A short note on the joint entropy of $n/2$-wise independence

Amey Bhangale∗† Aditya Potukuchi∗‡

September 5, 2017

Abstract

In this note, we prove a tight lower bound on the joint entropy of $n$ unbiased Bernoulli random variables which are $n/2$-wise independent.

For general $k$-wise independence, we give new lower bounds by adapting Navon and Samorodnitsky’s Fourier proof of the ‘LP bound’ on error correcting codes.

This counts as partial progress on a problem asked by Gavinsky and Pudlák in [3].

1 Introduction

In this note, we study the Shannon entropy of unbiased Bernoulli random variables that are $k$-wise independent. The Shannon entropy (or entropy) of a discrete random variable $X$, taking values in $Y$, is given by $H(X) = -\sum_{y \in Y} \Pr(X = y) \log(\Pr(X = y))$, where all logarithms are base 2. A joint distribution on $n$ unbiased, Bernoulli random variables $X = (X_1, \ldots, X_n)$ is said to be $k$-wise independent if for any set $S \subset [n]$ with $|S| \leq k$, and any string $a \in \{0, 1\}^k$, we have that $\Pr(X|_S = a) = \frac{1}{2^{|S|}}$, where $X|_S$ means $X$ restricted to the coordinates in $S$.

Bounded independence distributions spaces come up very naturally in the study of error correcting codes. Let $C$ be a binary linear code over $\mathbb{F}_2$ of dimension $k$, distance $d$, and length $n$, i.e., $C$ is (also) a linear subspace of $\mathbb{F}_2^n$ of dimension $k$. Let $M$ be the $(n - k) \times n$ parity check matrix for $C$ (i.e., $C = \text{nullspace}(M)$). It can be checked that every $d - 1$ columns of $M$ are linearly independent. So, the random variable $y^TM$, where $y$ is uniformly distributed in $\mathbb{F}_2^{n-k}$, is $(d - 1)$-wise independent. This connection can be used to construct $k$-wise independent sample spaces of small support. For $k = O(1)$, BCH codes give $k$-wise independent sample spaces of support size $O(n^k)$. And

∗Department of Computer Science, Rutgers University. Research supported in part by NSF grants CCF-1253886 and CCF-1540634.
†amey.bhangale@rutgers.edu
‡aditya.potukuchi@cs.rutgers.edu
for $k = n/2$, using the Hadamard code, one gets a sample space of support size $\leq \lceil \frac{2^n}{n+1} \rceil$. It can be shown that these sample spaces are optimal in support size.

The study of entropy of joint distributions with bounded dependence was first studied by Babai in [2]. In [3], Gavinsky and Pudlák prove asymptotically tight lower bounds on the joint entropy of $k$-wise independent (not necessarily Bernoulli) random variables for small values of $k$. They prove that such a distribution must have entropy at least $\log \left( n^k \right)$. This implies the previously stated lower bound on the size of the support, as it is more general (since $H(X) \leq \log |\text{supp}(X)|$). Here, we study the case when $k = \Theta(n)$ and in particular, we show asymptotically tight bounds when $k = n/2 - o(n)$. We state the results.

**Theorem 1.1.** Let $X$ be a joint distribution on unbiased Bernoulli random variables $(X_1, X_2, \ldots, X_n)$ which is $k-1$-wise independent, then

$$H(X) \geq n - nH\left(\frac{1}{2} - \sqrt{\frac{k}{n} \left(1 - \frac{k}{n}\right)}\right) + o(n).$$

Here, for a number $p \in (0, 1)$, we say $H(p)$ to mean $-p \log p - (1-p) \log(1-p)$, i.e., the entropy of a $p$-biased Bernoulli random variable. The case where $k = n/2$ is especially simple, and conveys most of the main idea, so we prove it separately in Section 3.

**Theorem 1.2.** Let $X$ be a joint distribution on unbiased Bernoulli random variables $(X_1, X_2, \ldots, X_n)$ which is $n/2$-wise independent, then $H(X) \geq n - \log(n+1)$.

Our proof follows the Navon and Samorodnitsky’s [5] approach to the the Linear Programming bound for error correcting codes (also known as the MRRW bound, [4]). This approach uses Fourier analysis and a covering argument. Our main observation is that these techniques essentially prove a lower bound on the Renyi entropy of any $k$-wise independent distribution, which then gives us a lower bound for the (Shannon) entropy.

2 Preliminaries

The (basically spectral) argument is stated in the language of Fourier analysis, as in [5]. Henceforth, for a random variable $Y = Y(x)$, we say $E_x[Y(x)]$ (or simply $E[Y]$) to mean the expected value of $Y$ when $x$ is drawn uniformly from $\{0, 1\}^n$. For a function $f : \{0, 1\}^n \to \mathbb{R}$, the Fourier decomposition of $f$ is given by

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S(x),$$

where $\chi_S(x) := (-1)^{\sum_{i \in S} x_i}$ and $\hat{f}(S) := E_x[f(x) \chi_S(x)]$.

For any two functions $f, g : \{0, 1\} \to \mathbb{R}$, we also have an inner product, given by
\[ \langle f, g \rangle = E_x[f(x)g(x)] \]

**Theorem 2.1** (Plancherel’s identity). For any \( f, g : \{0,1\}^n \to \mathbb{R} \),

\[ \langle f, g \rangle = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \hat{g}(S). \]

For \( f, g : \{0,1\}^n \to \mathbb{R} \), the convolution of \( f \) and \( g \) denoted by \( f \ast g \) is defined as:

\[ (f \ast g)(x) = E_y[f(y)g(y + x)]. \]

**Fact 2.2.** Let \( f, g : \{0,1\}^n \to \mathbb{R} \), then \( \int f \ast g(S) = \hat{f}(S) \cdot \hat{g}(S) \) for all \( S \subseteq [n] \).

Next, we define the Rényi entropy.

**Definition 2.3** (Rényi Entropy). For a random variable \( X \) supported on a finite set \( Y \), the Rényi Entropy, denoted by \( H_\alpha(X) \) is given by:

\[ H_\alpha(X) = -\log \left( \sum_{y \in Y} p(y)^{2\alpha} \right), \]

where \( p(y) = \Pr(X = y) \).

The following is a well known relation between entropy and the Rényi entropy:

**Fact 2.4.** For a random variable \( X \) of finite support size, \( H(X) \geq H_2(X) \)

**Proof.** Let \( p_1, \ldots, p_t \) be the (nonzero) probabilities on the support of \( X \). Since \( \log \) is a concave function, from Jensen’s Inequality, we have

\[ \log \left( \sum_{i \in [t]} p_i^2 \right) \geq \sum_{i \in [t]} p_i \log(p_i), \]

which proves this fact. \( \square \)

### 3 Entropy of \( n/2 \)-wise independent distributions

Here, we give the proof of Theorem 1.2

**Proof of Theorem 1.2.** Let \( f : \{0,1\}^n \to \mathbb{R}_{\geq 0} \) be the (normalized) probability density function of an \( n/2 \)-wise independent distribution of Bernoulli random variables \( X \). So \( \Pr[X = x] = \frac{f(x)}{E[f]} \) and \( E[f] = 1 \). Let \( A \) denote the adjacency matrix of the Hamming graph (containing \( 2^n \) vertices).
Let \( L : \{0, 1\}^n \to \mathbb{R} \) such that \( L(x) = 1 \) iff \( |x| = 1 \) and 0 otherwise. First, we observe that for any function \( f \), \( Af = L \ast f \). Also \( \hat{L}(S) = n - 2|S| \). We have,

\[
\langle Af, f \rangle = \langle L \ast f, f \rangle = \sum_{S \subseteq [n]} (\hat{L} \ast f)(S) \cdot \hat{f}(S) \quad \text{(By Plancherel’s theorem)}
\]

\[
= \sum_{S \subseteq [n]} \hat{L}(S) \cdot \hat{f}(S)^2
\]

\[
= \hat{L}(\emptyset)\hat{f}(\emptyset)^2 + \sum_{1 \leq |S| \leq n/2} \hat{L}(S) \cdot \hat{f}(S)^2 + \sum_{|S| > n/2} \hat{L}(S) \cdot \hat{f}(S)^2.
\]

We now use that fact that \( f \) is a normalized pdf of \( n/2 \)-wise independent distribution and hence \( \hat{f}(S) = 0 \) for all \( 1 \leq |S| \leq n/2 \). Thus, we can upper bound \( \langle Af, f \rangle \) as

\[
\langle Af, f \rangle = n\hat{f}(\emptyset)^2 + 0 + \sum_{S \subseteq [n], |S| > n/2} \hat{L}(S) \cdot \hat{f}(S)^2
\]

\[
\leq n\hat{f}(\emptyset)^2 - \sum_{S \subseteq [n], |S| > n/2} \hat{f}(S)^2
\]

\[
= n\hat{f}(\emptyset)^2 + 1 - \sum_{S \subseteq [n]} \hat{f}(S)^2
\]

\[
= n + 1 - \mathbb{E}[f^2].
\]

Since \( \langle Af, f \rangle \geq 0 \), we have \( \mathbb{E}[f^2] \leq n + 1 \). Let \( p_1, p_2, \ldots, p_t \) be the set of nonzero probabilities on the support of the distribution. We have that \( H_2(X) = -\log(\sum p_i^2) \leq n - \log(n + 1) \). By Fact 2.4 we have \( H(X) \geq n - \log(n + 1) \).

\[
\tag*{\Box}
\]

The bound obtained above is tight when \( n + 1 \) is a power of 2. In the usual way, we identify \( \{0, 1\}^n \) with \( \mathbb{F}_2^n \). The tight case is constructed from the Hadamard code. Let \( P \) be the parity check matrix of the Hadamard code, so \( Pv = 0 \) for codewords \( v \). It can be checked that the uniform distribution on the row space of \( P \) is \( n/2 \)-wise independent. Since this a uniform distribution on \( 2^n/n+1 \) points, we have the required bound.

### 4 Entropy of \( k \)-wise independent distributions where \( k = \Theta(n) \)

We carry over the notation from the previous section. For a subset \( B \subseteq \{0, 1\}^n \), define \( \lambda_B \) as

\[
\lambda_B = \max \left\{ \frac{\langle Af, f \rangle}{\langle f, f \rangle} \mid f : \{0, 1\}^n \to \mathbb{R}, \text{supp}(f) \subseteq B \right\}.
\]
For general $k - 1$-wise independent balanced Bernoulli distributions where $k = \Theta(n)$, we have the following approach: The main idea is that given a $k - 1$-wise independent distribution $X$ given by the density function $f$, we make another random variable $Z$, given by density function $g$ as follows: sample a point, according to $f$, and shift it to randomly to some point within a hamming ball of radius $r$. Formally, let $Y$ be a random variable that is distributed accordingly in the hamming ball of radius $r$. We have a new random variable $Z = X \oplus Y$. There are three useful facts about this distribution on $Z$: (1) The resulting distribution is also $k - 1$-wise independent. (2) $X$ and $Y$ determine the $Z$, so $H(X) + H(Y) \geq H(Z)$. (3) The resulting (normalized) probability distribution function $g$ is given by $f \ast \hat{d}$ where $\hat{d} : \{0, 1\}^n \rightarrow \mathbb{R}$ is the (normalized) distribution on the hamming ball of radius $r$ around the origin, that we will use to shift. The reason for this is given by the following lemma [5]:

**Lemma 4.1.** Let $B_r$ be a Hamming ball of radius $r$, then we have:

$$\lambda_{B_r} \geq 2\sqrt{r(n-r)} - o(n)$$

We omit the proof of the above lemma since we are going to use it exactly as is presented in [5]. Now we can choose the distribution $d$ as the normalized eigenfunction of the hamming ball, so we have that $d$ is a nonnegative function with $\mathbb{E}[d] = 1$, and $Ad \geq \lambda_{B_r} d$ (since $d$ is only supported on the Hamming ball of radius $r$). Denote $\lambda_r = \lambda_{B_r}$ for convenience.

Now, we are ready to give the proof of Theorem 1.1

**Proof of Theorem 1.1.** Let $f$ be the normalized probability density function of the $k - 1$-wise independent distribution. Let $g = f \ast \hat{d}$ where $B_r$ is the indicator function of the Hamming ball of radius $r$ (to be chosen later). The thing to note is that for $S \subseteq [n]$, since $\hat{g}(S) = \hat{f}(S)\hat{d}(S)$, we have that $\hat{g}(S) = 0$ for $|S| < k$. Again, we look at $\langle Ag, g \rangle$:

$$\langle Ag, g \rangle = \langle L \ast g, g \rangle$$
$$= \sum_{S \subseteq [n]} (L \ast \hat{g})(S) \cdot \hat{g}(S)$$
$$= \sum_{S \subseteq [n]} \hat{L}(S)\hat{g}^2(S)$$
$$\leq n\hat{g}^2(0) + 1 + (n - 2k) \sum_{S \subseteq [n]} \hat{g}^2(S)$$
$$= n + (n - 2k) \mathbb{E}[\hat{g}^2].$$

(1)
On the other hand, we have:

\[
\langle Ag, g \rangle = \langle L \ast d \ast f, g \rangle \\
\geq \langle \lambda_r d \ast f, g \rangle \\
= \lambda_r \langle g, g \rangle \\
= \lambda_r E[g^2].
\]

(2)

Combining (1) and (2), we have,

\[(\lambda_r - (n - 2k))E[g^2] \leq n.\]

We choose \(r\) such that \(\lambda_r \geq n - 2k + 1\), this gives us that \(E[g^2] \leq n\).

By Jensen’s Inequality, as before, \(H[Y] + H[X] \geq H[Z] \geq n - \log n\), giving us \(H[X] \geq n - \log n - H[Y]\). Since \(Y\) is supported on the hamming ball of radius \(r\), we just use the trivial bound \(H(Y) \leq \log \left(\binom{n}{r}\right) = n H\left(\frac{r}{n}\right) + o(n)\). The value \(r\) for our purpose is \(\frac{n}{2} - \sqrt{k(n - k)} + o(n)\) which, by Lemma 4.1 completes the proof. \(\square\)

Since the best known size lower bound goes by proving a lower bound on the \(\ell^2\) norm, it easily extends to entropy, which, by Jensen’s Inequality, is shown to be a ‘weaker’ quantity.

5 Acknowledgements

We would like to thank Swastik Kopparty for the many very helpful discussions, and suggestions for the writeup.

References

[1] N. Alon and J. H. Spencer, The Probabilistic Method, Third Edition, Wiley, 2010.

[2] L. Babai, Entropy Versus Pairwise Independence (Preliminary version) \(\text{http://people.cs.uchicago.edu/~laci/papers/13augEntropy.pdf}\), 2013.

[3] D. Gavinsky and P. Pudlák, On the joint entropy of d-wise-independent variables, Commentationes Mathematicae Universitatis Carolinae, 57, 3, pages 333–343,2016

[4] R. McEliece, E. Rodemich, H. Rumsey and L. Welch, New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities, IEEE Transactions on Information Theory, vol. 23, no. 2, pp. 157-166, Mar 1977.

[5] M. Navon, A. Samorodnitsky, Linear programming bounds for codes via a covering argument, Discrete and Computational Geometry, 41, 2, 2009.