subspaces.

2. Using Möbius inversion, we restate Equation (3) in terms of $|T_U|$ rather than $|\overline{T}_U|$.

2.1.1 Computing R_U

It is revealing to consider our treatment of X alongside a proof of the Central Limit Theorem (CLT) based on the moments method (e.g., [5]). In that proof, the k-th moment of a sum of random variables of expectation zero is expressed as a sum of expectations of degree-k monomials, just as in our Equation (2). These monomials are then grouped according to the relations between their factors. In the CLT proof, it is assumed that each tuple's non-repeating factors are independent, so monomials are grouped according to their degree sequence. Here, and

specifically in Equation (3), we need a more refined analysis that accounts for the linear matroid that is defined by the monomial's factors.

In the proof the the CLT there holds $\mathbb{E}(M) = 0$ for every monomial M that contains a degree-1 factor Y. This follows, since $\mathbb{E}(Y) = 0$ and the rest of the monomial is independent of Y. Something similar happens here too. If u does not participate in any linear relation with the other factors in its monomial, then Y_u can play a role analogous to that of Y. This intuition is captured by the following definition and proposition.

Definition 5. *Let $U \leq \mathbb{F}_2^k$ be a linear subspace. We say that its i-th coordinate is* sensitive *if $\mathcal{D}_{[k]\setminus\{i\}}(U) = \mathcal{D}(U) - 1$. We denote by $\mathrm{Sen}(U)$ the set of U's sensitive coordinates. Also, if $\mathrm{Sen}(U) = \varnothing$, we say that U is* robust.

It is not hard to see that equivalently, robustness means that every 1-co-dimensional coordinate-wise projection of U has the same dimension as U. Yet another description is that every system of linear equations that defines U must involve all coordinates.

Proposition 6. *For $U \leq \mathbb{F}_2^k$ it holds that*

1. *If U is robust then $R_U = \Theta\left(N^{-\mathcal{D}(U)\lambda}\right)$.*
2. *If U is not robust then $R_U = 0$.*

Proof. We use here the shorthand $\mathcal{D} = \mathcal{D}(U)$ and $\mathcal{D}_I = \mathcal{D}_I(U)$.

We start with the case of a robust U. Note that for every $I \subsetneq [k]$ there holds $\mathcal{D}_I \geq \mathcal{D} - k + |I| + 1$. For let us carry out the projection as $k - |I|$ steps of 1-co-dimensional projections. At each step the dimension either stays or goes down by one. But since U is robust, in the first step the dimension stays.

We claim that in the expression for R_U in Equation (4), the term $N^{-\lambda\mathcal{D}}$ that corresponds to $I = [k]$ dominates the rest of the sum. Indeed, each of the other $2^k - 1$ summands is $\pm\Theta(N^{-\lambda(\mathcal{D}+1)})$. Consequently, $R_U = \Theta(N^{-\lambda\mathcal{D}})$.

Let us consider next a non-robust U. Let j be a sensitive coordinate of U. If $I \subseteq [k] \setminus \{j\}$, then $\mathcal{D}_{I\cup\{j\}} = \mathcal{D}_I + 1$. Consequently:

$$\begin{aligned} R_U &= \sum_{I\subseteq[k]\setminus\{j\}} \left((-1)^{k-|I|} N^{-\lambda(\mathcal{D}_I+k-|I|)} + (-1)^{k-|I|-1} N^{-\lambda(\mathcal{D}_{I\cup\{j\}}+k-|I|-1)}\right) \\ &= \sum_{I\subseteq[k]\setminus\{j\}} \left((-1)^{k-|I|} N^{-\lambda(\mathcal{D}_I+k-|I|)} + (-1)^{k-|I|-1} N^{-\lambda(\mathcal{D}_I+k-|I|)}\right) = 0 \end{aligned}$$

□

2.1.2 From $|\overline{T}_U|$ to $|T_U|$

In order for Equation (3) to be expressed in terms of $|T_U|$ rather than $|\overline{T}_U|$ we can appeal to the Möbius inversion formula for vector spaces over a finite field (e.g., [14], Ch 3.10).

$$\begin{aligned}\mathbb{E}\left((X-\mathbb{E}(X))^k\right) &= \sum_{U\le\mathbb{F}_2^k} R_U \sum_{V\le U} (-1)^{\mathcal{D}(U)-\mathcal{D}(V)} \cdot 2^{\binom{\mathcal{D}(U)-\mathcal{D}(V)}{2}} |T_V| \\ &= \sum_{V\le\mathbb{F}_2^k} |T_V| \sum_{V\le U\le\mathbb{F}_2^k} R_U (-1)^{\mathcal{D}(U)-\mathcal{D}(V)} \cdot 2^{\binom{\mathcal{D}(U)-\mathcal{D}(V)}{2}}.\end{aligned}$$

Grouping the U's by their dimension $i = \mathcal{D}(U)$, we express the above as

$$\sum_{V\le\mathbb{F}_2^k} |T_V| (-1)^{\mathcal{D}(V)} \sum_{i=\mathcal{D}(V)}^{k} (-1)^i \cdot 2^{\binom{i-\mathcal{D}(V)}{2}} \sum_{\substack{V\le U\le\mathbb{F}_2^k \\ \mathcal{D}(U)=i}} R_U.$$

By Proposition 6, this sum can be further rewritten as

$$\Theta\left(\sum_{V\le\mathbb{F}_2^k} |T_V| (-1)^{\mathcal{D}(V)} \sum_{i=\mathcal{D}(V)}^{k} (-1)^i \cdot 2^{\binom{i-\mathcal{D}(V)}{2}} \cdot N^{-\lambda i} \cdot Z_{i,V}\right)$$

where

$$Z_{i,V} = |\{U \mid V \le U \le \mathbb{F}_2^k \ \wedge\ \mathcal{D}(U) = i \ \wedge\ U \text{ is robust}\}|.$$

Note that if V is non-robust then every $U \ge V$ is also non-robust. Hence, the outer sum terms corresponding to non-robust V's vanish. If V is robust, we claim that the inner sum is dominated by the term $i = \mathcal{D}(V)$ and that consequently

$$\mathbb{E}\left((X-\mathbb{E}(X))^k\right) = \Theta\left(\sum_{\substack{V\le\mathbb{F}_2^k \\ V \text{ robust}}} |T_V| \cdot N^{-\lambda\mathcal{D}(V)}\right). \tag{5}$$

Indeed, the number of i-dimensional subspaces containing V is given by the Gaussian binomial coefficient

$$\binom{k}{i-\mathcal{D}(V)}_2 = \frac{\prod_{j=k+1-(i-\mathcal{D}(V))}^{k}(2^j-1)}{\prod_{j=1}^{i-\mathcal{D}(V)}(2^j-1)} \le 4\cdot\frac{\prod_{j=k+1-(i-\mathcal{D}(V))}^{k} 2^j}{\prod_{j=1}^{i-\mathcal{D}(V)} 2^j} = 2^{2+(i-\mathcal{D}(V))(k-i)},$$

so the absolute value of the inner sum's i-term is at most

$$2^{\binom{i-\mathcal{D}(V)}{2}-\lambda ni+2+(i-\mathcal{D}(V))(k-i)} = 2^{2+(i-\mathcal{D}(V))(k-\frac{i+\mathcal{D}(V)+1}{2})-i\lambda n} \le 2^{-i(\lambda n+1-k)+2}.$$

In order to proceed we need to estimate the cardinalities $|T_V|$. As we show in Sections 3 and 4, at least for large enough k, Equation (5) is dominated by the term $V = \mathbb{V}^k$, the subspace of even-weight vectors.

3 The intersection of $\mathbb{V}^k$ and the γn-th layer

In this section we give tight estimates for $|T| = |T_{\mathbb{V}^k,n,\gamma}|$. As usual we assume that $0 < \gamma < \frac{1}{2}$ and γn is an even integer. We need the following terminology:

Definition 7. *Let $A_{k\times n}$ be a binary matrix.*

- *A row of A is said to satisfy the* row condition *if it weighs γn. If this holds for every row of A, we say that A satisfies the row condition.*
- *The* column condition *for A is that every column be of even weight.*
- *Recall that $T_{\mathbb{V}^k,n,\gamma}$ is the set of $k \times n$ binary matrices satisfying both the row and the column conditions.*

Our estimation of $|T|$ is based on an entropy argument (see [11] for a survey on the use of entropy in enumeration). We define a certain probability measure $\pi = \pi_{k,n,\gamma}$ on binary $k \times n$ matrices. We then show that the elements of T are highly typical for the distribution π, in the following sense: For every $A \in T$, a random matrix sampled from π is equal to A with probability exactly $2^{-h(\pi)}$. In particular, the restriction of π to T is uniform. Consequently,

$$|T| = \frac{\Pr_{A\sim\pi}(A \in T)}{\pi(A)} = \Pr_{A\sim\pi}(A \in T) \cdot 2^{h(\pi)}.$$

We then compute reasonably tight bounds on $\Pr_{A\sim\pi}(A \in T)$, yielding an estimation for $|T|$ in terms of $h(\pi)$.

In this distribution π, columns are chosen independently according to a distribution $P = P_{k,\gamma}$ that is supported on $\mathbb{V}^k$, and is invariant to permutations of the k coordinates. Naturally, we choose P so that for every i:

$$\Pr_{u\sim P}(u_i = 1) = \gamma. \tag{6}$$

Out of all distributions over $\mathbb{V}^k$ satisfying Equation (6), we seek one with maximal entropy, thus making our P as general as possible, in a sense. The theory of exponential families (E.g., [15] Chapter 3) provides a framework to describe and study maximum entropy distributions. However, we do not explicitly rely on this theory so that this paper remains self-contained.

Concretely, for some $1 > \alpha > 0$ and for every $u \in \mathbb{V}^k$ we define

$$P(u) = \frac{\alpha^{\|u\|}}{Z} \tag{7}$$

Here $Z = Z(k,\alpha) = \sum_{u\in\mathbb{V}^k} \alpha^{\|u\|}$. We claim that there is a unique $1 > \alpha > 0$ for which Condition (6) holds. First, note that

$$Z = \sum_{w \text{ is even}} \binom{k}{w}\alpha^w = \frac{(1+\alpha)^k + (1-\alpha)^k}{2}.$$

Also,

$$\Pr_{u\sim P}(u_i = 1) = \sum_{w \text{ is even}} \frac{\binom{k-1}{w-1}\alpha^w}{Z} = \alpha\frac{(1+\alpha)^{k-1} - (1-\alpha)^{k-1}}{(1+\alpha)^k + (1-\alpha)^k}$$

so that Equation (6) becomes

$$\alpha\frac{(1+\alpha)^{k-1} - (1-\alpha)^{k-1}}{(1+\alpha)^k + (1-\alpha)^k} = \gamma. \tag{8}$$

Denote the left side of this expression by $\gamma(k,\alpha)$.

Proposition 8. *Let $k \geq 2$. In the range $0 < \alpha < 1$ the function $\gamma(k,\alpha)$ increases from 0 to $\frac{1}{2}$.*

Proof. In the following, the sums are over even i, j and t:

$$\frac{\partial\gamma(k,\alpha)}{\partial\alpha} = \frac{\left(\sum_i i\binom{k-1}{i-1}\alpha^{i-1}\right)\left(\sum_j \binom{k}{j}\alpha^j\right) - \left(\sum_i \binom{k-1}{i-1}\alpha^i\right)\left(\sum_j j\binom{k}{j}\alpha^{j-1}\right)}{Z^2}.$$

Denoting $t = j + i$, the above equals

$$\frac{\sum_t \alpha^t \sum_i (2i-t)\binom{k-1}{i-1}\binom{k}{t-i}}{\alpha Z^2} = \frac{\sum_t \alpha^t \sum_i (2i-t)i\binom{k}{i}\binom{k}{t-i}}{k\alpha Z^2}.$$

Grouping the i and $t-i$ terms of the inner sum yields

$$\frac{\sum_t \alpha^t \sum_i (2i-t)^2\binom{k}{i}\binom{k}{t-i}}{2k\alpha Z^2},$$

which is clearly positive. □

It follows that the function $\gamma = \gamma(k,\alpha)$ has an inverse with respect to α, which we denote by $\alpha = \alpha(k,\gamma)$.

We summarize the new definitions pertaining to the distribution π.

Definition 9. *Let $k, n \in \mathbb{N}$.*

- *For* $0 < \alpha < 1$*, we define*

$$Z(k,\alpha) = \sum_{u\in\mathbb{V}^k} \alpha^{\|u\|} = \frac{(1+\alpha)^k + (1-\alpha)^k}{2}$$

and

$$\gamma(\alpha,k) = \sum_{\substack{u\in\mathbb{V}^k \\ u_1=1}} \frac{\alpha^{\|u\|}}{Z(k,\alpha)} = \alpha\frac{(1+\alpha)^{k-1} - (1-\alpha)^{k-1}}{(1+\alpha)^k + (1-\alpha)^k}.$$

- *If* $x \in (0, \frac{1}{2})$*, then* $\alpha(x,k) \in (0,1)$ *is the unique solution for* $\gamma(\alpha(x,k),k) = x$.

- $P_{k,\gamma}$ *is the distribution on* $\mathbb{V}^k$ *defined by*

$$P(u) = \frac{\alpha^{\|u\|}}{Z(k,\alpha)},$$

where $\alpha = \alpha(\gamma,k)$.

- $\pi_{k,n,\gamma}$ *is the distribution on binary* $k \times n$ *matrices in which the columns are sampled independently from the distribution* $P_{k,\gamma}$.

Proposition 10.

$$\alpha(k,\gamma) = \frac{\gamma}{1-\gamma} + O((1-2\gamma)^k)$$

for every fixed $\gamma \in (0, \frac{1}{2})$ *and* $k \to \infty$.

Proof. The proposition follows from the following inequality:

$$\gamma\left(k, \frac{\gamma_0}{1-\gamma_0}\right) \le \gamma_0 \le \gamma\left(k, \frac{\gamma_0+\epsilon}{1-\gamma_0}\right)$$

where $\epsilon = 2\gamma_0 \cdot \frac{(1-2\gamma_0)^{k-1}}{1-(1-2\gamma_0)^{k-1}}$.

The lower bound is easily verified, since

$$\gamma\left(k, \frac{\gamma_0}{1-\gamma_0}\right) = \gamma_0 \cdot \frac{1-(1-2\gamma_0)^{k-1}}{1+(1-2\gamma_0)^k}.$$

For the upper bound, our claim,

$$\gamma\left(k, \frac{\gamma_0+\epsilon}{1-\gamma_0}\right) = (\gamma_0+\epsilon)\frac{(1+\epsilon)^{k-1} - (1-2\gamma_0-\epsilon)^{k-1}}{(1+\epsilon)^k + (1-2\gamma_0-\epsilon)^k} \ge \gamma_0,$$

is equivalent by simple algebraic manipulation to

$$(1+\epsilon)^{k-1}\epsilon \ge (2\gamma_0+\epsilon)(1-2\gamma_0-\epsilon)^{k-1}.$$

To see that this last inequality holds, note that the l.h.s. is $\ge \epsilon$, and the r.h.s. is $\le (2\gamma_0+\epsilon)(1-2\gamma_0)^{k-1}$. Finally, the latter two expressions are identical due to the definition of ϵ. □

We next compute the entropies of the distributions we have just defined:

$$h(\pi) = nh(P)$$

where

$$\begin{aligned} h(P) &= -\sum_{u\in\mathbb{V}^k} \frac{\alpha^{\|u\|}}{Z} \log \frac{\alpha^{\|u\|}}{Z} = \log Z \cdot \sum_{u\in\mathbb{V}^k} \frac{\alpha^{\|u\|}}{Z} - \sum_{u\in\mathbb{V}^k} \frac{\|u\|\alpha^{\|u\|}}{Z} \log \alpha \\ &= \log Z - \mathbb{E}_{u\sim P}(\|u\|) \log \alpha = \log Z - k \Pr_{u\sim P}(u_1 = 1) \log \alpha = \log Z - k\gamma \log \alpha. \end{aligned}$$

To sum up:

$$h(\pi) = n(\log Z - k\gamma \log \alpha).$$

Definition 11. *For $k \in \mathbb{N}$ and $\gamma \in (0, \frac{1}{2})$, we denote*

$$F(k,\gamma) = \frac{h(\pi)}{n} = \log Z - k\gamma \log \alpha = \log\left((1+\alpha)^k + (1-\alpha)^k\right) - k\gamma \log \alpha - 1.$$

We next evaluate $\pi(A)$ for a matrix $A \in T$. Let $u_1, \ldots, u_n$ be the columns of A. Then

$$\pi(A) = \prod_{i=0}^{n} P(u_i) = \prod_{i=1}^{n} \frac{\alpha^{\|u_i\|}}{Z} = \frac{\alpha^{\|A\|}}{Z^n} = \frac{\alpha^{\gamma kn}}{Z^n} = 2^{-h(\pi)}.$$

Since π is constant on T, this yields an expression for $|T|$. Namely,

$$|T| = \frac{\Pr_{A\sim\pi}(A \in T)}{\pi(A)} = \Pr_{A\sim\pi}(A \in T) \cdot 2^{h(\pi)}. \tag{9}$$

This is complemented by the following Lemma.

Lemma 12. *Fix $\gamma \in (0, \frac{1}{2})$. Then, for every $k \geq 3$ and $n \in \mathbb{N}$, there holds*

$$\Pr_{A\sim\pi_{k,n,\gamma}}(A \in T) = n^{-\frac{k}{2}} \cdot 2^{\pm O(k)}.$$

We will prove Lemma 12 at the end of this section. Before doing so, we wish to explore its implications. Together with Equation (9), Lemma 12 allows us to conclude that

$$|T| = N^{F(k,\gamma) - \frac{k\log n}{2n} \pm O(\frac{k}{n})} \tag{10}$$

if $k \geq 3$.

For $k = 2$, a matrix in $|T|$ is defined by its first row, so

$$|T| = \binom{n}{\gamma n} = N^{h(\gamma) - \frac{\log n}{2n} + O(\frac{1}{n})}.$$

As we show later, $F(k,\gamma)$ has a linear (in k) asymptote. Consequently, the exponents in Equation (10) are dominated by the $F(k,\gamma)$ term. Thus, to understand $|T|$'s behavior we need to investigate F, which is what we do next.

3.1 Basic properties of $F(k,\gamma)$

We start with several simple observations about $F(k,\gamma)$.

Proposition 13. *For $\gamma \in (0,\frac{1}{2})$ there holds $F(2,\gamma) = h(\gamma)$. Also, $F(k,\gamma) \le k-1$ for all $k \ge 2$.*

Proof. For the first claim, note that $\gamma(2,\alpha) = \frac{\alpha^2}{1+\alpha^2}$ so $\alpha(2,\gamma) = \left(\frac{\gamma}{1-\gamma}\right)^{\frac{1}{2}}$. Hence

$$F(2,\gamma) = \log Z - 2\gamma\log\alpha = \log(1+\alpha^2) - \gamma\log(\alpha^2) = h(\gamma).$$

The second claim holds since $F(k,\gamma) = h(P)$ is the binary entropy of a distribution with support size 2^{k-1}. □

Next we develop an efficient method to calculate F to desirable accuracy. We recall (e.g., [3], p. 26) the notion *cross entropy* of D, E, two discrete probability distributions $H(D,E) := -\sum_i D(i)\log E(i)$. Recall also that $H(D,E) \ge h(D)$ with equality if and only if $D = E$. We apply this to $P = P_{k,\gamma}$, with $\alpha = \alpha(k,\gamma)$ and to Q, a distribution defined similarly according to Equation (7), but with some x in place of α. Then

$$\begin{aligned} F(k,\gamma) = h(P) \le H(P,Q) &= -\sum_u P(u)\log Q(u) = -\sum_u P(u)\log\frac{x^{\|u\|}}{Z(k,x)} \\ &= \log Z(k,x) - \sum_u P(u)\|u\|\cdot\log(x) = \log Z(k,x) - \mathbb{E}_{u\sim P}(\|u\|)\cdot\log(x) \\ &= \log Z(k,x) - \gamma k\log(x). \end{aligned} \tag{11}$$

Definition 14. *We Denote the r.h.s. of Equation (11) by*

$$g(k,\gamma,x) = \log Z(k,x) - \gamma k\log(x).$$

It follows that for an integer $k \ge 2$ and $\gamma \in (0,\frac{1}{2})$,

$$F(k,\gamma) = \min_{x\in(0,1)} g(k,\gamma,x) = \min_{x\in(0,\infty)} \log\left((1+x)^k + (1-x)^k\right) - \gamma k\log(x) - 1. \tag{12}$$

This minimum is attained at $x = \alpha(k,\gamma)$. Note that this expression allows us to conveniently compute F to desirable accuracy (see Figure 2). Also, we take Equation (12) as a definition for $F(k,\gamma)$ for all real positive k.

Proposition 15. *For an integer $k > 1$ and $0 < \gamma < \frac{1}{2}$, it holds that*

$$kh(\gamma) - 1 \le F(k,\gamma) \le kh(\gamma) + \log(1+(1-2\gamma)^k) - 1,$$

so,

$$F(k,\gamma) = kh(\gamma) - 1 + O((1-2\gamma)^k)$$

(see Figure 3).

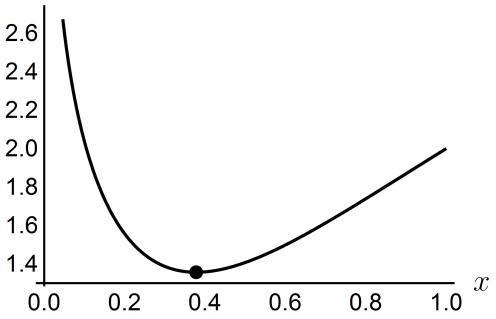

Figure 2: The function $g(3, \frac{1}{5}, x)$ and its minimum (see Equation (12)).

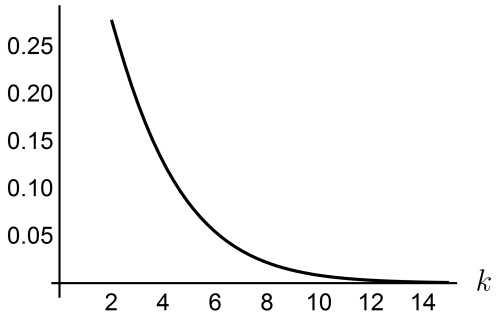

Figure 3: $F(k, \frac{1}{5}) - (k \cdot h(\frac{1}{5}) - 1)$. (See Proposition 15).

Proof. The upper bound follows from Equation (12) which yields

$$F(k,\gamma) \le g\left(k, \gamma, \frac{\gamma}{1-\gamma}\right) = kh(\gamma) + \log(1 + (1-2\gamma)^k) - 1.$$

We turn to proving the lower bound. Clearly,

$$g(k,\gamma,x) \ge \log((1+x)^k) - \gamma k \log(x) - 1.$$

The r.h.s. expression attains its minimum at $x = \frac{\gamma}{1-\gamma}$ and this minimum equals $kh(\gamma) - 1$. Equation (12) implies that this is a lower bound on $F(k,\gamma)$. □

3.2 Proof of Lemma 12

We turn to prove Lemma 12. It will be useful to view a vector $u \sim P$ as being generated in steps, with its i-th coordinate u_i determined in the i-th step. The following proposition describes the quantities involved in this process.

Proposition 16. *For $k \ge 2$ and $0 < \gamma < \frac{1}{2}$, let $u \in \mathbb{F}_2^k$ be a random vector sampled from P. For $0 \le i \le k$, let w_i denote the weight of the prefix vector $(u_1, \ldots, u_i)$. Then:*

1. *The distribution of the bit u_i conditioned on the prefix $(u_1, \ldots, u_{i-1})$ depends only on the parity of w_{i-1}.*
2. $$\Pr(u_i = 1 \mid w_{i-1} \textit{ is even}) = \alpha \cdot \frac{(1+\alpha)^{k-i} - (1-\alpha)^{k-i}}{(1+\alpha)^{k-i+1} + (1-\alpha)^{k-i+1}} \tag{13}$$
   and
   $$\Pr(u_i = 1 \mid w_{i-1} \textit{ is odd}) = \alpha \cdot \frac{(1+\alpha)^{k-i} + (1-\alpha)^{k-i}}{(1+\alpha)^{k-i+1} - (1-\alpha)^{k-i+1}}. \tag{14}$$

Proof. Fix a prefix $(u_1, \ldots, u_{i-1})$ of weight w_{i-1}. We sum over $x = \|u\| - w_i$ and $y = \|u\| - w_{i-1}$.

$$\Pr\left(u_i = 1 \mid u_1, \ldots, u_{i-1}\right) = \frac{\Pr\left(u_i = 1 \cap u_1, \ldots, u_{i-1}\right)}{\Pr\left(u_1, \ldots, u_{i-1}\right)} = \frac{\sum_{x \not\equiv w_{i-1} \bmod 2} \binom{k-i}{x} \frac{\alpha^{x+w_{i-1}+1}}{Z}}{\sum_{y \equiv w_{i-1} \bmod 2} \binom{k-i+1}{y} \frac{\alpha^{y+w_{i-1}}}{Z}}$$
$$= \alpha \frac{\sum_{x \not\equiv w_{i-1} \bmod 2} \binom{k-i}{x} \alpha^x}{\sum_{y \equiv w_{i-1} \bmod 2} \binom{k-i+1}{y} \alpha^y},$$

yielding the claim. □

We denote the r.h.s. of Equations (13) and (14) by $p_{0\to1,i} = p_{0\to1,i,k}$ and $p_{1\to0,i} = p_{1\to0,i,k}$, respectively. Also, for $0 \le i \le k$, let

$$e_i = e_{i,k} = \Pr_{u\sim P}(w_i \text{ is odd}).$$

Here are some useful facts about these terms. Equation (6) yields

$$\begin{aligned}\gamma &= \Pr_{u\sim P}(u_i = 1) = p_{0\to1,i} \cdot \Pr_{u\sim P}(w_{i-1} \text{ is even}) + p_{1\to0,i} \cdot \Pr_{u\sim P}(w_{i-1} \text{ is odd}) \\ &= p_{1\to0,i} e_{i-1} + p_{0\to1,i}(1 - e_{i-1}). \end{aligned} \tag{15}$$

By similar considerations, we have

$$e_i = e_{i-1} \cdot (1 - p_{1\to0,i}) + (1 - e_{i-1}) \cdot p_{0\to1,i}.$$

By combining these equations we find

$$p_{0\to1,i} \cdot (1 - e_{i-1}) = \frac{\gamma + (e_i - e_{i-1})}{2} \tag{16}$$

and

$$p_{1\to0,i} \cdot e_{i-1} = \frac{\gamma - (e_i - e_{i-1})}{2}. \tag{17}$$

We need some further technical propositions.

Proposition 17. *For every $\gamma \in (0, \frac{1}{2})$ there exists some $c = c(\gamma) > 0$ such that if $k \ge 3$ then*

$$e_{i,k}\ ,\ \ p_{0\to1,i,k}\ ,\ \ p_{1\to0,i,k} \in [c, 1-c]$$

for every $1 \le i \le k-1$.

Proof. It is not hard to see that both $p_{0\to1,i,k}$ and $p_{1\to0,i,k}$ are monotone in i. Therefore it suffices to check what happens for $i = 1$ and for $i = k-1$. For $i = k-1$ the two terms equal $\frac{\alpha^2}{1+\alpha^2}$ and $\frac{1}{2}$ respectively. Since α is bounded from 0 by Proposition 10, this yields the claim.

For $i = 1$ we note that $p_{0\to 1,1,k} = \gamma$.
It remains to consider $p_{1\to 0,1,k}$. Denote $x = \frac{1-\alpha}{1+\alpha}$ and note that x is bounded away from 1. This yields the bounds:

$$p_{1\to 0,1,k} = \frac{\alpha}{1+\alpha} \cdot \frac{1+x^{k-1}}{1-x^k} \geq \frac{\alpha}{1+\alpha} \cdot \frac{1-x}{1+x}$$

and

$$1 - p_{1\to 0,1,k} = \frac{1}{1+\alpha} \cdot \frac{1-x^{k-1}}{1-x^k} \geq \frac{1}{1+\alpha} \cdot \frac{1-x}{1+x}$$

We turn to deal with $e_{i,k}$. Denote $a = 1+\alpha$, $b = 1-\alpha$ and $r = k-i-1$. A bound on e_i follows from Equations (15) and (8) since

$$\begin{aligned} e_i &= \frac{\gamma - p_{0\to t,i+1}}{p_{1\to 0,i+1} - p_{0\to 1,i+1}} = \frac{\frac{a^{k-1}-b^{k-1}}{a^k+b^k} - \frac{a^{r-1}-b^{r-1}}{a^r+b^r}}{\frac{a^{r-1}+b^{r-1}}{a^r-b^r} - \frac{a^{r-1}-b^{r-1}}{a^r+b^r}} = \frac{(a^r-b^r)(a^{k-r}-b^{k-r})}{2(a^k+b^k)} \\ &= \frac{(1-x^r)(1-x^{k-r})}{2(1+x^k)} \geq \frac{(1-x)^2}{2(1+x)} \end{aligned}$$

and likewise,

$$1 - e_i = \frac{(1+x^r)(1+x^{k-r})}{2(1+x^k)} \geq \frac{(1-x)^2}{2(1+x)}.$$

□

The following simple and technical proposition will come in handy in several situations below. It speaks about an experiment where n balls fall randomly into r bins. An *outcome* of such an experiment is an r-tuple of nonnegative integers $a_1, \ldots, a_r$ with $\sum a_i = n$, where a_i is the number of balls at bin i at the end of the experiment.

Proposition 18. *Let $r \geq 2$ be an integer $\frac{1}{r} \geq c > 0$, and $p_1, \ldots, p_r \geq c$ with $\sum p_i = 1$. We drop randomly and independently n balls into r bins with probability p_i of falling into bin i. The probability of every possible outcome is at most $O\left(n^{-\frac{r-1}{2}}\right)$, where c, r are fixed and n grows.*

Proof. It is well known (e.g., [4] p. 171) that the most likely outcome of the above process $(a_1, \ldots, a_r)$, satisfies $np_i - 1 < a_i$ for every i and its probability is

$$\begin{aligned} \binom{n}{a_1, \ldots, a_r} \prod_{i=1}^r p_i^{a_i} &\leq \binom{n}{a_1, \ldots, a_r} \prod_{i=1}^r \left(\frac{a_i+1}{n}\right)^{a_i} = \binom{n}{a_1, \ldots, a_r} \prod_{i=1}^r \left(\frac{a_i}{n}\right)^{a_i} \cdot \left(1 + \frac{1}{a_i}\right)^{a_i} \\ &\leq e^r \cdot \binom{n}{a_1, \ldots, a_r} \prod_{i=1}^r \left(\frac{a_i}{n}\right)^{a_i}. \end{aligned}$$

By Stirling's apprximation for the multinomial term, the above is at most

$$O\left(\frac{\sqrt{n}}{\prod_{i=1}^{r}\sqrt{a_i}}\right) \leq O\left(\frac{\sqrt{n}}{\prod_{i=1}^{r}\sqrt{np_i - 1}}\right) \leq O\left(\frac{\sqrt{n}}{\sqrt{(cn-1)^r}}\right) \leq O\left(n^{-\frac{r-1}{2}}\right)$$

□

Proposition 19. *Let $a, c > 0$ be real and $n \in \mathbb{N}$. Consider a random variable $X \sim B(n,p)$ where $c \leq p \leq 1-c$. Let y be an integer such that $|y - pn| \leq a\sqrt{n}$. Then $\Pr(X = y) \geq \Omega\left(n^{-\frac{1}{2}}\right)$ for fixed a, c and $n \to \infty$.*

Proof. Let $q = 1 - p$, and let us denote $y = pn + x\sqrt{n}$, where $|x| \leq a$.

$$\Pr(X = y) = \binom{n}{y}p^y q^{n-y} = \binom{n}{y}\left(\frac{y}{n}\right)^y\left(\frac{n-y}{n}\right)^{n-y}\left(1 - \frac{x\sqrt{n}}{y}\right)^y\left(1 + \frac{x\sqrt{n}}{n-y}\right)^{n-y}$$

Expand into Taylor Series, using the fact that $|x|$ is bounded and $y = \Theta(n)$ to derive the following inequalities:

$$\left(1 - \frac{x\sqrt{n}}{y}\right)^y \geq \Omega\left(e^{-x\sqrt{n}}\right) \quad \text{and} \quad \left(1 + \frac{x\sqrt{n}}{n-y}\right)^{n-y} \geq \Omega\left(e^{x\sqrt{n}}\right).$$

The proposition now follows from Stirling's approximation, as

$$\binom{n}{y}\left(\frac{y}{n}\right)^y\left(\frac{n-y}{n}\right)^{n-y} \geq \Omega(n^{-\frac{1}{2}}).$$

□

We are now ready to prove the main lemma of this section.

Lemma 12. *Fix $\gamma \in (0, \frac{1}{2})$. Then, for every $k \geq 3$ and $n \in \mathbb{N}$, there holds*

$$\Pr_{A \sim \pi_{k,n,\gamma}}(A \in T) = n^{-\frac{k}{2}} \cdot 2^{\pm O(k)}.$$

Proof. Every binary $k \times n$ matrix A that is sampled from the distribution π satisfies the column condition, and we estimate the probability that the row condition holds.

By Proposition 17, there is some $c = c(\gamma) > 0$ so that $p_{0\to1,i}$, $p_{1\to0,i}$, e_i are in $[c, 1-c]$ for every $1 \leq i \leq k-1$.

We recall that A's columns are sampled independently and view A as being sampled row by row. Let b^i be the vector $A_1 + \ldots + A_{i-1} \bmod 2$. We want to observe how the ordered pairs $(\|b^i\|, \|A_i\|)$ evolve as i goes from 1 to k. By Proposition

16, this evolution depends probabilistically on $\|b^{i-1}\|$ and only on it. Namely, let s_i be the number of coordinates j where $b_j^{i-1} = 0$ and $A_{i,j} = 1$. Likewise t_i counts the coordinates j for which $b_j^{i-1} = A_{i,j} = 1$. It follows that $\|A_i\| = s_i + t_i$, and $\|b^i\| = \|b^{i-1}\| + s_i - t_i$, where $s_i \sim B(n - \|b^{i-1}\|, p_{0\to1,i})$ and $t_i \sim B(\|b^{i-1}\|, p_{1\to0,i})$ are independent binomial random variables.

Clearly $A \in T$ iff $\bigwedge_{i=1}^{k} D_i$, where D_i is the event that $\|A_i\| = \gamma n$.

We seek next an upper bound on $\Pr(A \in T)$.

$$\Pr(A \in T) = \Pr(\bigwedge_{i=1}^{k} D_i) = \prod_{i=1}^{k} \Pr(D_i \mid \bigwedge_{j=1}^{i-1} D_j)$$
$$\leq \left(\prod_{i=1}^{k-3} \max_w \Pr(D_i \mid \|b_{i-1}\| = w)\right) \cdot \max_w \Pr(D_{k-2} \wedge D_{k-1} \wedge D_k \mid \|b_{k-3}\| = w).$$

The inequality follows, since conditioned on $\|b_{i-1}\|$, the event D_i is independent of $D_1, \ldots, D_{i-1}$. We proceed to bound these terms. For $1 \leq i \leq k-3$,

$$\Pr(D_i \mid \|b_{i-1}\| = w) = \Pr(s_i + t_i = \gamma n \mid \|b_{i-1}\| = w).$$

If $w \geq \frac{n}{2}$, we condition on s_i and bound this expression from above by

$$\max_x \Pr(t_i = \gamma n - x \mid \|b_{i-1}\| = w \wedge s_i = x),$$

namely, the probability that a $B(w, p_{1\to0,i})$ variable takes a certain value. By Proposition 19 this is at most $O(w^{-\frac{1}{2}}) \leq O(n^{-\frac{1}{2}})$. When $w < \frac{n}{2}$ the same argument applies with reversed roles for t_i and s_i.

The last three rows of A require a separate treatment, since e.g., the last row is completely determined by the first $k-1$ rows. Let G be the matrix comprised of A's last three rows. Denote $\epsilon := b^{k-3}$, and let $w := \|\epsilon\|$. Again it suffices to consider the case $w \geq \frac{n}{2}$, and similarly handle the complementary situation. If $\epsilon_j = 1$, the j-th column in G must be one of the vectors $(1,0,0)^\intercal, (0,1,0)^\intercal, (0,0,1)^\intercal, (1,1,1)^\intercal$. Let a_1, a_2, a_3, a_4 denote the number of occurrences of each of these vectors respectively. There are $n-w$ indices j with $\epsilon_j = 0$, and a corresponding column of G must be one of the four even-weight vectors of length 3. We condition on the entries of these columns. Under this conditioning $a_i + a_4$ is determined by the row condition applied to row $k-3+i$, and clearly also $\sum_1^4 a_i = w$. This system of four linearly independent linear equations has at most one solution in nonnegative integers. To estimate how likely it is that this unique solution is reached, we view it as a w-balls and 4-bins experiment. The probability of each bin is a product of two terms from among $p_{0\to1,i}$, $1 - p_{0\to1,i}$, $p_{1\to0,i}$, $1 - p_{1\to0,i}$ where $i \in \{k-2, k-1\}$. Again, these probabilities are bounded away from 0. By Proposition 18 the probability of success is at most $O(n^{-\frac{3}{2}})$. Consequently, $\Pr(A \in T) \leq n^{-\frac{k}{2}} \cdot 2^{O(k)}$.

To prove a lower bound on $\Pr(A \in T)$, again we consider the rows one at a time. As before, it is easier to bound the probability of D_i by first conditioning on $\|b^{i-1}\|$. However, at present more care is needed, since letting the $\|b^i\|$'s take arbitrary values is too crude. Firstly, as long as the row conditions hold, necessarily $\|b^i\|$ is even. In addition, we monitor the deviation of $\|b^i\|$ from its expectation, which is $n \cdot e_i$. Accordingly, we define the following sets:

$$\text{For } 1 \le i \le k-2, \text{ let } \quad S_i := \{0 \le w \le n \mid \ |w - e_i \cdot n| \le \sqrt{n} \ \wedge \ w \text{ is even}\}.$$

The intuition is that the event $\|b^i\| \in S_i$ makes it likely that D_{i+1} holds, in which case it is also likely that $\|b^{i+1}\| \in S_{i+1}$. This chain of probabilistic implication yields our claim. To start, clearly $\|b^0\| \in S_0 := \{0\}$.

Now,

$$\begin{aligned}
&\Pr(A \in T) = \Pr\left(\bigwedge_{i=1}^{k} D_i\right) \ge \Pr\left(\bigwedge_{i=1}^{k} D_i \wedge \bigwedge_{i=1}^{k-2} \|b^i\| \in S_i\right) \\
&= \left(\prod_{i=1}^{k-2} \Pr\left((D_i \wedge \|b^i\| \in S_i) \ \mid \ \bigwedge_{j=1}^{i-1} (D_j \wedge \|b^j\| \in S_j)\right)\right) \cdot \Pr\left((D_{k-1} \wedge D_k) \ \mid \ \bigwedge_{j=1}^{k-2} (D_j \wedge \|b^j\| \in S_j)\right) \\
&\ge \left(\prod_{i=1}^{k-2} \min_{w \in S_{i-1}} \Pr((D_i \wedge \|b^i\| \in S_i) \ \mid \ \|b^{i-1}\| = w)\right) \cdot \min_{w \in S_{k-2}} \Pr((D_{k-1} \wedge D_k) \ \mid \ \|b^{k-2}\| = w).
\end{aligned}$$

It is in estimating these last terms that the assumption $\|b^i\| \in S_i$ becomes useful. We proceed to bound these terms, and claim the following:

1. $\min_{w \in S_{i-1}} \Pr((D_i \wedge \|b^i\| \in S_i) \ \mid \ \|b^{i-1}\| = w) \ge \Omega(\frac{1}{\sqrt{n}})$ for every $1 \le i \le k-2$.
2. $\min_{w \in S_{k-2}} \Pr((D_{k-1} \cap D_k) \ \mid \ \|b^{k-2}\| = w) \ge \Omega(\frac{1}{n})$.

It is clear that the above inequalities imply that $\Pr(A \in T) \ge n^{\frac{k}{2}} \cdot 2^{-O(k)}$, which proves the lemma.

Fix some $1 \le i \le k-2$ and let $w \in S_{i-1}$, and assume that D_i holds. Then

$$\|b^i\| - \|b^{i-1}\| \equiv s_i - t_i \equiv s_i + t_i \equiv \gamma n \equiv 0 \mod 2,$$

so that $\|b^i\|$ satisfies S_i's parity condition. Therefore

$$\Pr(D_i \wedge \|b^i\| \in S_i \ \mid \ \|b^{i-1}\| = w) = \Pr(D_i \wedge |\|b^i\| - \mathbb{E}(\|b^i\|)| \le \sqrt{n} \ \mid \ \|b^{i-1}\| = w)$$

Namely

$$\begin{aligned}
&\Pr(D_i \wedge \|b^i\| \in S_i \ \mid \ \|b^{i-1}\| = w) \\
&= \Pr(s_i + t_i = \gamma n \wedge |s_i - t_i - e_i \cdot n + w| \le \sqrt{n} \ \mid \ \|b^{i-1}\| = w). \qquad (18)
\end{aligned}$$

We want to express this last condition in terms of $x = s_i - t_i$, where clearly $s_i = \frac{\gamma n + x}{2}$ and $t_i = \frac{\gamma n - x}{2}$. Equation (18) means that $e_i \cdot n - w - \sqrt{n} \le x \le e_i \cdot n - w + \sqrt{n}$ and $x \equiv \gamma n \mod 2$. Summing over all such x's we have

$$\Pr(D_i \wedge \|b^i\| \in S_i \mid \|b^{i-1}\| = w) = \sum_x \Pr(s_i = \frac{\gamma n + x}{2}) \cdot \Pr(t_i = \frac{\gamma n - x}{2}). \quad (19)$$

Here $s_i \sim B(n - w, p_{0\to1,i})$ and $t_i \sim B(w, p_{1\to0,i})$. We use Proposition 19 to give lower bounds on a general term in Equation (19). To this end we show that $\frac{\gamma n + x}{2}$ and $\frac{\gamma n - x}{2}$ are close, respectively, to the means of s_i and t_i.

Since $w \in S_{i-1}$, we can write $w = e_{i-1} \cdot n + y$ where $|y| \le \sqrt{n}$. The bounds on x allow us to write $x = (e_i - e_{i-1})n - y + z$ for some $|z| \le \sqrt{n}$. By Equation (16),

$$\begin{aligned}
\left|\mathbb{E}(s_i) - \frac{\gamma n + x}{2}\right| &= \left|p_{0\to1,i} \cdot (n - w) - \frac{\gamma n + x}{2}\right| \\
&= \left|p_{0\to1,i} \cdot ((1 - e_{i-1})n - y) - \frac{(\gamma + e_i - e_{i-1})n - y + z}{2}\right| \\
&= \left|\frac{\gamma + (e_i - e_{i-1})}{2} n - p_{0\to1,i} \cdot y - \frac{(\gamma + e_i - e_{i-1})n - y + z}{2}\right| \\
&= \left|\frac{y - z}{2} - p_{0\to1,i} \cdot y\right| \le \sqrt{n}.
\end{aligned}$$

By Proposition 19, $\Pr(s_i = \frac{\gamma n + x}{2}) \ge \Omega(n^{-\frac{1}{2}})$. A similar proof, using Equation (17), shows that $\Pr(t_i = \frac{\gamma n - x}{2} \ge \Omega(n^{-\frac{1}{2}}))$. Thus, each of the $\Omega(\sqrt{n})$, summands in Equation (19) is at least $\Omega(n^{-1})$, so that

$$\Pr(D_i \wedge \|b^i\| \in S_i \mid \|b^{i-1}\| = w) \ge \Omega(n^{-\frac{1}{2}}).$$

We turn to proving a lower bound on $\min_{w \in S_{k-2}} \Pr((D_{k-1} \wedge D_k) \mid \|b^{k-2}\| = w)$. The column condition implies that $A_k = b^{k-1}$. Thus, for $w \in S_{k-2}$,

$$\begin{aligned}
&\Pr((D_{k-1} \wedge D_k) \mid \|b^{k-2}\| = w) = \Pr(D_{k-1} \wedge \|b^{k-1}\| = \gamma n \mid \|b^{k-1}\| = w) \\
&= \Pr(s_{k-1} + t_{k-1} = \gamma n \wedge s_{k-1} - t_{k-1} + w = \gamma n) \\
&= \Pr\left(s_{k-1} = \gamma n - \frac{w}{2}\right) \cdot \Pr\left(t_{k-1} = \frac{w}{2}\right),
\end{aligned}$$

where $s_{k-1} \sim B(n - w, p_{0\to1,k-1})$ and $t_{k-1} \sim B(w, p_{1\to0,k-1})$. Again, by applying Proposition 19 to s_{k-1} and t_{k-1}, we conclude that the above is at least $\Omega(n^{-1})$. □

4 Bounding $|T_V|$ in general

In this section we fix a robust subspace $V \le \mathbb{F}_2^k$ and bound its contribution to Equation (5). Let us sample, uniformly at random a matrix $A_{k\times n}$ in T_V. Since T_V

is invariant under column permutations, the columns of A are equally distributed. We denote this distribution on $\mathbb{F}_2^k$ by Q_V, and note that

$$\log |T_V| = h(A) \leq n \cdot h(Q_V).$$

To bound $h(Q_V)$ we employ the following strategy. Express V as the kernel of a $(k - \mathcal{D}(V)) \times k$ binary matrix B in reduced row echelon form. Suppose that $B_{i,j} = 1$. If $B_{i',j} = 0$ for every $i' < i$ we say that the coordinate j is *i-new*. Otherwise, j is said to be *i-old*. We denote the set of i-new coordinates by Δ_i. We have assumed that V is robust, so that $\bigcup_{i=1}^{k-\mathcal{D}} \Delta_i = [k]$, since $j \notin \bigcup_{i=1}^{k-\mathcal{D}} \Delta_i$ means that coordinate j is sensitive. Also B is in reduced row echelon form, so all Δ_i are nonempty.

Example. *The following $B_{3\times 7}$ corresponds to $k = 7$ and $\mathcal{D}(V) = 4$. In bold - the i-new entries in row i for $i = 1, 2, 3$.*

$$\begin{bmatrix} \mathbf{1} & 0 & 0 & \mathbf{1} & \mathbf{1} & 0 & 0 \\ 0 & \mathbf{1} & 0 & 1 & 0 & \mathbf{1} & \mathbf{1} \\ 0 & 0 & \mathbf{1} & 1 & 0 & 1 & 0 \end{bmatrix}$$

A vector v sampled from Q_V satisfies $Bv = 0$ and the expected value of each of its coordinates is $\mathbb{E}(v_i) = \gamma$. Consider v as generated in stages, with the coordinates in Δ_i determined in the i-th stage. We express v's entropy in this view:

$$h(Q_V) = h(v) = h(v_{\Delta_1}) + \sum_{i=2}^{k-\mathcal{D}(V)} h(v_{\Delta_i} \mid v_{\bigcup_{i'=1}^{i-1} \Delta_{i'}}). \tag{20}$$

We begin with the first term. Since Δ_1 is the support of B's first row and since $Bv = 0$, it follows that v_{Δ_1} has even weight. As we show in Lemma 23, the distribution P from Section 3 has the largest possible entropy for a distribution that is supported on even weight vectors with expectation γ per coordinate. Hence,

$$h(v_{\Delta_1}) \leq h(P_{|\Delta_1|,\gamma}) = F(|\Delta_1|, \gamma)$$

It takes more work to bound the other terms in Equation (20). Let $2 \leq i \leq k - \mathcal{D}(V)$. Before the i-th stage, v's i-old coordinates are already determined. Since the inner product $\langle B_i, v \rangle = 0$, the i-new coordinates of v have the same parity as its i-old coordinates. Hence $\|v_{\Delta_i}\|$'s parity is determined before this stage. Let $\delta_i = \Pr(\|v_{\Delta_i}\| \text{ is odd})$. Since conditioning reduces entropy

$$h(v_{\Delta_i} \mid v_{\bigcup_{i'=1}^{i-1} \Delta_{i'}}) \leq h(v_{\Delta_i} \mid \text{parity of } \|v_{\Delta_i}\|) = h(v_{\Delta_i}) - h(\delta_i).$$

We have already mentioned that Lemma 23 characterizes the max-entropy distribution on even-weight vectors with given per-coordinate expectation. We

actually do more, and find a maximum entropy distribution $P = P_{m,\gamma,\delta}$ on $\mathbb{F}_2^m$ satisfying

$$\Pr_{u\sim P}(u_i = 1) = \gamma \tag{21}$$

for every $1 \le i \le m$ and

$$\Pr_{u\sim P}(\|u\| \text{ is odd}) = \delta. \tag{22}$$

This distribution $P = P_{m,\gamma,\delta}$ extends something we did before, in that $P_{m,\gamma,0}$ coincides with $P_{m,\gamma}$ from Section 3

Since $v_{|\Delta_i|}$ also satisfies these conditions, this yields the bound $h(v_{\Delta_i} \mid v_{\bigcup_{i'=1}^{i-1}\Delta_{i'}}) \le F(|\Delta_i|,\gamma,\delta_i)$, where:

Definition 20. *For $m \in \mathbb{N}$, $\gamma \in (0,\frac{1}{2})$ and $\delta \in [0,1]$ we define*

$$F(m,\gamma,\delta) = h(P_{m,\gamma,\delta}) - h(\delta).$$

This generalizes Definition 11 since $F(m,\gamma) = F(m,\gamma,0)$.

We conclude that

$$\log|T_V| \le n\cdot h(Q_V) \le n\cdot\left(F(|\Delta_1|,\gamma) + \sum_{i=2}^{k-\mathcal{D}(V)} F(|\Delta_i|,\gamma,\delta_i)\right). \tag{23}$$

We determine next the distribution $P_{m,\gamma,\delta}$ and then return to the analysis of Equation (23).

4.1 The function $F(m,\gamma,\delta)$

As explained above we now find the max-entropy distribution satisfying Equations (21) and (22). The following proposition gives a necessary condition for the existence of such a distribution.

Proposition 21. *If there is a distribution satisfying conditions (21) and (22), then $\gamma \ge \gamma_{\min}$, where $\gamma_{\min} = \frac{\delta}{m}$.*

Proof. Let P be such a distribution and let $u \sim P$. By Equation (21), $\mathbb{E}(\|u\|) = \gamma m$. The lower bound on γ follows since each odd vector weighs at least 1 and thus

$$\delta = \Pr(\|u\| \text{ is odd}) \le \mathbb{E}(\|u\|).$$

□

Remark. *As we show soon, the condition in Proposition 21 is also sufficient.*

Let $m \geq 2$ and assume that m, γ, δ satisfy the strict inequalities $0 < \delta < 1$ and $\gamma_{\min} < \gamma$. We define the distribution $P = P_{m,\gamma,\delta}$ on $\mathbb{F}_2^m$ as follows:

$$P(u) = \begin{cases} \frac{\alpha^{\|u\|}}{Z} & \text{if } \|u\| \text{ is even} \\ \frac{\beta \cdot \alpha^{\|u\|}}{Z} & \text{if } \|u\| \text{ is odd} \end{cases} \tag{24}$$

where

$$Z = \sum_{u \in \mathbb{V}^m} \alpha^{\|u\|} + \beta \sum_{u \in \mathbb{D}^m} \alpha^{\|u\|} = \frac{(1+\beta)(1+\alpha)^m + (1-\beta)(1-\alpha)^m}{2}.$$

As we show there exist unique positive reals α, β for which Equations (21) and (22) hold. Note that

$$\Pr_{u \sim P}(\|u\| \text{ is odd}) = \frac{\beta\left((1+\alpha)^m - (1-\alpha)^m\right)}{2Z},$$

so Equation (22) is equivalent to

$$\beta = \frac{\delta}{1-\delta} \cdot \frac{(1+\alpha)^m + (1-\alpha)^m}{(1+\alpha)^m - (1-\alpha)^m},$$

showing in particular that α determines the value of β. Substituting the above into Equation (21) gives

$$\begin{aligned} \gamma = \Pr(u_i = 1) &= \alpha \frac{(1+\beta)(1+\alpha)^{m-1} + (1-\beta)(1-\alpha)^{m-1}}{2Z} \\ &= \alpha(1-\delta)\frac{(1+\alpha)^{m-1} - (1-\alpha)^{m-1}}{(1+\alpha)^m + (1-\alpha)^m} + \alpha\delta\frac{(1+\alpha)^{m-1} + (1-\alpha)^{m-1}}{(1+\alpha)^m - (1-\alpha)^m}. \end{aligned}$$

Denote the right side of this expression by $\gamma(m, \alpha, \delta)$. The following generalizes Proposition 8.

Proposition 22. *Let $m \geq 2$. In the range $1 > \alpha > 0$ the function $\gamma(m, \alpha, \delta)$ increases from $\gamma_{\min}$ to $\frac{1}{2}$.*

Proof. Clearly, it is enough to prove the proposition for $\delta = 0, 1$. The case $\delta = 0$ was dealt with in Proposition 8. The same argument works for $\delta = 1$ as well, since

$$\gamma = \alpha \frac{(1+\alpha)^{m-1} + (1-\alpha)^{m-1}}{(1+\alpha)^m - (1-\alpha)^m} = \frac{\sum_{i \text{ odd}} \binom{m-1}{i-1}\alpha^i}{\sum_{i \text{ odd}} \binom{m}{i}\alpha^i}.$$

□

Hence, $\gamma(m,\alpha,\delta)$ has an inverse with respect to α, which we denote $\alpha(m,\gamma,\delta)$. The uniqueness of α and β follows.

We can also define P at the extreme values $\delta \in \{0,1\}$ and $\gamma = \gamma_{\min}$ by taking limits in Equation (24). The limit $\alpha \to 0$ corresponds to $\gamma = \gamma_{\min}$ and $\beta \to 0$ resp. $\beta \to \infty$ to $\delta = 0$ or $\delta = \infty$. We still require, however, that $\gamma > 0$. E.g., if $\gamma = \gamma_{\min}$, P yields each weight 1 vector with probability $\frac{\delta}{m}$ and the weight 0 vector with probability $1-\delta$. Also, as already mentioned $P_{m,\gamma,0}$ coincides with $P_{m,\gamma}$ from Section 3.

We next compute P's entropy:

$$\begin{aligned} h(P) &= -\sum_{u\in\mathbb{V}^m} \frac{\alpha^{\|u\|}}{Z}\log\frac{\alpha^{\|u\|}}{Z} - \sum_{u\in\mathbb{D}^m}\frac{\beta\alpha^{\|u\|}}{Z}\log\frac{\beta\alpha^{\|u\|}}{Z} \\ &= \log Z - \delta\log\beta - \gamma m\log\alpha \\ &= h(\delta) + (1-\delta)\log\left((1+\alpha)^m + (1-\alpha)^m\right) + \delta\log\left((1+\alpha)^m - (1-\alpha)^m\right) \\ &\quad - \gamma m\log\alpha - 1 \end{aligned} \tag{25}$$

and recall that $F(m,\gamma,\delta) = h(P) - h(\delta)$. Consistency for the boundary cases $\delta \in \{0,1\}$ or $\gamma = \gamma_{\min}$ follows by continuity and passage to the limit. In particular, $F(m,\gamma,0) = F(m,\gamma)$. Also, let $F(m,\gamma,\delta) = -\infty$ for $\gamma < \gamma_{\min}$.

For $\gamma_{\min} < \gamma < \frac{1}{2}$ we also have the following generalization of Equation (12), which follows from the same argument:

$$F(m,\gamma,\delta) = \min_{x>0} g(m,\gamma,x,\delta) \tag{26}$$

where

$$g(m,\gamma,x,\delta) = (1-\delta)\log\left((1+x)^m + (1-x)^m\right) + \delta\log\left((1+x)^m - (1-x)^m\right) - \gamma m\log x - 1$$

with the minimum attained at $x = \alpha$.

We are now ready to show that P is the relevant max-entropy distribution.

Lemma 23. *Fix $m \geq 2$, $0 \leq \delta \leq 1$ and $\gamma_{\min} \leq \gamma < \frac{1}{2}$. The largest possible entropy of a $\mathbb{F}_2^m$-distribution satisfying Equations (21) and (22), is $h(P_{m,\gamma,\delta})$.*

Proof. Let $\mathcal{R}$ denote the polytope of $\mathbb{F}_2^m$-distributions that satisfy Conditions (21) and (22). Note that if $\gamma = \gamma_{\min}$ this polytope is reduced to a point, and the claim is trivial. We henceforth assume that $\gamma_{\min} < \gamma$, and seek a distribution $Q \in \mathcal{R}$ of maximum entropy. This distribution is unique, since the entropy function is strictly concave. Also, the value of $Q(u)$ depends only on $\|u\|$ for all $u \in \mathbb{F}_2^m$, since the optimum is unique and this maximization problem is invariant to permutation of coordinates in $\mathbb{F}_2^m$.

Let $a_i = Q(u)$ where $\|u\| = i$. We claim that

$$a_{i-2} \cdot a_{i+2} = a_i^2 \tag{27}$$

for every $2 \le i \le m-2$. Indeed, let $x, y, y', z \in \mathbb{F}_2^m$ be the indicator vectors for, respectively, the sets $\{3, \ldots, i\}$, $\{1, \ldots, i\}$, $\{3, \ldots, i+2\}$ and $\{1, \ldots, i+2\}$. Consider the distribution $Q + \theta$ where

$$\theta(u) = \begin{cases} \epsilon & \text{for } u = y, y' \\ -\epsilon & \text{for } u = x, z \\ 0 & \text{otherwise.} \end{cases}$$

Note that, if a_{i-2}, a_i, a_{i+2} are positive, $Q + \theta \in \mathcal{R}$ for $|\epsilon|$ small enough. Hence, by the optimality of Q,

$$0 = \nabla_\theta h(Q) = \log \frac{a_{i-2} a_{i+2}}{a_i^2},$$

yielding Equation (27).

We also want to rule out the possibility that exactly one side of Equation (27) vanishes. However, even if exactly one side vanishes, it is possible to increase $h(Q)$ by moving in the direction of either θ or $-\theta$.

A similar argument yields

$$a_i \cdot a_{i+3} = a_{i+1} \cdot a_{i+2} \tag{28}$$

for $0 \le i \le m-3$. Here, we take

$$\theta(u) = \begin{cases} \epsilon & \text{for } u = x, w \\ -\epsilon & \text{for } u = y, z \\ 0 & \text{otherwise.} \end{cases}$$

where x, y, z, w are the respective indicator vectors of $\{3, \ldots, i+2\}$, $\{3, \ldots, i+3\}$, $\{1, \ldots, i+2\}$ and $\{1, \ldots, i+3\}$.

Equation (27) and (28) imply that one of the following must hold:

1. $a_0, a_2, \ldots, a_{2\lfloor \frac{m}{2} \rfloor}$ and $a_1, a_3, \ldots, a_{2\lfloor \frac{m-1}{2} \rfloor + 1}$ are geometric sequences with the same positive quotient.
2. $a_0 = (1-\delta)$, $a_1 = \delta$ and $a_i = 0$ for every $i \ge 2$.
3. a_{m-1} and a_m are δ and $1-\delta$ according to m's parity, and $a_i = 0$ for all $i \le m-2$.

Case 2 corresponds to $\gamma = \gamma_{\min}$ and case 3 is impossible since $\gamma < \frac{1}{2}$, so we are left with case 1. If $0 < \delta < 1$, note that Q must satisfy Equation (24) for some positive α and β. By the uniqueness of these parameters, it follows that $Q = P$.

If $\delta = 0, 1$ then a_i vanishes for odd resp. even i's. Thus, Q satisfies Equation (24) with β going to 0 or ∞. □

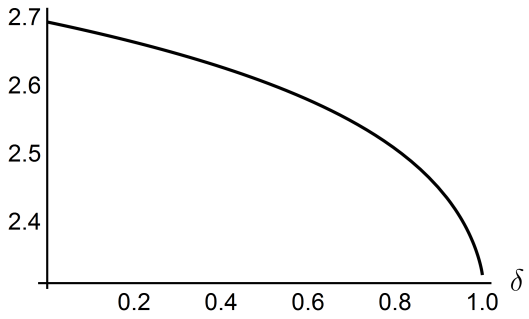

Figure 4: Illustration for Lemma 24 - $F(5, \frac{1}{5}, \delta)$

4.2 Properties of $F(m, \gamma, \delta)$

Our analysis of Equation (23) requires that we understand F's behavior in certain regimes.

Lemma 24. *If $m > 1$ is an integer, and $0 < \gamma < \frac{1}{2}$, then $F(m, \gamma, \delta)$ is a non-increasing function of δ (see Figure 4).*

Proof. If $\delta > \gamma m$, then $\gamma < \gamma_{\min}$ and $F(m, \gamma, \delta) = -\infty$. It suffices, therefore, to consider the range $0 \leq \delta < \gamma m$.

Let $0 \leq \delta < \delta' < \gamma m$ and let $\alpha = \alpha(m, \gamma, \delta)$. By Equations (26) and (25):

$$\begin{aligned} &F(m, \gamma, \delta') - F(m, \gamma, \delta) \leq g(m, \alpha, \delta') - F(m, \gamma, \delta) \\ &= (\delta' - \delta) \left(\log \left((1+\alpha)^m - (1-\alpha)^m \right) - \log \left((1+\alpha)^m + (1-\alpha)^m \right) \right) \leq 0 \end{aligned}$$

□

We now return to the case $\delta = 0$, and discuss the convexity of F in this regime.

Lemma 25. *For any $0 < \gamma < \frac{1}{2}$ the function $F(m, \gamma)$ is strictly convex in m for $m \geq 2$. (See Figure 3).*

Proof. Since γ is fixed throughout the proof, we can and will denote $F(m) = F(m, \gamma)$, $g(m, x) = g(m, \gamma, x)$. Also, $\alpha = \alpha(m, \gamma)$ is the value of x which minimizes $g(m, \gamma, x)$. This allows us to extend the definition of α to real m. Note that Equation (8) still holds in this extended setting, and that $1 > \alpha > 0$. In addition, $a = 1 + \alpha$ and $b = 1 - \alpha$.

Our goal is to show that for $m \geq 2$ there holds

$$\frac{\partial^2 F}{\partial m^2}(m, \alpha) \geq 0.$$

It follows from Equation (12) that

$$\frac{\partial g}{\partial x}(m,\alpha) = 0. \tag{29}$$

Taking the derivative w.r.t. m yields

$$\frac{\partial^2 g}{\partial x \partial m}(m,\alpha) + \frac{\partial^2 g}{\partial x^2}(m,\alpha)\frac{d\alpha}{dm} = 0. \tag{30}$$

Using Equation (29) we obtain:

$$\frac{\partial F}{\partial m} = \frac{\partial g}{\partial m}(m,\alpha) + \frac{\partial g}{\partial x}(m,\alpha)\frac{d\alpha}{dm} = \frac{\partial g}{\partial m}(m,\alpha).$$

Next,

$$\frac{\partial^2 F}{\partial m^2} = \frac{\partial^2 g}{\partial m^2}(m,\alpha) + \frac{\partial^2 g}{\partial m \partial x}(m,\alpha)\frac{d\alpha}{dm} = \frac{\partial^2 g}{\partial m^2}(m,\alpha) - \frac{\left(\frac{\partial^2 g}{\partial m \partial x}(m,\alpha)\right)^2}{\frac{\partial^2 g}{\partial x^2}(m,\alpha)}$$

where the second equality follows from Equation (30). The partial derivatives commute since g is smooth. We claim that $\frac{\partial^2 g}{\partial x^2} > 0$. To this end we refer to the definition of g in Equation (11) and take its derivative twice, then use the defining relation between γ and α (Equation (8)) to see that the sign of this derivative is the same as that of

$$\begin{aligned}
&(m-1)(a^{m-2}+b^{m-2})(a^m+b^m) - m(a^{m-1}-b^{m-1})^2 + \frac{(a^m+b^m)(a^{m-1}-b^{m-1})}{\alpha} \\
&> (m-1)(a^{m-2}+b^{m-2})(a^m+b^m) - m(a^{m-1}-b^{m-1})^2 + (a^m+b^m)(a^{m-1}-b^{m-1}) \\
&> 0.
\end{aligned}$$

Thus, to prove the lemma it suffices to show that

$$\frac{\partial^2 g}{\partial m^2}(m,\alpha)\frac{\partial^2 g}{\partial x^2}(m,\alpha) > \left(\frac{\partial^2 g}{\partial m \partial x}(m,\alpha)\right)^2$$

when $m \geq 2$.

We wish to show that $rs > t^2$, where

$$\begin{aligned}
r &= \ln 2 (a^m + b^m)^2 \frac{\partial^2 g}{\partial m^2}(m,\alpha) \\
s &= \ln 2 (a^m + b^m)^2 \frac{\partial^2 g}{\partial x^2}(m,\alpha) \\
t &= \ln 2 (a^m + b^m)^2 \frac{\partial^2 g}{\partial m \partial x}(m,\alpha)
\end{aligned}$$

We start with the first order derivatives

$$\frac{\partial g}{\partial m}(m,\alpha) = \frac{a^m \log a + b^m \log b}{a^m + b^m} - \gamma \log x$$

and

$$\frac{\partial g}{\partial x}(m,\alpha) = \frac{m(a^{m-1} - b^{m-1})}{a^m + b^m} - \frac{m\gamma}{x}.$$

Expand the second order derivatives with γ replaced according to Equation (8) to get

$$\begin{aligned}
r &= m(m-1)(a^{m-2} + b^{m-2})(a^m + b^m) - m^2(a^{m-1} - b^{m-1})^2 + \frac{m\gamma(a^m + b^m)^2}{\alpha^2} \\
&= m\left((m-1)(a^{m-2} + b^{m-2})(a^m + b^m) + \frac{(a^{m-1} - b^{m-1})(a^m + b^m)}{\alpha} - m(a^{m-1} - b^{m-1})^2\right) \\
&> m\left((m-1)(a^{m-2} + b^{m-2})(a^m + b^m) + (a^{m+2} + b^{m+2})(a^m + b^m) - m(a^{m-1} - b^{m-1})^2\right) \\
&= 4m^2 a^{m-2} b^{m-2}.
\end{aligned}$$

The inequality follows from $a^{m-1} - b^{m-1} - \alpha(a^{m-2} + b^{m-2}) = \frac{a+b}{2}(a^{m-2} - b^{m-2}) > 0$. Also

$$\begin{aligned}
s &= (a^m + b^m)(a^m(\log a)^2 + b^m(\log b)^2) - (a^m \log a + b^m \log b)^2 \\
&= a^m b^m (\log a - \log b)^2.
\end{aligned}$$

and

$$\begin{aligned}
t &= \left((m \log a + 1)a^{m-1} - (m \log b + 1)b^{m-1}\right)(a^m + b^m) \\
&\quad - m(a^{m-1} - b^{m-1})(a^m \log a + b^m \log b) - \frac{\gamma(a^m + b^m)^2}{\alpha} \\
&= 2ma^{m-1} b^{m-1} (\log a - \log b).
\end{aligned}$$

We therefore conclude that

$$rs > t^2$$

as claimed. □

The following corollary follows immediately from Lemma 25.

Corollary 26. *For every* $0 < \gamma < \frac{1}{2}$ *and every* $2 \le m \le m'$*, the holds*

$$F(m',\gamma) + F(m,\gamma) < F(m'+1,\gamma) + F(m-1,\gamma).$$

We also need the following result in order to bound $|T_V|$.

Proposition 27. *Let* $0 < \gamma < \frac{1}{2}$, $0 \le \delta \le 1$ *and* $m \ge 2$. *Then,*

$$F(1,\gamma,\delta) + F(m+1,\gamma) < F(2,\gamma) + F(m,\gamma).$$

Proof. Recall that $F(1,\gamma,\delta) \le 0$ and $F(2,\gamma) = h(\gamma)$. Thus, the claim follows from

$$F(m+1,\gamma) < F(m,\gamma) + h(\gamma).$$

This holds since F is strictly convex in m (Lemma 25) and since the limit slope of F is $h(\gamma)$ (Proposition 15). □

5 Derivation of the main theorems

We can now return to the beginning of Section 4 and complete our proof. Equation (5) can be restated as

$$\mathbb{E}\left((X - \mathbb{E}(X))^k\right) = \Theta\left(\sum_{\mathcal{D}=0}^{k-1} G_{\mathcal{D}}\right) \tag{31}$$

where

$$G_{\mathcal{D}} = N^{-\lambda\mathcal{D}} \sum_{\substack{V \le \mathbb{F}_2^k \\ \mathcal{D}(V)=\mathcal{D} \\ V \text{ robust}}} |T_V|. \tag{32}$$

We need to determine which term dominates Equation (31). We use the crude upper bound of $2^{\min(\mathcal{D},k-\mathcal{D})\cdot k}$ on the number of $\mathcal{D}$-dimensional linear subspaces V of $\mathbb{F}_2^k$. This bound follows by considering the smaller of the two: a basis for V or for its orthogonal complement.

We proceed to bound $|T_V|$ for a robust $\mathcal{D}$-dimensional subspace $V \le \mathbb{F}_2^k$. When $\mathcal{D} < \frac{k}{2}$, the trivial bound $\log|T_V| \le n \cdot h(Q_V) \le n\mathcal{D}h(\gamma)$ suffices. Indeed, a vector sampled from Q_V is determined by $\mathcal{D}$ of its bits, each of which has entropy $h(\gamma)$. It follows that

$$G_{\mathcal{D}} \le N^{\mathcal{D}(h(\gamma)-\lambda)+\frac{k\mathcal{D}}{n}} \tag{33}$$

for $\mathcal{D} < \frac{k}{2}$.

To deal with the range $\mathcal{D} \ge \frac{k}{2}$ we return to the notations of Equation (23),

$$\frac{\log(|T_V|)}{n} \le F(m_1,\gamma) + \sum_{i=2}^{k-\mathcal{D}} F(m_i,\gamma,\delta_i) \tag{34}$$

where $m_i = |\Delta_i|$ and $\sum_{i=1}^{k-\mathcal{D}} m_i = k$.

Lemma 24 yields $F(m_i, \gamma, \delta_i) \leq F(m_i, \gamma)$. By repeatedly applying Corollary 26 and Proposition 27 we get the upper bound

$$\frac{\log(|T_V|)}{n} \leq F(2(\mathcal{D}+1)-k, \gamma)+(k-\mathcal{D}-1)F(2) = F(2(\mathcal{D}+1)-k, \gamma)+(k-\mathcal{D}-1)h(\gamma).$$

Hence,

$$\begin{aligned}\log G_{\mathcal{D}} &\leq -\lambda\mathcal{D}n + (k-\mathcal{D})k + n(F(2(\mathcal{D}+1)-k, \gamma) + (k-\mathcal{D}-1)h(\gamma)) \\ &= n\left(F(2(\mathcal{D}+1)-k, \gamma) - (k-1)\lambda + (k-\mathcal{D}-1)(h(\gamma)+\lambda)\right) + (k-\mathcal{D})k. \end{aligned} \tag{35}$$

Our bounds on $G_{\mathcal{D}}$ are in fact tight up to a polynomial factor in n (but perhaps exponential in k). This follows from the existence of certain large terms in Equation (32). For $\mathcal{D} < \frac{k}{2}$, pick any map φ from $\{\mathcal{D}+1, \ldots, k\}$ *onto* $\{1, \ldots, \mathcal{D}\}$. Consider the space V that is defined by the equations $v_i = v_{\varphi(i)}$ for every $k \geq i > \mathcal{D}$. It is clear that the space V is robust. For $\mathcal{D} \geq \frac{k}{2}$, consider the contribution of the term corresponding to

$$V = \left\{ u \in \mathbb{F}_2^k \mid \sum_{i=1}^{t} u_i = 0 \wedge u_{t+1} = u_{t+2} \wedge u_{t+3} = u_{t+4} \wedge \ldots \wedge u_{k-1} = u_k \right\},$$

where $t = 2(\mathcal{D}+1) - k$.

We turn to use these bounds to compute X's central moments. We consider two cases, according the value of γ.

5.1 Moments of even order

Let k be even. By Lemma 25 and Proposition 15, there is a positive integer $k_0 = k_0(\gamma, \lambda)$ such that

$$\left\{ 2 \leq m \in \mathbb{N} \mid F(m, \gamma) - (m-1)\lambda > \frac{m}{2}(h(\gamma) - \lambda) \right\} = \{k_0, k_0+1, k_0+2, \ldots\}$$

We claim that the sum in Equation (31) is dominated by either $G_{\frac{k}{2}}$ or G_{k-1} depending on whether $k < k_0$ or $k \geq k_0$.

5.1.1 When $k < k_0$

Since $k_0 = k_0(\gamma, \lambda)$ does not depend of n, and since $k < k_0$ there is only a bounded number of $\mathbb{F}_2^k$-subspaces. We wish to compute the term $G_{\mathcal{D}} = G_{\frac{k}{2}}$. We show that in this case, the sum in Equation (32) is dominated by spaces of the form

$$V = \left\{ v \in \mathbb{F}_2^k \mid v_{i_1} = v_{j_1} \wedge v_{i_2} = v_{j_2} \wedge \cdots \wedge v_{i_{\frac{k}{2}}} = v_{j_{\frac{k}{2}}} \right\}, \tag{36}$$

where the pairs $\{i_1, j_1\}, \ldots, \{i_{\frac{k}{2}}, j_{\frac{k}{2}}\}$ form a partition of $[k]$. Clearly, for such a space V, a matrix in T_V is defined by $\frac{k}{2}$ of its rows, so

$$|T_V| = \binom{n}{\gamma n}^{\frac{k}{2}}.$$

If $U \leq \mathbb{F}_2^k$ is robust, of dimension $\frac{k}{2}$, and not of this form (36), then at least one of its associated m_i's (see Equation (34)) equals 1. By repeated application of Proposition 27, it follows that

$$|T_U| \leq N^{\frac{k}{2}F(2,\gamma)-\Omega(1)} = N^{\frac{k}{2}h(\gamma)-\Omega(1)},$$

which, as claimed, is exponentially negligible relative to $|T_V|$. The number of subspaces of the form (36) is $k!!$, whence

$$G_{\frac{k}{2}} = k!!\binom{n}{\gamma n}^{\frac{k}{2}} N^{-\lambda\frac{k}{2}}\left(1 + N^{-\Omega(1)}\right) = N^{\frac{k}{2}(h(\gamma)-\lambda)-\frac{k\log n}{4n}+O(\frac{k}{n})}.$$

We turn to show that $G_{\mathcal{D}} = o(G_{k/2})$ for any $\mathcal{D} \neq \frac{k}{2}$. For $\mathcal{D} < \frac{k}{2}$ this follows from Equation (33). For $\mathcal{D} > \frac{k}{2}$, due to Lemma 25, the r.h.s. of Equation (35) is strictly convex in $\mathcal{D}$, and therefore attains its maximum at $\mathcal{D} = \frac{k}{2}$ or $\mathcal{D} = k-1$. Since $k < k_0$, the former holds.[2]

Equation (31) yields

$$\mathbb{E}\left((X - \mathbb{E}(X))^k\right) = k!!\binom{n}{\gamma n}^{\frac{k}{2}} N^{-\lambda\frac{k}{2}}(1 + o(1)).$$

5.1.2 When $k \geq k_0$

Note that $\mathbb{V}^k$ is the one and only $(k-1)$-dimensional robust subspace of $\mathbb{F}_2^k$. Hence, by Equation (10),

$$G_{k-1} = N^{-\lambda\mathcal{D}}|T_{\mathbb{V}^k}| = N^{F(k,\gamma)-(k-1)\lambda-\frac{k\log n}{2n}+O(\frac{k}{n})}.$$

We next show that the sum in Equation (31) is dominated by this term. By Proposition 15 and Equations (35) and (33),

$$G_{\mathcal{D}} \leq N^{\mathcal{D}\cdot(h(\gamma)-\lambda)-1+O((1-2\gamma)^k)+\frac{(k-\mathcal{D})k}{n}}$$

[2]It is possible that the r.h.s. of Equation (35) attains the same value with $\mathcal{D} = \frac{k}{2}$ and $\mathcal{D} = k-1$. Note that $G_{\frac{k}{2}}$ still dominates in this case, due to polynomial factors

for all $0 \le \mathcal{D} \le k-2$. Consequently,

$$\frac{G_{\mathcal{D}}}{G_{k-1}} \le N^{(k-1-\mathcal{D})(\lambda-h(\gamma))+O((1-2\gamma)^k)+\frac{k\log n}{2n}+\frac{(k-\mathcal{D})k}{n}}.$$

For large enough k, this is at most $N^{-\Omega(k-\mathcal{D})}$, so

$$\mathbb{E}\left((X-\mathbb{E}(X))^k\right) = G_{k-1}(1-o(1)) = N^{F(k,\gamma)-(k-1)\lambda-\frac{k\log n}{2n}+O(\frac{k}{n})} \tag{37}$$

It is left to show that Equation (37) holds for *all* $k \ge k_0$, but this follows again from the convexity of F. Namely, since $k \ge k_0$, the r.h.s. of Equation (35) is strictly maximized by $\mathcal{D} = k-1$, whence $G_{\mathcal{D}} = o(G_{k-1})$ for $\frac{k}{2} \le \mathcal{D} < k-1$. For $\mathcal{D} < \frac{k}{2}$, this inequality follows from $G_{\mathcal{D}} < G_{\frac{k}{2}}$.

We are now ready to state our main theorem:

Theorem 2. *For every $\gamma < \frac{1}{2}$ and $0 < \lambda < h(\gamma)$ and for every even integer $k \le o(\frac{n}{\log n})$, the expectation $\mathbb{E}((X-\mathbb{E}(X))^k)$ is the larger of the two expressions*

$$k!!\binom{n}{\gamma n}^{\frac{k}{2}} N^{-\lambda\frac{k}{2}}(1+o(1)) \qquad \textit{and}$$
$$N^{F(k,\gamma)-(k-1)\lambda-\frac{k\log n}{2n}+O(\frac{k}{n})}.$$

There is an integer $k_0 = k_0(\gamma,\lambda) \ge 3$ such that the former term is the larger of the two when $k < k_0$ and the latter when $k \ge k_0$.

5.2 Moments of odd order

We turn to the case of odd $k > 2$. The arguments that we used to derive the moments of even order hold here as well, with a single difference, as we now elaborate.

The role previously held by $G_{\frac{k}{2}}$ is now be taken by either

$$G_{\frac{k-1}{2}} = \Theta\left(N^{\frac{k-1}{2}(h(\gamma)-\lambda)-\frac{(k-1)\log n}{4}}\right)$$

or

$$G_{\frac{k+1}{2}} = \Theta\left(N^{\frac{k-3}{2}(h(\gamma)-\lambda)+F(3,\gamma)-2\lambda-\frac{(k+1)\log n}{4}}\right).$$

These asymptotics are for bounded k. Which of these two terms is larger depends on whether $F(3,\gamma) > (h(\gamma)-\lambda)$. This yields our main theorem for moments of odd order.

Theorem 3. *For every $\gamma < \frac{1}{2}$ and $0 < \lambda < h(\gamma)$ and for every odd integer $3 \le k \le o(\frac{n}{\log n})$, the expectation $\mathbb{E}((X - \mathbb{E}(X))^k)$ is the larger of the two expressions*

$$\Theta\left(N^{\frac{k-3}{2}(h(\gamma)-\lambda)-\lambda-\frac{(k-1)\log n}{4}} \cdot N^{\max(h(\gamma),F(3,\gamma)-\lambda-\frac{\log n}{2n})}\right) \quad \text{and}$$

$$N^{F(k,\gamma)-(k-1)\lambda-\frac{k\log n}{2n}+O(\frac{k}{n})}.$$

There is an integer $k_1 = k_1(\gamma, \lambda)$ such that the former term is the larger of the two when $k < k_1$ and the latter when $k \ge k_1$.

5.3 Normalized moments

In this section we return to a theorem stated in the introduction. While it is somewhat weaker than our best results, we hope that it is more transparent and may better convey the spirit of our main findings. Recall that

$$\mathrm{Var}(X) = \binom{n}{\gamma n} N^{-\lambda}(1 + o(1)).$$

Consider the variable $\frac{X}{\sqrt{\mathrm{Var}(X)}}$. By the same convexity arguments as above, its odd moments of order up to k_0 are $o_n(1)$. This yields the following result.

Theorem 1. *Fix $\gamma < \frac{1}{2}$ and $0 < \lambda < h(\gamma)$, let $X = X_{n,\gamma,\lambda}$, and let*

$$k_0 = \min\left\{m \mid F(m,\gamma) - (m-1)\lambda > \frac{m}{2}(h(\gamma) - \lambda)\right\}.$$

Then, for $2 \le k \le o(\frac{n}{\log n})$,

$$\frac{\mathbb{E}((X - \mathbb{E}(X))^k)}{\mathrm{Var}(X)^{\frac{k}{2}}} = \begin{cases} o(1) & \text{if } k \text{ is odd and } < k_0 \\ (1 + o(1)) \cdot k!! & \text{if } k \text{ is even and } < k_0 \\ N^{F(k,\gamma)-\frac{k}{2}h(\gamma)-(\frac{k}{2}-1)\lambda-\frac{k\log n}{4n}+O(\frac{k}{n})} & \text{if } k \ge k_0 \end{cases}$$

6 Discussion

6.1 Extensions and refinements

Throughout this paper, we have limited γ to the range $(0, \frac{1}{2})$. What about $\gamma > \frac{1}{2}$? The function $F(k, \gamma, \delta)$ can be naturally extended to $\gamma \in (\frac{1}{2}, 1)$ and it satisfies the following obvious identity that follows by negating all bits in the underlying distribution.

$$F(m,\gamma,\delta) = \begin{cases} F(m, 1-\gamma, \delta) & \text{if } m \text{ is even} \\ F(m, 1-\gamma, 1-\delta) & \text{if } m \text{ is odd.} \end{cases}$$

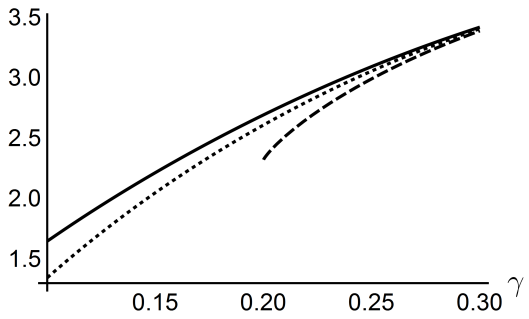

Figure 5: . Illustration for Section 6.1 - Extending F to $\gamma \in (\frac{1}{2}, 1)$. **Solid:** $F(5, \gamma)$ **Dashed:** $F(5, \gamma, 1) = F(5, 1-\gamma)$ **Dotted:** $5h(\gamma) - 1$

In particular, when $\gamma > \frac{1}{2}$ and m is odd, F is increasing rather than decreasing in δ. Also, Lemma 25 is no longer valid. In fact, $F(m, \gamma)$ is larger than the linear function $m \cdot h(\gamma) - 1$ when m is even, but smaller than it when m is odd (see Figure 5 for an example of the odd case).

It can be shown that Theorem 2 still holds in this range, but the odd moments are more complicated. The dominant term in Equation (31) is no longer necessarily a product of $\mathbb{V}^m$ spaces. Rather, it may be a $(k-2)$-dimensional space, the exact parameters of which are determined by γ.

We illustrate this unexpected additional complexity with a numerical example. Consider the following two 7-dimensional subspaces of $\mathbb{F}_2^9$:

$$U = \left\{ u \in \mathbb{F}_2^9 \mid \sum_{i=1}^{8} u_i = 0 \wedge u_9 = u_8 \right\}$$

and

$$V = \left\{ u \in \mathbb{F}_2^9 \mid \sum_{i=1}^{3} u_i = \sum_{i=4}^{8} u_i = \sum_{i=7}^{9} u_i \right\}.$$

For most values of γ there holds $|T_U| > |T_V|$, but for $\gamma > 0.9997$ the opposite inequality holds.

We believe that further analysis along the lines of the present paper may yield these odd moments as well.

Similar phenomena occur when γn is odd. Due to parity considerations, T_V is empty when there is an odd weight vector that is orthogonal to V. It turns out that computing the moments in this case comes down to essentially the same problem as the one described above for $\gamma > \frac{1}{2}$.

We next discuss the possible range of k. Namely, which moments we know. We are presently restricted to $k \leq o(\frac{n}{\log n})$, but it is conceivable that with some additional work the same conclusions can be shown to hold for all $k \leq o(n)$. The current bound arises in our analysis of the expression $\frac{G_{\mathcal{D}}}{G_{k-1}}$ in Equation (37).

Our lower bound on G_{k-1} includes a factor of $N^{-\frac{k \log n}{2n}}$, which is absent from our upper bound on $G_{\mathcal{D}}$. Lemma 12 can presumably be adapted to work for *general* robust subspaces, thereby improving this upper bound, thus yielding the same conclusions for k up to $o(n)$.

Pushing k to the linear range $k \geq \Omega(n)$ is likely a bigger challenge, since many basic ingredients of our approach are no longer valid. If $k > (1-\lambda)n+1$, we expect our code to have dimension smaller than $k-1$, whereas our main theorems show that the k-th moment of X is dominated by $(k-1)$-dimensional subsets of the (γn)-th layer of $\mathbb{F}_2^k$. Concretely, for $k \geq \Omega(n)$, our derivation of Equation (37) would fail, since the term $\frac{(k-\mathcal{D})k}{n}$ is no longer negligible. It is interesting to understand which terms dominate these very high moments.

The above discussion about large k is also related to the way that we sample random linear subspaces C in this paper. In our model there is a negligible probability that $\dim(C) > (1-\lambda)n$. This can be avoided by opting for another natural choice, viz. to sample C uniformly at random from among the $(1-\lambda)$-dimensional subspaces of $\mathbb{F}_2^n$. The effect of this choice manifests itself already in Proposition 1. This effect is negligible when $\mathcal{D} \ll (1-\lambda)n$, but becomes significant as $\mathcal{D}$ grows, e.g., under the alternative definition $\Pr(Y \subseteq C) = 0$ whenever $\dim(Y) > (1-\lambda)n$. Presumably, X's moments of order $\Theta(n)$ are sensitive to this choice of model.

There is further potential value to improving Lemma 12. A reduction in its error term would have interesting implications for the range $\frac{n}{\log n} \gg k > \frac{\log n}{-\log(1-2\gamma)}$. As things stand now, the difference between the upper and lower estimates in Proposition 15 is smaller than the error term in our estimates for the moments and yields

$$N^{kh(\gamma)-1-(k-1)\lambda-\frac{k \log n}{2n}+O(\frac{k}{n})}.$$

as our best estimate for the k-th moment. Reducing the error term in Lemma 12 may significantly improve several of our results.

Since the original submission of this paper, extensions of our techniques have turned out to be useful in other ongoing lines of research. One such line concerns Gallager's classic construction of *LDPC codes* [6]. Gallager's codes are a more structured variant of the generic random linear codes with which we deal in the current paper. Hence, it is not surprising that our methods apply to them as well. Our approach also seems helpful in analyzing the *list-decodability* parameters of certain codes, namely, for bounding the number of codewords contained in a ball of some given radius (see e.g., [7] lec. 9 for the exact definition).

Finally, we note that much of our analysis, at the very least the part contained in Sections 2 and 3, can be naturally generalized from the binary regime to random linear codes over any finite field.

6.2 Open problems

The long-term goal of this research is to understand the distribution of the random variable X. In particular, it would be interesting to understand the large deviation probabilities of this variable. Although our computation of X's moments is a step in this direction, we still do not yet have a clear view of this distribution. In particular, since all but boundedly many of X's normalized moments tend to infinity, there is no obvious way to apply moment convergence theorems.

Taking an even broader view, let us associate with a linear code C the probability measure μ on $[0, 1]$, with the CDF

$$f(x) = |C|^{-1} \cdot |\{u \in C \mid \|u\| \leq nx\}|.$$

We are interested in the typical behavior of this measure when C is chosen at random. In this context, our random variable X corresponds to the PDF of μ at the point γ. Note that μ is typically concentrated in the range $\frac{1}{2} \pm O(n^{-\frac{1}{2}})$, so that our questions correspond to large deviations in μ.

Many further problems concerning μ suggest themselves. What can be said about correlations between μ's PDF at two or more different points? Also, clearly, μ is binomial in expectation, but how far is it from this expectation in terms of moments, CDF, or other standard measures of similarity? We believe that the framework developed in this paper can be used to tackle these questions.

References

[1] Anja Becker, Antoine Joux, Alexander May, and Alexander Meurer, *Decoding random binary linear codes in $2^{n/20}$: How 1+1=0 improves information set decoding*, IACR Cryptology ePrint Archive **2012** (2012), 26.

[2] Elwyn R. Berlekamp, Robert J. McEliece, and Henk C. A. van Tilborg, *On the inherent intractability of certain coding problems (corresp.)*, IEEE Trans. Information Theory **24** (1978), no. 3, 384–386.

[3] Thomas M. Cover and Joy A. Thomas, *Elements of information theory 2nd edition (Wiley series in telecommunications and signal processing)*, Wiley-Interscience, July 2006.

[4] William Feller, *An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd Edition*, 3rd ed., Wiley, January 1968.

[5] Yuval Filmus, *Two proofs of the central limit theorem*, June 2010, `http://www.cs.toronto.edu/~yuvalf/CLT.pdf`.

[6] Robert G. Gallager, *Low-density parity-check codes*, IRE Trans. Information Theory **8** (1962), no. 1, 21–28.

[7] Venkatesan Guruswami, *Lecture notes in introduction to coding theory*, `http://www.cs.cmu.edu/~venkatg/teaching/codingtheory`, January 2010.

[8] R. J. McEliece, *A Public-Key Cryptosystem Based On Algebraic Coding Theory*, Deep Space Network Progress Report **44** (1978), 114–116.

[9] Robert J. McEliece, Eugene R. Rodemich, Howard Rumsey Jr., and Lloyd R. Welch, *New upper bounds on the rate of a code via the delsarte-macwilliams inequalities*, IEEE Trans. Information Theory **23** (1977), no. 2, 157–166.

[10] Michael Navon and Alex Samorodnitsky, *Linear programming bounds for codes via a covering argument*, Discrete & Computational Geometry **41** (2008), no. 2, 199.

[11] Jaikumar Radhakrishnan, *Entropy and counting*, `http://www.tcs.tifr.res.in/~jaikumar/Papers/EntropyAndCounting.pdf`, 2001.

[12] Oded Regev, *On lattices, learning with errors, random linear codes, and cryptography*, J. ACM **56** (2009), no. 6, 34:1–34:40.

[13] C. E. Shannon, *A mathematical theory of communication*, Bell System Technical Journal **27** (1948), no. 3, 379–423.

[14] Richard P. Stanley, *Enumerative combinatorics: Volume 1*, 2nd ed., Cambridge University Press, New York, NY, USA, 2011.

[15] Martin J. Wainwright and Michael I. Jordan, *Graphical models, exponential families, and variational inference*, Foundations and Trends in Machine Learning **1** (2008), no. 1-2, 1–305.