Entropy Versus Pairwise Independence

László Babai
University of Chicago
laci ** cs * uchicago * edu

Preliminary version, August 22, 2013

Abstract

We give lower bounds on the joint entropy of n pairwise independent random variables. We show that if the variables have no dominant value (their min-entropies are bounded away from zero) then this joint entropy grows as $\Omega(\log n)$. This rate of growth is known to be best possible. If k-wise independence is assumed, we obtain an optimal $\Omega(k \log n)$ lower bound for not too large k. We also show that the joint entropy of an arbitrary family of pairwise independent random variables grows as $\Omega(\min(L, \sqrt{\log(2+L)})$ where L is the sum of the entropies of the variables in the family. We expect that the $\sqrt{\log}$ in this expression can be improved to log. We also prove a tight $\Omega(\log \log n)$ lower bound on the joint entropy of n balanced Bernoulli trials with pairwise correlation bounded away from 1.

1 Introduction

1.1 The main results

$H(X)$ denotes the Shannon entropy of the discrete random variable X, i. e., $H(X) = -\sum_x P(X = x) \log P(X = x)$. (All logarithms are to the base 2.) Let $X_1, \dots, X_n$ be random variables and $X = (X_1, \dots, X_n)$. It is well known that if the X_i are independent then $H(X) = \sum_{i=1}^n H(X_i)$.

The purpose of this note is to study the effect on the joint entropy of relaxing the independence assumption to *pairwise independence.*

The effect can be dramatic. For instance, if the X_i are independent balanced Bernoulli trials (success probability 1/2) then $H(X_i) = 1$ and $H(X) = n$. On the other hand, it is well known (at least since 1965 [12]) that pairwise independence can be achieved with logarithmic joint entropy:

Proposition 1.1. *For $n = 2^k - 1$ there exist n pairwise independent balanced Bernoulli trials over a uniform probability space of size $n+1$. In particular, their joint entropy is* $\log(n+1)$.

We believe that this bound should be optimal.

Conjecture 1.2. *Let $X_1, \dots, X_n$ be pairwise independent balanced Bernoulli trials. Then*

$$H(X_1, \dots, X_n) \ge \log(n+1).$$

Our main result confirms this conjecture up to a constant factor. In fact, this lower bound holds even for the min-entropy.

Let $H_{\min}(X)$ denote the *min-entropy* of the random variable X, defined as $H_{\min}(X) = -\log(\max_x P(X = x))$. Clearly $H(X) \geq H_{\min}(X)$ and equality holds exactly when all outcomes of X are equally likely.

A *Bernoulli trial* is a random variable that takes only two values, "success" and "failure." A Bernoulli trial is *balanced* if the success probability is $1/2$.

We consider pairwise independent variables with no dominant value (i. e., with min-entropy bounded away from 0). The following is our main result.

Theorem 1.3. *Let $X_1, \ldots, X_n$ be pairwise independent random variables with $H_{\min}(X_i) \geq -\log(1-p)$ where $0 < p \leq 1/2$ (no value is taken with probability greater than $1-p$). Then*

$$H_{\min}(X_1, \ldots, X_n) > \left\lfloor \frac{\log n - 2}{3\left(1 + \log \frac{1-p^*}{p^*}\right)} \right\rfloor . \tag{1}$$

where $p^* = \min(1/3, p)$. In particular, for $4n^{-1/3} \leq p \leq 1/2$ we obtain

$$H_{\min}(X_1, \ldots, X_n) = \Omega\left(\frac{\log n}{\log(1/p)}\right) \tag{2}$$

where the Ω notation hides a positive absolute constant.

If the X_i are Bernoulli trials then we can take $p^* = p$. In particular, if the X_i are pairwise independent balanced Bernoulli trials then $H_{\min}(X_1, \ldots, X_n) > (1/3)\log n - (5/3)$.

The proof is based on a result of possibly independent interest: a family of n pairwise independent Bernoulli trials always includes a logarithmic number of variables that behave nearly like fully independent variables, assuming the min-entropies of these Bernoulli trials are bounded away from zero. The exact statement follows.

Let $Y_1, \ldots, Y_n$ be non-constant random variables and let X_i denote the normalized version of Y_i. Using terminology that goes back to Umesh Vazirani's thesis [20] (cf. [16, 3]), we say that the Y_i are ϵ*-biased* if for all nonempty $I \subseteq \{1, \ldots, n\}$ we have $|E(X_I)| \leq \epsilon$ where $X_I = \prod_{i \in I} X_i$.

Theorem 1.4. *Let $X_1, \ldots, X_n$ be pairwise independent Bernoulli trials with success probabilities between p and $1-p$ where $0 < p \leq 1/2$. Then for every ℓ there exist $i_1, \ldots, i_\ell$ $(1 \leq i_j \leq n)$ such that $X_{i_1}, \ldots, X_{i_\ell}$ are ϵ_ℓ-biased, where*

$$\epsilon_\ell = \frac{\left(\sqrt{2}\left(\frac{1-p}{p}\right)\right)^{\ell-1}}{\sqrt{n}}. \tag{3}$$

Note that for constant p, the value ϵ_ℓ remains small up to $\ell = \Omega(\log n)$ steps.

We also study the growth of the joint entropy of families of variables that have a large fraction of their entropy concentrated on a small mass. We find that in this case we get close to additivity.

We say that a subset Ψ of the sample space is *stable* w. r. to the random variable X if it is of the form $\Psi = X^{-1}(U)$ for some subset U of the range of X. We define the Ψ-portion of the entropy of X as

$$H_\Psi(X) = -\sum_{x \in U} P(X = x) \log P(X = x). \tag{4}$$

Definition 1.5. The *entropy-concentration* of X is the minimum mass $P(\Psi)$ of an X-stable set Ψ such that $H_\Psi \geq H(X)/2$. We denote this quantity by $c(X)$.

We note that for non-constant variables $c(X) \le 2/3$.

Proposition 1.6. *Let $X = (X_1, \dots, X_n)$ where the X_i are pairwise independent random variables. Assume $\sum_{i=1}^n c(X_i) \le 1/2$. Then $H(X) \ge (1/4)\sum_{i=1}^n H(X_i)$.*

In particular, we obtain near-additivity when the sum of the entropies is bounded.

Proposition 1.7. *Let $X = (X_1, \dots, X_n)$ where the X_i are pairwise independent random variables. Let $L = \sum_{i=1}^n H(X_i)$. Then $H(X) \ge \min(3/16, L/4)$.*

Combining the results stated, we obtain an unconditional lower bound on the joint entropy as a function of the sum of entropies.

Theorem 1.8. *Let $X_1, \dots, X_n$ be pairwise independent random variables; let $L = \sum_{i=1}^n H(X_i)$. Then $H(X) = \Omega(\min(L, \sqrt{\log(2+L)})$.*

Conjecture 1.9. *Let $X_1, \dots, X_n$ be pairwise independent random variables; let $L = \sum_{i=1}^n H(X_i)$. Then $H(X) = \Omega(\min(L, \log(2+L))$.*

Relaxing the condition of pairwise independence to correlation bounded away from 1, we obtain a tight $\Omega(\log\log n)$ lower bound on the joint entropy of n balanced Bernoulli trials (Propositions 7.1 and 7.2).

1.2 k-wise independence

The variables $X_1, \dots, X_n$ are *k-wise independent* if every k of them are independent. For fixed k, such variables have long been known to be realizable over a uniform probability space of size $O(n^k)$ [10] (1974) and therefore with joint entropy $\lesssim k \log n$ (a result rediscovered by a number of authors in the theory of computing, cf. [1] for the history). For balanced Bernoulli trials, a sample space of size $O(n^{\lfloor k/2 \rfloor})$ suffices, as shown in [1] using BCH codes.

It was also shown in [1] and (for balanced Bernoulli trials) in [9] that if $X_1, \dots, X_n$ are k-wise independent non-constant random variables then the size of the sample space is at least $n^{\lfloor k/2 \rfloor}$.

We extend this lower bound to the entropy context.

Theorem 1.10. *Let $X_1, \dots, X_n$ be k-wise independent random variables with $H_{\min}(X_i) \ge -\log(1-p)$ where $0 < p \le 1/2$ (no value is taken with probability greater than $1-p$). Let $t = \lfloor k/2 \rfloor$. Then*

$$H_{\min}(X_1, \dots, X_n) > \left\lfloor \frac{\log N - 2}{3\left(1 + \log \frac{1-p^*}{p^*}\right)} \right\rfloor . \tag{5}$$

where $N = \binom{n}{t}$ and $p^* = \min(1/3, 1-(1-p)^t)$. In particular, for $k \le \sqrt{n}$ and $pk \ge 1/2$ we obtain

$$H_{\min}(X_1, \dots, X_n) = \Omega(k \log n) \tag{6}$$

and for $k \le \sqrt{n}$ and $10n^{-(k-1)/6} \le pk \le 1/2$ we obtain

$$H_{\min}(X_1, \dots, X_n) = \Omega\left(\frac{k \log n}{\log(1/pk)}\right) \tag{7}$$

where the Ω notation hides a positive absolute constant.

Proof. We adapt a trick from [1]. Set $t = \lfloor k/2 \rfloor$. Consider the $N = \binom{n}{t}$ variables $Y_I = (X_i \mid i \in I)$ where $I \subseteq [n]$ and $|I| = t$. If the X_i are k-wise independent then the Y_I are pairwise independent. It is also clear that $H_{\min}(Y_I) \geq -t \log(1-p)$ (Y_I takes no value with probability greater than $(1-p)^t$ because of t-wise independence). So we can apply Theorem 1.3 with N in the role of n and $1-(1-p)^t$ in the role of p. We omit the elementary calculations. Finally we observe that given that the Y_I are determined by the X_i, we have $H_{\min}(X_1, \ldots, X_n) \geq H_{\min}(Y_I \mid I \subseteq [n], |I| = k)$. □

The $k < \sqrt{n}$ cutoff was arbitrary; the same statement holds with $k < n^{1-c}$ for any constant $c > 0$. But the main interest is in small values of k (constant or $O(\log n)$).

We note that for $k < \sqrt{n}$ and for k-wise independent balanced Bernoulli trials, our lower bound $\Omega(k \log n)$ is tight within an absolute constant factor.

It is known that considerably smaller sample spaces realize *nearly k-wise independence* [5, 2, 16, 3]. The study of the entropy version of this concept would be next on the agenda.

1.3 Motivation

Pairwise independence has been ubiquitous in algorithms as well as in complexity theory (cf. [13]) at least since the introduction of universal hashing [7] and the classic paper by Karp and Wigderson [11] that used pairwise independent random variables over a small sample space to derandomize algorithms.

The bigger context is that we view *randomness as a resource.* This point of view has been explicitly stated in dozens if not hundreds of papers.

Entropy is the *measure of randomness*, implicit in the above statement, and explicit in the sizable literature that studies randomness-extraction from "weakly random" sources (see e. g. [17]). More recently we have witnessed a move from viewing "size" as the central resource to viewing "entropy" as a key resource. For instance, this information-theoretic point of view has transformed communication complexity, a classical area of complexity theory, over the past decade (cf. e g., [8, 4, 6]). A similar movement is observable in combinatorics, especially arithmetic combinatorics [19, 18, 14], a fast-moving area of mathematical with close links to the theory of computing (cf. [21]); sets are replaced by distributions and the size measure by entropy.

All these developments would appear to make our question, the entropy content of pairwise independent variables, one of the most natural foundational questions for this transition and I expected to *easily* find a body of literature on the subject.

To my surprise, my limited search did not turn up a single paper dealing explicitly with this type of question, either in the information theory literature or in the theory of computing literature. Of course this is no proof that such works do not exist. Another possibility is that the results might be *implicit* in some general inequalities like those in [15]. So far, I have not seen an indication of this either.

Although at present I only have vague notions of potential applications of entropy lower bounds of the type described in this paper (the work started from a combinatorial application I cannot at present connect to complexity theory, see the Appendix), I am convinced that before long, such lower bounds will make their way into complexity theory.

2 Preliminaries

We consider discrete probability spaces. All our random variables are assumed to have finite second moment and finite entropy.

Some more notation. $E(X)$ and $\mathrm{Var}(X)$ denote the expected value and the variance, resp., of the random variable X. We write $[n] = \{1, \dots, n\}$. For a sequence $W_1, \dots, W_n$ of variables and for $I \subseteq [n]$, we write $W_I = \prod_{i \in I} W_i$.

We define the *dominance* of the random variable X as

$$\mathrm{dom}(X) = \max_x P(X = x). \tag{8}$$

Note that

$$H(X) \geq H_{\min}(X) = -\log \mathrm{dom}(X). \tag{9}$$

In particular, if $\mathrm{dom}(X) \leq 1/2$ then $H(X) \geq 1$.

Our main result speaks of variables with dominance $\leq 1 - p$. For fixed p this means the variables have min-entropy bounded away from zero.

Next we reduce Theorem 1.3 to the case when all the X_i are Bernoulli trials. We begin with a well-known fact.

Observation 2.1. *Let X be a random variable, f a function, and $Y = f(X)$. Then $H(X) \geq H(Y)$ and $H_{\min}(X) \geq H_{\min}(Y)$.*

Lemma 2.2. *Let X be a random variable with dominance $\leq 1 - p$ where $0 < p \leq 1/2$. Then there exists a function f such that the variable $Y = f(X)$ is a Bernoulli trial with success probability between p^* and $1 - p^*$ where $p^* = \min(1/3, p)$. (Of course if X is Bernoulli trial then we can set $Y = X$ and $p^* = p$.)*

Proof. Let $x_1, \dots, x_k$ be the distinct values taken by X; let $p_i = P(X = x_i)$. For $S \subseteq [k]$ let $P_S = \sum_{i \in S} p_i$. Let $T \subseteq [k]$ be a subset that minimizes the value $1/2 - P_T$ subject to the constraint $P_T \leq 1/2$. Set $Y = -1$ if $X \in \{x_i \mid i \in T\}$ and $Y = 1$ otherwise.

We need to show that $P_T \geq p^*$.

If $\max_i p_i \geq 1/2$ then clearly $P_T = d \geq p^*$.

If $\max_i p_i \leq 1/2$ then it is easy to see that $P_T \geq 1/3 \geq p^*$. $\square$

Reduction of Theorem 1.3 to the case of Bernoulli trials. Use the Lemma to find functions f_i such that $Y_i := f_i(X_i)$ is a Bernoulli trial with success probability between p^* and $1 - p^*$. The reduction is completed by Observation 2.1. $\square$

We review some terminology. We say that a random variable X is *normalized* if $E(X) = 0$ and $\mathrm{Var}(X) = 1$. The *normalized version* of a non-constant random variable Y is the normalized variable $X = (Y - E(Y))/\mathrm{Var}(Y)$.

Consider the probability space (Ω, P) and the (in our case finite-dimensional) Hilbert space $L^2(\Omega, P)$. This is the space of random variables under the inner product $\langle X, Y \rangle = E(XY)$. We note that a family of pairwise independent normalized random variables forms an orthonormal system in this space. The following statement, variably referred to as Plancherel's or Bessel's inequality, is immediate.

Proposition 2.3. *If $X_1, \dots, X_n$ are pairwise independent normalized random variables and U is an arbitrary random variable then*

$$E(U^2) \geq \sum_{i=1}^{n} (E(X_i U))^2. \tag{10}$$

3 Proof of the main theorem

We restate Theorem 1.3 for the case of Bernoulli trials.

Theorem 3.1. *Let $Y_1, \dots, Y_n$ be pairwise independent Bernoulli trials with success probabilities between p and $1-p$, where $0 < p \le 1/2$. Then*

$$H_{\min}(Y_1, \dots, Y_n) > \left\lfloor \frac{\log n - 2}{3\left(1 + \log \frac{1-p}{p}\right)} \right\rfloor. \tag{11}$$

In particular, if $p = 1/2$ (pairwise independent balanced Bernoulli trials) then $H_{\min}(Y_1, \dots, Y_n) > (1/3)\log n - (5/3)$.

We prove this by demonstrating that there are at least a logarithmic number among the variables that have small bias.

Lemma 3.2. *Let $X_1, \dots, X_n$ be pairwise independent normalized random variables and let $U_1, \dots, U_k$ be arbitrary random variables with finite second moments. Then there exists i such that $(E(X_i U_j))^2 \le (k/n)E(U_j^2)$ holds simultaneously for all j.*

Proof. Set $\alpha_{ij} = E(X_i U_j)$. Let n_j denote the number of those i for which $\alpha_{ij}^2 > (k/n)E(U_j^2)$. Then, by eq. (10), we have $n_i < n/k$. Therefore $\sum_{j=0}^{k} n_i < n$ and hence there is an i such that $\alpha_{ij}^2 \le (k/n)E(U_j^2)$ for all j. □

Remark 3.1. The quantity k/n in the Lemma can be replaced by $k/(n+1)$ if $(E(U_j))^2 > \frac{k}{n+1}E(U_j^2)$ holds for at least one j. Indeed, let X_0 be the constant random variable taking the value 1. Then $X_0, \dots, X_n$ are still an orthonormal system and the same argument as above selects an appropriate i. If X_0 is selected then for all j we have $(E(X_0 U_j))^2 \le \frac{k}{n+1}E(U_j^2)$, contrary assumption.

Proof of Theorem 1.4. Let p_i denote the success probability of Y_i; we may assume $p \le p_i \le 1/2$. Let X_i be the normalized version of Y_i; so $X_i = \sqrt{(1-p_i)/p_i}$ with probability p_i and $X_i = -\sqrt{p_i/(1-p_i)}$ with probability $1-p_i$. We need to prove the statement for the X_i.

We prove it by induction on ℓ. It holds for $\ell = 1$.

Suppose $\ell \ge 2$ and $i_1, \dots, i_{\ell-1}$ have already been found such that the corresponding X_i are $\epsilon_{\ell-1}$-biased. Let us write $U_j = X_{i_j}$ $(j \le \ell - 1)$. Let us apply Lemma 3.2 to the family $\{U_J \mid J \subseteq [\ell-1]\}$ (so $k = 2^{\ell-1}$). Let X_{i_ℓ} be the X_i selected by the Lemma. We need to verify that $|E(U_J X_{i_\ell})| \le \epsilon_\ell$ for all $J \subseteq [\ell-1]\}$.

In applying the Lemma, we set $k = \ell - 1$ and we estimate U_J trivially as $|U_J| \le ((1-p)/p)^{(\ell-1)/2}$. This gives the desired bound. □

Lemma 3.3. *Let $Y_1, \dots, Y_\ell$ be an ϵ-biased family of Bernoulli trials where Y_i has success probability p_i and $p \le p_i \le 1/2$ for all i. Then $H_{\min}(Y_1, \dots, Y_\ell) > -\log(2^{-\ell} + ((1-p)/p)^{(\ell-1)/2}\epsilon)$. In particular, if $\epsilon \le (p/(1-p))^{(\ell-1)/2}2^{-\ell}$ then $H_{\min}(X_1, \dots, X_\ell) > \ell - 1$.*

Proof. Let X_i be the normalized version of Y_i. Let $\alpha_i = \sqrt{(1-p_i)/p_i}$ so $P(X_i = \alpha_i) = p_i$ and $P(X_i = -1/\alpha_i) = 1 - p_i$.

Let $\delta = (\delta_1, \dots, \delta_\ell)$ where $\delta_i \in \{\pm 1\}$. Let A_δ denote the event that for each i, the outcome of Y_i is "success" if $\delta_i = 1$ and "failure" if $\delta_i = -1$. In other words, A_δ represente the event that $(\forall i)(X_i = \delta_i \alpha_i^{\delta_i})$. Let θ_δ denote the $(0,1)$-indicator variable for A_δ. Then

$$\theta_\delta = 2^{-\ell} \prod_{i=1}^{\ell} (1 + \delta_i \alpha_i^{-\delta_i} X_i) = 2^{-\ell} \sum_{I \subseteq [m]} \Delta_I X_I \tag{12}$$

where $\Delta_I = \prod_{i\in I} \delta_i \alpha_i^{-\delta_i}$. Note that $|\Delta_I| \le \alpha_I \le ((1-p)/p)^{|I|/2}$.

Now $P(A_\delta) = E(\theta_\delta) = 2^{-\ell} \sum_{I\subseteq[m]} E(\Delta_I X_I) = 2^{-\ell} + R$ where R comprises all the terms with $I \ne \emptyset$, so

$$|R| \le 2^{-\ell} \sum_{I\subseteq[\ell], I\ne\emptyset} |\Delta_I||E(X_I)| < ((1-p)/p)^{(\ell-1)/2}\epsilon. \tag{13}$$

We infer that $P(A_\delta) < 2^{-\ell} + ((1-p)/p)^{(\ell-1)/2}\epsilon$ and therefore

$$H_{\min}(X_1,\dots,X_\ell) = -\log(\max_\delta P(A_\delta)) > -\log(2^{-\ell} + ((1-p)/p)^{(\ell-1)/2}\epsilon).$$

□

Proof of Theorem 1.3. Select the largest ℓ such that $((1-p)/p)^{(\ell-1)/2}\epsilon_\ell \le 2^{-\ell}$ where ϵ_ℓ is defined by eq. (3). So

$$\ell = 1 + \left\lfloor \frac{\log n - 2}{3\left(1 + \log\frac{1-p}{p}\right)} \right\rfloor \tag{14}$$

Apply Theorem 1.4 to select $X_{i_1},\dots,X_{i_\ell}$ that are ϵ_ℓ-biased. Now apply Lemma 3.3 to this family to conclude that

. $$H_{\min}(X_1,\dots,X_n) \ge H_{\min}(X_{i_1},\dots,X_{i_\ell}) > \ell - 1.$$ □

4 For small sets, pairwise independence is almost as good as independence

We show that pairwise independence suffices for the contributions of small subsets to the joint entropy nearly to add up.

Let Ω be the sample space and X a random variable over Ω. For $\omega \in \Omega$ let $\rho_X(\omega) = \log(1/P(X = X(\omega))$. We note that

$$H(X) = E(\rho_X). \tag{15}$$

For $\Psi \subseteq \Omega$, let $H_\Psi(X)$ denote the contribution of Ψ to the entropy of X, defined as

$$H_\Psi(X) = \sum_{\omega\in\Psi} P(\omega)\rho_X(\omega) = E(\theta_\Psi \rho_X) \tag{16}$$

where θ_Ψ denotes the $(0,1)$-indicator variable for Ψ. This agrees with the definition (4).

Lemma 4.1. *Let X and Y be independent random variables. Let $\Psi \subseteq \Omega$ be a Y-stable set. Then $H_\Psi(X) = P(\Psi)H(X)$.*

Proof. Since Ψ is Y-stable, its indicator variable θ_Ψ and X are independent. Therefore $H_\Psi(X) = E(\theta_\Psi\rho_X) = E(\theta_\Psi)E(\rho_X) = P(\Psi)H(X)$. □

Let $X = (X_1,\dots,X_n)$. Then

$$\rho_{X_i} \le \rho_X \tag{17}$$

because the partition of Ω to atomic stable sets of X is a refinement of the corresponding partition for X_i.

Observation 4.2. *Let the X_i be random variables and $X = (X_1, \ldots, X_n)$. If $\Psi_1, \ldots, \Psi_n \subseteq \Omega$ are disjoint X-stable sets then $H(X) \geq \sum_{i=1}^n H_{\Psi_i}(X_i)$.*

This is immediate from inequality (17).

Lemma 4.3. *Let $X_1, \ldots, X_n$ be pairwise independent random variables and $X = (X_1, \ldots, X_n)$. For each i, let $\Psi_i \subseteq \Omega$ be an X_i-stable set; let $q_i = P(\Psi_i)$. If $\sum_{i=1}^n q_i \leq 1/2$ then $H(X) \geq (1/2) \sum_{i=1}^n H_{\Psi_i}(X_i)$.*

Proof. Let $\Pi_i = \bigcup_{j:j \neq i} \Psi_j$ Let $\Delta_i = \Psi_i \setminus \Pi_i$. The Δ_i are disjoint. So the statement will follow from Obs. 4.2 and the following.

Claim. $H_{\Delta_i}(X_i) \geq (1/2) H_{\Psi_i}(X_i)$.

Proof. By Lemma 4.1, for $j \neq i$ we have $H_{\Psi_i \cap \Psi_j}(X_i) = q_j H_{\Psi_i}(X_i)$. Therefore $H_{\Psi_i \cap \Pi_i}(X_i) \leq \sum_{j \neq i} H_{\Psi_i \cap \Psi_j}(X_i) = \sum_{j \neq i} q_j H_{\Psi_i}(X_i) \leq (1/2) H_{\Psi_i}(X_i)$. We conclude that $H_{\Delta_i}(X_i) = H_{\Psi_i}(X_i) - H_{\Psi_i \setminus \Pi_i}(X_i) \geq (1/2) H_{\Psi_i}(X_i)$. □

□

5 When the total entropy is bounded

In this section we prove Prop. 1.7.

Lemma 5.1. *Let $\mathrm{dom}(X) = 1 - c$. Let Ψ be the set on which X does not take its dominant value (so $P(\Psi) = c$). Then $H_\Psi(X) \geq H(X)/2$.*

Proof. Clearly $H_\Psi(X) \geq c \log(1/c)$ so all we need to show is that for $c \leq 1/3$ we have $c \log(1/c) \geq (1-c) \log(1/(1-c))$. This is true for $c = 1/3$ and only becomes more true as c decreases. □

Observation 5.2. *Let $d = 1 - \mathrm{dom}(X)$. If $d \leq 1/2$ then $H(X) \geq d \log(1/d)$. If $d \geq 1/2$ then $H(X) \geq 1$.*

Proof of Prop. 1.7. Let $c_i = 1 - \mathrm{dom}(X_i)$.

First assume $L \leq 1/2$. In particular, for all i we have $H(X_i) \leq 1/2$ and therefore, by Obs. 5.2, $c_i \leq d$.

By Lemma 5.1, at least half the entropy of X_i is concentrated on an X_i-stable set $\Psi_i \subseteq \Omega$ with $P(\Psi_i) = c_i$. For $i \in J$ we have $H(X_i) \geq c_i \log(1/c_i) \geq 2c_i$, so $L \geq 2 \sum_{i \in J} c_i$; therefore $\sum c_i \leq L/2 \leq 1/4$. It follows by Lemma 4.3 that

$$H(X) \geq (1/2) \sum H_{\Psi_i}(X_i) \geq (1/4) \sum H(X_i). \tag{18}$$

Now consider the case $L > 1/2$. Sort the variables in non-increasing order of their entropies, so $H(X_1) \geq H(X_2) \geq \ldots$. Since X_1 and X_2 are independent, we have $H(X) \geq H(X_1) + H(X_2)$, so if $H(X_1) + H(X_2) \geq 1/4$ then we are done. Assume $H(X_1) + H(X_2) < \min\{1/4, L/4\}$. In particular, $H(X_1) \leq L/4$ and $H(X_i) \leq 1/8$ for all $i \geq 2$. It also follows that $n \geq 4$ (because $H(X_1) + H(X_2) + H(X_3) < 1/4 + 1/8 + 1/8 = 1/2$).

Select $I \subseteq \{1, \ldots, n\}$ so as to maximize $L_I := \sum_{i \in I} H(X_i)$ subject to the constraint $L_I \leq 1/2$. We claim that $L_I \geq 3/8$. Indeed assume $L_I < 3/8$. If some $i \geq 2$ does not belong to I, then adding i to I would increase S_I by at most $1/8$, which is impossible by assumption. So $I = \{2, \ldots, n\}$. Therefore $L = L_I + L(X_1) < 3/8 + L/4$ and therefore $L < 1/2$, a contradiction.

Let us now apply the bound we already proved for $L \leq 1/2$ to the set $\{X_i \mid i \in I\}$. We obtain that $H(X) \geq L_I/4 \geq 3/16$. □

Remark 5.1. Essentially the same proof yields the following: For every $\epsilon > 0$ there exists $\delta > 0$ such that if $L := \sum_{i=1}^n H(X_i) \le \delta$ then $H(X) \ge (1-\epsilon)L$.

6 Growth as a function of the sum

In this section we prove Theorem 1.8.

Proof. Let $L = \sum_{i=1}^n H(X_i)$ and $L_k = \sum_{i=k}^n H(X_i)$ (so $L = L_1$).

We assume $n \ge 2$. First we observe that

$$H(X) \ge \max_{i \ne j}(H(X_i) + H(X_j)) \ge 2L/n. \tag{19}$$

This proves the statement for bounded n, so henceforth we assume that n is greater than an appropriate large constant C_0.

Moreover, if L is bounded then the statement follows from Prop. 1.7, so henceforth we assume that L is greater than an appropriate constant C_1. Since for $L \ge 2$ we have $\log(2+L) = \Theta \log(L)$, we shall prove $H(X) = \Omega(\sqrt{L})$ for $L \ge C_1$.

If $L \ge n \log n$ then $H(X) \ge 2L/n > \log L$ and again we are done. Henceforth we assume $L < n \log n$.

Moreover, if $\max_i H(X_i) \ge \sqrt{\log L}$ then we are done since $H(X) \ge \max_i H(X_i)$. Henceforth we assume $\max_i H(X_i) < \sqrt{\log L}$.

Let $q_i = \min(1/2, 1 - \mathrm{dom}(X_i))$. Let us sort the X_i so that $q_1 \ge q_2 \ge \cdots \ge q_n$. Let $Q = \sum_{i=1}^n q_i$ and let $Q_k = \sum_{i=k}^n q_i$ (so $Q = Q_1$).

Let $m = \max\{j \mid \log(1/q_j) < \sqrt{\log L}\}$.

If $m \ge \sqrt{L}$, we shall apply eq. (2) from Theorem 1.3 to $X' = (X_1, \ldots, X_m)$. We need to verify that the condition $q_m > 4m^{-1/3}$ is satisfied. Indeed, we have $\log(1/q_m) < \sqrt{\log L}$ and therefore $q_m > 2^{-\sqrt{\log L}}$. For $L \ge 6584$ (which holds when C_1 is chosen appropriately), the righ-hand side is greater than $4L^{-1/6} \ge 4m^{-1/3}$, as desired. So, by (2) we have

$$H(X) \ge H(X') = H(X_1, \ldots, X_m) = \Omega\left(\frac{\log m}{\log(1/q_m)}\right) \ge \Omega(\sqrt{\log L}) \tag{20}$$

and we are done.

Suppose now that $m < \sqrt{L}$. In this case, $\sum_{i=1}^m H(X_i) \le m \max_i H(X_i) < L/2$, so $L_{m+1} \ge L/2$. Recall also that for all $j \ge m+1$ we have $\log(1/q_j) \ge \sqrt{\log L}$.

We have $H(X_i) \ge q_i \log(1/q_i)$, so $L \ge L_{m+1} \ge Q_{m+1} \log(1/q_{m+1}) \ge Q_{m+1}\sqrt{\log L}$. It follows that $Q_{m+1} \le L/\sqrt{\log L}$.

Let us split the set $\{m+1, \ldots, n\}$ into $s = \lceil 3L/\sqrt{\log L}\rceil$ disjoint blocks $M_1, \ldots, M_s$ such that sum Q_{m+1} be split nearly evenly among the subsums $\sum_{i \in M_j} q_i$. So we shall have $\sum_{i \in M_j} q_i \le 1/2$ for each j. At the same time, for at least one value j we shall have $\sum_{i \in M_j} H(X_i) \ge L/(2s) \ge (1/6)\sqrt{\log L}$. Fix this j and consider the joint distribution $Y = (X_i \mid i \in M_j)$ corresponding to block M_j. By Lemma 4.3 we have $H(Y) \ge (1/2)\sum_{i \in M_j} H(X_i)$. Putting these all together, we conclude that $H(X) \ge H(Y) \ge (1/2)\sum_{i \in M_j} H(X_i) \ge (1/12)\sqrt{\log L}$. □

7 Appendix: Relaxing pairwise independence: loglog bounds

This section arose out of a combinatorial application and provided the starting point of this work.

Consider a family of n balanced Bernoulli trials. If we relax the condition of pairwise independence to pairwise small correlation, we can further reduce the joint entropy to $O(\log\log n)$. In this section we show that this rate of growth is also optimal.

Proposition 7.1. *For every $\epsilon > 0$ there exists $C > 1$ such that for all sufficiently large k there exist $n \geq C^k$ balanced Bernoulli trials X_i with pairwise correlation $\leq \epsilon$ in absolute value over a uniform probability space with $2k$ elements. As a consequence, these n variables will have joint entropy $\log(2k) = O(\log\log n)$.*

The correlation between X and Y is defined as $(E(XY) - E(X)E(Y))/\operatorname{Var}(X)\operatorname{Var}(Y)$. This quantity does not change under normalization so we may assume the X_i are normalized, i. e., they are ± 1-variables with zero expectation, so their pairwise correlation is $E(X_iX_j)$.

Proof. Our sample space will be $[2k]$. We use the probabilistic method to show the existence of the desired family of random variables $X_1, \ldots, X_n$.

Let us choose the k-subsets $A_i \subset [2k]$ uniformly and independently among the $\binom{2k}{k}$ possible choices $(i = 1, \ldots, n)$. Let X_i be the ± 1-indicator variable of A_i. Then $E(X_iX_j) = (2/k)|A_i \cap A_j| - 1$. Now

$$P(|A_i \cap A_j| = \ell) = \frac{\binom{k}{\ell}^2}{\binom{2k}{k}}. \tag{21}$$

So

$$P(|E(X_iX_j)| > \epsilon) = 2\sum_{\ell < k(1-\epsilon)/2} \frac{\binom{k}{\ell}^2}{\binom{2k}{k}} < 4\sqrt{k}\mathrm{e}^{-k\epsilon^2/8}. \tag{22}$$

Let us choose any constant c in the interval $0 < c < \epsilon^2/16$ and let $n = \mathrm{e}^{ck}$. Then, by the union bound, with high probability, none of the $\binom{n}{2}$ pairs X_i, X_j will have correlation greater than ϵ in absolute value. $\square$

We show that the $O(\log\log n)$ upper bound on the entropy is best possible by proving a matching $\Omega(\log\log n)$ lower bound under even weaker conditions.

Proposition 7.2. *Let $X_1, \ldots, X_n$ be random ± 1 variables. Assume for every $i \neq j$ we have $E(X_iX_j) \leq 1 - c$ for some $c > 0$. Then*

$$H(X_1, \ldots, X_n) \geq (c/2)\log\log n.$$

Note that we do not assume that our variables are balanced.

Proof. Let A_j denote the atoms of the Boolean algebra generated by the sets $X_i^{-1}(1)$. Let $p_j = P(A_j)$, so $\sum_j p_j = 1$ and

$$H(X_1, \ldots, X_n) = \sum_j p_j \log(1/p_j). \tag{23}$$

Let us fix a value $\epsilon > 0$ and let us split the sum (23) into $\sum_1 + \sum_2$ where $\sum_1$ includes all terms with $p_j \geq \epsilon$ and $\sum_2$ includes the rest. Let $B = \bigcup_1 A_j$ be the union of the atoms of probability $\geq \epsilon$; so $P(B) = \sum_1 p_j$.

There are at most $1/\epsilon$ atoms in B and therefore the restriction $X_i|_B$ can only be one of $2^{1/\epsilon}$ functions. Assume $n > 2^{1/\epsilon}$, i. e., $\epsilon > 1/\log n$. Then, by the pigeon hole principle, there exist $i \neq \ell$ such that $X_i|_B = X_\ell|_B$. It follows that $E(X_iX_\ell) \geq 2P(B) - 1$ and therefore $P(B) \leq 1 - c/2$.

Consequently $P(\overline{B}) \geq c/2$ and $\sum_2 p_j \log(1/p_j) \geq (c/2)\log(1/\epsilon)$. This inequality holds for every $\epsilon > 1/\log n$, so it also holds for $\epsilon = 1/\log n$. We conclude that $\sum_2 p_j \log(1/p_j) \geq (c/2)\log\log n$. $\square$

References

[1] Noga Alon, László Babai, and Alon Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *J. Algorithms*, 7:567–583, 1986.

[2] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple construction of almost k-wise independent random variables. *Random Structures and Algorithms*, 3:289–304, 1992.

[3] Yossi Azar, Rajeev Motwani, and Joseph Naor. Approximating arbitrary probability distributions using small sample spaces. *Combinatorica*, 18:151–171, 1998.

[4] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *42nd ACM STOC*, pages 67–76, 2010. www.cs.princeton.edu/∼mbraverm/pmwiki/uploads/directsum.pdf.

[5] Bonnie Berger and John Rompel. Simulating $(\log^c n)$-wise independence in nc. *J. ACM*, 38:1026–1046, 1991.

[6] Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein. From information to exact communication. *45th ACM STOC*, pages 151–160, 2013. ECCC.

[7] J. Lawrence Carter and Mark N. Wegman. Universal classes of hash functions. *J. Computer and System Sciences*, 18:143–154, 1979. DOI:10.1016/0022-0000(79)90044-8.

[8] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *42nd IEEE FOCS*, pages 270–278, 2001.

[9] Benny Chor, Oded Goldreich, Johan Håstad, Joel Friedman, Steven Rudich, and Roman Smolensky. The bit extraction problem or t-resilient functions. In *26th IEEE FOCS*, pages 396–407, 1985.

[10] A. Joffe. On a set of almost deterministic k-independent random variables. *Ann. Probability*, 2:161–162, 1974.

[11] Richard M. Karp and Avi Wigderson. A fast parallel algorithm for the maximal independent set problem. In *16th ACM STOC*, pages 266–272, 1984.

[12] Henry Oliver Lancaster. Pairwise statistical independence. *Ann. Math. Stat.*, 36:1313–1317, 1965.

[13] Michael Luby and Avi Wigderson. *Pairwise Independence and Derandomization*, volume 1, issue 4 of *Found. and Trends in Theor. Comp. Sci.* Now Publ., 2006. DOI:10.1561/0400000009.

[14] Mokshay Madiman, Adam W. Marcus, and Prasad Tetali. Information-theoretic inequalities in additive combinatorics. In *Proc. IEEE Inform. Theory Workshop*, 2010. www.stat.yale.edu/∼mm888/Pubs/2010/ITW-addcomb10.pdf.

[15] Mokshay Madiman and Prasad Tetali. Information inequalities for joint distributions, with interpretations and applications. *IEEE Trans. Information Theory*, 56(6):2699–2713, 2010.

[16] Joseph Naor and Moni Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM J. Comput.*, 22:838–856, 1993.

[17] Anup Rao. A 2-source almost-extractor for linear entropy. In *12th Int. Worksh. on Randomization and Computation (RANDOM'08)*, pages 549–556. Springer, 2008.

[18] Terry Tao. An entropy Plünnecke–Ruzsa inequality. http://terrytao.wordpress.com/2009/10/27/an-entropy-plunnecke-ruzsa-inequality/.

[19] Terry Tao. Sumset and inverse sumset theory for Shannon entropy. *Combinat. Probab. Comput.*, 19(4):603–639, 2010.

[20] Umesh Vazirani. *Randomness, Adversaries, and Computation.* PhD thesis, UC Berkeley, 1986.

[21] Emanuele Viola. *Selected Results in Additive Combinatorics: An Exposition.* Number 3 in Graduate Surveys. Theory of Computing Library, 2011. www.theoryofcomputing.org/libfiles/gradsurveys.html.