III Entropy Methods in Combinatorics

Ishan Nath, Michaelmas 2024

Based on Lectures by Prof. Timothy Gowers

March 18, 2025

Contents

1 The Khinchin (Shannon) Axioms for Entropy **2**

2 A Special Case of Sidarenko's Conjecture **10**

3 Brigman's Theorem **12**

4 Shearer's Lemma and Applications **15**

5 Isoperimetric Inequalities **19**

6 The Union-Closed Conjecture **23**

7 Entropy in Additive Combinatorics **28**
7.1 Conditional Distances 33

8 A Proof of Marton's Conjecture in $\mathbb{F}_2^n$ **35**

Index **44**

1 The Khinchin (Shannon) Axioms for Entropy

The *entropy* of a discrete random variable X is a quantity $H[X]$ that takes real vales and has the following properties:

(i) If X is uniform on $\{0, 1\}$, then $H[X] = 1$ (normalization).

(ii) If $Y = f(X)$ for some bijection f, then $H[Y] = H[X]$ (invariance).

(iii) If X takes values in a set A, B is disjoint from A, Y takes values in $A \cup B$ and for all $a \in A$,

$$\mathbb{P}(Y = a) = \mathbb{P}(X = a),$$

then $H[X] = H[Y]$ (extendability).

(iv) If X takes values in a finite set A and Y is uniformly distributed in A, then $H[X] \le H[Y]$ (maximality).

(v) H depends continuously on X with respect to the total variation distance, defined as

$$\sup_E |\mathbb{P}(X \in E) - \mathbb{P}(Y \in E)| .$$

(continuity)

For the last axiom we need a definition.

Definition 1.1. Let X and Y be random variables. The *conditional entropy* $H[X|Y]$ of X given Y is

$$\sum_y \mathbb{P}(Y = y) H[X|Y = y].$$

(vi) $H[(X, Y)] = H[X, Y] = H[Y] + H[X|Y]$ (additivity).

Lemma 1.1. *If X and Y are independent, then*

$$H[X, Y] = H[X] + H[Y].$$

Proof: We look at

$$H[X|Y] = \sum_y \mathbb{P}(Y = y) H[X|Y = y].$$

Since X and Y are independent, the distribution of X is unaffected by knowing Y, so $H[X|Y = y] = H[X]$ for all y, which gives the result.

Note we are implicitly using the invariance principle.

Corollary 1.1. *If $X_1, \dots, X_n$ are independent, then*

$$H[X_1, \dots, X_n] = H[X_1] + \dots + H[X_n].$$

Proof: Use lemma 1.1, and induction.

Lemma 1.2 (Chain rule)**.** *Let $X_1, \dots, X_n$ be random variables. Then*

$$H[X_1, \dots, X_n] = H[X_1] + H[X_2|X_1] + H[X_3|X_1, X_2] + \dots + H[X_n|X_1, \dots, X_{n-1}].$$

Proof: The case $n = 2$ is additivity. In general,

$$H[X_1, \dots, X_n] = H[X_1, \dots, X_{n-1}] + [H_n|X_1, \dots, X_{n-1}].$$

We are done by induction.

Lemma 1.3. *If $Y = f(X)$, then $H[X, Y] = H[X]$. Also, $H[Z|X, Y] = H[Z|X]$.*

Proof: The map $g : x \mapsto (x, f(x))$ is a bijection, and $(X, Y) = g(X)$. So the first statement follows by invariance. For the second,

$$H[Z|X, Y] = H[Z, X, Y] - H[X, Y] = H[Z, X] - H[X] = H[Z|X],$$

using the first part.

Lemma 1.4. *If X takes only one value, then $H[X] = 0$.*

Proof: X and X are independent, therefore by lemma 1.1 and invariance,

$$H[X] = H[X, X] = 2H[X].$$

So $H[X] = 0$.

Proposition 1.1. *If X is uniformly distributed on a set of size 2^n, then $H[X] = n$.*

Proof: Let $X_1, \dots, X_n$ be independent random variables uniformly distributed on $\{0, 1\}$. By corollary 1.2 and normalization,

$$H[X_1, \dots, X_n] = H[X_1] + \dots + H[X_n] = n.$$

But $(X_1, \dots, X_n)$ is uniformly distributed on $\{0, 1\}^n$, so by invariance the result follows.

Proposition 1.2. *Let X be uniformly distributed on a set A of size n. Then*

$$H[X] = \log n.$$

Proof: Let r be a positive integer, and let $X_1, \dots, X_r$ be independent copies of X. Then $(X_1, \dots, X_r)$ is uniform on A^r, and

$$H[X_1, \dots, X_r] = rH[X].$$

Now pick k such that $2^k \le n^r \le 2^{k+1}$. Then by invariance and maximality, and the entropy of a random variable on 2^k elements,

$$k \le rH[X] \le k+1.$$

So, we find that

$$\frac{k}{r} \le \log n \le \frac{k+1}{r} \implies \frac{k}{r} \le H[X] \le \frac{k+1}{r}.$$

Since we can approximate $\log n$ as close as possible, we find $H[X] = \log n$.

Theorem 1.1 (Khinchin)**.** *If H satisfies the Khinchin axioms, and X takes values in a finite set A, then*

$$H[X] = \sum_{a \in A} p_a \log\left(\frac{1}{p_a}\right),$$

where $p_a = \mathbb{P}(X = a)$.

Here we use the convention that if $p_a = 0$, then $p_a \log p_a = 0$.

Proof: First we do the case when all p_a are rational. Pick $n \in \mathbb{N}$ such that $p_a = m_a/n$.

Let Z be uniform on $[n]$, and let $(E_a \mid a \in A)$ be a partition of $[n]$ into sets with $|E_a| = m_a$. By invariance, we may assume that

$$X = a \iff Z \in E_a.$$

Then,

$$\begin{aligned} \log n = H[Z] = H[Z,X] &= H[X] + H[Z|X] \\ &= H[X] + \sum_{a\in A} p_a H[Z|X=a] \\ &= H[X] + \sum_{a\in A} p_a \log(m_a) \\ &= H[X] = \sum_{a\in A} p_a(\log p_a + \log n) \\ \implies H[X] = -\sum_{a\in A} p_a \log p_a. \end{aligned}$$

Corollary 1.2. *Let X and Y be random variables. Then $H[X] \geq 0$ and $H[X|Y] \geq 0$.*

This is an immediate consequence of the formula for entropy.

Corollary 1.3. *If $Y = f(X)$, then*

$$H[Y] \leq H[X].$$

Proof: Use the previous corollary

$$H[X] = H[X,Y] = H[Y] + H[X|Y],$$

but $H[X|Y] \geq 0$.

Proposition 1.3 (Subadditivity)**.** *Let X and Y be random variables. Then*

$$H[X,Y] \leq H[X] + H[Y].$$

Proof: Note that for any two random variables X and Y,

$$\begin{aligned} H[X,Y] \leq H[X] + H[Y] &\iff H[X|Y] \leq H[X] \\ &\iff H[Y|X] \leq H[Y]. \end{aligned}$$

This ought to be obvious, but it is not quite the case. Observe that $H[X|Y] \leq$

$H[X]$ if X is uniform on a finite set. This is because

$$\begin{aligned} H[X|Y] &= \sum_y \mathbb{P}(Y=y) H[X|Y=y] \\ &\leq \sum_y \mathbb{P}(Y=y) H[X] \\ &= H[X], \end{aligned}$$

where we use maximality. By the equivalence noted above, we also know that $H[X|Y] \leq H[X]$ if Y is uniform.

Let $p_{ab} = \mathbb{P}((X,Y) = (a,b))$, and assume that all p_{ab} are rational. Pick n such that we can write $p_{ab} = m_{ab}/n$, with each m_{ab} an integer. Partition $[n]$ into sets E_{ab} each of size m_{ab}. Let Z be uniform of $[n]$, and without loss of generality write $(X,Y) = (a,b) \iff Z \in E_{ab}$.

Let $E_b = \bigcup_a E_{ab}$ for each b. So $Y = b \iff Z \in E_b$. Define a random variable W as follows: if $Y = b$, then $W \in E_b$ is uniformly distributed in E_b and is independent of X.

So W and X are conditionally independent given Y, and W is uniform on $[n]$. Then,

$$H[X|Y] = H[X|Y,W] = H[X|W] \leq H[X],$$

as W is uniform. By continuity, we get the result for general probabilities.

Corollary 1.4. $H[X] \geq 0$ *for every* X.

Proof: Without using the formula,

$$0 = H[X|X] \leq H[X].$$

Corollary 1.5. *Let* $X_1, \ldots, X_n$ *be random variables. Then*

$$H[X_1, \ldots, X_n] \leq H[X_1] + \cdots + H[X_n].$$

Proposition 1.4 (Submodularity)**.** *Let* X, Y, Z *be random variables. Then,*

$$H[X|Y,Z] \leq H[X|Z].$$

Proof: Either use non-negativity of entropy and the fact (Y, Z) determines Z (cannot do this because the proof of this uses submodularity!), or

$$\begin{aligned} H[X|Y,Z] &= \sum_z \mathbb{P}(Z=z)H[X|Y,Z=z] \\ &\le \sum_z \mathbb{P}(Z=z)H[X|Z=z] = H[X|Z]. \end{aligned}$$

Submodularity can be expressed in several equivalent ways. Expanding using subadditivity,

$$H[X,Y,Z] - H[Y,Z] \le H[X,Z] - H[Z],$$

or

$$H[X,Y,Z] \le H[X,Z] + H[Y,Z] - H[Z],$$

or

$$H[X,Y,Z] + H[Z] \le H[X,Z] + H[Y,Z].$$

Lemma 1.5. *Let X, Y, Z be random variables with $Z = f(Y)$. Then*

$$H[X|Y] \le H[X|Z].$$

Proof: Use submodularity:

$$\begin{aligned} H[X|Y] = H[X,Y] - H[Y] &= H[X,Y,Z] - H[Y,Z] \\ &\le H[X,Z] - H[Z] = H[X|Z]. \end{aligned}$$

Lemma 1.6. *Let X, Y, Z be random variables with $Z = f(X) = g(Y)$. Then,*

$$H[X,Y] + H[Z] \le H[X] + H[Y].$$

Proof: Again, use submodularity:

$$H[X,Y,Z] + H[Z] \le H[X,Z] + H[Y,Z],$$

which implies the result since Z depends on X and Y.

Lemma 1.7. *Let X take values in a finite set A, and let Y be uniform on A. Then if $H[X] = H[Y]$, then X is uniform.*

Proof: Let $p_a = \mathbb{P}(X = a)$. Then

$$H[X] = \sum p_a \log(1/p_a) = |A|\mathbb{E}_{a \in A} p_a \log(1/p_a).$$

The function $x \mapsto x \log(1/x)$ is strictly concave on $[0, 1]$, so by Jensen's inequality, this is at most

$$|A|(\mathbb{E}_a p_a) \log(1/\mathbb{E}_a p_a) = \log(|A|) = H[X].$$

Equality holds if and only if $a \mapsto p_a$ is constant, i.e. X is uniform.

Corollary 1.6. *If $H[X, Y] = H[X] + H[Y]$, then X and Y are independent.*

Proof: We will go through the proof of subadditivity, and check when the equality holds.

Suppose that X is uniform on A. Then

$$\begin{aligned} H[X|Y] &= \sum_y \mathbb{P}(Y = y) H[X|Y = y] \\ &\le \le \sum_y \mathbb{P}(Y = y) H[X] = H[X], \end{aligned}$$

with equality if and only if $H[X|Y = y]$ is uniform on A for all y by the previous lemma, which implies that X and Y are independent.

At the last stage of the proof, we introduced W and said

$$H[X|Y] = H[X|Y, W] = H[X|W] \le H[X].$$

Since W is uniform, equality holds if and only if X and W are independent, which implies (since Y depends on W) that X and Y are independent.

Definition 1.2. Let X and Y be random variables. The *mutual information* $I[X : Y]$ is

$$H[X] + H[Y] - H[X, Y].$$

This can be rewritten as

$$H[X] - H[X|Y] = H[Y] - H[Y|X].$$

Subadditivity is equivalent to the statement that $I[X : Y] \ge 0$, and the previous corollary implies that $I[X : Y] = 0$ if and only if X and Y are independent.

Note that

$$H[X,Y] = H[X] + H[Y] - I[X:Y].$$

Definition 1.3. Let X, Y and Z be random variables. The *conditional mutual information* of X and Y given Z, denoted by $I[X:Y|Z]$ is

$$\begin{aligned}\sum_z \mathbb{P}(Z=z) I[X|Z=z : Y|Z=z] &= \sum_z \mathbb{P}(Z=z)(H[X|Z=z] \\ &\qquad + H[Y|Z=z] - H[X,Y|Z=z]) \\ &= H[X|Z] + H[Y|Z] - H[X,Y|Z] \\ &= H[X,Z] + H[Y,Z] - H[X,Y,Z] - H[Z].\end{aligned}$$

Submodularity is equivalent to the statement that $I[X:Y|Z] \geq 0$.

2 A Special Case of Sidarenko's Conjecture

Let G be a bipartite graph with vertex sets X and Y (finite), and density α, defined to be $|E(G)|/|X||Y|$. Let H be another (small) bipartite graph with vertex sets U and V, and m edges.

Now let $\phi : U \to X$ and $\psi : V \to Y$ be random functions. We say that (ϕ, ψ) is a *graph homomorphism* if $\phi(x)\psi(y) \in E(G)$, for every $xy \in E(H)$.

Sidarenko conjectured that for every G, H,

$$\mathbb{P}((\phi, \psi) \text{ is a homomorphism}) \geq \alpha^m.$$

This is what we expect when G is random, and is not hard to prove when H is $K_{r,s}$.

We are going to prove the theorem when $H = P_3$.

Theorem 2.1. *Sidarenko's conjecture is true if H is a path of length* 3.

Proof: We want to show that if G is a bipartite graph of density α with vertex sets X, Y of size m and n, and we choose $x_1, x_2 \in X$, $y_1, y_2 \in Y$ independent and at random, then

$$\mathbb{P}(x_1y_1, x_2y_1, x_2y_2 \in E(G)) \geq \alpha^3.$$

It would be enough to let P be a P_3 chosen uniformly at random, and show that $H[P] \geq \log(a^3m^2n^2)$. This is a trivial rephrasing, and is not useful.

Instead, we shall define a different random variable, taking values in the set of all P_3's.

To do this, let (X_1, Y_1) be a random edge of G, with $X_1 \in X, Y_1 \in Y$. Now let X_2 be a random neighbour of Y_1, and Y_2 be a random neighbour of X_2.

It will be enough to prove that $H[X_1, Y_1, X_2, Y_2] \geq \log(a^3m^2n^2)$. We can choose X_1Y_1 in three equivalent ways:

- Pick an edge uniformly at random.
- Pick a vertex x with probability proportional to its degree $d(x)$, and then a random neighbour y of x.
- The same with x and y exchanged.

This shows that $Y_1 = y$ with probability proportional to $d(y)$, so X_2Y_1 is a uniform edge. This also means that X_2Y_2 is uniform in $E(G)$. Therefore,

$$\begin{aligned}
H[X_1, Y_1, X_2, Y_2] &= H[X_1] + H[Y_1|X_1] + H[X_2|X_1, Y_1] + H[Y_2|X_1, Y_1, X_2] \\
&= H[X_1] + H[Y_1|X_1] + H[X_2|Y_1] + H[Y_2|X_2] \\
&= H[X_1] + H[X_1, Y_1] - H[X_1] \\
&\qquad + H[X_2, Y_1] - H[Y_1] + H[X_2, Y_2] - H[X_2] \\
&= 3H[U_{E(G)}] - H[Y_1] - H[X_2] \\
&\geq 3H[U_{E(G)}] - H[U_Y] - H[U_X] \\
&= 3\log(\alpha mn) - \log n - \log m = \log(\alpha^3 m^2 n^2).
\end{aligned}$$

So we are done by maximality.

An alternative finish is as follows: let X', Y' be uniform in X and Y and independent of each other, and X_1, Y_1, X_2, Y_2. Then

$$\begin{aligned}
H[X_1, Y_1, X_2, Y_2, X', Y'] &= H[X_1, Y_1, X_2, Y_2] + H[U_X] + H[U_Y] \\
&\geq 3H[U_{E(G)}].
\end{aligned}$$

So by maximality,

$$|P_3| \times |X| \times |Y| \geq |E(G)|^3.$$

3 Brigman's Theorem

Let A be an $n \times n$ matrix over, say $\mathbb{R}$. The *permanent* of A, $\mathrm{per}(A)$ is

$$\sum_{\sigma \in S_n} \prod_{i=1}^{n} A_{i\sigma(i)},$$

i.e. the determinant without the sign.

Let G be a bipartite graph with vertex sets X, Y of size n. Given $(x, y) \in X \times Y$, let

$$A_{xy} = \begin{cases} 1 & xy \in E(G), \\ 0 & xy \notin E(G), \end{cases}$$

i.e. A is the bipartite adjacency matrix of G. This is not quite the adjacency matrix as we do not care about the X to X connections.

This matrix is not well-defined as we can reorder the rows and columns, but no matter how we choose an ordering, we find that $\mathrm{per}(A)$ is the number of perfect matchings in G.

Brigman's theorem concerns how large $\mathrm{per}(A)$ can be if A is a 01-matrix and the sum of the entries in the i'th row is d_i.

Let G be a disjoint union of $K_{a_i a_i}$, for $i = 1, \dots, k$, with $a_1 + \cdots + a_k = n$. Then the number of perfect matchings in G is

$$\prod_{i=1}^{k} a_i!.$$

Theorem 3.1 (Brigman)**.** *Let G be a bipartite graph with vertex sets X, Y of size n. Then the number of perfect matchings in G is at most*

$$\prod_{x \in X} (d(x)!)^{1/d(x)}.$$

Proof: The following is a proof by Radhakrishnan.

Each matching corresponds to a bijection $\sigma : X \to Y$ such that $x\sigma(x) \in E(G)$ for every x.

Let σ be chosen uniformly from all such bijections. Then

$$H[\sigma] = H[\sigma(x_1)] + H[\sigma(x_2)|\sigma(x_1)] + \cdots + H[\sigma(x_n)|\sigma(x_1), \ldots, \sigma(x_{n-1})],$$

where $x_1, \ldots, x_n$ is some enumeration of X. Then,

$$H[\sigma(x_1)] \le \log d(x_1),$$
$$H[\sigma(x_2)|\sigma(x_1)] \le \mathbb{E}_\sigma \log d^\sigma_{x_1}(x_2),$$

where we introduce

$$d^\sigma_{x_1}(x_2) = |N(x_1) \setminus \{\sigma(x_1)\}|.$$

In general, we have

$$H[\sigma(x_i)|\sigma(x_1), \ldots, \sigma(x_{i-1})] \le \mathbb{E}_\sigma \log^\sigma_{x_1,\ldots,x_{i-1}}(x_i),$$

where

$$d^\sigma_{x_1,\ldots,x_{i-1}}(x_i) = |N(x_i) \setminus \{\sigma(x_1), \ldots, \sigma(x_{i-1})\}|.$$

The key idea is to regard $x_1, \ldots, x_n$ as a random enumeration of X, and take the average.

For each $x \in X$, define the *contribution* of x to be

$$\log(d^\sigma_{x_1,\ldots,x_{i-1}}(x_i)),$$

where $x_i = x$.

We shall now fix σ.

Let the neighbours of x be $y_1, \ldots, y_k$. Then one of the y_h will be $\sigma(x)$. We can write

$$d^\sigma_{x_1,\ldots,x_{i-1}}(x_i) = d(x) - \left|\{j \mid \sigma^{-1}(y_j) \text{ comes earlier than } x = \sigma^{-1}(y_h)\}\right|.$$

When we average, all positions of $\sigma^{-1}(y_h)$ are equally likely, so the average position of x is

$$\frac{1}{d(x)}(\log d(x) + \log(d(x)-1) + \cdots + \log(1)) = \frac{1}{d(x)}\log(d(x)!).$$

By linearity of expectation,

$$H[\sigma] \le \sum_{x \in X} \frac{1}{d(x)} \log(d(x)!),$$

so the number of matchings is at most

$$\prod_{x \in X} (d(x)!)^{1/d(x)}.$$

Definition 3.1. Let G be a graph with $2n$ vertices. A *one-factor* in G is a collection of n disjoint edges.

Theorem 3.2 (Kahn, Lovász)**.** *Let G be a graph with $2n$ vertices. Then the number of one-factors in G is at most*

$$\prod_{x \in V(G)} (d(x)!)^{1/2d(x)}.$$

If the graph happens to be bipartite, this agrees with Brigman's theorem.

Proof: Proof by Alon and Friedman.

Let $\mathcal{M}$ be the set of one-factors of G, and let (M_1, M_2) be a uniform random elements of $\mathcal{M}^2$.

For each M_1, M_2, the union $M_1 \cup M_2$ is a collection of disjoint edges and even cycles that covers all the vertices of G. Call such a union a *cover* of G by edges and even cycles.

If we are given such a cover, then the number of pairs (M_1, M_2) that could give rise to it is exactly 2^k, where k is the number of even cycles in the cover.

Now build a bipartite graph G_2 out of G. G_2 has two vertex sets V_1, V_2, both copies of $V(G)$. Join $x \in V_1$ to $y \in V_2$ if $xy \in E(G)$.

By Brigman's theorem, the number of perfect matchings in G_2 is at most

$$\prod_{x \in V(G)} (d(x)!)^{1/d(x)}.$$

Each matching gives a permutation of $V(G)$, σ such that $x\sigma(x) \in E(G)$ for every $x \in V(G)$.

Each such σ has a cycle decomposition, and each cycle gives a cycle in G. So σ gives a cover of $V(G)$ by isolated vertices, edges and cycles.

Given such a cover with k cycles, each cycle can be directed in two ways, so the number of σ that give rise to it is equal to 2^k, where k is the number of cycles.

So there is an injection from $\mathcal{M}^2$ to the set of matchings of G_2, since every cover by edges and even cycles is a cover by vertices, edges and cycles. So

$$|\mathcal{M}|^2 \leq \prod_{x \in V(G)} (d(x)!)^{1/d(x)}.$$

4 Shearer's Lemma and Applications

Given a random variable $X = (X_1, \dots, X_n)$ and a subset $A \subseteq [n]$, say $a = \{a_1, \dots, a_k\}$ with $a_1 < a_2 < \cdots < a_k$, write X_A for the random variable

$$X_A = (X_{a_1}, X_{a_2}, \dots, X_{a_k}).$$

Lemma 4.1 (Shearer)**.** *Let $X = (X_1, \dots, X_n)$ be a random variable and let $\mathcal{A}$ be a family of subsets of $[n]$ such that every $i \in [n]$ belongs to at least r of the sets $A \in \mathcal{A}$. Then,*

$$H[X_1, \dots, X_n] \le \frac{1}{r} \sum_{A \in \mathcal{A}} H[X_A].$$

Proof: For each $a \in [n]$, write

$$X_{<a} = (X_1, \dots, X_{a-1}).$$

For each $A \in \mathcal{A}$,

$$\begin{aligned} H[X_A] &= H[X_{a_1}] + H[X_{a_2} | X_{a_1}] + \cdots + H[X_{a_k} | X_{a_1}, \dots, X_{a_{k-1}}] \\ &\ge H[X_{a_1} | X_{<a_1}] + H[X_{a_2} | X_{<a_2}] + \cdots + H[X_{a_k} | X_{<a_k}] \\ &= \sum_{a \in A} H[X_a | X_{<a}]. \end{aligned}$$

Therefore,

$$\sum_{A \in \mathcal{A}} H[X_A] \ge r \sum_{a=1}^{n} H[X_a | X_{<a}] = rH[X].$$

An alternative version:

Lemma 4.2. *Let $X = (X_1, \dots, X_n)$ be a random variable, and let $A \subseteq [n]$ be a random subset of $[n]$ according to some probability distribution.*

Suppose that for each $i \in [n]$,

$$\mathbb{P}(i \in A) \ge \mu.$$

Then,

$$H[X] \le \mu^{-1} \mathbb{E}_A H[X_A].$$

Proof: As before,

$$H[X_A] \geq \sum_{a \in A} H[X_a | X_{<a}].$$

So,

$$\begin{aligned} \mathbb{E}_A H[X_A] &\geq \mathbb{E}_a \sum_{a \in A} H[X_a | X_{<a}] \\ &\geq \mu \sum_{a=1}^{n} H[X_a | X_{<a}] = \mu H[X]. \end{aligned}$$

Let $E \subseteq \mathbb{Z}^n$ and let $A \subseteq [n]$. Then we write $P_A E$ for $A = \{a_1, \dots, a_k\}$ for the set of all $u \in \mathbb{Z}^A$ such that there exists $v \in \mathbb{Z}^{[n] \setminus A}$ such that $[u, v] \in E$, where $[u, v]$ is u suitably intertwined with v.

Corollary 4.1. *Let $E \subseteq \mathbb{Z}^n$ and let $\mathcal{A}$ be a family of subsets of $[n]$ such that every $i \in [n]$ is contained in at least r sets $A \in \mathcal{A}$. Then,*

$$|E| \leq \prod_{A \in \mathcal{A}} |P_A E|^{1/r}.$$

Proof: Let X be a uniform random element of E. Then by Shearer's,

$$H[X] \leq \frac{1}{r} \sum_{A \in \mathcal{A}} H[X_A].$$

But X_A takes values in $P_A E$, so

$$H[X_A] \leq \log |P_A E| \implies \log |E| \leq \frac{1}{r} \sum_{A} \log |P_A E|.$$

If $\mathcal{A} = \{[n] \setminus \{i\} \mid i = 1, \dots, n\}$, we get

$$|E| \leq \prod_{i=1}^{n} |P_{[n] \setminus \{i\}} E|^{1/n-1}.$$

This is the discrete Loomis-Whitney theorem.

Theorem 4.1. *Let G be a graph with m edges. Then G has at most $(2m)^{3/2}/6$ triangles.*

This is basically sharp for complete graphs.

Proof: Let (X_1, X_2, X_3) be a random triple of vertices such that X_1X_2, X_1X_3 and X_2X_3 are all edges. Let t be the number of triangles in G.

By Shearer's,

$$\log(6t) = H[X_1, X_2, X_3] \leq \frac{1}{2}(H[X_1, X_2] + H[X_1, X_3] + H[X_2, X_3]).$$

Each $H[X_i, X_j]$ is supported in the set of edges of G, given a direction. So

$$\frac{1}{2}(H[X_1, X_2] + H[X_1, X_3] + H[X_2, X_3]) \leq \frac{3}{2}\log(2m).$$

Definition 4.1. Let X be a set of size n, and $\mathcal{G}$ be a set of graphs with vertex set X. $\mathcal{G}$ is *triangle-intersecting* if $G_1 \cap G_2$ contains a triangle, for all $G_1, G_2 \in \mathcal{G}$.

Theorem 4.2. *If $|V| = n$, then a triangle-intersecting family of graphs with vertex set V has size at most*

$$2^{\binom{n}{2}-2}.$$

Proof: Let $\mathcal{G}$ be triangle-intersecting family and X be chosen uniformly from $\mathcal{G}$.

We write $V^{(2)}$ for the set of (unordered) pairs of elements of V, and we think of any $G \in \mathcal{G}$ as a function from $V^{(2)}$ to $\{0, 1\}$. Define

$$X = (X_e \mid e \in V^{(2)}).$$

For each $R \subseteq V$, let G_R be the graph $K_R \cup K_{V \setminus R}$.

We shall look at the projection X_{G_R}, which we can think of as taking values in the set $\{G \cap G_R \mid G \in \mathcal{G}\} = \mathcal{G}_R$.

Note that if $G_1, G_2 \in \mathcal{G}$ and $R \subseteq [n]$, then $G_1 \cap G_2 \cap G_R \neq \emptyset$, since $G_1 \cap G_2$ contains a triangle, which must intersect G_R by pigeon-hole principle.

Thus $\mathcal{G}_R$ is an intersecting family, so it has size at most $2^{|E(\mathcal{G}_R)|-1}$.

By alternative Shearer, and noticing that if we pick R at random then each $e \in G_R$ with probability $1/2$,

$$\begin{aligned} H[X] &\leq 2\mathbb{E}_R H[X_{G_R}] \leq 2\mathbb{E}_R(|E(\mathcal{G}_R)| - 1) \\ &= 2\left(\frac{1}{2}\binom{n}{2} - 1\right) = \binom{n}{2} - 2, \end{aligned}$$

by linearity of expectation (each edge is present in half of the $\mathcal{G}_R$).

5 Isoperimetric Inequalities

Definition 5.1. Let G be a graph, and $A \subseteq V(G)$. The *edge boundary* ∂A of A is the set of edges xy such that $x \in A$, $y \notin A$.

If $G = \mathbb{Z}^n$ or $\{0,1\}^n$ and $i \in [n]$, then the i'th boundary $\partial_i A$ is the set of edges $xy \in \partial A$ such that $x - y = \pm e_i$.

Theorem 5.1 (Edge-isoperimetric inequality)**.** *Let $A \subseteq \mathbb{Z}^n$ be a finite set. Then*

$$|\partial A| \geq 2n|A|^{(n-1)/n}.$$

Proof: By the discrete Loomis-Whitney inequality,

$$\begin{aligned}|A| &\leq \prod_{i=1}^n |P_{[n]\setminus\{i\}}A|^{1/(n-1)} = \left(\prod_{i=1}^n |P_{[n]\setminus\{i\}}A|^{1/n}\right)^{n/(n-1)} \\ &\leq \left(\frac{1}{n}\sum_{i=1}^n |P_{[n]\setminus\{i\}}A|\right)^{n/(n-1)}.\end{aligned}$$

But $|\partial_i A| \geq 2|P_{[n]\setminus\{i\}A}|$ since each fibre contributes at least 2. So,

$$|A| \leq \left(\frac{1}{2n}\sum_{i=1}^n |\partial_i A|\right)^{n/(n-1)} = \left(\frac{1}{2n}|\partial A|\right)^{n/(n-1)}.$$

Theorem 5.2 (Edge-isoperimetric inequality in the cube)**.** *Let $A \subseteq \{0,1\}^n$. Then*

$$|\partial A| \geq |A|(n - \log|A|).$$

Proof: Let X be a uniformly random element of A, and write $X = (X_1, \ldots, X_n)$. Write $X_{\setminus i}$ for $(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$.

By Shearer's inequality,

$$\begin{aligned}H[X] &\leq \frac{1}{n-1}\sum_{i=1}^n H[X_{\setminus i}] = \frac{1}{n-1}\sum_{i=1}^n \big(H[X] - H[X_i|X_{\setminus i}]\big) \\ &\implies \sum_{i=1}^n H[X_i|X_{\setminus i}] \leq H[X].\end{aligned}$$

But,

$$H[X_i|X_{\setminus i} = u] = \begin{cases} 1 & |P^{-1}_{[n]\setminus\{i\}}(u)| = 2, \\ 0 & |P^{-1}_{[n]\setminus\{i\}}(u)| = 1. \end{cases}$$

The number of points of the second kind is exactly $|\partial_i A|$. So,

$$H[X_i|X_{\setminus i}] = 1 - \frac{|\partial_i A|}{|A|}.$$

So,

$$H[X] \geq \sum_{i=1}^{n} \left(1 - \frac{|\partial_i A|}{|A|}\right) = n - \frac{|\partial A|}{|A|}.$$

Also $H[X] = \log|A|$, so we are done.

Definition 5.2. Let $\mathcal{A}$ be a family of sets of size d. The *lower shadow* $\partial\mathcal{A}$ is

$$\{B \mid |B| = d-1, \exists A \in \mathcal{A}, B \subseteq A\}.$$

Theorem 5.3 (Kruskal-Katona)**.** *If* $|\mathcal{A}| = \binom{t}{d}$ *for some real number* t*, then* $|\partial\mathcal{A}| \geq \binom{t}{d-1}$.

Here we do not restrict ourselves to integer t; t may be any real number

Proof: Let $X = (X_1, \ldots, X_d)$ be a random ordering of the elements of a uniformly random $A \in \mathcal{A}$. Then

$$H[X] = \log\left(d!\binom{t}{d}\right).$$

Note that $(X_1, \ldots, X_{d-1})$ is an ordering of the elements of some $B \in \partial\mathcal{A}$, so

$$H[X_1, \ldots, X_{d-1}] \leq \log\left((d-1)!|\partial\mathcal{A}|\right).$$

It is enough to show that

$$H[X_1, \ldots, X_{d-1}] \geq \log\left((d-1)!\binom{t}{d-1}\right).$$

Note that

$$H[X_1, \ldots, X_d] = H[X_1] + H[X_2|X_1] + \cdots + [X_d|X_1, \ldots, X_{d-1}].$$

We want a lower bound on this entropy. Our strategy will be to obtain a lower bound for $H[X_k|X_{<k}]$ in terms of $H[X_{k+1}|X_{<k+1}]$. We shall prove that

$$2^{H[X_k|X_{<k}]} \geq 2^{H[X_{k+1}|X_{<k+1}]} + 1$$

for all k. Let T be chosen independently of $X_1, \ldots, X_{k-1}$, where $T = \mathrm{Ber}(1-p)$. Given $X_1, \ldots, X_{k-1}$, let

$$X^* = \begin{cases} X_{k+1} & T = 0, \\ X_k & T = 1. \end{cases}$$

Note that X_k and X_{k+1} have the same distribution given $(X_1, \ldots, X_{k-1})$, so X^* does as well. Then

$$\begin{aligned} H[X_k|X_1, \ldots, X_{k-1}] &= H[X^*|X_1, \ldots, X_{k-1}] \geq H[X^*|X_1, \ldots, X_k] \\ &= H[X^*, T|X_1, \ldots, X_k] \\ &= H[T|X_1, \ldots, X_k] + H[X^*|T, X_1, \ldots, X_k] \\ &= H[T] + pH[X_{k+1}|X_1, \ldots, X_k] \\ &\qquad + (1-p)H[X_k|X_1, \ldots, X_k] \\ &= h(p) + ps, \end{aligned}$$

where $h(x) = -(x \log x + (1-x)\log(1-x))$ is the *binary entropy function*, and $s = H[X_{k+1}|X_1, \ldots, X_k]$.

It turns out that this is maximized when $p = 2^s/(2^s+1)$, whence the bound is

$$\frac{2^s}{2^s+1}(\log(2^s+1) - \log 2^s) + \frac{\log(2^s+1)}{2^s+1} + \frac{s2^s+1}{2^s+1} = \log(2^s+1).$$

Let $r = 2^{H[X_d|X_1, \ldots, X_{d-1}]}$. Then,

$$\begin{aligned} H[X] &= H[X_1] + H[X_2|X_1] + \cdots + H[X_d|X_1, \ldots, X_{d-1}] \\ &\geq \log r + \log(r+1) + \cdots + \log(r+d-1) \\ &= \log\left(\frac{(r+d-1)!}{(r-1)!}\right) = \log\left(d!\binom{r+d-1}{d}\right). \end{aligned}$$

Since we known $H[X] = \log(d!\binom{t}{d})$, it follows that

$$r + d - 1 \leq t \implies r \leq t + 1 - d.$$

It follows that

$$\begin{aligned}
H[X_1, \ldots, X_{d-1}] &= \log\left(d!\binom{t}{d}\right) - \log r \\
&\geq \log\left(d!\frac{t!}{d!(t-d)!(t+1-d)}\right) \\
&= \log\left((d-1)!\binom{t}{d-1}\right).
\end{aligned}$$

6 The Union-Closed Conjecture

Let $\mathcal{A}$ be a finite family of sets. We say that $\mathcal{A}$ is *union closed* if $A \cup B \in \mathcal{A}$ whenever $A \in \mathcal{A}$ and $B \in \mathcal{A}$.

The following is an unproven conjecture.

Union-Closed Conjecture: If $\mathcal{A}$ is a non-empty union-closed family then there exists some x that belongs to at least $\frac{1}{2}|\mathcal{A}|$ sets in $\mathcal{A}$.

However, the following is proven.

Theorem 6.1 (Gilmer)**.** *There exists $c > 0$ such that if $\mathcal{A}$ is a union-closed family, then there exists x that belongs to at least $c|\mathcal{A}|$ of the sets in $\mathcal{A}$.*

The constant c given in the original paper was around $1/100$, but the bound could be improved to $(3 - \sqrt{5})/2$, which is the natural barrier to this approach.

In fact this constant is the best if we change our problem to look only at almost-union closed family, i.e. families in which $A \cup B \in \mathcal{A}$ for almost-all $A, B \in \mathcal{A}$. Let

$$\mathcal{A} = [n]^{(pn)} \cup [n]^{(\geq (2p - p^2 - o(1))n}.$$

With high probability, if A, B are random elements of $[n]^{(pn)}$, then $|A \cup B| \geq (2p - p^2 - o(1))n$. If $1 - (2p - p^2 - o(1)) = p$, then almost all of $\mathcal{A}$ is in $[n]^{(pn)}$, i.e.

$$1 - 3p + p^2 = 0 \implies p = \frac{3 - \sqrt{5}}{2}.$$

If we want to prove this theorem, it is natural to let A, B be independent uniformly random elements of $\mathcal{A}$, and to consider $H[A \cup B]$. Since $\mathcal{A}$ is union closed $A \cup B \in \mathcal{A}$, so $H[A \cup B] \leq \log |\mathcal{A}|$.

Now we would like to get a lower bound for $H[A \cup B]$ assuming that no x belongs to more than $p|\mathcal{A}|$ sets in $\mathcal{A}$.

Lemma 6.1. *Suppose that $c > 0$ is such that*

$$h(xy) \geq c(xh(y) + yh(x))$$

for every $x, y \in [0, 1]$. Let $\mathcal{A}$ be a family of sets such that every element belongs to fewer than $p|\mathcal{A}|$ members of $\mathcal{A}$. Then

$$H[A \cup B] > c(1 - p)(H[A] + H[B]).$$

Proof: We think of A and B as characteristic functions, i.e. indicator functions for each element of $|\mathcal{A}|$. Write $A_{<k}$ for $(A_1, \ldots, A_{k-1})$. By the chain rule it is enough to prove that for every k that

$$H[(A \cup B)_k | (A \cup B)_{<k}] > c(1-p)[H[A_k|A_{<k} + H[B_k|B_{<k}]).$$

By submodularity,

$$H[(A \cup B)_k | (A \cup B)_{<k}] \geq H[(A \cup B)_k | A_{<k}, B_{<k}].$$

For each $u, v \in \{0,1\}^{k-1}$, we write

$$p(u) = \mathbb{P}[A_k = 0 | A_{<k} = u], \qquad q(v) = \mathbb{P}[B_k = 0 | B_{<k} = v].$$

Then,

$$H[(A \cup B)_k | A_{<k} = u, B_{<k} = v] = H[A_k \cup B_k | A_{<k}, B_{<k}] = h(p(u)q(v)),$$

which by hypothesis is at least

$$c(p(u)h(q(v)) + q(v)h(p(u))).$$

So,

$$\begin{aligned} H[(A \cup B)_k | (A \cup B)_{<k}] \geq c \sum_{u,v} & \mathbb{P}(A_{<k} = u)\mathbb{P}(B_{<k} = v) \\ & \times (p(u)h(q(v)) + q(v)h(p(u))), \end{aligned}$$

but

$$\sum_k \mathbb{P}(A_{<k} = u)\mathbb{P}(A_k = 0 | A_{<k} = u) = \mathbb{P}(A_k = 0) \geq 1 - p,$$

and

$$\sum_v \mathbb{P}(B_{<k} = v)h(q(v)) = \sum_v \mathbb{P}(B_{<k} = v)H[B_k | B_{<k} = v] = H[B_k | B_{<k}],$$

so this expands as

$$\begin{aligned} & c(P(A_k = 0)H[B_k|B_{<k}] + P(B_k = 0)H[A_k|A_{<k}]) \\ & > c(1-p)(H[A_k|A_{<k}] + H[B_k|B_{<k}]), \end{aligned}$$

as required.

This shows that if $\mathcal{A}$ is union closed, then $c(1-p) \leq 1/2$, so $p \geq 1 - 1/2c$. This is non-trivial as long as $c > 1/2$, and we will obtain $c = 1/(\sqrt{5}-1)$.

To show this inequality, we start by proving the diagonal case, i.e. when $x = y$.

Lemma 6.2 (Boppana)**.** *For every* $x \in [0,1]$,

$$h(x^2) \geq \phi x h(x),$$

for $\phi = (\sqrt{5}+1)/2$.

Proof: Write ψ for $\phi^{-1} = (\sqrt{5}-1)/2$. Then $\psi^2 = 1-\psi$, so

$$h(\psi^2) = h(1-\psi) = h(\psi) \implies h(\psi^2) = \phi\psi h(\psi),$$

so equality holds when $x = \psi$, and as well when $x = 0$ or 1.

Our first fact will be

$$\begin{aligned}
\ln 2 h(x) &= -x\ln x - (1-x)\ln(1-x), \\
\ln 2 h'(x) &= -\ln x - 1 + \ln(1-x) + 1 = \ln(1-x) - \ln x, \\
\ln 2 h''(x) &= -\frac{1}{x} - \frac{1}{1-x}, \\
\ln 2 h'''(x) &= \frac{1}{x^2} - \frac{1}{(1-x)^2}.
\end{aligned}$$

We also introduce

$$\begin{aligned}
f(x) &= h(x^2) - \phi x h(x), \\
f'(x) &= 2xh'(x^2) - \phi h(x) - \phi x h'(x), \\
f''(x) &= 2h'(x^2) + 4x^2 h''(x^2) - 2\phi h'(x) - \phi x h''(x), \\
f'''(x) &= 12xh''(x^2) + 8x^3 h'''(x^2) - 3\phi h''(x) - \phi x h'''(x) \\
&= \frac{-12x}{x^2(1-x^2)} + \frac{8x^3(1-2x^2)}{x^4(1-x^2)^2} + \frac{3\phi}{x(1-x)} - \frac{\phi x(1-2x)}{x^2(1-x)^2} \\
&= \frac{-12}{x(1-x^2)} + \frac{8(1-2x^2)}{x(1-x^2)^2} + \frac{3\phi}{x(1-x)} - \frac{\phi(1-2x)}{x(1-x)^2} \\
&= \frac{-12(1-x^2) + 8(1-2x^2) + 3\phi(1-x)(1+x)^2 - \phi(1-2x)(1+x)^2}{x(1-x)^2(1+x)^2}.
\end{aligned}$$

This is zero if and only if

$$\begin{aligned}
&-12 + 12x^2 + 8 - 16x^2 + 3\phi(1 + x - x^2 - x^3) - \phi(1 - 3x^2 - 2x^3) \\
&\qquad = -\phi x^3 - 4x^2 + 3\phi x + (2\phi - 4) = 0.
\end{aligned}$$

The numerator of $f'''(x)$ is a cubic with negative leading coefficient and constant term, so it has at least one negative root. Hence it has at most two roots in $(0,1)$. It follows (using Rolle's theorem) that f has at most five roots in $[0,1]$, up to multiplicity.

But $f'(0) = -\phi h(0) = 0$, so f has a double root at 0.

Using $\psi^2 + \psi = 1$, note

$$\begin{aligned} f'(\psi) &= 2\psi(\log\psi - 2\log\psi) + \phi(\psi\log\psi + 2(1-\psi)\log\psi) - (2\log\psi - \log\psi) \\ &= -2\psi\log\psi + \log\psi + 2\phi\log\psi - 2\log\psi - \log\psi \\ &= \log\psi(-\psi + \phi - 1) = 0. \end{aligned}$$

Moreover $f(1) = 0$. So f is either non-negative on all of $[0,1]$ or non-positive. If x is small, then

$$\begin{aligned} f(x) &= -x^2\log x^2 - (1-x^2)\log(1-x^2) + \phi x(x\log x(1-x)\log(1-x)) \\ &= 2x^2\log\frac{1}{x} - \phi x^2\log\frac{1}{x} + \mathcal{O}(x^2), \end{aligned}$$

so there is x with $f(x) > 0$.

Lemma 6.3. *The function*

$$f(x,y) = \frac{h(x,y)}{xh(y) + yh(x)}$$

is minimized on $(0,1)^2$ *at a point where* $x = y$.

Proof: We can extend f continuously to the boundary by setting $f(x,y) = 1$ whenever x or y is 0 or 1. To see this, note first that this is easy if neither x nor y is 0.

If either x or y is small, then

$$\begin{aligned} h(xy) &= -xy(\log x + \log y) + \mathcal{O}(xy), \\ xh(y) + yh(x) &= -x(y\log y + \mathcal{O}(y)) - y(x\log x + \mathcal{O}(x)) \\ &= h(xy) + \mathcal{O}(xy), \end{aligned}$$

so this also tends to 1. One can also check that $f(1/2,1/2) < 1$, so f is minimized somewhere in $(0,1)^2$.

Let (x^*, y^*) be a minimum with $f(x^*, y^*) = \alpha$. For convenience, let

$$g(x) = \frac{f(x)}{x},$$

and note that

$$f(x, y) = \frac{g(xy)}{g(x) + g(y)},$$

and also that

$$g(xy) - \alpha(g(x) + g(y)) \geq 0,$$

with equality at (x^*, y^*). The partial derivatives of the left hand side are both 0 at x^*, y^*, so

$$\begin{aligned} y^* g'(x^* y^*) - \alpha g'(x^*) &= 0, \\ x^* g'(x^* y^*) - \alpha g'(y^*) &= 0. \end{aligned}$$

So multiplying, we find

$$x^* g'(x^*) = y^* g'(y^*).$$

It is enough to prove that $xg'(x)$ is an injection:

$$\begin{aligned} g'(x) &= \frac{h'(x)}{x} - \frac{h(x)}{x^2}, \\ xg'(x) &= h'(x) - \frac{h(x)}{x} \\ &= \log(1-x) - \log x + \frac{x \log x + (1-x)\log(1-x)}{x} \\ &= \frac{\log(1-x)}{x}. \end{aligned}$$

This is injective as $\log(1-x)$ is concave. Or we can differentiate again.

Combining this with lemma 6.1, we get that

$$h(xy) \geq \frac{\phi}{2}(xh(y) + yh(x)),$$

and so we can take

$$p = 1 - \frac{1}{\phi} = 1 - \frac{\sqrt{5}-1}{2} = \frac{3-\sqrt{5}}{2}.$$

7 Entropy in Additive Combinatorics

We shall need two simple results from additive combinatorics due to Imre Ruzsa.

Let G be an abelian group, and let $A, B \subseteq G$. The *sumset* $A+B$ is the set

$$A + B = \{x + y \mid x \in A, y \in B\},$$

and the *difference set* $A - B$ is the set

$$A - B = \{x - y \mid x \in A, y \in B\}.$$

We write $2A$ for $A + A$, $3A$ for $A + A + A$, and so on.

The *Ruzsa distance* $d(A, B)$ is defined to be

$$\frac{|A - B|}{|A|^{1/2}|B|^{1/2}}.$$

Lemma 7.1 (Ruzsa Triangle Inequality)**.** *$d(A, C) \le d(A, B)d(B, C)$.*

Proof: This is equivalent to the statement that

$$|A - C||B| \le |A - B||B - C|.$$

For each $x \in A - C$, pick $a(x) \in A$, $c(x) \in C$ such that $a(x) = c(x) = x$. Define a map $\phi : (A - C) \times B \to (A - B, B - C)$ by

$$\phi(x, b) = (a(x) - b, b - c(x)).$$

Adding the coordinates of $\phi(x, b)$ gives x, so we can calculate $a(x)$ and $c(x)$ from $\phi(x, b)$, and hence b. So ϕ is an injection.

Lemma 7.2 (Ruzsa Covering Lemma)**.** *Let G be an abelian group, and let A and B be finite subsets of G. Then A can be covered by at most*

$$\frac{|A + B|}{|B|}$$

translates of $B - B$.

Proof: Let $\{x_1, \dots, x_k\}$ be a maximal subset of A, such that the sets $x_i + B$

are disjoint. Then if $a \in A$, then there exists i such that

$$(a + B) \cap (x_i + B) \neq 0.$$

So $a \in x_i + B - B$. So A can be covered by k translated of $B - B$. But

$$|B|k = |\{x_1, \ldots, x_k\} + B| \leq |A + B|.$$

Let X, Y be discrete random variables taking values in an abelian group. What is $X + Y$, when X and Y are independent? For each z, writing p_x and q_y for $\mathbb{P}(X = x)$ and $\mathbb{P}(Y = y)$,

$$\begin{aligned}
\mathbb{P}(X + Y = z) &= \sum_{x+y=z} \mathbb{P}(X = x)\mathbb{P}(Y = y) \\
&= \sum_{x+y=z} p_x q_y = p * q(z),
\end{aligned}$$

the convolutions of the functions $p(x) = p_x$ and $q(y) = q_y$. So sums of independent random variables correspond to convolutions.

Definition 7.1. Let G be an abelian group and let X, Y be G-valued random variables. Then the (entropic) *Ruzsa distance* $d[X; Y]$ is

$$H[X' - Y'] - \frac{1}{2}H[X] - \frac{1}{2}H[Y],$$

where X' and Y' are independent copies of X and Y.

Lemma 7.3. *If A, B are finite subsets of G and X, Y are uniform on A, B respectively, then*

$$d[X; Y] \leq \log d(A, B).$$

Proof: Without loss of generality X and Y are independent. Then

$$\begin{aligned}
d[X, Y] &= H[X - Y] - \frac{1}{2}H[X] - \frac{1}{2}H[Y] \\
&\leq \log |A - B| - \frac{1}{2}\log |A| - \frac{1}{2}\log |B| = \log d(A, B).
\end{aligned}$$

Lemma 7.4. *Let X, Y be G-valued random variables. Then*

$$H[X + Y] \geq \max\{H[X], H[Y]\} - I[X : Y].$$

Proof: By subadditivity,

$$\begin{aligned} H[X+Y] \geq H[X+Y|Y] &= H[X+Y,Y] - H[Y] \\ &= H[X,Y] - H[Y] \\ &= H[X] + H[Y] - H[Y] - I[X:Y] \\ &= H[X] - I[X:Y]. \end{aligned}$$

By symmetry, we get the other inequality, and we an take the maximum.

Corollary 7.1. *$H[X-Y] \geq \max\{H[X], H[Y]\} - I[X:Y]$.*

Corollary 7.2. *If X, Y are G-valued random variables, then*

$$d[X,Y] \geq 0.$$

Proof: Without loss of generality, X and Y are independent. Then $I[X : Y] = 0$, so

$$H[X-Y] \geq \max\{H[X], H[Y]\} \geq \frac{1}{2}(H[X] + H[Y]).$$

Lemma 7.5. *If X and Y are G-valued random variables, then $d[X;Y] = 0$ if and only if there is some (finite) subgroup H of G such that X and Y are uniform on cosets of H.*

Proof: If X and Y are uniform on $x+H$ and $y+H$, then $X' - Y'$ is uniform on $x - y + H$, so

$$H[X' - Y'] = H[X] = H[Y],$$

giving $d[X;Y] = 0$.

Conversely, suppose that X and Y are independent and

$$H[X-Y] = \frac{1}{2}(H[X] + H[Y]).$$

Since we have equality in the proof of the lemma, it follows that

$$H[X-Y|Y] = H[X-Y].$$

Therefore, $X - Y$ and Y are independent. So for every $z \in A - B$ and for every $y_1, y_2 \in B$,

$$\mathbb{P}(X - Y = z | Y = y_1) = \mathbb{P}(X - Y = z | Y = y_2),$$

where A and B are the supports of X and Y. So

$$\mathbb{P}(X = y_1 + z) = \mathbb{P}(X = y_2 + z),$$

for all $y_1, y_2 \in B$. So p_x is constant on $z + B$, and in particular $z + B \subseteq A$. By symmetry, $A - z \subseteq B$, so $A = B + z$ for all $z \in A - B$.

So for every $x \in A$, $y \in B$, $A = B + x - y$, so $A - x = B - y$. So $A - x$ is the same for every $x \in A$. Therefore $A - x = A - A$ for all $x \in A$. It follows that $A - A + A - A = (A - x) - (A - x) = A - A$, so it a closed subset containing inverses under addition, hence a subgroup.

Moreover $A = A - A + x$, hence a coset of $A - A$. Since $B = A + x$, B is also a coset.

Recall that if Z is a function of X and is a function of Y, then

$$H[X, Y] + H[Z] \leq H[X] + H[Y].$$

Lemma 7.6 (Entropic Ruzsa Triangle Inequality). *Let X, Y, Z be G-valued random variables. Then,*

$$d[X; Z] \leq d[X; Y] + d[Y; Z].$$

Proof: We must show that

$$H[X - Z] - \frac{1}{2}H[X] - \frac{1}{2}H[Z] \leq H[X - Y] - \frac{1}{2}H[X] - \frac{1}{2}H[Y] + H[Y - Z] - \frac{1}{2}H[Y] - \frac{1}{2}H[Z],$$

or that

$$H[X - Z] + H[Y] \leq H[X - Y] + H[Y - Z].$$

Since $X - Z$ depends on $(X - Y, Y - Z)$ and on (X, Z),

$$H[X - Y, Y - Z, X, Z] + H[X - Z] \leq H[X - Y, Y - Z] + H[X, Z],$$

i.e.

$$H[X, Y, Z] + H[X - Z] \leq H[X, Z] + H[X - Y, Y - Z].$$

So by independence and subadditivity, we get the lemma.

Lemma 7.7 (Submodularity for Sums). *If X, Y, Z are independent G-valued* *random variables, then*

$$H[X+Y+Z]+H[Z] \leq H[X+Z]+H[Y+Z].$$

Proof: $X+Y+Z$ is a function of $(X+Z,Y)$ and of $(X,+Z)$ so

$$H[X+Z,Y,X,Y+Z]+H[X+Y+Z] \leq H[X+Z,Y]+H[Y,X+Z],$$

or by rewriting,

$$H[X,Y,Z]+H[X+Y+Z] \leq H[X+Z]+H[Y]+H[X]+H[Y+Z].$$

By independence and cancellations, we ge the desired inequality.

Lemma 7.8. *Let G be an abelian group, and let X be a G-valued random variable. Then*

$$d[X;-X] \leq 2d[X;X].$$

Proof: Let X_1, X_2, X_3 be independent copies of X. Then

$$\begin{aligned} d[X;-X] &= H[X_1+X_2]-\frac{1}{2}H[X_1]-\frac{1}{2}H[X_2] \leq H[X_1+X_2-X_3]-H[X] \\ &\leq H[X_1-X_3]+H[X_2-X_3]-H[X_3]-H[X] \\ &= 2d[X;X], \end{aligned}$$

as X_1, X_2, X_3 are all copies of X.

Corollary 7.3. *Let X and Y be G-valued random variables. Then*

$$d[X;-Y] \leq 5d[X;Y].$$

Proof: We have, by using the Ruzsa triangle inequality,

$$\begin{aligned} d[X;-Y] &\leq d[X;Y]+d[Y;-Y] \\ &\leq d[X;Y]+2d[Y;Y] \leq d[X;Y]+2(d[Y;X]+d[X;Y]) \\ &= 5d[X;Y]. \end{aligned}$$

7.1 Conditional Distances

Definition 7.2. Let X, Y, U, V be G-valued random variables. Then the *conditional distance* is

$$d[X|U; Y|V] = \sum_{u,v} \mathbb{P}(U = u)\mathbb{P}(V = v)d[X|U = u; Y|V = v].$$

The next definition is not completely standard.

Let X, Y, U be G-valued random variables. Then the *simultaneous conditional distance* of X to Y given U is

$$d[X; Y||U] = \sum_{u} \mathbb{P}(U = u)d[X|U = u; Y|U = u].$$

We say that X', Y' are *conditionally independent trials* of X and Y given U if X' is distributed like X, Y' is distributed like Y, and for each $u \in U$, $X'|U = u$ is distributed like $X|U = u$, $Y'|U = u$ is distributed like $Y|U = u$ and $X'|U = u$ and $Y'|U = u$ are independent. Then,

$$d[X; Y||U] = H[X' - Y'|U] - \frac{1}{2}H[X'|U] - \frac{1}{2}H[Y'|U],$$

as can be seen directly from the formula.

Lemma 7.9 (Entropic BSG Theorem). *Let A and B be G-valued random variables. Then*

$$d[A; B||A + B] \leq 3I[A : B] + 2H[A + B] - H[A] - H[B].$$

Proof: We have

$$d[A; B||A + B] = H[A' - B'|A + B] - \frac{1}{2}H[A'|A + B] - \frac{1}{2}H[B'|A + B],$$

where A' and B' are conditionally independent given $A + B$. Now

$$\begin{aligned} H[A'|A + B] &= H[A|A + B] = H[A, A + B] - H[A + B] \\ &= H[A, B] - H[A + B] \\ &= H[A] + H[B] - I[A : B] - H[A + B]. \end{aligned}$$

Similarly, $H[B'|A + B]$ is the same, so

$$\frac{1}{2}H[A'|A + B] + \frac{1}{2}H[B'|A + B]$$

is also the same. Also

$$H[A' - B'|A + B] \leq H[A' - B'].$$

Let (A_1, B_1) and (A_2, B_2) be conditionally independent trials of (A, B), given $A + B$. Then,

$$H[A' - B'] = H[A_1 - B_2].$$

By submodularity,

$$\begin{aligned}
H[A_1 - B_2] &= H[A_1 - B_2, A_1] + H[A_1 - B_2, B_1] - H[A_1 - B_2, A_1, B_1]. \\
H[A_1 - B_2, A_1] &= H[A_1, B_2] \leq H[A_1] + H[B_2] = H[A] + H[B], \\
H[A_1 - B_2, B_1] &= H[A_2 - B_1, B_1] = H[A_2, B_1] \leq H[A] + H[B]. \\
H[A_1 - B_2, A_1, B_1] &= H[A_1, B_1, A_2, B_2] \\
&= H[A_1, B_2, A_2, B_2|A + B] + H[A + B] \\
&= 2H[A, B|A + B] + H[A + B] \\
&= 2H[A, B] - H[A + B] \\
&= 2H[A] + 2H[B] - 2I[A : B] - H[A + B].
\end{aligned}$$

Adding or subtracting all these terms gives the required inequality.

8 A Proof of Marton's Conjecture in $\mathbb{F}_2^n$

We shall prove the following theorem.

Theorem 8.1 (Green, Manners, Tao, Gi)**.** *There is a polynomial p with the following property: If $n \in \mathbb{N}$ and $A \subseteq \mathbb{F}_2^n$ is such that $|A+A| \leq C|A|$, then there is a subspace $H \subseteq \mathbb{F}_2^n$ of size at most $|A|$ such that A is contained in at most $p(C)$ translates of H.*

Equivalently, there exists $K \subseteq \mathbb{F}$, $|K| \leq p(C)$ such that $A \subseteq K + H$.

In fact, we shall prove the following statement.

Theorem 8.2 (EPFR)**.** *Let $G = \mathbb{F}_2^n$ and let X, Y be G-valued random variables. Then there exists a subgroup H of G such that*

$$d[X; U_H] + d[U_H; Y] \leq \alpha d[X, Y],$$

where U_H is the uniform distribution on H and α is an absolute constant.

We will show EPFR implies the Marton's conjecture proof.

Lemma 8.1. *Let X be a discrete random variable, and write p_x for $\mathbb{P}(X = x)$. Then there exists x such that $p_x \geq 2^{-H[X]}$.*

Proof: If not, then

$$H[X] = \sum_x p_x \log\left(\frac{1}{p_x}\right) > H[X] \sum_x p_x = H[X].$$

Proposition 8.1. *EPFR implies theorem 8.1.*

Proof: Let $A \subseteq \mathbb{F}_2^n$, and $|A+A| \leq C|A|$. Let X and Y be independent copies of U_A. Then by EPFR, there exists a subgroup H such that

$$d[X; U_H] + d[U_H; Y] \leq \alpha d[X, Y],$$

so

$$d[X; U_H] \leq \frac{\alpha}{2} d[X; Y].$$

But,

$$\begin{aligned} d[X; Y] &= H[U_A + U_A] - H[U_A] \leq \log(C|A|) - \log|A| \\ &= \log C. \end{aligned}$$

So,

$$d[X, U_H] \leq \frac{\alpha \log C}{2},$$

hence

$$\begin{aligned} H[X + U_H] &\leq \frac{1}{2}H[X] + \frac{1}{2}H[U_H] + \frac{\alpha \log C}{2} \\ &= \frac{1}{2}\log|A| + \frac{1}{2}\log|H| + \frac{\alpha \log C}{2}. \end{aligned}$$

Therefore, by the previous lemma there exists z such that

$$\mathbb{P}(X + U_H = z) \geq |A|^{-1/2}|H|^{-1/2}C^{-\alpha/2}.$$

But,

$$\mathbb{P}(X + U_H = z) = \frac{|A \cap (z + H)|}{|A||H|}.$$

So there exists $z \in G$ such that

$$|A \cap (z + H)| \geq C^{-\alpha/2}|A|^{1/2}|H|^{1/2}.$$

Let $B = A \cap (z + H)$. By the Ruzsa covering lemma, we can cover A by at most most $\frac{|A+B|}{|B|}$ translates of $B + B$. Since $B \subseteq z + H$, $B + B \subseteq H + H = H$, so A can be covered by at most $\frac{|A+B|}{|H|}$ translates of H.

But $|A + B| \leq |A + A| \leq C|A|$. So

$$\frac{|A + B|}{|B|} \leq \frac{C|A|}{C^{-\alpha/2}|A|^{1/2}|H|^{1/2}} = C^{\alpha/2+1}\frac{|A|^{1/2}}{|H|^{1/2}}.$$

Since B is contained in $z + H$,

$$|H| \geq C^{-\alpha/2}|A|^{1/2}|H|^{1/2} \implies |H| \geq C^{-\alpha}|A|,$$

so we find

$$C^{\alpha/2+1}\frac{|A|^{1/2}}{|H|^{1/2}} \leq C^{\alpha+1}.$$

If $|H| \leq |A|$, then we are done. Otherwise, since $B \subseteq A$,

$$|A| \geq C^{-\alpha/2}|A|^{1/2}|H|^{1/2} \implies |H| \leq C^{\alpha}|A|.$$

Pick a subgroup H' of H of size between $\frac{|A|}{2}$ and $|A|$. Then H is a union of at most $2C^{\alpha}$ translates of H', and A is a union of at most $2C^{2\alpha+1}$ translates of H'.

Now we reduce further. We shall prove the following statement.

Theorem 8.3 (EPFR'). *There is a constant $\eta > 0$ such that if X and Y are any two $\mathbb{F}_2^n$-valued random variables with $d[X;Y] > 0$, then there exist $\mathbb{F}_2^n$-valued random variables U and V such that*

$$d[U;V] + \eta(d[U;X] + d[V;Y]) < d[X,Y].$$

Proposition 8.2. *EPFR' implies EPFR.*

Proof: By compactness, we can find U and V such that

$$\tau_{X,Y}[U;V] = d[U;V] + \eta(d[U;X] + d[V;Y])$$

is minimized. If $d[U;V] \neq 0$, then we can apply EPFR', to show there exists Z and W such that

$$\tau_{U,V}[Z;W] < d[U;V].$$

But then,

$$\begin{aligned}\tau_{X,Y}[Z;W] &= d[Z;W] + \eta(d[Z;X] + d[W;Y]) \\ &\leq d[Z;W] + \eta(d[Z;U] + d[W;V]) + \eta(d[U;X] + d[V;Y]) \\ &< d[U;V] + \eta(d[U,X] + d[V;Y]) = \tau_{X,Y}[U;V].\end{aligned}$$

It follows that $d[U;V] = 0$. So there exists H such that U and V are uniform on cosets of H, so

$$\eta(d[U_H, X] + d[U_H, Y]) < d[X,Y],$$

which gives EPFR with constant $\alpha = \eta^{-1}$.

Definition 8.1. We write $\tau_{X,Y}[U|Z;V|W]$ for

$$\sum_{z,w} \mathbb{P}(Z = z)\mathbb{P}(W = w)\tau_{X,Y}[U|Z = z; V|W = w],$$

and $\tau_{X,Y}[U;V||Z]$ for

$$\sum_{z} \mathbb{P}(Z = z)\tau_{X,Y}[U|Z = z; V|Z = z].$$

Remark. If we can prove EPFR' for conditioned random variable, then by averaging we get it for some pair of random variables, e.g. of the form $U|Z = z$, $V|W = w$.

Lemma 8.2 (Fibring Lemma)**.** *Let G and H be abelian groups, and let $\phi : G \to H$ be a homomorphism. Let X and Y be G-valued random variables. Then*

$$d[X;Y] = d[\phi(X);\phi(Y)] + d[X|\phi(X);Y|\phi(Y)] + I[X-Y : \phi(X),\phi(Y)|\phi(X)-\phi(Y)].$$

Proof: We will follow our noses:

$$\begin{aligned} d[X;Y] - H[X-Y] - \frac{1}{2}H[X] - \frac{1}{2}H[Y] \\ &= H[\phi(X)-\phi(Y)] + H[X-Y|\phi(X)-\phi(Y)] - \frac{1}{2}H[\phi(x)] \\ &\quad - \frac{1}{2}H[X|\phi(X)] - \frac{1}{2}H[\phi(Y)] - \frac{1}{2}H[\phi(Y)|Y] \\ &= d[\phi(X);\phi(Y)] + d[X|\phi(X);Y|\phi(Y)] + H[X-Y|\phi(X)-\phi(Y)] \\ &\quad - H[X-Y|\phi(X),\phi(Y)]. \end{aligned}$$

But this last line of the expression equals

$$\begin{aligned} H[X-Y|\phi(X)-\phi(Y)] - H[X-Y|\phi(X),\phi(Y),\phi(X)-\phi(Y)] \\ = I[X-Y : \phi(X),\phi(Y)|\phi(X)-\phi(Y)]. \end{aligned}$$

We shall be interested in the following special case.

Corollary 8.1. *Let $G = \mathbb{F}_2^n$, and let X_1, X_2, X_3 and X_4 be independent G-valued random variables. Then,*

$$\begin{aligned} d[(X_1,X_2);(X_3,X_4)] &= d[X_1;X_3] + d[X_2;X_4] \\ &= d[X_1+X_2, X_3+X_4] + d[X_1|X_1+X_2;X_3|X_3+X_4] \\ &\quad + I[X_1+X_3, X_2+X4 : X_1+X_2, X_3+X_4|X_1+X_2+X_3+X_4]. \end{aligned}$$

This is true by applying the fibring lemma with $X = (X_1,X_2)$, $Y = (X_3,X_4)$ and $\phi(x,y) = x+y$.

We shall now set $W = X_1+X_2+X_3+X_4$.

Recall that entropic BSG says that

$$d[X;Y||X+Y] \le 3I[X:Y] + 2H[X+Y] - H[X] - H[Y].$$

Equivalently,

$$I[X:Y] \ge \frac{1}{3}\left(d[X,Y||X+Y] + H[X] + H[Y] - 2H[X+Y]\right).$$

Applying this to the information term in this previous corollary, we get that it is at least

$$\frac{1}{3}\bigg(d[X_1 + X_3, X_2 + X_4; X_1 + X_2, X_3 + X_4||X_2 + X_3, W] \\ + H[X_1 + X_3, X_2 + X_4|W] + H[X_1 + X_2, X_3 + X_4|W] \\ - 2H[X_2 + X_3, X_2 + X_3|W]\bigg).$$

This simplifies to

$$\frac{1}{3}\bigg(d[X_1 + X_3, X_2 + X_4; X_1 + X_2, X_3 + X_4||X_2 + X_3, W] \\ + H[X_1 + X_3|W] + H[X_1 + X_2|W] - 2H[X_2 + X_3|W]\bigg).$$

We also have the inequality

$$d[X_1; X_3] + d[X_2; X_4] \geq d[X_1 + X_2; X_3 + X_4] + d[X_1|X_1 + X_2; X_3|X_3 + X_4] \\ + \frac{1}{3}\bigg(d[X_1 + X_2; X_1 + X_3||X_2 + X_3, W] + H[X_1 + X_2|W] \\ + H[X_1 + X_3|W] - 2H[X_2 + X_3|W]\bigg).$$

Apply this to (X_1, X_2, X_3, X_4), (X_1, X_2, X_4, X_3) and (X_1, X_4, X_3, X_2) and add. We look at the first entropy terms. We get

$$2H[X_1 + X_2|W] + H[X_1 + X_4|W] + H[X_1 + X_3|W] + H[X_1 + X_4|W] \\ + H[X_1 + X_3|W] - 2H[X_2 + X_3|W] - 2H[X_2 + X_4|W] \\ - 2H[X_1 + X_2|W] = 0,$$

where we made heavy use of the observation that if i, j, k, l are some permutation of $1, 2, 3, 4$, then

$$H[X_i + X_j|W] = H[X_k + X_l|W].$$

This allows us to replace, for example

$$d[X + 1 + X_2, X_3 + X_4; X_1 + X_3, X_2 + X_4||X_2 + X_3|W]$$

by

$$d[X_1 + X_2; X_1 + X_3||X_2 + X_3, W].$$

Therefore, we get the following inequality as well.

Lemma 8.3.

$$\begin{aligned} d[X_1; X_3] + 2d[X_2; X_4] + d[X_1; X_4] + d[X_2; X_3] &\geq 2d[X_1 + X_2; X_3 + X_4] \\ &+ d[X_1 + X_4; X_2 + X_3] + 2d[X_1|X_1 + X_2; X_3|X_3 + X_4] \\ &+ d[X_1|X_1 + X_4; X_2|X_2 + X_3] + \frac{1}{3}\Big(d[X_1 + X_2; X_1 + X_3||X_2 + X_3, W] \\ &+ d[X_1 + X_2; X_3 + X_4||X_2 + X_4, W] + d[X_1 + X_4; X_1 + X_3||X_3 + X_4, W] \Big). \end{aligned}$$

Now let X_1, X_2 be copies of X, and Y_1, Y_2 copies of Y and apply the previous lemma to (X_1, X_2, Y_1, Y_2) to get the following.

Lemma 8.4. *Let X_1, X_2, Y_1, Y_2 be as above. Then,*

$$\begin{aligned} 6d[X, Y] \geq 2d[X_1 + X_2; Y_1 + Y_2] &+ d[X_1 + Y_2; X_2 + Y_1] + 2d[X_1|X_1 + X_2; Y_1|Y_1 + Y_2] \\ &+ d[X_1|X_1 + Y_1; X_2|X_2 + Y_2] + \frac{2}{3}d[X_1 + X_2; X_1 + Y_1||X_2 + Y_1, X_1 + Y_2] \\ &+ \frac{1}{3}d[X_1 + Y_1; X_1 + Y_2||X_1 + X_2, Y_1 + Y_2]. \end{aligned}$$

Recall that we want (U, V) such that

$$T_{X,Y}(U, V) = d[U; V] + \eta(d[U; X] + d[V; Y]) < d[X; Y].$$

This lemma gives us a collections of distances (some conditional), at least one of which is at most $\frac{6}{7}d[X; Y]$. So it will be enough to show that for all of them, we get

$$d[U; X] + d[V; Y] \leq Cd[X; Y]$$

for some absolute constant C. Then we can take $\eta \leq \frac{1}{7C}$.

Definition 8.2. We say that (U, V) is *C-relevant* to (X, Y) if

$$d[U; X] + d[V; Y] \leq Cd[X; Y].$$

Lemma 8.5. (Y, X) *is 2-relevant to* (X, Y).

Proof: Trivial.

$$d[Y; X] + d[X; Y] = 2d[X; Y].$$

Lemma 8.6. *Let U, V, X be independent $\mathbb{F}_2^n$-valued random values. Then,*

$$d[U + V, X] \leq \frac{1}{2}\left(d[U; X] + d[V; X] + d[U; V]\right).$$

Proof: Apply submodularity at $(*)$:

$$\begin{aligned}
d[U+V;X] &= H[U+V;X] - \frac{1}{2}H[U+V] - \frac{1}{2}H[X] \\
&= H[U+V+X] - H[U+V] + \frac{1}{2}H[U+V] - \frac{1}{2}H[X] \\
&\overset{(*)}{\leq} \frac{1}{2}H[U+X] - \frac{1}{2}H[U] + \frac{1}{2}H[V+X] - \frac{1}{2}H[V] \\
&\quad + \frac{1}{2}H[U+V] - \frac{1}{2}H[X] \\
&= \frac{1}{2}(d[U;X] + d[V:X] + d[U;V]).
\end{aligned}$$

Corollary 8.2. *If (U,V) is C-relevant to (X,Y) and U_1, U_2, V_1, V_2 are copies of U, V, then (U_1+U_2, V_1+V_2) is $2C$-relevant to (X,Y).*

Proof: We have

$$\begin{aligned}
d[U_1+U_2;X] + d[V_1+V_2;Y] &\overset{\text{LIO}}{\leq} \frac{1}{2}(2d[U;X] + d[U;U] + 2d[V;Y] + d[V;V]) \\
&\overset{\triangle}{\leq} 2(d[U;X] + d[V;Y]) \leq 2Cd[X:Y].
\end{aligned}$$

Corollary 8.3. *(X_1+X_2, Y_1+Y_2) is 4-relevant to (Y,X).*

Proof: (X,Y) is 2-relevant to (Y,X), and we can use the previous corollary.

Corollary 8.4. *If (U,V) is C-relevant to (X,Y), then $(U+V, U+V)$ is $(2C+1)$-relevant to (X,Y).*

Proof: By the lemma on $d[U+V;X]$,

$$\begin{aligned}
d[U+V;X] &\leq \frac{1}{2}\Big(d[U;X] + d[V;X] + d[U;V]\Big) \\
&\leq \frac{1}{2}\Big(d[U;X] + d[V;Y] + d[X;Y] + d[U;X] + d[X;Y] + d[V;Y]\Big) \\
&= d[U;X] + d[V;Y] + d[X;Y].
\end{aligned}$$

The same holds for $d[U+V;Y]$.

Lemma 8.7. *Let U, V, X be independent $\mathbb{F}_2^n$-valued random variables. Then*

$$d[U|U+V;X] \leq \frac{1}{2}\left(d[U;X] + d[V;X] + d[U;V]\right).$$

Proof:

$$\begin{aligned} d[U|U+V;X] &= H[U+X|U+V] - \frac{1}{2}H[U|U+V] - \frac{1}{2}H[X] \\ &\leq H[U+X] - \frac{1}{2}H[U] - \frac{1}{2}H[V] + \frac{1}{2}H[U+V] - \frac{1}{2}H[X]. \end{aligned}$$

This comes from $H[A|B] \leq H[A]$ and from the definition of conditional entropy of $H[U|U+V]$, using U, V are independent.

But, $d[U|U+V;X] = d[V|U+V;X]$, so it is also at most

$$H[V+X] - \frac{1}{2}H[U] - \frac{1}{2}H[V] + \frac{1}{2}H[U+V] - \frac{1}{2}H[X].$$

Arranging the two inequalities gives the result.

Corollary 8.5. *Let U, V be independent random variables and suppose that (U, V) is C-relevant to (X, Y). Then,*

(i) $(U_1|U_1+U_2, V_1|V_1+V_2)$ *is $2C$-relevant to* (X, Y).

(ii) $(U|U_1+V_1, U_2|U_2+V_2)$ *is $2(C+1)$-relevant to* (X, Y).

Proof: Use the previous lemma. Then as soon as it is used, we are in exactly the situation when we were bounding the relevance of (U_1+U_2, V_1+V_2) and (U_1+V_1, U_2+V_2).

It remains to tackle the last two terms in the big lemma. For the penultimate term, we need to bound

$$d[X_1+X_2|X_2+Y_1, X_1+Y_2; X] + d[X_1+Y_1|X_2+Y_1, X_1+Y_2; Y].$$

But the first term of this is at most (by lemma 8.6):

$$\begin{aligned}
&\frac{1}{2}\Big(d[X_1|X_2+Y_1, X_1+Y_2; X] \\
&\quad + d[X_2|X_2+Y_1, X_1+Y_2; X] + d[X_1; X_2||X_2+Y_1, X_1+Y_2]\Big) \\
&\quad \leq d[X_1|X_1+Y_2; X] + d[X_2|X_2+Y_1; X] \\
&\quad = 2d[X|X+Y; X].
\end{aligned}$$

Then we can use lemma 8.7 and similarly for the other term.

Index

C-relevant, 40

binary entropy function, 21

conditional distance, 33
conditional entropy, 2
conditional mutual information, 9
conditionally independent trials, 33
contribution, 13

difference set, 28

edge boundary, 19
entropy, 2

graph homomorphism, 10

lower shadow, 20

mutual information, 8

one-factor, 14

permanent, 12

Ruzsa distance, 28, 29

simultaneous conditional distance, 33
sumset, 28

triangle-intersecting, 17

union closed, 23