Entropy Methods in Combinatorics

Daniel Naylor

April 4, 2025

Contents

Lecture 1

1 The Khinchin (Shannon?) axioms for entropy

Note. In this course, "random variable" will mean "discrete random variable" (unless otherwise specified).
All logarithms will be base 2 (unless otherwise specified).

Definition (Entropy)**.** The *entropy* of a discrete random variable X is a quantity $\mathbf{H}[X]$ that takes real values and has the following properties:

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

(ii) *Invariance*: If X takes values in A, Y takes values in B, f is a bijection from A to B, and for every $a \in A$ we have $\mathbb{P}[X = a] = \mathbb{P}[Y = f(a)]$, then $\mathbf{H}[Y] = \mathbf{H}[X]$.

(iii) *Extendability*: If X takes values in a set A, and B is disjoint from A, Y takes values in $A \cup B$, and for all $a \in A$ we have $\mathbb{P}[Y = a] = \mathbb{P}[X = a]$, then $\mathbf{H}[Y] = \mathbf{H}[X]$.

(iv) *Maximality*: If X takes values in a finite set A and Y is uniformly distributed in A, then $\mathbf{H}[X] \leq \mathbf{H}[Y]$.

(v) *Continuity*: $\mathbf{H}$ depends continuously on X with respect to total variation distance (defined by the distance between X and Y is $\sup_E |\mathbb{P}[X \in E] - \mathbb{P}[Y \in E]|$).

For the last axiom we need a definition:
Let X and Y be random variables. The *conditional entropy* $\mathbf{H}[X \mid Y]$ of X given Y is

$$\sum_y \mathbb{P}[Y = y]\mathbf{H}[X \mid Y = y].$$

(vi) *Additivity*: $\mathbf{H}[X, Y] = \mathbf{H}[Y] + \mathbf{H}[X \mid Y]$.

Lemma 1.1. Assuming that:

- X and Y are independent random variables

Then

$$\mathbf{H}[X, Y] = \mathbf{H}[X] + \mathbf{H}[Y].$$

Proof. $\mathbf{H}[X \mid Y] = \sum_y \mathbb{P}[Y = y]\mathbf{H}[X \mid Y = y]$.

Since X and Y are independent, the distribution of X is unaffected by knowing Y (so by invariance, $\mathbf{H}[X \mid Y = y] = \mathbf{H}[X]$), so

$$\mathbf{H}[X \mid Y = y] = \mathbf{H}[X]$$

for all y, which gives the result. □

Corollary 1.2. If $X_1, \ldots, X_n$ are independent, then

$$\mathbf{H}[X_1, \ldots, X_n] = \mathbf{H}[X_1] + \cdots + \mathbf{H}[X_n].$$

Proof. Lemma 1.1 and obvious induction. □

Lemma 1.3 (Chain rule)**.** Assuming that:

- $X_1, \ldots, X_n$ are random variables

Then

$$\mathbf{H}[X_1, \ldots, X_n] = \mathbf{H}[X_1] + \mathbf{H}[X_2 \mid X_1] + \mathbf{H}[X_3 \mid X_1, X_2] + \cdots + \mathbf{H}[X_n \mid X_1, \ldots, X_{n-1}].$$

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

$$\mathbf{H}[X_1, \ldots, X_n] = \mathbf{H}[X_1, \ldots, X_{n-1}] + \mathbf{H}[X_n \mid X_1, \ldots, X_{n-1}]$$

so we are done by induction. □

Lemma 1.4. Assuming that:

- $Y = f(X)$

Then

$$\mathbf{H}[X, Y] = \mathbf{H}[X].$$

Also,

$$\mathbf{H}[Z \mid X, Y] = \mathbf{H}[Z \mid X].$$

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 statement:

$$\begin{aligned} \mathbf{H}[Z \mid X, Y] &= \mathbf{H}[Z, X, Y] - \mathbf{H}[X, Y] && \text{(by additivity)} \\ &= \mathbf{H}[Z, X] - \mathbf{H}[X] && \text{(by first part)} \\ &= \mathbf{H}[Z \mid X] && \text{(by additivity)} \end{aligned}$$

□

Lemma 1.5. Assuming that:

- X takes only one value

Then $\mathbf{H}[X] = 0$.

Proof. X and X are independent. Therefore, by Lemma 1.1, $\mathbf{H}[X, X] = 2\mathbf{H}[X]$. But by invariance, $\mathbf{H}[X, X] = \mathbf{H}[X]$. So $\mathbf{H}[X] = 0$. □

Proposition 1.6. Assuming that:

- X is uniformly distributed on a set of size 2^n

Then $\mathbf{H}[X] = n$.

Proof. Let $X_1, \dots, X_n$ be independent random variables uniformly distributed on $\{0, 1\}$. By Corollary 1.2 and normalisation, $\mathbf{H}[X_1, \dots, X_n] = n$. But $(X_1, \dots, X_n)$ is uniformly distributed on $\{0, 1\}^n$, so by invariance, the result follows. □

Lecture 2

Proposition 1.7. Assuming that:

- X is uniformly distributed on a set A of size n

Then $\mathbf{H}[X] = \log n$.

Reminder: log here is to the base 2 (which is the convention for this course).

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

$$\mathbf{H}[X_1, \dots, X_r] = r\mathbf{H}[X].$$

Now pick k such that $2^k \leq n^r \leq 2^{k+1}$. Then by invariance, maximality, and Proposition 1.6, we have that

$$k \leq r\mathbf{H}[X] \leq k + 1.$$

So

$$\frac{k}{r} \leq \log n \leq \frac{k+1}{r} \implies \frac{k}{r} \leq \mathbf{H}[X] \leq \frac{k+1}{r} \qquad \forall k, r$$

Therefore, $\mathbf{H}[X] = \log n$ as claimed. □

Notation. We will write $p_a = \mathbb{P}[X = a]$.
We will also use the notation $[n] = \{1, 2, \dots, n\}$.

Theorem 1.8 (Khinchin)**.** Assuming that:

- H satisfies the Khinchin axioms
- X takes values in a finite set A

Then

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

Proof. First we do the case where all p_a are rational (and then can finish easily by the continuity axiom).

Pick $n \in \mathbb{N}$ such that for all a, there is some $m_a \in \mathbb{N} \cup \{0\}$ such that $p_a = \frac{m_a}{n}$.

Let Z be uniform on $[n]$. Let $(E_a : a \in A)$ 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

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

Hence

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

By continuity, since this holds if all p_a are rational, we conclude that the formula holds in general. □

Corollary 1.9. Assuming that:

- X and Y random variables

Then $\mathbf{H}[X] \geq 0$ and $\mathbf{H}[X \mid Y] \geq 0$.

Proof. Immediate consequence of Theorem 1.8. □

Corollary 1.10. Assuming that:

- $Y = f(X)$

Then $\mathbf{H}[Y] \leq \mathbf{H}[X]$.

Proof. $\mathbf{H}[X] = \mathbf{H}[X, Y] = \mathbf{H}[Y] + \mathbf{H}[X \mid Y]$. But $\mathbf{H}[X \mid Y] \geq 0$. □

Proposition 1.11 (Subadditivity)**.** Assuming that:

- X and Y be random variables

Then $\mathbf{H}[X, Y] \leq \mathbf{H}[X] + \mathbf{H}[Y]$.

Proof. Note that for any two random variables X, Y we have

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

Next, observe that $\mathbf{H}[X \mid Y] \leq \mathbf{H}[X]$ if X is uniform on a finite set. That is because

$$\begin{aligned} \mathbf{H}[X \mid Y] &= \sum_y \mathbb{P}[Y = y]\mathbf{H}[X \mid Y = y] \\ &\leq \sum_y \mathbb{P}[Y = y]\mathbf{H}[X] && \text{(by maximality)} \\ &= \mathbf{H}[X] \end{aligned}$$

By the equivalence noted above, we also have that $\mathbf{H}[X \mid Y] \leq \mathbf{H}[X]$ if Y is uniform.

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

Let $E_b = \cup_a E_{ab}$ for each b. So $Y = b \iff Z \in E_b$. Now define a random variable W as follows: If $Y = b$, then $W \in E_b$, but then W is uniformly distributed in E_b and independent of X (or Z if you prefer).

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

Then

$$\begin{aligned} \mathbf{H}[X \mid Y] &= \mathbf{H}[X \mid Y, W] && \text{(by conditional independence)} \\ &= \mathbf{H}[X \mid W] && \text{(as } W \text{ determines } Y\text{)} \\ &\leq \mathbf{H}[X] && \text{(as } W \text{ is uniform)} \end{aligned}$$

By continuity, we get the result for general probabilities. □

Corollary 1.12. Assuming that:

- X a random variable

Then $\mathbf{H}[X] \geq 0$.

Proof (Without using formula). By Subadditivity, $\mathbf{H}[X \mid X] \leq \mathbf{H}[X]$. But $\mathbf{H}[X \mid X] = 0$. □

Corollary 1.13. Assuming that:

- $X_1, \ldots, X_n$ are random variables

Then

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

Proof. Induction using Subadditivity. □

Proposition 1.14 (Submodularity)**.** Assuming that:

- X, Y, Z are random variables

Then

$$\mathbf{H}[X \mid Y, Z] \leq \mathbf{H}[X \mid Z].$$

Proof. Calculate:

$$\begin{aligned}\mathbf{H}[X \mid Y, Z] &= \sum_z \mathbb{P}[Z = z]\mathbf{H}[X \mid Y, Z = z] \\ &\leq \sum_z \mathbb{P}[Z = z]\mathbf{H}[X \mid Z = z] \\ &= \mathbf{H}[X \mid Z]\end{aligned}$$

□

Submodularity can be expressed in many ways.

Expanding using additivity gives the following inequalities:

$$\begin{aligned}\mathbf{H}[X, Y, Z] - \mathbf{H}[Y, Z] &\leq \mathbf{H}[X, Z] - \mathbf{H}[Z] \\ \mathbf{H}[X, Y, Z] &\leq \mathbf{H}[X, Z] + \mathbf{H}[Y, Z] - \mathbf{H}[Z] \\ \mathbf{H}[X, Y, Z] + \mathbf{H}[Z] &\leq \mathbf{H}[X, Z] + \mathbf{H}[Y, Z]\end{aligned}$$

Lecture 3

Lemma 1.15. Assuming that:

- X, Y, Z random variables

- $Z = f(Y)$

Then

$$\mathbf{H}[X \mid Y] \leq \mathbf{H}[X \mid Z].$$

Proof.

$$\begin{aligned}
\mathbf{H}[X \mid Y] &= \mathbf{H}[X, Y] - \mathbf{H}[Y] \\
&= \mathbf{H}[X, Y, Z] - \mathbf{H}[Y, Z] \\
&\leq \mathbf{H}[X, Z] - \mathbf{H}[Z] && \text{(Submodularity)} \\
&= \mathbf{H}[X \mid Z]
\end{aligned}$$

□

Lemma 1.16. Assuming that:

- X, Y, Z random variables
- $Z = f(X) = g(Y)$

Then

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

Proof. Submodularity says:

$$\mathbf{H}[X, Y, Z] = \mathbf{H}[Z] \leq \mathbf{H}[X, Z] + \mathbf{H}[Y, Z]$$

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

Lemma 1.17. Assuming that:

- X takes values in a finite set A
- Y is uniform on A
- $\mathbf{H}[X] = \mathbf{H}[Y]$

Then X is uniform.

Proof. Let $p_a = \mathbb{P}[X = a]$. Then

$$\begin{aligned}
\mathbf{H}[X] &= \sum_{a \in A} p_a \log\left(\frac{1}{p_a}\right) \\
&= |A| \sum_{a \in A} p_a \log\left(\frac{1}{p_a}\right)
\end{aligned}$$

The function $x \mapsto x \log \frac{1}{x}$ is concave on $[0,1]$. So, by Jensen's inequality this is at most

$$|A|(\mathbb{E}_a p_a) \log\left(\frac{1}{\mathbb{E}_a p_a}\right) = \log(|A|) = \mathbf{H}[Y].$$

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

Corollary 1.18. Assuming that:

- X, Y random variables
- $\mathbf{H}[X,Y] = \mathbf{H}[X] + \mathbf{H}[Y]$

Then X and Y are independent.

Proof. We go through the proof of Subadditivity and check when equality holds.

Suppose that X is uniform on A. Then

$$\begin{aligned}\mathbf{H}[X \mid Y] &= \sum_y \mathbb{P}[Y=y]\mathbf{H}[X \mid Y=y] \\ &\leq \mathbf{H}[X]\end{aligned}$$

with equality if and only if $\mathbf{H}[X \mid Y=y]$ is uniform on A for all y (by Lemma 1.17), which implies that X and Y are independent.

At the last stage of the proof we used

$$\mathbf{H}[X \mid Y] = \mathbf{H}[X \mid Y, W] = \mathbf{H}[X \mid W] \leq \mathbf{H}[X]$$

where W was uniform. So equality holds only if X and W are independent, which implies (since Y depends on W) that X and Y are indpendent. □

Definition (Mutual information)**.** Let X and Y be random variables. The *mutual information* $\mathbf{I}[X:Y]$ is

$$\begin{aligned}\mathbf{H}[X] + \mathbf{H}[Y] - \mathbf{H}[X,Y] &= \mathbf{H}[X] - \mathbf{H}[X \mid Y] \\ &= \mathbf{H}[Y] - \mathbf{H}[Y \mid X]\end{aligned}$$

Subadditivity is equivalent to the statement that $\mathbf{I}[X:Y] \geq 0$ and Corollary 1.18 implies that $\mathbf{I}[X:Y] = 0$ if and only if X and Y are independent.

Note that

$$\mathbf{H}[X,Y] = \mathbf{H}[X] + \mathbf{H}[Y] - \mathbf{I}[X:Y].$$

Definition (Conditional mutual information)**.** Let X, Y and Z be random variables. The *conditional mutual information* of X and Y given Z, denoted by $\mathbf{I}[X : Y|Z]$ is

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

Submodularity is equivalent to the statement that $\mathbf{I}[X : Y \mid Z] \geq 0$.

2 A special case of Sidorenko's conjecture

Let G be a bipartite graph with vertex sets X and Y (finite) and density α (defined to be $\frac{|E(G)|}{|X||Y|}$). Let H be another (think of it as '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. Say that (ϕ, ψ) is a *homomorphism* if $\phi(x)\phi(y) \in E(G)$ for every $xy \in E(H)$.

Sidorenko conjectured that: for every G, H, we have

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

Not hard to prove when H is $K_{r,s}$. Also not hard to prove when H is $K_{2,2}$ (use Cauchy Schwarz).

Theorem 2.1. Sidorenko'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 $x_1, x_2 \in X$, $y_1, y_2 \in Y$ independently at random, then

$$\mathbb{P}[x_1y_1, x_2y_2, x_3y_3 \in E(G)] \geq \alpha^3.$$

It would be enough to let P be a P3 chosen uniformly at random and show that $\mathbf{H}[P] \geq \log(\alpha^3 m^2 n^2)$.

Instead we shall define a *different* random variable taking values in the set of all P3s (and then apply maximality).

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 let Y_2 be a random neighbour of X_2.

It will be enough to prove that

$$\mathbf{H}[X_1, Y_1, X_2, Y_2] \geq \log(\alpha^3 m^2 n^2).$$

Lecture 4

We can choose X_1Y_1 in three equivalent ways:

(1) Pick an edge uniformly from all edges.

(2) Pick a vertex x with probability proportional to its degree $d(x)$, and then pick a random neighbour y of x.

(3) Same with x and y exchanged.

It follows that $Y_1 = y$ with probability $\frac{d(y)}{|E(G)|}$, so X_2Y_1 is uniform in $E(G)$, so $X_2 = x'$ with probability $\frac{d(x')}{|E(G)|}$, so X_2Y_2 is uniform in $E(G)$.

Therefore,

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

So we are done my maximality.

Alternative finish (to avoid using log!):

Let X', Y' be uniform in X, Y and independent of each other and X_1, Y_1, X_2, Y_2. Then:

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

So by maximality,

$$\#P_3 s \times |X| \times |Y| \geq |E(G)|^3. \qquad \square$$

3 Brégman's Theorem

Definition (Permanent of a matrix)**.** Let A be an $n \times n$ matrix over $\mathbb{R}$. The *permanent* of A, denoted $\text{per}(A)$, is

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

i.e. "the determinant without the signs".

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

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

ie A is the bipartite adjacency matrix of G.

Then $\text{per}(A)$ is the number of perfect matchings in G.

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

Let G be a disjoint union of $K_{a_i a_i}$s for $i = 1, \ldots, k$, 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 (Bregman)**.** Assuming that:

- G a bipartite graph with vertex sets X, Y of size n

Then the number of perfect matchings in G is at most

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

Proof (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.

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

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

Then

$$\mathbf{H}[\sigma(x_1)] \leq \log d(x_1)$$
$$\mathbf{H}[\sigma(x_2) \mid \sigma(x_1)] \leq \mathbb{E}_\sigma \log d^\sigma_{x_1}(x_2)$$

where

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

In general,

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

where

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

Lecture 5

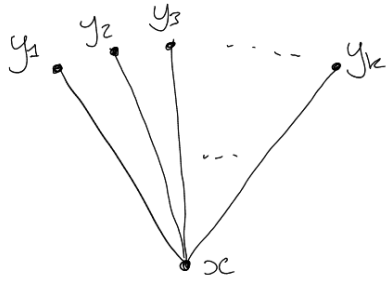
Key idea: we now regard $x_1, \ldots, x_n$ as a *random* enumeration of X and take the average.

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

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

where $x_i = x$ (note that this "contribution" is a random variable rather than a constant).

We shall now fix σ. Let the neighbours of x be $y_1, \ldots, y_k$.



Then one of the y_j will be $\sigma(x)$, say y_h. Note that $d^\sigma_{x_1,\ldots,x_{i-1}}(x_i)$ (given that $x_i = x$) is

$$d(x) - |\{j : \sigma^{-1}(y_j) \text{ comes earlier than } x = \sigma^{-1}(y_h)\}|.$$

All positions of $\sigma^{-1}(y_h)$ are equally likely, so the average contribution 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,

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

so the number of matchings is at most

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

Definition (1-factor)**.** Let G be a graph with $2n$ vertices. A 1-*factor* in G is a collection of n disjoint edges.

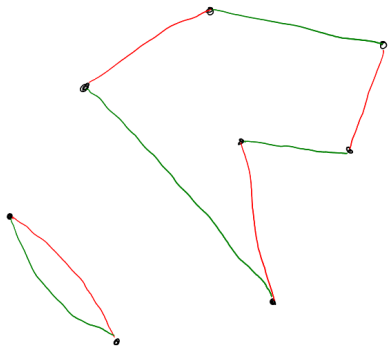
Theorem 3.2 (Kahn-Lovasz)**.** Assuming that:

- G a graph with $2n$ vertices

Then the number of 1-factors in G is at most

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

Proof (Alon, Friedman). Let $\mathcal{M}$ be the set of 1-factors of G, and let (M_1, M_2) be a uniform random element 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.

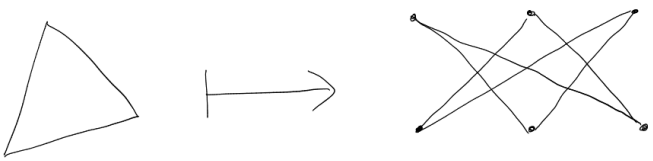


Call such a union a *cover of G by edges and even cycles.*

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

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

For example:



By Bregman, the number of perfect matchings in G_2 is $\leq \prod_{x \in V(G)} (d(x)!)^{\frac{1}{d(x)}}$. Each matching gives a permutation σ of $V(G)$, 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 edge can be directed in two ways, so the number of σ that give rise to is is 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

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

4 Shearer's lemma and applications

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

Lemma 4.1 (Shearer)**.** Assuming that:

- $X = (X_1, \ldots, X_n)$ a random variable
- $\mathcal{A}$ 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

$$\mathbf{H}[X_1, \ldots, X_n] \leq \frac{1}{r} \sum_{A \in \mathcal{A}} \mathbf{H}[X_A].$$

Proof. For each $a \in [n]$, write $X_{<a}$ for $(X_1, \ldots, X_{a-1})$.

For each $A \in \mathcal{A}$, $A = \{a_1, \ldots, a_k\}$ with $a_1 < \cdots < a_k$, we have

$$\begin{aligned}
\mathbf{H}[X_A] &= \mathbf{H}[X_{a_1}] + \mathbf{H}[X_{a_2} \mid X_{a_1}] + \cdots + \mathbf{H}[X_{a_k} \mid X_{a_1}, \ldots, X_{a_k}] \\
&\geq \mathbf{H}[X_{a_1} \mid X_{<a_1}] + \mathbf{H}[X_{a_2} \mid X_{<a_2}] + \cdots + \mathbf{H}[X_{a_k} \mid X_{<a_k}] \qquad \text{(Lemma 1.15)} \\
&= \sum_{a \in A} \mathbf{H}[X_a \mid X_{<a}]
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sum_{A \in \mathcal{A}} \mathbf{H}[X_A] &\geq \sum_{A \in \mathcal{A}} \sum_{a \in A} \mathbf{H}[X_a \mid X_{<a}] \\
&\geq r \sum_{a=1}^{n} \mathbf{H}[X_a \mid X_{<a}] \\
&= r\mathbf{H}[X]
\end{aligned}$$

□

Lecture 6

Alternative version:

Lemma 4.2 (Shearer, expectation version)**.** Assuming that:

- $X = (X_1, \ldots, X_n)$ a random variable
- $A \subset [n]$ a randomly chosen subset of $[n]$, according to some probability distribution (don't need any independence conditions!)
- for each $i \in [n]$, $\mathbb{P}[i \in A] \geq \mu$

Then

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

Proof. As before,

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

So

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

□

Definition (P_A)**.** Let $E \subset \mathbb{Z}^n$ and let $A \subset [n]$. Then we write P_AE 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 (i.e. $u \cup v$ as functions).

Corollary 4.3. Assuming that:

- $E \subset \mathbb{Z}^n$
- $\mathcal{A}$ a family of subsets of $[n]$ such that every $i \in [n]$ is contained at least r sets $A \in \mathcal{A}$

Then

$$|E| \leq \prod_{A\in\mathcal{A}}|P_AE|^{\frac{1}{r}}.$$

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

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

But X_A tkaes values in P_AE, so

$$\mathbf{H}[X_A] \leq \log|P_AX,$$

so

$$\log|E| \leq \frac{1}{r}\sum_{A}\log|P_AE|.$$

□

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

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

This case is the discrete Loomis-Whitney theorem.

Theorem 4.4. Assuming that:

- G a graph with m edges

Then G has at most $\frac{(2m)^{\frac{3}{2}}}{6}$ triangles.

Is this bound natural? Yes: if $m = \binom{n}{2}$, and we consider a complete graph on n vertices, then we get approximately $\frac{(2m)^{\frac{2}{3}}}{6}$ triangles.

Proof. Let (X_1, X_2, X_3) be a random ordered triangle (without loss of generality G has a triangle so that this is possible).

Let t be the number of triangles in G. By Shearer,

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

Each edge $\mathbf{H}[X_i, X_j]$ is supported in the set of edges G, given a direction, i.e.

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

Definition. Let X be a set of size n and let $\mathcal{G}$ be a set of graphs with vertex set X. Then $\mathcal{G}$ is Δ-*intersecting* (read as "triangle-intersecting") if for all $G_1, G_2 \in \mathcal{G}$, $G_1 \cap G_2$ contains a triangle.

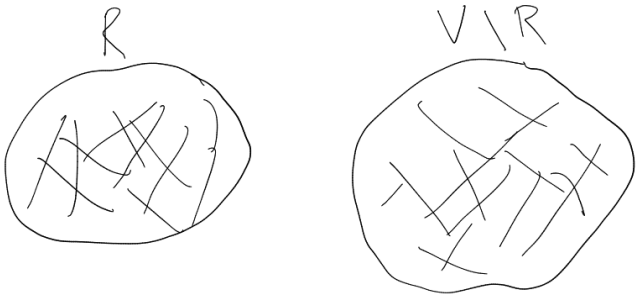
Theorem 4.5. Assuming that:

- $|V| = n$
- $\mathcal{G}$ a Δ-intersecting family with vertex set V

Then $\mathcal{G}$ has size at most $2^{\binom{n}{2}-2}$.

Proof. Let X be chosen uniformly at random from $\mathcal{G}$. We write $V^{(2)}$ for the set of (unordered) pairs of elements of V. Think of any $G \in \mathcal{G}$ as a function from $V^{(2)}$ to $\{0, 1\}$. So $X = (X_e : e \in V^{(2)})$.

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



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

Note that if $G_1, G_2 \in \mathcal{G}$, $R \subset [n]$, 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 Pigeonhole Principle.

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

$$\begin{aligned}
\mathbf{H}[X] &\leq 2\mathbb{E}_R \mathbf{H}[X_{G_R}] && \text{(since each } e \text{ belongs to } G_R \text{ with probability } 1/2) \\
&\leq 2\mathbb{E}_R(|E(G_R)| - 1) \\
&= 2\left(\frac{1}{2}\binom{m}{2} - 1\right) \\
&= \binom{n}{2} - 2
\end{aligned}$$

□

Lecture 7

Definition (Edge-boundary)**.** Let G be a graph and let $A \subset V(G)$. The *edge-boundary* ∂A of A is the set of edges xy such that $y \notin A$.
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 $x - y = \pm e_i$, i.e. $\partial_i A$ consists of deges pointing in direction i.

Theorem 4.6 (Edge-isoperimetric inequality in Z^n)**.** Assuming that:

- $A \subset \mathbb{Z}^n$ a finite set

Then $|\partial A| \geq 2n|A|^{\frac{n-1}{n}}$.

Proof. By the discrete Loomis-Whitney inequality,

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

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

So

$$\begin{aligned}|A| &\leq \left(\frac{1}{2n}\sum_{i=1}^{n} |\partial_i A|\right)^{\frac{n}{n-1}} \\ &= \left(\frac{1}{2n}|\partial A|\right)^{\frac{n}{n-1}}\end{aligned}$$

□

Theorem 4.7 (Edge-isoperimetric inequality in the cube)**.** Assuming that:

- $A \subset \{0,1\}^n$ (where we take the usual graph)

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

Proof. Let X be a uniform 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,

$$\begin{aligned}\mathbf{H}[X] &\leq \frac{1}{n-1}\sum_{i=1}^{r} \mathbf{H}[X_{\setminus i}] \\ &= \frac{1}{n-1}\sum_{i=1}^{n} \mathbf{H}[X] - \mathbf{H}[X_i \mid X_{\setminus i}]\end{aligned}$$

Hence

$$\sum_{i=1}^{n} \mathbf{H}[X_i \mid X_{\setminus i}] \leq \mathbf{H}[X].$$

Note

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

The number of points of the second kind is $|\partial_i A|$, so $\mathbf{H}[X_i \mid X_{\setminus i}] = 1 - \frac{|\partial_i A|}{|A|}$. So

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

Also, $\mathbf{H}[X] = \log |A|$. So we are done. □

Definition (Lower shadow)**.** Let $\mathcal{A}$ be a family of sets of size d. The *lower shadow* $\partial\mathcal{A}$ is $\{B : |B| = d-1, \exists A \in \mathcal{A}, B \subset A\}$.

Notation. Let $h(x) = x \log \frac{1}{x} + (1-x) \log \frac{1}{1-x}$ (for $x \in [0,1]$).

Theorem 4.8 (Kruskal-Katona)**.** Assuming that:

- $|\mathcal{A}| = \binom{t}{d} = \frac{t(t-1)\cdots(t-d-1)}{d!}$ for some real number t

Then $|\partial\mathcal{A}| \geq \binom{t}{d-1}$.

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

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

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

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

So it's enough to show

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

Also,

$$\mathbf{H}[X] = \mathbf{H}[X_1, \ldots, X_{d-1}] + \mathbf{H}[X_d \mid X_1, \ldots, X_{d-1}]$$

and

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

We would like an *upper bound* for $\mathbf{H}[X_d] X_{<d}$. Our strategy will be to obtain a *lower bound* for $\mathbf{H}[X_k \mid X_{<k}]$ in terms of $\mathbf{H}[X_{k+1} \mid X_{<k+1}]$. We shall prove that

$$2^{\mathbf{H}[X_k \mid X_{<k}]} \geq 2^{\mathbf{H}[X_{k+1} \mid X_{<k+1}]} + 1 \qquad \forall k.$$

Let T be chosen independently of $X_1, \dots, X_{k-1}$ with

$$T = \begin{cases} 0 & \text{probability } p \\ 1 & \text{probability } 1-p \end{cases}$$

(p will be chosen and optimised later).

Given $X_1, \dots, X_{k-1}$, 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 $X_1, \dots, X_{k-1}$), so X^* does as well. Then

$$\begin{aligned}
\mathbf{H}[X_k \mid X_{<k}] &= \mathbf{H}[X^* \mid X_{<k}] \\
&\geq \mathbf{H}[X^* \mid X_{\leq k}] && \text{(Submodularity)} \\
&= \mathbf{H}[X^*, T \mid X_{\leq k}] && (X_{\leq k} \text{ and } X^* \text{ determine } T) \\
&= \mathbf{H}[T \mid X_{\leq k}] + \mathbf{H}[X^* \mid T, X_{\leq k}] && \text{(additivity)} \\
&= \mathbf{H}[T] + p\mathbf{H}[X_{k+1} \mid X_1, \dots, X_k] \\
&\quad + (1-p)\mathbf{H}[X_k \mid X_1, \dots, X_k] \\
&= h(p) + ps
\end{aligned}$$

Lecture 8

where $h(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$ and $s = \mathbf{H}[X_{k+1} \mid X_1, \dots, X_k]$.

It turns out that this is maximised when $p = \frac{2^s}{2^s+1}$. Then we get

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

This proves the claim.

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

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

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

$$r + d - 1 \leq t, \qquad r \leq t + 1 - d.$$

It follows that

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

□

5 The union-closed conjecture

Definition (Union-closed)**.** Let $\mathcal{A}$ be a (finite) family of sets. Say that $\mathcal{A}$ is *union closed* if for any $A, B \in \mathcal{A}$, we have $A \cup B \in \mathcal{A}$.

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

Theorem (Justin 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 $\mathcal{A}$.

Justin Gilmer's constant was about $\frac{1}{100}$.

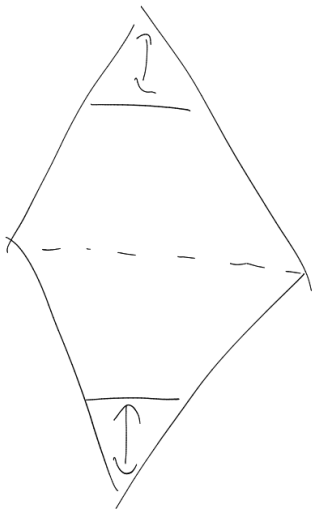
His method has a "natural barrier" of $\frac{3-\sqrt{5}}{2}$.

We will briefly and "informally" discuss this.

A reason for this is that if we weaken the property union-closed to "almost union-closed" (if we pick two elements randomly, then with high probability the union is in the family), then $\frac{3-\sqrt{5}}{2}$ is the right bound.

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 $[n]^{(pn)}$.



One of the roots of the quadratic $1 - 3p + p^2 = 0$ is $p = \frac{3-\sqrt{5}}{2}$.

If we want to prove Justin Gilmer's Theorem, it is natural to let A, B be independent uniformly random elements of $\mathcal{A}$ and to consider $\mathbf{H}[A \cup B]$. Since $\mathcal{A}$ is union-closed, $A \cup B \in \mathcal{A}$, so $\mathbf{H}[A \cup B] \leq \log |\mathcal{A}|$. Now we would like to get a lower bound for $\mathbf{H}[A \cup B]$ assuming that no x belongs to more than $p|\mathcal{A}|$ sets in $\mathcal{A}$.

$$h(xy) \geq c(xh(y) + yh(x)), \qquad h(x^2) \geq 2cxh(x).$$

Lecture 9

Lemma 5.1. Assuming that:

- $c > 0$ is such that
$$h(xy) \geq c(xh(y) + yh(x))$$
for every $x, y \in [0, 1]$
- $\mathcal{A}$ is a family of sets such that every element (of $\bigcup \mathcal{A}$) belongs to fewer than $p|\mathcal{A}|$ members of $\mathcal{A}$

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

Proof. Think of A, B as characteristic functions. Write $A_{<k}$ for $(A_1, \ldots, A_{k-1})$ etc. By the Chain rule it is enough to prove for every k that

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

By Submodularity,

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

For each $u, v \in \{0,1\}^{k-1}$ write $p(u) = \mathbb{P}(A_k = 0 \mid A_{<k} = u)$, $q(v) = \mathbb{P}(B_k = 0 \mid B_{<k} = v)$.

Then

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

which by hypothesis is at least

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

So

$$\mathbf{H}[(A \cup B)_k \mid (A \cup B)_{<k}] \geq c \sum_{u,v} \mathbb{P}(A_{<k} = u)\mathbb{P}(B_{<k} = v)\left(p(u)h(q(v)) + q(v)h(p(u))\right).$$

But

$$\sum_u \mathbb{P}(A_{<k} = u)\mathbb{P}(A_k = 0 \mid A_{<k} = u) = \mathbb{P}(A_k = 0)$$

and

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

Similarly for the other term, so the RHS equals

$$c(\mathbb{P}(A_k = 0)\mathbf{H}[B_k \mid B_{<k}] + \mathbb{P}(B_k = 0)\mathbf{H}[A_k \mid A_{<k}]),$$

which by hypothesis is greater than

$$c(1-p)(\mathbf{H}[A_k \mid A_{<k}] + \mathbf{H}[B_k \mid B_{<k}])$$

as required. □

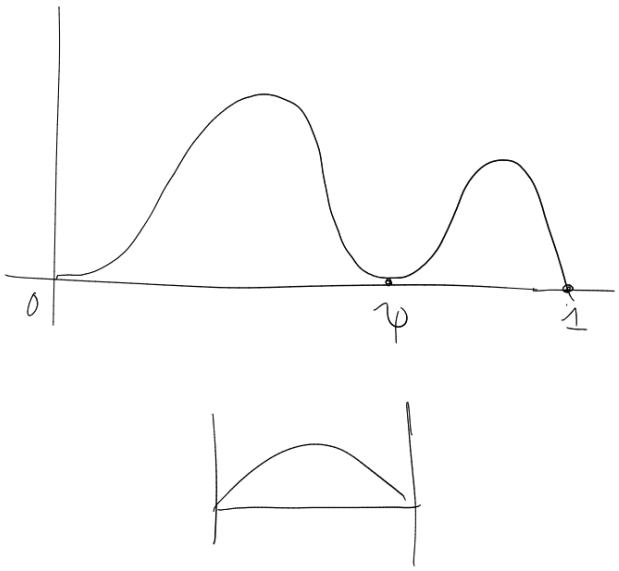
This shows that if $\mathcal{A}$ is union-closed, then $c(1-p) \leq \frac{1}{2}$, so $p \geq 1 - \frac{1}{2c}$. Non-trivial as long as $c > \frac{1}{2}$.

We shall obtain $\frac{1}{\sqrt{5}-1}$. We start by proving the diagonal case – i.e. when $x = y$.

Lemma 5.2 (Boppana)**.** For every $x \in [0,1]$,

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

Proof. Write ψ for $\phi^{-1} = \frac{\sqrt{5}-1}{2}$. Then $\psi^2 = 1 - \psi$, so $h(\psi^2) = h(1-\psi) = h(\psi)$ and $\phi\psi = 1$, so $h(\psi^2) = \phi\psi h(\psi)$. Equality also when $x = 0, 1$.



Toolkit:

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

Let $f(x) = h(x^2) - \phi x h(x)$. Then

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

So

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

This is zero if and only if

$$-12 + 12x^2 + 8 - 16x^2 + 3\phi(1 + x - x^2 - x^3) - \phi(1 - 3x^2 - 2x^3) = 0$$

which simplifies to

$$-\phi x^3 - 4x^2 + 3\phi x - 4 + 2\phi = 0.$$

Lecture 10

Since this is a cubic with negative leading coefficient and constant term, it has a negative root, so 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

$$f'(x) = 2x(\log(1 - x^2) - \log x^2) + \phi(x \log x + (1 - x) \log(1 - x)) - \phi x(\log(1 - x) - \log x).$$

So $f'(0) = 0$, so f has a double root at 0.

We can also calculate (using $\psi^2 + \psi = 1$):

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

So there's a double root at ψ.

Also, note $f(1) = 0$.

So f is either non-negative on all of $[0, 1]$ or non-positive on all of $[0, 1]$.

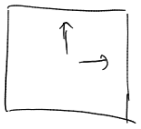
If x is small,

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

so there exists x such that $f(x) > 0$. □

Lemma 5.3. The function $f(x,y) = \frac{h(xy)}{xh(y)+yh(x)}$ is minimised 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 it's valid if neither x nor y is 0.



If either x or y is small, then

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

So it tends to 1 again.

One can check that $f\left(\frac{1}{2}, \frac{1}{2}\right) < 1$, so f is minimised somewhere in $(0,1)^2$.

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

Let $g(x) = \frac{h(x)}{x}$ and note that

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

Also,

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

with equality at (x^*, y^*). So the partial derivatives of LHS are both 0 at (x^*, y^*).

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

So $x^* g'(x^*) = y^* g'(y^*)$. So it's enough to prove that $xg'(x)$ is an injection. $g'(x) = \frac{h'(x)}{x} - \frac{h(x)}{x^2}$, so

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

Differentiating gives

$$\frac{-1}{x(1-x)} - \frac{\log(x-1)}{x^2} = \frac{-x-(1-x)\log(1-x)}{x^2(1-x)}.$$

The numerator differentiates to $-1+1+\log(1-x)$, which is negative everywhere. Also, it equals 0 at 0. So it has a constant sign. □

Combining this with Lemma 5.2, we get that

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

This allows us to take $1-\frac{1}{\phi} = 1-\frac{\sqrt{5}-1}{2} = \frac{3-\sqrt{5}}{2}$.

6 Entropy in additive combinatorics

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

Definition (Sum set / difference set / etc)**.** Let G be an abelian group and let $A, B \subset G$.
The *sumset* $A + B$ is the set $\{x + y : x \in A, y \in B\}$.
The *difference set* $A - B$ is the set $\{x - y : x \in A, y \in B\}$.
We write $2A$ for $A + A$, $3A$ for $A + A + A$, etc.

Definition (Ruzsa distance)**.** The *Ruzsa distance* $d(A, B)$ is

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

Lemma 6.1 (Ruzsa triangle inequality)**.** $d(A, C) \leq d(A, B) + d(B, C)$.

Proof. This is equivalent to the statement

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

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

$$\begin{aligned} \phi : (A - C) \times B &\to (A - B, B - C) \\ (x, b) &\mapsto (a(x) - b, b - c(x)) \end{aligned}$$

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

Lemma 6.2 (Ruzsa covering lemma)**.** Assuming that:

- G an abelian group
- A, B finite subsets of G

Then A can be covered by at most $\frac{|A+B|}{|B|}$ translates of $B - B$.

Proof. Let $\{x_1, \ldots, x_k\}$ be a maximal subset of A such that the sets $x_i + B$ are disjoint.

Then if $a \in A$, there exists i such that $(a + B) \cap (x_i + B) \neq \emptyset$. Then $a \in x_i + B - B$.

So A can be covered by k translates of $B - B$. But

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

Lecture 11

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, $\mathbb{P}(X+Y=z) = \sum_{x+y=z} \mathbb{P}(X=x)\mathbb{P}(Y=y)$. Writing p_x and q_y for $\mathbb{P}(X=x)$ and $\mathbb{P}(Y=y)$ respectively, this givesim $\sum_{x+y=z} p_x q_y = p * q(z)$ where $p(x) = p_x$ and $q(y) = z_y$.

So, sums of independent random variables $\leftrightarrow$ convolutions.

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

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

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

Lemma 6.3. Assuming that:

- A, B are finite subsets of G
- X, Y are uniformly distributed on A, B respectively

Then

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

Proof. Without loss of generality X, Y are indepent. Then

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

□

Lemma 6.4. Assuming that:

- X, Y are G-valued random variables

Then

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

Proof.

$$\begin{aligned}\mathbf{H}[X+Y] &\geq \mathbf{H}[X+Y \mid Y] && \text{(by Subadditivity)}\\ &= \mathbf{H}[X+Y, Y] - \mathbf{H}[Y]\\ &= \mathbf{H}[X, Y] - \mathbf{H}[Y]\\ &= \mathbf{H}[X] + \mathbf{H}[Y] - \mathbf{H}[Y] - \mathbf{I}[X:Y]\\ &= \mathbf{H}[X] - \mathbf{I}[X:Y]\end{aligned}$$

By symmetry we also have

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

□

Corollary. Assuming that:

- X, Y are G-valued random variables

Then:

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

Corollary 6.5. Assuming that:

- X, Y are G-valued random variables

Then

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

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

$$\begin{aligned}\mathbf{H}[X-Y] &\geq \max\{\mathbf{H}[X], \mathbf{H}[Y]\}\\ &\geq \frac{1}{2}(\mathbf{H}[X] + \mathbf{H}[Y])\end{aligned}$$

□

Lemma 6.6. Assuming that:

- X, 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.

$\Leftarrow$ If X, Y are uniform on $x+H$, $y+H$, then $X'-Y'$ is uniform on $x-y+H$, so

$$\mathbf{H}[X'-Y'] = \mathbf{H}[X] = \mathbf{H}[Y].$$

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

$\Rightarrow$ Suppose that X, Y are independent and $\mathbf{H}[X-Y] = \frac{1}{2}(\mathbf{H}[X] + \mathbf{H}[Y])$.

From the first line of the proof of Lemma 6.4, it follows that $\mathbf{H}[X-Y \mid Y] = \mathbf{H}[X-Y]$. Therefore, $X-Y$ and Y are independent. So for every $z \in A-B$ and every $y_1, y_2 \in B$,

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

where $A = \{x : p_x \neq 0\}$, $B = \{y : q_y \neq 0\}$, i.e. for all $y_1, y_2 \in B$,

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

So p_x is constant on $z+B$.

In particular, $A \supset z+B$.

By symmetry, $B \supset A-z$.

So $A = B+z$ for any $z \in A-B$. 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 every $x \in A$.

It follows that

$$A-A+A-A = (A-x)-(A-x) = A-A.$$

So $A-A$ is a subgroup. Also, $A = A-A+c$, so A is a coset of $A-A$. $B = A+z$, so B is also a coset of $A-A$. $\square$

Recall Lemma 1.16: If $Z = f(X) = g(Y)$, then:

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

Lemma 6.7 (The entropic Ruzsa triangle inequality)**.** Assuming that:

- X, Y, Z are G-valued random variables

Then

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

Proof. We must show that (assuming without loss of generality that X, Y and Z are independent) that

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

i.e. that

$$\mathbf{H}[X-Z] + \mathbf{H}[Y] \leq \mathbf{H}[X-Y] + \mathbf{H}[Y-Z]. \qquad (*)$$

Since $X-Z$ is a function of $(X-Y, Y-Z)$ and is also a function of (X,Z), we get using Lemma 1.16 that

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

This is the same as

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

By independence, cancelling common terms and Subadditivity, we get $(*)$. $\square$

Lemma 6.8 (Submodularity for sums)**.** Assuming that:

- X, Y, Z are independent G-valued random variables

Then

$$\mathbf{H}[X+Y+Z]+\mathbf{H}[Z]\leq\mathbf{H}[X+Z]+\mathbf{H}[Y+Z].$$

Proof. $X+Y+Z$ is a function of $(X+Z,Y)$ and also a function of $(X,Y+Z)$. Therefore (using Lemma 1.16),

$$\mathbf{H}[X+Z,YX,Y+Z]+\mathbf{H}[X+Y+Z]\leq\mathbf{H}[X+Z,Y]+\mathbf{H}[X,Y+Z].$$

Hence

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

By independence and cancellation, we get the desired inequality. □

Lecture 12

Lemma 6.9. Assuming that:

- G an abelian group
- X 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

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

(as X_1,X_2,X_3 are all copies of X). □

Corollary 6.10. Assuming that:

- X and Y are G-valued random variables

Then

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

Proof.

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

□

Conditional Distances

Definition (Conditional distance)**.** Let X, Y, U, V be G-valued random variables (in fact, U and V don't have to be G-valued for the definition to make sense). Then the *conditional distance* is

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

The next definition is not completely standard.

Definition (Simultaneous conditional distance)**.** Let X, Y, U be G-valued random variables. The *simultaneous conditional distance of X to Y given U* is

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

We say that X', Y' are *conditionally independent trials* of X, Y given U if:

- X' is distributed like X.
- Y' is distributed like Y.
- For each $u \in U$, $X' \mid U = u$ is distributed like $X \mid U = u$,
- For each $u \in U$, $Y' \mid U = u$ is distributed like $Y \mid U = u$.
- $X' \mid U = u$ and $Y' \mid U = u$ are independent.

Then

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

(as can be seen directly from the formula).

Lemma 6.11 (The entropic BSG theorem)**.** Assuming that:

- A and B are G-valued random variables

Then

$$d[A; B \parallel A + B] \le 3\mathbf{I}[A : B] + 2\mathbf{H}[A + B] - \mathbf{H}[A] - \mathbf{H}[B].$$

Remark. The last few terms look like $2d[A; -B]$. But they aren't equal to it, because A and B aren't (necessarily) independent!

Proof.

$$d[A; B \parallel A + B] = \mathbf{H}[A' - B' \mid A + B] - \frac{1}{2}\mathbf{H}[A' \mid A + B] - \frac{1}{2}\mathbf{H}[B' \mid A + B]$$

where A', B' are conditionally independent trials of A, B given $A + B$. Now calculate

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

Similarly, $\mathbf{H}[B' \mid A + B]$ is the same, so $\frac{1}{2}\mathbf{H}[A' \mid A + B] + \frac{1}{2}\mathbf{H}[B' \mid A + B]$ is also the same.

$$\mathbf{H}[A' - B' \mid A + B] \leq \mathbf{H}[A' - B'].$$

Let (A_1, B_1) and (A_2, B_2) be conditionally independent trials of (A, B) given $A+B$. Then $\mathbf{H}[A'-B'] = \mathbf{H}[A_1 - B_2]$. By Submodularity,

$$\begin{aligned}
\mathbf{H}[A_1 - B_2] &\leq \mathbf{H}[A_1 - B_2, A] + \mathbf{H}[A_1 - B_2, B_1] - \mathbf{H}[A_1 - B_2, A_1, B_1] \\
\mathbf{H}[A_1 - B_2, A_1] &= \mathbf{H}[A_1, B_2] \\
&\leq \mathbf{H}[A_1] + \mathbf{H}[B_2] \\
&= \mathbf{H}[A] + \mathbf{H}[B] \\
\mathbf{H}[A_1 - B_2, B_1] &= \mathbf{H}[A_2 - B_1, B_1] && \text{(since } A_1 + B_1 = A_2 + B_2\text{)} \\
&= \mathbf{H}[A_2, B_1] \\
&\leq \mathbf{H}[A] + \mathbf{H}[B]
\end{aligned}$$

Finally,

$$\begin{aligned}
\mathbf{H}[A_1 - B_2, A_1, B_1] &= \mathbf{H}[A_1, B_1, A_2, B_2] \\
&= \mathbf{H}[A_1, B_1, A_2, B_2 \mid A + B] + \mathbf{H}[A + B] \\
&= 2\mathbf{H}[A, B]A + B + \mathbf{H}[A + B] && \text{(by conditional independence of } (A_1, B_1) \text{ and } (A_2, B_2)\text{)} \\
&= 2\mathbf{H}[A, B] - \mathbf{H}[A + B] \\
&= 2\mathbf{H}[A] + 2\mathbf{H}[B] - 2\mathbf{I}[A : B] - \mathbf{H}[A + B]
\end{aligned}$$

Adding or subtracting as appropriate all these terms gives the required inequality. □

Lecture 13

7 A proof of Marton's conjecture in $\mathbb{F}_2^n$

We shall prove the following theorem.

Theorem 7.1 (Green, Manners, Tao, Gowers)**.** There is a polynomial p with the following property: If $n \in \mathbb{N}$ and $A \subset \mathbb{F}_2^n$ is such that $|A+A| \leq C|A|$, then there is a subspace $H \subset \mathbb{F}_2^n$ of size at most $|A|$ such that A is contained in the union of at most $p(C)$ translates of H. (Equivalently, there exists $K \subset \mathbb{F}_2^n$, $|K| \leq p(C)$ such that $A \subset K + H$).

This is known as "Polynomial Freiman–Ruzsa".

In fact, we shall prove the following statement.

Theorem 7.2 (Entropic Polynomial Freiman–Ruzsa)**.** There exists an absolute constant α satisfying the following: Let $G = \mathbb{F}_2^n$ and let X, Y be G-valued random variables. Then there exists a subsgroup 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.

Lemma 7.3. Assuming that:

- X a discrete random variable (and write p_x for $\mathbb{P}(X = x)$)

Then there exists x such that $p_x \geq 2^{-\mathbf{H}[X]}$.

Proof. If not, then

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

contradiction. □

Proposition 7.4. Theorem 7.2 implies Theorem 7.1.

Proof. Let $A \subset \mathbb{F}_2^n$, $|A+A| \leq C|A|$. Let X and Y be independent copies of U_A. Then by Theorem 7.2, there exists H (a subgroup) such that

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

so

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

But

$$\begin{aligned} d[X;Y] &= \mathbf{H}[U_A - U'_A] - \mathbf{H}[U_A] \\ &= \mathbf{H}[U_A + U'_A] - \mathbf{H}[U_A] \qquad \text{(characteristic 2)} \\ &\leq \log(C|A|) - \log|A| \\ &= \log C \end{aligned}$$

So $d[X;U_H] \leq \frac{\alpha \log C}{2}$. Therefore

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

Therefore, by Lemma 7.3, there exists z such that

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

But

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

(using characteristic 2). So there exists $z \in G$ such that

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

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

But using $B \subset A$,

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

So

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

Since B is contained in $z + H$,

$$|H| \geq C^{-\frac{\alpha}{2}}|A|^{\frac{1}{2}}|H|^{\frac{1}{2}}$$

so $|H| \geq C^{-\alpha}|A|$, so

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

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

$$|A| \geq C^{-\frac{\alpha}{2}}|A|^{\frac{1}{2}}|H|^{\frac{1}{2}}$$

so $|H| \leq C^{\alpha}|A|$.

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', so 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 7.5 (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 exists $\mathbb{F}_2^n$-valued random variables U and V such that

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

Lecture 14

Proposition 7.6. EPFR′$(\eta) \implies$ EPFR(η^{-1}).

Proof. By compactness we can find U, V such that

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

is minimised. If $d[U;V] \neq 0$ then by EPFR′(η) there exist Z, W such that $\tau_{U,V}[Z;W] < d[U;V]$.

But then

$$\begin{aligned}\tau_{X,Y}[Z;W] &= d[Z;W] + \eta(d[Z;X] + d[W;Y]) \\ &\leq d[Z;W] + \eta(d[Z;U] + d[W;V]) + \eta(d[U;X] + d[V;Y]) \\ &\quad \text{(by The entropic Ruzsa triangle inequality)} \\ &< d[U;V] + \eta(d[U;X] + d[V;Y]) \\ &= \tau_{X,Y}[U;V]\end{aligned}$$

Contradiction.

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 us EPFR(η^{-1}). □

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

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

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

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

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

Lemma 7.7 (Fibring lemma)**.** Assuming that:

- G and H are abelian groups
- $\phi : G \to H$ a homomorphism
- let X, Y be G-valued random variables.

Then

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

Proof.

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

But the last line of this expression equals

$$\mathbf{H}[X-Y \mid \phi(X)-\phi(Y)] - \mathbf{H}[X-Y \mid \phi(X), \phi(Y), \phi(X)-\phi(Y)] = \mathbf{I}[X-Y : \phi(X), \phi(Y) \mid \phi(X)-\phi(Y)].$$

□

We shall be interested in the following special case.

Corollary 7.8. Assuming that:

- $G = \mathbb{F}_2^n$ and X_1, X_2, X_3, X_4 are independent G-valued random variables

Then

$$\begin{aligned} d[(X_1,X_2);(X_3,X_4)] &= d[X_1;X_3] + d[X_2;X_4] \\ &= d[X_1+X_2;X_3+X_4] + d[X_1 \mid X_1+X_2; X_3 \mid X_3+X_4] \\ &\quad + \underbrace{\mathbf{I}[X_1+X_3, X_2+X_4 : X_1+X_2, X_3+X_4 \mid X_1+X_2+X_3+X_4]}_{(*)} \end{aligned}$$

Proof. Apply Lemma 7.7 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 Lemma 6.11 says

$$d[X; Y \parallel X + Y] \leq 3\mathbf{I}[X : Y] + 2\mathbf{H}[X + Y] - \mathbf{H}[X] - \mathbf{H}[Y].$$

Equivalently,

$$\mathbf{I}[X : Y] \geq \frac{1}{3}(d[X; Y \parallel X + Y] + \mathbf{H}[X] + \mathbf{H}[Y] - 2\mathbf{H}[X + Y]).$$

Applying this to the information term $(*)$, we get that it is at least

$$\begin{aligned}&\frac{1}{3}(d[X_1 + X_3, X_2 + X_4; X_1 + X_2, X_3 + X_4 \parallel X_2 + X_3, W] + \mathbf{H}[X_1 + X_3, X_2 + X_4 \mid W] \\ &\quad + \mathbf{H}[X_1 + X_2, X_3 + X_4 \mid W] - 2\mathbf{H}[X_2 + X_3, X_2 + X_3 \mid W])\end{aligned}$$

which simplifies to

$$\begin{aligned}&\frac{1}{3}(d[X_1 + X_3, X_2 + X_4; X_1 + X_2, X_3 + X_4 \parallel X_2 + X_3, W] + \mathbf{H}[X_1 + X_3 \mid W] \\ &\quad + \mathbf{H}[X_1 + X_2 \mid W] - 2\mathbf{H}[X_2 + X_3 \mid W])\end{aligned}$$

Lecture 15 So Corollary 7.8 now gives us:

$$\begin{aligned}d[X_1; X_3] + d[X_2; X_4] \geq\ & d[X_1 + X_2; X_3 + X_4] + d[X_1 \mid X_1 + X_2; X_3 \mid X_4] \\ & \frac{1}{3}(d[X_1 + X_2; X_1 + X_3 \parallel X_2 + X_3, W] \\ & + \mathbf{H}[X_1 + X_2 \mid W] + \mathbf{H}[X_1 + X_3 \mid W] - \mathbf{H}[X_2 + X_3 \mid W])\end{aligned}$$

Now 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 first at the entropy terms. We get

$$\begin{aligned}&2\mathbf{H}[X_1 + X_2 \mid W] + \mathbf{H}[X_1 + X_4 \mid W] + \mathbf{H}[X_1 + X_3 \mid W] + \mathbf{H}[X_1 + X_4 \mid W] + \mathbf{H}[X_1 + X_2 \mid W] \\ &\quad - 2\mathbf{H}[X_1 + X_2 \mid W] - 2\mathbf{H}[X_2 + X_4 \mid W] - 2\mathbf{H}[X_1 + X_2 \mid W] \\ &= 0\end{aligned}$$

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

$$\mathbf{H}[X_i + X_j \mid W] = \mathbf{H}[X_k + X_l \mid W].$$

This also allowed use e.g. to replace

$$d[X_1 + X_2, X_3 + X_4; X_1 + X_3, X_2 + X_4 \parallel X_2 + X_3, W]$$

by

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

Therefore, we get the following inequality:

Lemma 7.9.

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

Proof. Above. □

Now let X_1, X_2 be copies of X and Y_1, Y_2 copies of Y and apply Lemma 7.9 to (X_1, X_2, Y_1, Y_2) (all independent), to get this.

Lemma 7.10. Assuming that:

- X_1, X_2, Y_1, Y_2 satisfy: X_1 and X_2 are copies of X, Y_1 and Y_2 are copies of Y, and all of them are independent

Then

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

OR? TODO: figure out which is correct

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

Proof. Use above. □

Recall that we want (U, V) such that

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

Lemma 7.10 gives us a collection of distances (some conditioned), 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 $\eta < \frac{1}{7C}$.

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

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

Lemma 7.11. (Y, X) is 2-relevant to (X, Y).

Proof. $d[Y;X] + d[X;Y] = 2d[X;Y]$. □

Lemma 7.12. Assuming that:

- U, V, X be independent $\mathbb{F}_2^n$-valued random variables

Then

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

Proof.

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

□

Corollary 7.13. Assuming that:

- (U, V) is C-relevant to (X, Y)
- U_1, U_2, V_1, V_2 are independent copies of U, V

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

Proof.

$$\begin{aligned}
&d[U_1 + U_2; X] + d[V_1 + V_2; Y] \\
&\leq \frac{1}{2}(2d[U; V] + d[U; U] + 2d[V; Y] + d[V; V]) && \text{(by Lemma 7.12)} \\
&\leq 2(d[U; X] + d[V; Y]) && \text{(by The entropic Ruzsa triangle inequality)} \\
&\leq 2Cd[X; Y]
\end{aligned}$$

□

Corollary 7.14. $(X_1 + X_2, Y_1 + Y_2)$ is 4-relevant to (Y, X).

Proof. (X, Y) is 2-relevant to (Y, X), so by Corollary 7.13 we're done. □

Corollary. Assuming that:

- (U, V) is C-relevant to (X, Y)

Then $(U + V, U + V)$ is $(3C + 2)$-relevant to (X, Y).

Proof. By Lemma 7.12,

$$\begin{aligned}
d[U + V; X] + d[U + V; Y] &\leq \frac{1}{2}\big(d[U; X] + d[V; X] + d[U; Y] + d[V; Y] + 2d[U; V]\big) \\
&\leq \frac{1}{2}(2d[U; X] + 4d[U; V] + 2d[V; Y]) \\
&\leq \frac{1}{2}(6d[U; X] + 6d[V; Y] + 4d[X; Y])
\end{aligned}$$

□

Lecture 16

Corollary 7.15. Assuming that:

- (U, V) is C-relevant to (X, Y)

Then $(U + V, U + V)$ is $2(C + 1)$-relevant to (X, Y).

Proof.

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

Similarly for $d[U+V;Y]$. □

Lemma 7.16. Assuming that:

- U, V, X are independent $\mathbb{F}_2^n$-valued random variables

Then

$$d[U \mid U+V;X] \le \frac{1}{2}(d[U;X]+d[V;X]+d[U;V]).$$

Proof.

$$\begin{aligned} d[U \mid U+V;X] &\le \mathbf{H}[U+X \mid U+V] - \frac{1}{2}\mathbf{H}[U \mid U+V] - \frac{1}{2}\mathbf{H}[X] \\ &\le \mathbf{H}[U+X] - \frac{1}{2}\mathbf{H}[U] - \frac{1}{2}\mathbf{H}[V] + \frac{1}{2}\mathbf{H}[U+V] - \frac{1}{2}\mathbf{H}[X] \end{aligned}$$

But $d[U \mid U+V;X] = d[V \mid U+V;X]$, so it's also

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

Averaging the two inequalities gives the result (as earlier). □

Corollary 7.17. Assuming that:

- U, V are independent random variables
- (U, V) is C-relevant to (X, Y)

Then

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

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

Proof. Use Lemma 7.16. Then as soon as it is used, we are in exactly the situation we were in when 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 Lemma 7.10. For the fifth term we need to bound

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

But first term of this is at most (by Lemma 7.12)

$$\frac{1}{2}(d[X_1; X_2 + Y_1, X_1 + Y_2]X + d[X_2 \mid X_1 + Y_1, X_1 + Y_2; X] + d[X_1; X_2 \parallel X_2 + Y_1, X_1 + Y_2]).$$

By The entropic Ruzsa triangle inequality and independence, this is at most

$$\begin{aligned} &\leq d[X_1 \mid X_1 + Y_2; X] + d[X_2 \mid X_2 + Y_1; X] \\ &= 2d[X \mid X + Y; X] \end{aligned}$$

Now we can use Lemma 7.16, and similarly for the other terms.

In this way, we get that the fifth and sixth terms have relevances bounded above by λC for an absolute constant λ.

Index