ENTROPY METHODS

TERENCE TAO, VAN VU

1. Introduction

In this expository article we introduce the notion of an *entropy* $\mathbf{H}(X)$ of a discrete variable X taking a finite number of values; informally, this captures the amount of information one learns when one evaluates the value of X. Our emphasis here shall be a little different than that for the probabilistic method, in which the focus is on complex-valued or real-valued random variables, and on their expectation and variance; here, the range of the random variable will play almost no role, and expectation and variance will not appear at all in this discussion. On the other hand, the concepts of determinism and independence will still play major roles in this discussion.

Definition 1.1. *Let X be a discrete random variable. We define the entropy $\mathbf{H}(X)$ of X to be the quantity*[1]

$$\mathbf{H}(X) := -\sum_k \mathbf{P}(X=k) \log \mathbf{P}(X=k)$$

with the convention that $0 \log 0 = 0$; note that only finitely many terms in the series will be non-zero since X is discrete. If Y is another discree random variable, and (X, Y) is the joint random variable, we abbreviate $H((X,Y))$ as $H(X,Y)$; similarly we define $H(X,Y,Z)$, etc.

Remark 1.2. *Clearly $\mathbf{H}(X)$ depends only on the probability distribution of X, thus $\mathbf{H}(X) = \mathbf{H}(X')$ whenever $[X] = [X']$. For similar reasons, $\mathbf{H}(X)$ depends only on the σ-algebra $\mathcal{B}_X$ generated by X, thus $\mathbf{H}(X) = \mathbf{H}(Y)$ whenever $\mathcal{B}_X = \mathcal{B}_Y$. Indeed, we can define an entropy of a probability distribution or of a σ-algebra, for instance $\mathbf{H}(\mathcal{B}) := -\sum_A \mathbf{P}(A) \log \mathbf{P}(A)$ where A ranges over all atoms of $\mathcal{B}$; thus $\mathbf{H}(X) = \mathbf{H}(\mathcal{B}_X)$ for any discrete random variable X.*

A basic estimate is

Lemma 1.3 (Jensen's inequality)**.** *Let X be a discrete random variable taking values in a finite set K. Then we have*

$$0 \leq \mathbf{H}(X) \leq \log|K|. \tag{1}$$

We have $\mathbf{H}(X) = \log|K|$ if and only if X has the uniform distribution on K, and $\mathbf{H}(X) = 0$ if and only if X is deterministic. In particular, a random variable must take at least $\exp(\mathbf{H}(X))$ values.

[1] In some texts the entropy is defined using the logarithm base two intead of the natural logarithm. This does not make a substantial difference to any of the entropy inequalities we will derive here, though.

We leave the proof of this lemma as an exercise (Exercise 2). The quantity $\mathbf{H}(X)$ is related to the asymptotic probability distribution of a large number of trials $X_1, \ldots, X_M$ of the same random variable X:

Lemma 1.4 (Concentration of measure)**.** *Let X be a random variable taking values in a finite set K. Let M be a large integer, and let $X_1, \ldots, X_M$ be independent trials of X. Then there exists a set $K_{X,M} \subset K^M$ of cardinality*

$$|K_{X,M}| = e^{(\mathbf{H}(X)+o(1))M} \tag{2}$$

such that

$$\mathbf{P}((X_1, \ldots, X_M) \in K_{X,M}) = 1 - o(1) \tag{3}$$

and

$$\mathbf{P}((X_1, \ldots, X_M) = \vec{k}) = e^{(-\mathbf{H}(X)+o(1))M} \tag{4}$$

for all $\vec{k} \in K_{X,M}$. Here we use $o(1)$ to denote any expression which goes to zero as $M \to \infty$ for fixed X.

One can obtain much better control on the $o(1)$ errors, of course, but we will not need to do so here. The above lemma asserts, roughly speaking, one should think of the joint variable $(X_1, \ldots, X_M)$ as approximately having a uniform distribution on $\Omega := e^{M\mathbf{H}(X)}$ elements; this can be viewed as a "thermodynamic" definition $\mathbf{H}(X) = \frac{1}{M}\log\Omega$ of the entropy. The approximate uniformity of distribution here means that we do not have to invoke tools such as the pigeonhole principle or the Cauchy-Schwarz inequality in the analysis of entropy; these tools come in some sense "for free", being subsumed in the *Shannon entropy inequalities* which we shall describe below.

Proof. Let $\omega(M)$ be any function of M which goes to ∞ as $M \to \infty$. For each $k \in K$, let $Y_k := \sum_{m=1}^M \mathbf{I}(X_m = k)$, i.e. Y_k counts the number of trials of X which take the value of k. From Hoeffding's inequality we see that

$$\mathbf{P}(|Y_k - M\mathbf{P}(X = k)| \geq \omega(M)M^{1/2}) = o(1)$$

for all $k \in K$. Since the cardinality $|K|$ of K does not depend on M, we thus have

$$\mathbf{P}(|Y_k - M\mathbf{P}(X = k)| \leq \omega(M)M^{1/2} \text{ for all } k \in K) = o(1).$$

Thus if we define

$$K_{X,M} := \{(k_1, \ldots, k_M) \in K^M : |\sum_{m=1}^M \mathbf{I}(k_m = k) - M\mathbf{P}(X = k)| \leq \omega(M)M^{1/2} \text{ for all } k \in K\}$$

then we have (3). It thus remains to prove (4), since (2) will then follow since the total probability must equal 1. But by independence we have for any $\vec{k} =$

$(k_1, \ldots, k_M) \in K_{X,M}$

$$\begin{aligned}\log \mathbf{P}((X_1, \ldots, X_M) = \vec{k}) &= \sum_{m=1}^{M} \log \mathbf{P}(X_m = k_m) \\ &= \sum_{k \in K} (\sum_{m=1}^{M} \mathbf{I}(k_m = k)) \log \mathbf{P}(X = k) \\ &= \sum_{k \in K} (M\mathbf{P}(X = k) + O(\omega(M)M^{1/2}) \log \mathbf{P}(X = k) \\ &= -M\mathbf{H}(X) + O(\omega(M)M^{1/2}),\end{aligned}$$

and (4) follows by choosing ω to be suitably slowly growing in M. □

For any discrete random variable $X : \Omega \to K$, and any event $A \subset \Omega$ of positive probability, we can define the *conditional entropy* $\mathbf{H}(X|A)$ as the entropy of the random variable $(X|A)$, thus

$$\mathbf{H}(X|A) = -\sum_{k \in K} \mathbf{P}(X = k|A) \log \mathbf{P}(X = k|A).$$

If $Y : \Omega \to L$ is another discrete random variable, we define the conditional entropy $\mathbf{H}(X|Y)$ as

$$\mathbf{H}(X|Y) := \sum_{l \in L} \mathbf{H}(X|Y = l) P(Y = l). \tag{5}$$

Informally, this is the expected amount of information one learns from X assuming that one already knows the value of Y. Observe that this quantity is automatically non-negative.

We now record some fundamental inequalities concerning entropy.

Theorem 1.5 (Shannon entropy inequalities)**.** *Let X, Y, Z, W be discrete random variables on a probability space Ω.*

- *(Bayes identity) We have*

$$\mathbf{H}(X, Y) = \mathbf{H}(X|Y) + \mathbf{H}(Y) = \mathbf{H}(Y|X) + \mathbf{H}(X) \tag{6}$$

  and more generally

$$\mathbf{H}(X, Y|Z) = \mathbf{H}(X|Y, Z) + \mathbf{H}(Y|Z) = \mathbf{H}(Y|X, Z) + \mathbf{H}(X|Z). \tag{7}$$

- *(Independence) If X and Y are independent, then*

$$\mathbf{H}(X, Y) = \mathbf{H}(X) + \mathbf{H}(Y); \quad \mathbf{H}(X|Y) = \mathbf{H}(X); \quad \mathbf{H}(Y|X) = \mathbf{H}(Y). \tag{8}$$

  More generally, if X and Y are independent conditioning on Z, then

$$\mathbf{H}(X, Y|Z) = \mathbf{H}(X|Z) + \mathbf{H}(Y|Z); \quad \mathbf{H}(X|Y, Z) = \mathbf{H}(X|Z); \quad \mathbf{H}(Y|X, Z) = \mathbf{H}(Y, Z). \tag{9}$$

- *(Monotonicity) If Y is determined by X, then*

$$\mathbf{H}(Y) \leq \mathbf{H}(Y) + \mathbf{H}(X|Y) = \mathbf{H}(X) \text{ and } \mathbf{H}(Y|X) = 0. \tag{10}$$

  In particular, if X and Y determine each other, then $\mathbf{H}(X) = \mathbf{H}(Y)$. More generally, if Y is determined by X, Z, then

$$\mathbf{H}(Y|Z) \leq \mathbf{H}(Y|Z) + \mathbf{H}(X|Y, Z) = \mathbf{H}(X|Z) \text{ and } \mathbf{H}(Y|X, Z) = 0. \tag{11}$$

Thus if X and Y determine each other given Z, then $\mathbf{H}(X|Z) = \mathbf{H}(Y|Z)$. Similarly, if Z is determined by W, then

$$\mathbf{H}(X|W) \leq \mathbf{H}(X|Z) \leq \mathbf{H}(X). \tag{12}$$

- *(Sub-additivity) If X and Y determine W, then*

$$\mathbf{H}(W) \leq \mathbf{H}(X,Y) \leq \mathbf{H}(X)+\mathbf{H}(Y). \tag{13}$$

More generally, if X,Y,Z determine W, then

$$\mathbf{H}(W|Z) \leq \mathbf{H}(X,Y|Z) \leq \mathbf{H}(X|Z)+\mathbf{H}(Y|Z). \tag{14}$$

- *(Submodularity) If X determines Z, that Y determines Z, and (X,Y) determines W, then*

$$\mathbf{H}(Z)+\mathbf{H}(W) \leq \mathbf{H}(X)+\mathbf{H}(Y). \tag{15}$$

We remark that all of these inequalities correspond well to one's intuitive notion of information. The deepest and most powerful inequality here is the submodularity inequality, which shall use often.

Proof. Let K, L be the range of X,Y respectively. Expanding out $\mathbf{H}(X,Y)$, we obtain

$$\begin{aligned}\mathbf{H}(X,Y) &= \sum_{k\in K}\sum_{l\in L}\mathbf{P}(X=k,Y=l)\log\mathbf{P}(X=k,Y=l)\\ &= \sum_{l\in L}P(Y=l)(\sum_{k\in K}\mathbf{P}(X=k|Y=l)(\log\mathbf{P}(X=k|Y=l)+\log\mathbf{P}(Y=l))\\ &= \sum_{l\in L}P(Y=l)\mathbf{H}(X|Y=l)+\sum_{l\in L}P(Y=l)\log\mathbf{P}(Y=l)\\ &= \mathbf{H}(X|Y)+\mathbf{H}(Y);\end{aligned}$$

from this and symmetry we obtain (6). The generalization (7) then follows immediately by substituting in the formula $\mathbf{H}(X|Y) = \mathbf{H}(X,Y)-\mathbf{H}(Y)$ to eliminate all the conditional expectations.

If X and Y are independent, then $(X|Y=l)$ has the same distribution as X for all l in the range of Y, and hence $\mathbf{H}(X|Y=l) = \mathbf{H}(X)$. Taking expectations over l we obtain $\mathbf{H}(X|Y) = \mathbf{H}(X)$, and (8) then follows from (6) and symmetry. If instead X and Y are independent conditioning on Z, then $(X|Z=m)$ is independent of $(Y|Z=m)$, hence $\mathbf{H}(X,Y|Z=m) = \mathbf{H}(X|Z=m)+\mathbf{H}(Y|Z=m)$ for all m in the range of Z. Taking expectations over m we obtain (9).

If Y is determined by X, then X and (X,Y) generate the same σ-algebra, and hence $\mathbf{H}(X) = \mathbf{H}(X,Y)$, and (10) follows from (6) and the positivity of conditional entropy. The claim (11) then follows by conditioning on $Z=m$ and then taking expectations over m, as before.

Now let X and Z be arbitrary, and let K and M be the ranges of X and Z. Let f denote the function $f(x) := -x\log x$, thus

$$\mathbf{H}(X|Z) = \sum_{m\in M}\mathbf{P}(Z=m)\sum_{k\in K}f(\mathbf{P}(X=k|Z=m)).$$

But since f is concave on $0 \leq x \leq 1$, we have

$$\sum_{m\in M}\mathbf{P}(Z=m)f(\mathbf{P}(X=k|Z=m)) \leq f(\sum_{m\in M}\mathbf{P}(Z=m)\mathbf{P}(X=k|Z=m)) = f(\mathbf{P}(X=k)),$$

and summing over all k we obtain $\mathbf{H}(X|Z) \leq \mathbf{H}(X)$. The remaining claim in (12) can be proven by conditioning on $Z = m$ and taking expectations on m as before.

The first inequality in (13) follows from (10); the second inequality follows from (6) and (12). One can then deduce (14) from (13) by conditioning on $Z = m$ and taking expectations as before.

Finally, (15) follows directly from (14) and (10). □

From these Shannon inequalities we can compute the entropy of multiple (conditionally) independent trials:

Corollary 1.6. *Let $X_1,\ldots,X_N$ be random variables, and let $X_1',\ldots,X_N'$ be independent trials of $X_1,\ldots,X_N$, then*

$$\mathbf{H}(X_1',\ldots,X_N') = \mathbf{H}(X_1)+\ldots+\mathbf{H}(X_N).$$

More generally, if $X_1,\ldots,X_N,Y_1,\ldots,Y_N$ are random variables, with $[Y_1] = \ldots = [Y_N]$, and $X_1',\ldots,X_N',Y$ are conditionally independent trials of $X_1,\ldots,X_N$ conditioning on $Y_1 = \ldots = Y_N = Y$, then

$$\mathbf{H}(X_1',\ldots,X_N'|Y) = \mathbf{H}(X_1|Y_1)+\ldots+\mathbf{H}(X_N|Y_N)$$

or equivalently

$$\mathbf{H}(X_1',\ldots,X_N') = \mathbf{H}(X_1,Y_1)+\ldots+\mathbf{H}(X_N,Y_N)-(N-1)\mathbf{H}(Y).$$

In particular, if each of the Y_n are determined by X_n, then

$$\mathbf{H}(X_1',\ldots,X_N') = \mathbf{H}(X_1)+\ldots+\mathbf{H}(X_N)-(N-1)\mathbf{H}(Y).$$

(1) If $\mathcal{B}$ is a σ-algebra, show that $\mathbf{H}(\mathcal{B}) = \mathbf{E}(X)$, where X be the random variable $X(x) := -\log \mathbf{P}(\mathcal{B}(x))$.
(2) Prove Lemma 1.3. (Hint: use the fact that the function $-x\log x$ is concave when $0 \leq x \leq 1$.)
(3) If $\mathcal{B}$ is a σ-algebra, show that $\mathcal{B}$ contains at least $\exp(\mathbf{H}(\mathcal{B}))$ atoms.
(4) Let X,Y be discrete random variables on a probability space Ω. Show that $\mathbf{H}(X,Y) = \mathbf{H}(X)+\mathbf{H}(Y)$ (or equivalently, that $\mathbf{H}(X|Y) = \mathbf{H}(X)$ or $\mathbf{H}(Y|X) = \mathbf{H}(Y)$) if and only if X and Y are independent. Show that $\mathbf{H}(X,Y) = \mathbf{H}(X)$ if and only if Y is determined by X. This should be compared with the inequalites $\max(\mathbf{H}(X),\mathbf{H}(Y)) \leq \mathbf{H}(X,Y) \leq \mathbf{H}(X)+\mathbf{H}(Y)$ which are true for general pairs of random variables X,Y, and emphasizes the extent to which determinism and independence are opposite extremes.
(5) Use the concentration of measure lemma to give an alternate proof of (13).
(6) Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two σ-algebras on a finite probability space Ω. Use the submodularity inequality (15) to deduce that

$$\mathbf{H}(\mathcal{B}_1 \vee \mathcal{B}_2)+\mathbf{H}(\mathcal{B}_1 \wedge \mathcal{B}_2) \leq \mathbf{H}(\mathcal{B}_1)+\mathbf{H}(\mathcal{B}_2);$$

this may help explain the term "submodularity". In particular we have $\mathbf{H}(\mathcal{B}_1 \vee \mathcal{B}_2) \leq \mathbf{H}(\mathcal{B}_1)+\mathbf{H}(\mathcal{B}_2)$.
(7) (Uniqueness of entropy) Suppose for every finite set K and any discrete random variable X taking values in K, we assigned a non-negative real number $\tilde{\mathbf{H}}(X)$ which depended only on the probability distribution of X (so we have $\tilde{\mathbf{H}}(X) = \tilde{\mathbf{H}}(X')$ whenever $[X] = [X']$), such that $\tilde{H}$ obeyed the identities (6) and (10), where $\tilde{\mathbf{H}}(X|Y)$ is defined as in (5). Suppose also that for any fixed K, the dependence of $\tilde{\mathbf{H}}(X)$ on the probabilities $\mathbf{P}(X = k)$ is (jointly) continuous. Show that there exists a non-negative constant $c \geq 0$

Additive sets A, B	Discrete random variable X, Y
(Log-)cardinality $\log \lvert A \rvert$	Entropy $\mathbf{H}(X)$
Complete sum set $A \pm B$	$X \pm Y$ with X, Y independent
Partial sum set $A \pm_G B$	$X \pm Y$ with X, Y correlated
Cartesian product $A \times B$	Joint variable (X, Y)
Tensor power $A^{\oplus N}$	N independent trials $X_1, \dots, X_N$
Subset A' of A	Conditioning $(X \vert \mathcal{B})$ of X
Ruzsa distance $d(A, B)$	Ruzsa distance $d(X, Y)$
Functional relationship $f : A \to B$	Y is determined by X
(Log-)multiplicity of $f : A \to B$	Conditional entropy $\mathbf{H}(X \vert Y)$
Additive energy $\log \frac{\lvert A \rvert^2 \lvert B \rvert^2}{E(A,B)}$	Renyi 2-entropy $\mathbf{H}_2(X+Y)$

Figure 1. A partial dictionary between sums of additive sets, and discrete random variables.

such that $\tilde{\mathbf{H}}(X) = c\mathbf{H}(X)$ for all random variables X. (Hint: Normalize $\tilde{\mathbf{H}}(X) = \mathbf{H}(X)$ when X is the uniform distribution on $\{0,1\}$. Then show that $\tilde{\mathbf{H}}(X) = \mathbf{H}(X)$ whenever X is the uniform distribution on A and $|A|$ is a power of 2. Then, show that $\tilde{\mathbf{H}}(X) = \mathbf{H}(X)$ whenever all the probabilities $\mathbf{P}(X = x)$ are either zero or a negative power of 2. Then conclude using the continuity.)

2. Entropy sum set estimates

Let us now work in a fixed additive group G. There is a strong analogy between additive sets A in G and discrete random variables X in G; see Figure 2. For instance, to every additive set A one can associate the uniform distribution X_A on A, and $\mathbf{H}(X_A) = \log|A|$. In the converse direction we have Lemma 1.4, which shows that any random variable X, when sampled a large number M of times, becomes approximately uniform on some large additive set $K_{X,M} \subset G^M$ of cardinality roughly $\exp(M\mathbf{H}(X))$. Note however that since one has to discard an error of $o(1)$ to achieve this, the analogy between the two theories, while mostly accurate, is not completely perfect. However, there are certain areas of additive combinatorics where there is a rigorous equivalence between the sumset-theoretic and entropy-theoretic formulations of a problem; see Section 3.

For finite sets we have $|A \pm B| \leq |A||B|$. The entropy analogue of this is

Lemma 2.1. *Let X, Y be discrete random variables in an additive group G, and let $\pm$ be a sign. Then*

$$\mathbf{H}(X \pm Y) \leq \mathbf{H}(X) + \mathbf{H}(Y).$$

If furthermore X and Y are independent, then

$$\max(\mathbf{H}(X), \mathbf{H}(Y)) \leq \mathbf{H}(X \pm Y).$$

Proof. The first inequality follows immediately from sub-additivity (13). For the second bound, we observe

$$\mathbf{H}(X \pm Y) \geq \mathbf{H}(X \pm Y|Y) = \mathbf{H}(X|Y) = \mathbf{H}(X)$$

where the first inequality is monotonicity (12), the second follows since $X \pm Y$ and X determine each other conditioning on Y, and the third is independence (8). From this and symmetry we obtain the claim. □

We can now give the analogue of Ruzsa distance for discrete random variables.

Definition 2.2 (Ruzsa distance)**.** *Let X, Y be discrete random variables taking values in an additive group G. We define the Ruzsa distance $d(X,Y)$ between X, Y to be*

$$d(X,Y) := \mathbf{H}(X_1 - Y_1) - \frac{1}{2}\mathbf{H}(X_1) - \frac{1}{2}\mathbf{H}(Y_1)$$

where X_1 and Y_1 are independent trials of X and Y respectively.

It is clear from the definition that $d(X,Y)$ depends only on the individual distributions of X and Y, and not on the joint distribution, since we have enforced independence of the trials. This is the natural analogue of Ruzsa distance, but for random variables rather than sets. One piece of evidence in this direction is in Exercise 1; another is in the following analogue of the Ruzsa triangle inequality.

Lemma 2.3 (Ruzsa triangle inequality, entropy version)**.** *The Ruzsa distance $d(X,Y)$ is non-negative, symmetric, and obeys the triangle inequality*

$$d(X,Z) \leq d(X,Y) + d(Y,Z)$$

for all discrete random variables X, Y, Z taking values in an additive group G.

Proof. The symmetry is clear, while the non-negativity follows from Lemma 2.1. It remains to prove the triangle inequality, which is equivalent to showing that

$$\mathbf{H}(X_1 - Z_1) + \mathbf{H}(Y_1) \leq \mathbf{H}(X_1 - Y_1) + \mathbf{H}(Y_1 - Z_1)$$

where X_1, Y_1, Z_1 are independent trials of X, Y, Z respectively. But if we define the joint random variables

$$\begin{aligned}
\mathbf{X} &:= (X_1, Z_1)\\
\mathbf{Y} &:= (X_1 - Y_1, Y_1 - Z_1)\\
\mathbf{Z} &:= X_1 - Z_1\\
\mathbf{W} &:= (X_1, Y_1, Z_1)
\end{aligned}$$

we see that $\mathbf{X}$ determines $\mathbf{Z}$, that $\mathbf{Y}$ determines $\mathbf{Z}$, and $(\mathbf{X}, \mathbf{Y})$ determines $\mathbf{W}$. Applying the submodularity inequality (15) we then have

$$\mathbf{H}(\mathbf{Z}) + \mathbf{H}(\mathbf{W}) \leq \mathbf{H}(\mathbf{X}) + \mathbf{H}(\mathbf{Y}).$$

But by independence (8) and sub-additivity (13) we have

$$\mathbf{H}(\mathbf{X}) = \mathbf{H}(X_1) + \mathbf{H}(Z_1); \quad \mathbf{H}(\mathbf{Y}) \leq \mathbf{H}(X_1 - Y_1) + \mathbf{H}(Y_1 - Z_1); \quad \mathbf{H}(\mathbf{W}) = \mathbf{H}(X_1) + \mathbf{H}(Y_1) + \mathbf{H}(Z_1)$$

and the claim follows. □

We can establish a number of other entropy inequalities in a similar spirit, using elementary arithmetic identities and entropy inequalities such as the submodularity inequality to obtain estimates similar to standard sum set estimates. While the proofs here may seem quite different from the proofs of the sumset estimates, they are at heart the same proof, because they both ultimately rely on the same elementary arithmetic identities. For instance, we have the following analogue of the set inequality $|A+B| \leq |A-B|^3/|A||B|$.

Proposition 2.4. *Let X,Y be discrete random variables in an ambient group G. Then we have*

$$d(X,-Y) \leq 3d(X,Y)$$

or in other words, if X_1,Y_1 are independent trials of X,Y, then

$$\mathbf{H}(X_1+Y_1) \leq 3\mathbf{H}(X_1-Y_1)-\mathbf{H}(X_1)-\mathbf{H}(Y_1).$$

Proof. Without loss of generality we may assume X,Y are independent (since the claim only depends on the individual distributions of X,Y and not on the joint distribution). It thus suffices to show that $\mathbf{H}(X_1+Y_1) \leq 3\mathbf{H}(X-Y)-\mathbf{H}(X)-\mathbf{H}(Y)$.

Let (X_2,Y_2) and (X_3,Y_3) be conditionally independent trials of (X,Y) conditioning on $X-Y$, thus $X_2-Y_2 = X_3-Y_3$, and hence

$$X_1+Y_1 = (X_1-Y_3)-(X_2-Y_1)+X_3+Y_2.$$

Thus if we define the joint random variables

$$\begin{aligned}\mathbf{X} &:= (X_1-Y_3,X_2-Y_1,X_3,Y_2)\\ \mathbf{Y} &:= (X_1,Y_1)\\ \mathbf{Z} &:= X_1+Y_1\\ \mathbf{W} &:= (X_1,X_2,X_3,Y_1,Y_2,Y_3)\end{aligned}$$

we see that $\mathbf{X}$ determines $\mathbf{Z}$, that $\mathbf{Y}$ determines $\mathbf{Z}$, and $(\mathbf{X},\mathbf{Y})$ determines $\mathbf{W}$. Applying the submodularity inequality (15) we then have

$$\mathbf{H}(\mathbf{Z})+\mathbf{H}(\mathbf{W}) \leq \mathbf{H}(\mathbf{X})+\mathbf{H}(\mathbf{Y}).$$

But by independence (8) and Corollary 1.6 we have

$$\mathbf{H}(\mathbf{Y}) = \mathbf{H}(X)+\mathbf{H}(Y); \quad \mathbf{H}(\mathbf{W}) = 3\mathbf{H}(X)+3\mathbf{H}(Y)-\mathbf{H}(X-Y)$$

while from subadditivity (13), and the observation that (X_1,Y_3) and (X_2,Y_1) have the same distribution as (X,Y), we have

$$\mathbf{H}(\mathbf{X}) \leq 2\mathbf{H}(X-Y)+\mathbf{H}(X)+\mathbf{H}(Y).$$

Combining these inequalities, the claim follows. □

Another result in additive combinatorics says that if A and B are subsets of an additive group then there exists a large sumset S of $A+B$ such that $|nS| \leq O_n(|A+B|^{2n+1}/|A|^n|B|^n)$ for all n. We have an entropy version of this fact:

Proposition 2.5. *Let X,Y be independent discrete random variables taking values in an additive group G, let $S := X+Y$, and let $S_0,\dots,S_n$ be $n+1$ independent trials of S. Then for any integer $n \geq 0$ we have*

$$\mathbf{H}(S_0+\ldots+S_n) \leq (2n+1)\mathbf{H}(X+Y)-n\mathbf{H}(X)-n\mathbf{H}(Y).$$

Proof. Without loss of generality we may assume $S_i = X_i+Y_i$ for $0 \leq i \leq n$, where (X_i,Y_i) are independent trials of (X,Y). Then

$$S_0+\ldots+S_n = (Y_0+X_1)+\ldots+(Y_{n-1}+X_n)+(Y_n+X_0).$$

Observe that the random variables $Y_0+X_1,\dots,Y_{n-1+X_n},Y_n+X_0,X_0,Y_0,S_1,\dots,S_n$ will determine all of the X_i, Y_i. Thus if we set

$$\begin{aligned}\mathbf{X} &:= (Y_0+X_1,\dots,Y_{n-1}+X_n,Y_n+X_0)\\ \mathbf{Y} &:= (X_0,Y_0,S_1,\dots,S_n)\\ \mathbf{Z} &:= S_0+\dots+S_n\\ \mathbf{W} &:= (X_0,\dots,X_n,Y_0,\dots,Y_n)\end{aligned}$$

we see that $\mathbf{X}$ determines $\mathbf{Z}$, that $\mathbf{Y}$ determines $\mathbf{Z}$, and $(\mathbf{X},\mathbf{Y})$ determines $\mathbf{W}$. Applying the submodularity inequality (15) we then have

$$\mathbf{H}(\mathbf{Z})+\mathbf{H}(\mathbf{W}) \leq \mathbf{H}(\mathbf{X})+\mathbf{H}(\mathbf{Y}).$$

However, by subadditivity (13) we have $\mathbf{H}(\mathbf{X}) \leq (n+1)\mathbf{H}(X+Y)$ and $\mathbf{H}(\mathbf{Y}) \leq \mathbf{H}(X)+\mathbf{H}(Y)+n\mathbf{H}(X+Y)$, while from independence (8) we have $\mathbf{H}(\mathbf{W}) = (n+1)\mathbf{H}(X)+(n+1)\mathbf{H}(Y)$. The claim follows. □

From this and the Ruzsa triangle inequality, one can obtain the inequality

$$\mathbf{H}(X_1+\dots+X_n-X_1'-\dots-X_{n'}'+Y_1+\dots+Y_m-Y_1'-\dots-Y_{m'}') \leq H(X)+C(n+n'+m+m')d(X,X) \tag{16}$$

whenever n,n',m,m' and the random variables in the left-hand side are independent trials of X and Y; see Exercise 4. Another consequence is that

$$d(X_1-X_2,X_1-X_2) <= 2d(X,X) \tag{17}$$

whenever X_1,X_2 are independent trials of X.

The additive energy has an information theoretic counterpart relating to the Renyi 2-entropy; see Exercise 12 or [4].

There are analogues of the Balog-Szemerédi-Gowers theorem for random variables. The main point of the Balog-Szemerédi-Gowers theorem was to pass from cardinality control on a partial sum set to cardinality control on a complete sum set; here, the idea will be to pass from entropy control of a sum of *correlated* random variables to entropy control of a sum of (conditionally) *independent* random variables. We first give an entropy version of a weak Balog-Szemerédi theorem.

Lemma 2.6 (Weak Balog-Szemerédi-Gowers theorem, entropy version)**.** *Let (X,Y) be a pair of discrete random variables, not necessarily independent, taking values in $G\times G$. Let (X_1,Y), (X_2,Y) be conditionally independent trials of (X,Y) conditioning on Y. Then*

$$\mathbf{H}(X_1-X_2|Y) \leq 2\mathbf{H}(X+Y)+2\mathbf{H}(X)+\mathbf{H}(Y)-2\mathbf{H}(X,Y).$$

In particular, if we have $\mathbf{H}(X,Y) \geq \mathbf{H}(X)+\mathbf{H}(Y)-\log K$ and $\mathbf{H}(X+Y) \leq \frac{1}{2}\mathbf{H}(X)+\frac{1}{2}\mathbf{H}(Y)+\log K'$ then

$$\mathbf{H}(X_1|Y) = \mathbf{H}(X_2|Y) \geq \mathbf{H}(X)-\log K$$

(thanks to (6)) and

$$\mathbf{H}(X_1-X_2|Y) \leq \mathbf{H}(X)+2\log K+2\log K'.$$

Proof. We view (X_1,X_2,Y) as a single random variable taking values in G, and create two conditionally independent trials (X_1,X_2,Y) and (X_1,X_2,Y') of this variable conditioning on (X_1,X_2), thus by Corollary 1.6

$$\begin{aligned}\mathbf{H}(X_1,X_2,Y,Y') &= 2\mathbf{H}(X_1,X_2,Y)-\mathbf{H}(X_1,X_2)\\ &\geq 2(2\mathbf{H}(X,Y)-\mathbf{H}(Y))-(\mathbf{H}(X_1)+\mathbf{H}(X_2))\\ &= 4\mathbf{H}(X,Y)-2\mathbf{H}(Y)-2\mathbf{H}(X).\end{aligned}$$

Now we use submodularity again, introducing

$$\begin{aligned}\mathbf{X} &:= (X_1+Y',X_2+Y',Y)\\ \mathbf{Y} &:= (X_1,X_2,Y)\\ \mathbf{Z} &:= (X_1-X_2,Y)\\ \mathbf{W} &:= (X_1,X_2,Y,Y').\end{aligned}$$

Observe that $\mathbf{X}$ determines $\mathbf{Z}$, that $\mathbf{Y}$ determines $\mathbf{Z}$, and $(\mathbf{X},\mathbf{Y})$ determines $\mathbf{W}$, and thus

$$\mathbf{H}(\mathbf{Z})+\mathbf{H}(\mathbf{W}) \leq \mathbf{H}(\mathbf{X})+\mathbf{H}(\mathbf{Y}).$$

We have already computed that $\mathbf{H}(\mathbf{W}) \geq 4\mathbf{H}(X,Y)-2\mathbf{H}(Y)-2\mathbf{H}(X)$; also from Corollary 1.6 we have

$$\mathbf{H}(\mathbf{Y}) = 2\mathbf{H}(X,Y)-\mathbf{H}(Y)$$

and by Bayes' formula (6)

$$\mathbf{H}(\mathbf{Z}) = \mathbf{H}(X_1-X_2|Y)+\mathbf{H}(Y).$$

Finally, by subadditivity (13) we have

$$\mathbf{H}(\mathbf{X}) \leq 2\mathbf{H}(X+Y)+\mathbf{H}(Y)$$

since (X_1,Y') and (X_2,Y') both have the same distribution as (X,Y). Combining all these inequalities, the claim follows. □

Now we give the counterpart to the full Balog-Szemerédi-Gowers theorem.

Theorem 2.7 (Balog-Szemerédi-Gowers theorem, entropy version)**.** *Let (X,Y) be a pair of discrete random variables, not necessarily independent, taking values in $G\times G$. Let (X_1,Y), (X_2,Y) be conditionally independent trials of (X,Y) conditioning on Y, and then let (X_1',X_2',Y_1) and (X_1',Y_2) be conditionally independent trials of (X_1,X_2,Y) and (X,Y) conditioning on $X_1=X$. Then*

$$\mathbf{H}(X_2'|X_1',Y_1) = \mathbf{H}(X|Y); \quad \mathbf{H}(Y_2|X_1',Y_1) = \mathbf{H}(Y|X)$$

and

$$\mathbf{H}(X_2'+Y_2|X_1',Y_1) \leq 3\mathbf{H}(X+Y)+3\mathbf{H}(X)+3\mathbf{H}(Y)-4H(X,Y).$$

Furthermore, X_2' and Y_2 are conditionally independent, conditioning on X_1',Y_1. In particular, if $\mathbf{H}(X,Y) \geq \mathbf{H}(X)+\mathbf{H}(Y)-\log K$ and $\mathbf{H}(X+Y) \leq \frac{1}{2}\mathbf{H}(X)+\frac{1}{2}\mathbf{H}(Y)+\log K'$ then

$$\mathbf{H}(X_2'|X_1',Y_1) \geq \mathbf{H}(X)-\log K; \quad \mathbf{H}(Y_2|X_1',Y_1) \geq \mathbf{H}(Y)-\log K; \quad \mathbf{H}(X_2'+Y_2|X_1',Y_1) \leq \frac{1}{2}\mathbf{H}(X)+\frac{1}{2}\mathbf{H}(Y)+3\log K'+4\log K.$$

Proof. By construction, if one conditions on X_1' then Y_2 and (X_2',Y_1) are conditionally independent, which implies that X_2' and Y_2 are conditionally independent conditioning on X_1',Y_1. Also, since (X_1',X_2',Y_1) has the same distribution as (X_1,X_2,Y), we have

$$\mathbf{H}(X_2'|X_1',Y_1') = \mathbf{H}(X_2|X_1,Y) = \mathbf{H}(X_2|Y) = \mathbf{H}(X|Y)$$

since X_1 is independent of (X_2, Y). Similarly, since Y_2 and Y_1 are independent conditioning on X_1', we have (by (9))

$$\mathbf{H}(Y_2|X_1', Y_1) = \mathbf{H}(Y_2|X_1') = \mathbf{H}(Y|X).$$

It remains to establish the bound on $\mathbf{H}(X_2'+Y_2|X_1', Y_1)$. By Lemma 2.6 we already have

$$\mathbf{H}(X_1'-X_2'|Y_1) \leq 2\mathbf{H}(X+Y)+2\mathbf{H}(X)+\mathbf{H}(Y)-2\mathbf{H}(X,Y).$$

We then apply the submodularity inequality with

$$\begin{aligned}\mathbf{X} &:= (X_2', Y_2, Y_1)\\ \mathbf{Y} &:= (X_1'-X_2', X_1'+Y_2, Y_1)\\ \mathbf{Z} &:= (X_2'+Y_2, Y_1)\\ \mathbf{W} &:= (X_1', X_2', Y_1, Y_2);\end{aligned}$$

it is clear that $\mathbf{X}$ determines $\mathbf{Z}$, that $\mathbf{Y}$ determines $\mathbf{Z}$, and $(\mathbf{X}, \mathbf{Y})$ determine $\mathbf{W}$, so by (15) we have

$$\mathbf{H}(\mathbf{Z})+\mathbf{H}(\mathbf{W}) \leq \mathbf{H}(\mathbf{X})+\mathbf{H}(\mathbf{Y}).$$

From sub-additivity we have $\mathbf{H}(\mathbf{X}) \leq \mathbf{H}(X_2', Y_1)+\mathbf{H}(Y_2) = \mathbf{H}(X,Y)+\mathbf{H}(Y)$ and

$$\begin{aligned}\mathbf{H}(\mathbf{Y}) &\leq \mathbf{H}(X_1'-X_2'|Y_1)+\mathbf{H}(Y_1)+\mathbf{H}(X_1'+Y_2)\\ &\leq (2\mathbf{H}(X+Y)+2\mathbf{H}(X)+\mathbf{H}(Y)-2\mathbf{H}(X,Y))+\mathbf{H}(Y)+\mathbf{H}(X+Y)\end{aligned}$$

while from two applications of Corollary 1.6 we have

$$\begin{aligned}\mathbf{H}(\mathbf{W}) &= \mathbf{H}(X_1', X_2', Y_1)+\mathbf{H}(X_1', Y_2)-\mathbf{H}(X_1')\\ &= (2\mathbf{H}(X,Y)-\mathbf{H}(Y))+\mathbf{H}(X,Y)-\mathbf{H}(X)\end{aligned}$$

and from (6) we have $\mathbf{H}(\mathbf{Z}) = \mathbf{H}(X_2'+Y_2|Y_1)+\mathbf{H}(Y)$. Combining these inequalities gives the result. □

As the above examples show, there is certainly a strong analogy between sum set estimates, and entropy inequalities for sums of random variables. However the latter theory has not been developed nearly as thoroughly as the former; for instance we do not have a "Freiman theorem" which classifies the random variables X for which $d(X,X)$ or $d(X,-X)$ is small. Nor do we have an analogue of the Plünnecke inequalities for random variables.

(1) Let A, B be additive sets with ambient group G, and let X_A and X_B be uniform distributions on A and B respectively. Show that $d(X_A, X_B) \leq d(A,B)$. (Hint: use (1)).
(2) Let X, Y be discrete random variables taking values in a group G. Show that $|\mathbf{H}(X)-\mathbf{H}(Y)| \leq \frac{1}{2}d(X,Y)$.
(3) Let X, Y be independent random variables, and let (X_1, Y_1) and (X_2, Y_2) be conditionally independent trials of (X,Y) conditioning on $X+Y$, thus we have $Y_2 = X_1-X_2+Y_1$. Show that $\mathbf{H}(X_1, X_2, Y_1|Y_2) = 2\mathbf{H}(X)+\mathbf{H}(Y)-\mathbf{H}(X+Y)$. (This result is analogous to Ruzsa's covering lemma.)
(4) Deduce (16) and (17) from Proposition 2.5. What constant C can you get for (16)?

(5) Let X, Y be independent discrete random variables taking values in an additive group G, let Y_1, Y_2, Y_3, Y_4 be independent trials of Y, and let X_5, X_6 be independent trials of X. Show that

$$\mathbf{H}(Y_1 - Y_2 - Y_3 + Y_4) \leq 4\mathbf{H}(X+Y) + \mathbf{H}(X_5 - X_6) - 4\mathbf{H}(X),$$

which is the counterpart of the set inequality $|2A-2A| \leq \frac{|A+B|^4|B-B|}{|B|^4}$. Hint: let (X_3, Y_3) and (X_5, Y_5) be conditionally independent trials of (X, Y) conditioning on $X+Y$, let (X_4, Y_4) and (X_6, Y_6) be another pair of conditionally independent trials of (X, Y) conditioning on $X+Y$ which are independent of the preceding random variables, and let Y_1, Y_2 be two independent trials of Y which are independent of all preceding variables. Exploit the submodularity inequality applied to the random variables

$$\begin{aligned} \mathbf{X} &:= (X_3 + Y_1, X_5 - Y_6, X_4 + Y_2, Y_5, Y_6) \\ \mathbf{Y} &:= (Y_1, Y_2, Y_3, Y_4) \\ \mathbf{Z} &:= Y_1 - Y_2 + Y_3 + Y_4 \\ \mathbf{W} &:= (X_3, X_4, X_5, X_6, Y_1, Y_2, Y_3, Y_4, Y_5, Y_6). \end{aligned}$$

taking advantage of the identity

$$Y_1 - Y_2 - Y_3 + Y_4 = (X_3 + Y_1) - (X_5 - X_6) - (X_4 + Y_2) - Y_5 + Y_6.$$

(6) Suppose that X, Y are discrete random variables in an ambient group G such that $d(X, Y) = 0$. Show that there exists a finite subgroup H in G, and elements $x, y \in G$, such that X is the uniform distribution on $x + H$ and Y is the uniform distribution on $y + H$. (Hint: Use Exercise 4 from Section 1).

(7) Develop entropy analogues of non-commutative sumset theory.

(8) [Imre Ruzsa, private communication] Let X, Y be independent discrete random variables taking values in an additive group G, and suppose that $\mathbf{H}(X+Y) > \mathbf{H}(X-Y)$. Let $0 < \alpha < \mathbf{H}(X)$, $0 < \beta < \mathbf{H}(Y)$ be real numbers chosen so that $\mathbf{H}(X-Y) < \alpha + \beta < \mathbf{H}(X+Y)$ (this is always possible, thanks to Lemma 2.1). Let n be a large number, and let $(X_1, \ldots, X_n), (Y_1, \ldots, Y_n) \in G^n$ be independent trials of X and Y respectively. let $A, B \subseteq G^n$ be chosen by letting A be the values of $\lfloor e^{\alpha n} \rfloor$ independent trials of $(X_1, \ldots, X_n)$, and letting B be the values of $\lfloor e^{\beta n} \rfloor$ independent trials of $(Y_1, \ldots, Y_n)$ (independent of the trials used to select A). Show that with probability $1 - o(1)$, one has $|A| = e^{(\alpha - o(1))n}$, $|B| = e^{(\beta - o(1))n}$, $|A+B| = e^{(\alpha+\beta-o(1))n}$, but $|A-B| \leq e^{(\alpha+\beta-\varepsilon+o(1))n}$ for some $\varepsilon > 0$ depending only in α, β, X, Y. Thus $|A+B|$ is close to $|A||B|$ but $|A-B|$ is significantly smaller than A, B. (Hint: the elements $a \in A$, $b \in B$ for which a does not have the statistics of X, b does not have the statistics of Y, $a+b$ does not have the statistics of $X+Y$, or $a-b$ does not have the statistics of $X-Y$ all give negligible contributions thanks to Lemma 1.4. To get the lower bound on $|A+B|$, estimate the multiplicity of $A+B$ at points which have the statistics of $X+Y$).

(9) [Imre Ruzsa, private communication] By modifying the arguments in Exercise 8, give an example of an arbitrarily large additive set A such that $|A+A| = |A|^{2-o(1)}$ but $|A-A| \leq |A|^{2-\varepsilon+o(1)}$ for some absolute constant

$\varepsilon > 0$, or such that $|A-A| = |A|^{2-o(1)}$ but $|A+A| \leq |A|^{2-\varepsilon+o(1)}$ for some absolute constant $\varepsilon > 0$.

(10) If X is a discrete random variable, we define the *Renyi* 2*-entropy* $\mathbf{H}_2(X)$ to be the quantity $\mathbf{H}_2(X) := -\log_2 \sum_k \mathbf{P}(X = k)^2$. Show that $\mathbf{H}(X) \leq \mathbf{H}_2(X)$, with equality if and only if X has a uniform distribution. The Renyi 2-entropy is also known as the *extension entropy*, and the quantity $\sum_k \mathbf{P}(X = k)^2$ is also known as the *collision probability.*

(11) [4] Let X, Y be independent variables taking values in an additive group. Show that $\mathbf{H}_2(X-Y) = \mathbf{H}_2(X+Y) \leq \frac{1}{2}\mathbf{H}_2(X-X) + \frac{1}{2}\mathbf{H}_2(Y-Y)$.

(12) Let A and B be additive sets in an ambient group G, and let X, Y be random variables selected uniformly and independently from A and B respectively. Show that $\mathbf{H}_2(X+Y) = \mathbf{H}_2(X-Y) = \log \frac{|A|^2|B|^2}{E(A,B)}$.

3. Entropy inequalities of Kakeya and Erdös distance type

One striking development in recent years is the discovery that entropy sum set inequalities for *non-independent* pairs of variables X, Y can be of great utility in some classical problems in incidence geometry, in particular the Kakeya problem and the Erdös distance problem. We phrase these two problems as follows.

Conjecture 3.1 (Geometric Kakeya conjecture for finite fields). *Let F be a finite field and let $d \geq 1$, thus F^d is a vector space over the field F. Let $E \subseteq F^d$ be a set such that given any $v \in F^d$ there exists an $x \in F^d$ such that the line $\{x+tv : t \in F\}$ is completely contained in F. Then we have $|E| \geq c_{\varepsilon,d}|F|^{d-\varepsilon}$ for every $\varepsilon > 0$, where $c_{\varepsilon,d} > 0$ is a quantity that does not depend on F or E, provided that the characteristic of F is sufficiently large depending on ε. (It may even be possible to take $\varepsilon = 0$; no counterexample to this claim is known).*

Conjecture 3.2 (Geometric Erdös distance conjecture). *Let P be a set of n points in the plane $\mathbf{R}^2$, and let $\Delta(P) := \{|x-y| : x, y \in P\}$ be the set of distances. Then $|\Delta(P)| \geq c_\varepsilon n^{2-\varepsilon}$ for every $\varepsilon > 0$, where $c_\varepsilon > 0$ is a quantity that does not depend on n or P.*

These conjectures remain open, but in both cases some partial bounds are known. For instance, using geometric methods one can obtain the lower bounds of $|E| \geq c_d|F|^{(d+2)/2}$ for the Kakeya conjecture and of $|\Delta(P)| \geq cn^{4/5}$ for the Erdös distance problem. But the only known improvement to these bounds have come from employing methods from additive combinatorics (with the exception of the finite field Kakeya problem in four dimensions, in which algebraic geometry has been used to obtain a slight additional improvement). In particular, the following two conjectures have been identified:

Conjecture 3.3 (Entropy Kakeya conjecture). *Given any $\varepsilon > 0$ there exist numbers $r_1, \ldots, r_k \in (\mathbf{Q}\backslash\{-1\}) \cup \{\infty\}$ such that one has the entropy inequality*

$$\mathbf{H}(X-Y) \leq (1+\varepsilon) \max_{1 \leq i \leq k} \mathbf{H}(X+r_iY) \tag{18}$$

for all discrete random variables X, Y (not necessarily independent) in an finite vector space F^{d-1}, provided that the characteristic of F is sufficiently large depending on ε (in particular, so that multiplication by r_i will make sense). Here we adopt the convention that $\mathbf{H}(X+\infty Y) = \mathbf{H}(Y)$.

Conjecture 3.4 (Entropy Erdös distance conjecture)**.** *Given any $\varepsilon > 0$ there exists a $d \geq 3$ such that one has the entropy inequality*

$$\mathbf{H}(X_1,\dots,X_d) \leq (2+\varepsilon) \max_{1\leq i<j\leq d} \mathbf{H}(X_i+X_j)$$

whenever $X_1,\dots,X_d$ are discrete random variables (not necessarily independent) in the real line $\mathbf{R}$, with the property that $X_1,\dots,X_d$ is determined by any of the X_i (thus $\mathbf{H}(X_1,\dots,X_d|X_i) = 0$ for all $1 \leq i \leq d$).

A version of the Entropy Kakeya conjecture is implicit in [6] and [15], but the entropy formulation given above is due to Ruzsa (private communication). A version of the Entropy Erdös conjecture is implicit in [25], but the entropy formulation given above is due to [13]. For a more systematic study of these conjectures (and in particular a surprising connection to the theory of finite groups) see [14].

The connection between the entropy and geometric versions of these conjectures is the following. It was implicitly noted in [6] that if the entropy Kakeya conjecture is true for some $\varepsilon > 0$, then the geometric Kakeya conjecture is true with ε replaced by $\frac{\varepsilon}{\varepsilon+1}(d-1)$; in particular the full entropy Kakeya conjecture would imply the full geometric Kakeya conjecture. In [29] (and implicitly in [25]) it was shown that if the entropy Erdös distance conjecture were true for some $\varepsilon > 0$, then the geometric Erdös distance conjecture would be true with ε replaced by any quantity greater than $1 - \frac{4}{5-1/(2-\varepsilon)}$. In particular, the full entropy Erdös distance conjecture would imply the geometric Erdös distance conjecture for all $\varepsilon > 1/9 = 0.111\dots$ (currently this is only known for $\varepsilon > 3/22$, see [13]). We sketch the connection between the two Kakeya conjectures Exercises 1, 2, and the connection between the two Erdös conjectures in ???.

For now, we give some partial results on these two entropy conjectures, beginning with the entropy Kakeya conjecture. We will always assume that the characteristic of the field F is large enough that one can make sense of multiplication by the rational numbers involved. Observe that for any distinct $r, r' \in Q\backslash\{-1\}\cup\{\infty\}$, that $X-Y$ is completely determined by $X+rY$ and $X+r'Y$, and hence we have

$$\mathbf{H}(X-Y) \leq \mathbf{H}(X+rY) + \mathbf{H}(X+r'Y).$$

This establishes the entropy Kakeya conjecture for $\varepsilon \geq 1$. Now we lower the value of ε.

Proposition 3.5. [15] *Let X,Y be two discrete random variables (not necessarily independent) taking values in an additive group where every non-zero element has order at least 3. Then*

$$3\mathbf{H}(X,Y) + \mathbf{H}(X-Y) \leq 2\mathbf{H}(X) + 2\mathbf{H}(Y) + 2\mathbf{H}(X+Y) + \mathbf{H}(X+2Y).$$

Since $\mathbf{H}(X-Y) \leq \mathbf{H}(X,Y)$, this establishes the entropy Kakeya conjecture for $\varepsilon \geq 1 - 1/4$.

Proof. Let (X,Y), (X,Y') be two conditionally independent trials of (X,Y) conditioning on X. We then apply the submodularity inequality to the random variables

$$\begin{aligned}
\mathbf{X} &:= (Y, X+2Y)\\
\mathbf{Y} &:= (X+Y, X+Y')\\
\mathbf{Z} &:= Y-(X+2Y') = (X+Y)-2(X+Y')\\
\mathbf{W} &:= (X,Y,Y')
\end{aligned}$$

to conclude

$$\mathbf{H}(Y-X-2Y')+\mathbf{H}(X,Y,Y') \leq \mathbf{H}(Y)+\mathbf{H}(X+2Y)+2\mathbf{H}(X+Y).$$

But by Corollary 1.6 we have $\mathbf{H}(X,Y,Y') = 2\mathbf{H}(X,Y)-\mathbf{H}(X)$. Also, since $X-Y$ is clearly determined by $Y-X-2Y'$ and Y' we have $\mathbf{H}(X-Y) \leq \mathbf{H}(Y-X-2Y')+\mathbf{H}(Y)$. Inserting these bounds in the above inequality, the claim follows. □

In a similar vein, we have the following estimate which only establishes the entropy Kakeya conjecture for $\varepsilon \geq 1-1/6 = 0.833\ldots$ but does not require the random variable $X+2Y$. This is actually fairly sharp; see Exercise 5. The estimate also has some application to partial sum and difference sets, see Exercise 4.

Proposition 3.6. [15] *Let X,Y be two discrete random variables (not necessarily independent) taking values in an additive group G. Then*

$$5\mathbf{H}(X,Y)+\mathbf{H}(X-Y) \leq 4\mathbf{H}(X)+4\mathbf{H}(Y)+3\mathbf{H}(X+Y).$$

Proof. As before we let (X,Y) and (X,Y') be two conditionally independent trials of (X,Y) conditioning on X. Applying the submodularity inequality to the random variables

$$\begin{aligned}\mathbf{X} &:= (X+Y, X+Y')\\ \mathbf{Y} &:= (Y,Y')\\ \mathbf{Z} &:= Y-Y'\\ \mathbf{W} &:= (X,Y,Y')\end{aligned}$$

we conclude that

$$\mathbf{H}(Y-Y')+\mathbf{H}(X,Y,Y') \leq 2\mathbf{H}(Y)+2\mathbf{H}(X+Y). \tag{19}$$

Next, we let (X,Y,Y') and (X'',Y'',Y') be two conditionally independent trials of (X,Y,Y') conditioning on Y' and on $X+Y = X''+Y''$. Then by Corollary 1.6

$$\mathbf{H}(X,Y,Y',X'',Y'') = 2\mathbf{H}(X,Y,Y')-\mathbf{H}(Y)-\mathbf{H}(X+Y).$$

On the other hand, we split (using (12))

$$\mathbf{H}(X,Y,Y',X'',Y'') = \mathbf{H}(X'',Y',X''-Y',X,Y'',Y-Y') \leq \mathbf{H}(X'',Y'|X''-Y')+\mathbf{H}(X''-Y',X,Y'',Y-Y').$$

But from the identity

$$X''-Y' = X-Y''+(Y-Y')$$

we see that $X''-Y'$ is determined by X, Y'', and $Y-Y'$, and thus

$$\mathbf{H}(X''-Y',X,Y'',Y-Y') = \mathbf{H}(X,Y'',Y-Y') \leq \mathbf{H}(X)+\mathbf{H}(Y)+\mathbf{H}(Y-Y').$$

Combining these inequalities we obtain

$$2\mathbf{H}(X,Y,Y')-\mathbf{H}(Y)-\mathbf{H}(X+Y) \leq \mathbf{H}(X,Y|X-Y)+\mathbf{H}(X)+\mathbf{H}(Y)+\mathbf{H}(Y-Y').$$

Inserting this into (19) we obtain

$$3\mathbf{H}(X,Y,Y') \leq \mathbf{H}(X)+4\mathbf{H}(Y)+3\mathbf{H}(X+Y)+\mathbf{H}(X,Y|X-Y).$$

From Corollary 1.6 we have $\mathbf{H}(X,Y,Y') = 2\mathbf{H}(X,Y)-\mathbf{H}(X)$, while $\mathbf{H}(X,Y|X-Y) = \mathbf{H}(X,Y)-\mathbf{H}(X-Y)$. The claim follows. □

For some further progress on the entropy Kakeya conjecture, see [16]; the best known result is that this conjecture is true for all $\varepsilon > 0.67513\ldots$.

Now we consider the entropy Erdös conjecture. Observe that if X_1, X_2, X_3 are real valued, then (X_1, X_2, X_3) are determined by X_1+X_2, X_2+X_3, and X_3+X_1, thus

$$\mathbf{H}(X_1,X_2,X_3) \leq \mathbf{H}(X_1+X_2)+\mathbf{H}(X_2+X_3)+\mathbf{H}(X_3+X_1).$$

This establishes the entropy Erdös conjecture when $\varepsilon = 1$ (this was first observed implicitly in [25]). By increasing $d = 3$ to $d = 5$ one can improve ε to $3/4$:

Proposition 3.7. [29] *Let X_1, X_2, X_3, X_4, X_5 be random variables taking values in an additive group where all non-zero elements have order at least three, and such that $(X_1, X_2, X_3, X_4, X_5)$ is determined by X_i for each $1 \leq i \leq 5$. Then*

$$\mathbf{H}(X_1,X_2,X_3,X_4,X_5) \leq (3-\frac{1}{4}) \max_{1\leq i<j\leq 5} \mathbf{H}(X_i+X_j).$$

Proof. Let us normalize $\max_{1\leq i<j\leq 5} \mathbf{H}(X_i+X_j) = \log N$. By (14) we have

$$\mathbf{H}(X_1+X_3,X_1+X_4,X_1+X_5|X_1-X_2) \leq \sum_{j=3,4,5} \mathbf{H}(X_1+X_j|X_1-X_2).$$

Observe that X_1+X_4, X_1+X_5 determine X_4-X_5, hence

(20)
$$\mathbf{H}(X_1+X_3,X_1+X_4,X_1+X_5,X_4-X_5|X_1-X_2) \leq \sum_{j=3,4,5} \mathbf{H}(X_1+X_j|X_1-X_2).$$

By (12) we have

$$\mathbf{H}(X_1+X_3,X_1+X_4,X_1+X_5,X_3+X_4,X_4-X_5|X_1-X_2) \leq \mathbf{H}(X_1+X_3,X_1+X_4,X_1+X_5,X_4-X_5|X_1-X_2)+\mathbf{H}(X_3+X_4|X_4-$$

But observe that from $X_1+X_3, X_1+X_4, X_3+X_4$ one can already determine X_1, X_3, X_4, which then determine all the X's by hypothesis. Thus

$$\mathbf{H}(X_1,X_2,X_3,X_4,X_5|X_1-X_2) \leq H(X_1+X_3,X_1+X_4,X_1+X_5,X_4-X_5|X_1-X_2)+\mathbf{H}(X_3+X_4|X_4-X_5).$$

Combining this with (20) we obtain

$$\mathbf{H}(X_1,X_2,X_3,X_4,X_5) \leq \sum_{j=3,4,5} \mathbf{H}(X_1+X_j|X_1-X_2)+\mathbf{H}(X_1-X_2)+\mathbf{H}(X_3+X_4|X_4-X_5).$$

Now we observe that

$$\mathbf{H}(X_3+X_4|X_4-X_5) = \mathbf{H}(X_3+X_4,X_3+X_5|X_4-X_5) \leq \mathbf{H}(X_3+X_4)+\mathbf{H}(X_3+X_5)-\mathbf{H}(X_4-X_5) \leq 2\log N-\mathbf{H}(X_4-X_5)$$

and similarly for $\mathbf{H}(X_1+X_j|X_1-X_2)$, thus

$$\mathbf{H}(X_1,X_2,X_3,X_4,X_5) \leq 8\log N - 2\mathbf{H}(X_1-X_2) - \mathbf{H}(X_4-X_5).$$

On the other hand, we have

$$\mathbf{H}(X_1,X_2,X_3,X_4,X_5) = \mathbf{H}(X_1) \leq \mathbf{H}(X_1+X_2)+\mathbf{H}(X_1-X_2) \leq \log N+\mathbf{H}(X_1-X_2)$$

and similarly for $\mathbf{H}(X_4-X_5)$. Thus we have

$$\mathbf{H}(X_1,X_2,X_3,X_4,X_5) \leq 11\log N - 3\mathbf{H}(X_1,X_2,X_3,X_4,X_5)$$

and the claim follows. □

The constant $3-\frac{1}{4}$ in the above proposition was improved further to $3-\frac{2}{7}$ in [13] by using the technique of conditionally independent trials as in Proposition 3.6 and Proposition 3.5. Combining these arguments with an additional iteration argument, the arithmetic Erdös conjecture was established in [13] for all $\varepsilon > \frac{4-e}{10-3e} = 0.69464\ldots$.

(1) (Ruzsa, private communication) Suppose that $\varepsilon, r_1, \ldots, r_k, F, d-1$ are such that the entropy inequality (18) holds. Let $A, B \subset F^{d-1}$ be additive sets. Establish the inequality

$$|A -_G B| \leq (\max_{1 \leq i \leq k} |A +_G r_i \cdot B|)^{1+\varepsilon} \tag{21}$$

where by slight abuse of notation we have $A +_G r \cdot B := \{a + rb : (a,b) \in G\}$ and one adopts the convention that $|A +_G \infty \cdot B| = |\{b : (a,b) \in G\}|$. (Hint: first refine G so that the map $(a,b) \mapsto a-b$ is injective on G, and then consider the uniform distribution on G). Conversely, if the inequality (21) holds for all A, G, B and all dimensions d, establish the entropy inequality (18) for the same values of $\varepsilon, r_1, \ldots, r_k, F$. (Hint: use Lemma 1.4 to convert the random variables into approximately uniform distributions, and eliminate the $o(1)$ losses using a tensor power trick).

(2) Suppose that $\varepsilon, r_1, \ldots, r_k, F, d$ are such that the entropy inequality (18) holds. Let $E \subseteq F^d$ be as in the geometric Kakeya conjecture, and for each $t \in F$ let $E_t := \{x \in F^{d-1} : (x,t) \in E\}$ be the "slice" of E at height t. Let $t_i \in F$ be such that $r_i = t_i/(1-t_i)$. Using Exercise 1, establish the inequality

$$|F|^{d-1} \leq (\max_{1 \leq i \leq k} |E_{t_i+s}|)^{1+\varepsilon}$$

for any $s \in F$. Conclude in particular that

$$|E| \geq |F|^{d - \frac{\varepsilon}{\varepsilon+1}(d-1)}.$$

(3) [29] Let $\varepsilon > 0$ and $d \geq 3$ be such that the entropy Erdös distance conjecture holds. Let G be a subset of $\mathbf{R}^d$ such that the co-ordinate maps $(x_1, \ldots, x_d) \mapsto x_i$ are injective on G for each $1 \leq i \leq d$, and such that

$$|\{x_i + x_j : (x_1, \ldots, x_d) \in G\}| \leq N$$

for all $1 \leq i < j \leq d$ and some $N \geq 1$. Conclude that $|G| \leq N^{2+\varepsilon}$.

(4) Using Proposition 3.6, establish the inequality

$$|A -_G B| \leq \frac{|A|^4 |B|^4 |A +_G B|^3}{|G|^5}$$

for any additive sets A, B in an ambient group Z, and any non-empty $G \subseteq A \times B$. If all non-zero elements in Z have order at least three, establish the inequality

$$|A -_G B| \leq \frac{|A|^2 |B|^2 |A +_G B|^2 |A +_G 2 \cdot B|}{|G|^3}.$$

For a proof of these inequalities which does not require entropy, see [15].

(5) (Ruzsa, private communication). Let F be a finite field, let $(X,Y) \in F \times F$ be the uniform distribution on $\{(0,0), (0,1), (1,0)\}$. Show that

$$\mathbf{H}(X-Y) = (1+\varepsilon)\max(\mathbf{H}(X), \mathbf{H}(Y), \mathbf{H}(X+Y))$$

where $\varepsilon := \frac{\log 27}{\log 27/4} - 1 = 0.726\ldots$.

(6) Assume the Erdös distance conjecture (Conjecture 3.2). Conclude that for any additive set A of real numbers, one has $|f(A-A)+f(A-A)| \geq c_\varepsilon |A|^{2-\varepsilon}$ for all $\varepsilon > 0$, where $f : \mathbf{R} \to \mathbf{R}$ is the squaring function $f(x) := x^2$. This statement, which is thus weaker than the Erdös distance conjecture, remains open.

References

[1] M. Ajtai, V. Chvátal, M. Newborn, and E. Szemerdi, *Crossing-free subgraphs*, Annals of Discrete Mathematics **12** (1982), 9–12.

[2] B. Aronov, J. Pach, M. Sharir, G. Tardos, *Distinct distances in three and higher dimensions*, Combin. Probab. Comput. **13** (2004), no. 3, 283–293.

[3] A. Balog, E. Szemerédi, *A statistical theorem of set addition*, Combinatorica, **14** (1994), 263–268.

[4] B. Barak, R. Impagliazzo, A. Wigderson, *Extracting randomness using few independent sources*, preprint.

[5] J. Beck, *On the lattice property of the plane and some problems of Dirac, Motzkin, and Erdos in combinatorial geometry*, Combinatorica **3** (1983), 281–297.

[6] J. Bourgain, *On the dimension of Kakeya sets and related maximal inequalities*, GAFA **9** (1999), 256–282.

[7] F. Chung, *The number of distinct distances determined by n points in the plane*, J. Combin. Theory Ser. A **36** (1984), 342–354.

[8] F. Chung, E. Szemerédi, W. Trotter, *The number of different distances determined by a set of n points in the Euclidean plane*, Discrete Computational Geom. **7** (1992), 1–11.

[9] K. Clarkson, H. Edelsbrunner, L. Gubias, M. Sharir, E. Welzl, *Combinatorial complexity bounds for arrangements of curves and spheres*, Discrete Comput. Geom. **5** (1990), 99–160.

[10] P. Erdős, *On sets of distances of n points*, Amer. Math. Monthly **53** (1946), 248–250.

[11] J. Garibaldi, Erdős distance problem for convex metrics, UCLA Ph.D. Thesis.

[12] A. Iosevich, *Curvature, combinatorics, and the Fourier transform*, Not. Amer. Math. Soc. **48** (2001), 577–583.

[13] N. Katz, G. Tardos, *A new entropy inequality for the Erdős distance problem*, Towards a theory of geometric graphs, (ed. J. Pach), Contemporary Mathematics **342** (2004), 119–126.

[14] N. Katz, *On arithmetic combinatorics and finite groups*, preprint.

[15] N. Katz, T. Tao, *Bounds on arithmetic progressions, and applications to the Kakeya conjecture*, Math. Res. Letters **6** (1999), 625–630.

[16] N. Katz, T. Tao, *New bounds for Kakeya problems*, Journal d'Analyse de Jerusalem, **87** (2002), 231–263.

[17] N. Katz, T. Tao, *Some connections between the Falconer and Furstenburg conjectures*, New York J. Math. **7** (2001), 148–187.

[18] G. Mockenhaupt, T. Tao, *Kakeya and restriction phenomena for finite fields*, Duke Math. J. **121** (2004), 35-74

[19] L. Moser, *On the different distances determined by n points*, Amer. Math. Monthly **59** (1952), 85–91.

[20] J. Pach, *Crossing numbers*, Discrete and computational geometry (Tokyo, 1998), 267–273, Lecture Notes in Comput. Sci., 1763, Springer, Berlin, 2000.

[21] J. Pach, G. Tardos, *Isosceles triangles determined by a planar point set*, Graphs and Combinatorics **18** (2002), 769–779.

[22] J. Pach and G. Tóth, *Graphs drawn with few crossings per edge*, Combinatorica 17 (1997), 427–439.

[23] J. Solymosi, V. Vu, *Distinct distances in high dimensional homogeneous sets*, preprint.

[24] J. Solymosi, V. Vu, *Near optimal bound for the distinct distances problem in high dimensions*, preprint.

[25] J. Solymosi, C. D. Tóth, *Distinct distances in the plane*, Discrete Comput. Geom. **25** (4) (2001), 629–634.

[26] J. Solymosi, G. Tardos, C.D. Tóth, *The k most frequent distances in the plane*, preprint.

[27] L. Székely, *Crossing numbers and hard Erdős problems in discrete geometry*, Combin. Probab. Comput. **6** (1997), 353–358.

[28] E. Szemerédi, W. T. Trotter Jr., *Extremal problems in discrete geometry*, Combinatorica **3** (1983), 381–392.

[29] G. Tardos, *On distinct sums and distinct distances*, preprint.

[30] T. Tao, *Finite field analogues of the Erdős, Falconer, and Furstenburg problems*, unpublished.

[31] C. Toth, *The Szemeredi-Trotter theorem in the complex plane*, to appear.