Entropy and additive combinatorics

by

Marcel K. Goh

16 January 2024

Note. This is a gentle introduction to the notion of entropy as it is used in additive combinatorics, with a view towards understanding the new proof of the polynomial Freiman–Ruzsa conjecture by W. T. Gowers, B. Green, F. Manners, and T. Tao. The preliminary portion of these notes are largely transcribed from lectures given by W. T. Gowers.

1. The Khintchine–Shannon axioms

Let X be a discrete random variable. Its entropy $\mathbf{H}\{X\}$ is a real number (or ∞) that measures the "information content" of X. For example, if X is a constant random variable, then $\mathbf{H}\{X\}$ should be zero (we do not gain any information from knowing the value of X), and if X is uniformly distributed on $\{0,1\}^n$, then $\mathbf{H}\{X\}$ should be proportional to n, since X is determined by n bits of information. It satisfies the following axioms, which are sometimes called the Khinchine–Shannon axioms.

a) (*Invariance.*) If X takes values in A, Y takes values in B, $\phi : A \to B$ is a bijection, and $\mathbf{P}\{Y = \phi(a)\} = \mathbf{P}\{X = a\}$ for all $a \in A$, then $\mathbf{H}\{X\} = \mathbf{H}\{Y\}$.

b) (*Extensibility.*) If X takes values in A and Y takes values in B for a set B such that $A \subseteq B$, and furthermore $\mathbf{P}\{Y = a\} = \mathbf{P}\{X = a\}$ for all $a \in A$, then $\mathbf{H}\{X\} = \mathbf{H}\{Y\}$.

c) (*Continuity.*) The quantity $\mathbf{H}\{X\}$ depends continuously on the probabilities $\mathbf{P}\{X = a\}$.

d) (*Maximisation.*) Over all possible random variables X taking values in a finite set A, the quantity $\mathbf{H}\{X\}$ is maximised for the uniform distribution.

e) (*Additivity.*) For X taking values in A and Y taking values in B, we have the formula

$$\mathbf{H}\{X, Y\} = \mathbf{H}\{X\} + \mathbf{H}\{Y \mid X\},$$

where $\mathbf{H}\{X, Y\} = \mathbf{H}\big\{(X, Y)\big\}$ and

$$\mathbf{H}\{Y \mid X\} = \sum_{x \in A} \mathbf{P}\{X = x\}\, \mathbf{H}\{Y \mid X = x\}.$$

For now, we shall take it on faith that there really exists a function on random variables satisfying these axioms. Later on, when we prove this, we will

find that actually, the axioms only define entropy up to a multiplicative constant, so we shall add the following axiom.

f) (*Normalisation.*) If X is uniformly distributed on $\{0,1\}$, then $\mathbf{H}\{X\} = \log_2(e)$.

It is actually more common (for obvious reasons, especially in computer science) for one to set the entropy of a uniform random variable on $\{0,1\}$ to 1, but we shall follow the convention of Gowers et al., since the eventual goal of these notes is to understand their proof of the polynomial Freiman–Ruzsa conjecture.

Notationally, we would expect that $\mathbf{H}\{Y \mid X\} = \mathbf{H}\{Y\}$ if X and Y are independent. This is the first proposition we will carefully prove using the axioms above.

Proposition 1. *Let X and Y be independent random variables. Then $\mathbf{H}\{Y \mid X\} = \mathbf{H}\{Y\}$ and consequently $\mathbf{H}\{X,Y\} = \mathbf{H}\{X\} + \mathbf{H}\{Y\}$.*

Proof. Suppose X takes values in a finite set A. Then for all $x \in A$, the distribution of Y and Y given that $X = x$ is the same, so

$$\mathbf{H}\{Y \mid X\} = \sum_{x \in A} \mathbf{P}\{X = x\}\, \mathbf{H}\{Y \mid X = x\} = \sum_{x \in A} \mathbf{P}\{X = x\}\, \mathbf{H}\{Y\} = \mathbf{H}\{Y\}.$$

The second version of the statement follows from the additivity axiom. ∎

We will sometimes use the notation X^n to denote the vector $(X_1, \dots, X_n)$ where the X_i are independent copies of the random variable X.

Corollary 2. *For a discrete random variable X, we have $\mathbf{H}\{X^n\} = n\,\mathbf{H}\{X\}$.*

Proof. Perform induction on n and use the previous proposition. ∎

In the case where X is the uniform distribution on $\{0,1\}$, and $X_1, \dots, X_n$ are independent and distributed as X, we can additionally use the normalisation axiom to conclude that

$$\mathbf{H}\{X^n\} = \mathbf{H}\{X_1, \dots, X_n\} = \sum_{i=1}^{n} \mathbf{H}\{X_i\} = \log_2(e) \cdot n,$$

by induction.

A similar induction on general random variables $X_1, \dots, X_n$ gives the following statement, often known as the *chain rule*.

Proposition 3 (*Chain rule*). *Let $X_1, \dots, X_n$ be random variables. Then*

$$\mathbf{H}\{X_1, \dots, X_n\} = \mathbf{H}\{X_1\} + \mathbf{H}\{X_2 \mid X_1\} + \cdots + \mathbf{H}\{X_n \mid X_1, \dots, X_{n-1}\}. \quad ∎$$

Next we establish the intuitive fact that the entropy of a uniform random variable supported on a set A is at most the entropy of a uniform random variable supported on a superset B of A.

Proposition 4. *Let $A \subseteq B$ with B finite, let X be uniformly distributed on A, and let Y be uniformly distributed on B. Then $\mathbf{H}\{X\} \leq \mathbf{H}\{Y\}$, with equality if and only if $A = B$.*

Proof. By the extensibility axiom, $\mathbf{H}\{X\}$ is not affected if we regard X as a function taking values in B. Then by the maximisation axiom, $\mathbf{H}\{X\} \leq \mathbf{H}\{Y\}$, since Y is uniform on B. If $A = B$, then it is clear that $\mathbf{H}\{X\} = \mathbf{H}\{Y\}$, since X and Y are the same random variable. On the other hand, say $|A| = m$ and $|B| = n$ with $m < n$. If $m = 1$, then by the previous proposition we have $\mathbf{H}\{X\} = 0$, and by normalisation and invariance, $\mathbf{H}\{Y\} = \log_2(e)$. ∎

If Y is a random variable such that $Y = f(X)$ for some random variable Y and some function f, then we say that Y is *determined by* X or X *determines* Y. We want to show that $\mathbf{H}\{Y\} \leq \mathbf{H}\{X\}$, which reflects the idea that we get more information from X than from Y. This, rather annoyingly, seems to require two steps.

Lemma 5. *If $Y = f(X)$ then $\mathbf{H}\{X\} = \mathbf{H}\{Y\} + \mathbf{H}\{X \mid Y\}$.*

Proof. There is a a bijection between values x taken by Y and values $\big(x, f(x)\big)$ taken by (X, Y), so we have

$$\mathbf{H}\{X\} = \mathbf{H}\{X, Y\} = \mathbf{H}\{Y\} + \mathbf{H}\{X \mid Y\}$$

by additivity. ∎

We are now done if we can show that entropy is nonnegative. This can also be derived straight from the axioms. The following proof is due to S. Eberhard.

Proposition 6. *Let X be a discrete random variable taking values in a finite set A. Then $\mathbf{H}\{X\} \geq 0$.*

Proof. First we will work in the case where there exists n such that $\mathbf{P}\{X = a\}$ is a multiple of $1/n$ for all $a \in A$. Let Y be uniformly distributed on $[n]$ and let $\{E_a\}_{a \in A}$ be a partition of $[n]$ such that $|E_a| = n\,\mathbf{P}\{X = a\}$ for all $a \in A$, and let $Z = a$ if $Y \in E_a$. This definition makes Z and X identically distributed, so $\mathbf{H}\{Z\} = \mathbf{H}\{X\}$ by the invariance axiom, and it suffices to prove $\mathbf{H}\{Z\} \geq 0$.

Since Z is determined by Y, we have $\mathbf{H}\{Y\} = \mathbf{H}\{Z\} + \mathbf{H}\{Y \mid Z\}$ by the previous lemma. Furthermore, for every $a \in A$, the conditional entropy $\mathbf{H}\{Y \mid Z = a\}$ is uniformly distributed on a set of size at most n, whereas is uniformly distributed on n, so by Proposition 4, we have $\mathbf{P}\{Y \mid Z = a\} \leq \mathbf{H}\{Y\}$. This implies that $\mathbf{H}\{Y\} \geq \mathbf{H}\{Y \mid Z\}$, which in turn gives us $\mathbf{H}\{Z\} \geq 0$. ∎

Corollary 7. *If $Y = f(X)$ then $\mathbf{H}\{X\} \geq \mathbf{H}\{Y\}$.* ∎

Now we show that a random variable has zero entropy if and only if it is constant. This reflects the idea that the variables from which we get no information are those which take the same value no matter what.

Proposition 8. *Let X be a discrete random variable. Then $\mathbf{H}\{X\} = 0$ if and only if it takes exactly one value.*

Proof. First suppose that X takes only one value. Let a be the value of X such that $\mathbf{P}\{X = a\} = 1$. Then (X, X) equals (a, a) with probability 1 as well, so $\mathbf{H}\{X\} = \mathbf{H}\{X, X\}$ by the invariance axiom. But it can easily be checked that X and (X, X) are independent (we have

$$\mathbf{P}\big\{X = a, (X, X) = (a, a)\big\} = \mathbf{P}\{X = a\}\,\mathbf{P}\big\{(X, X) = (a, a)\big\}$$

for instance), so $\mathbf{H}\{X, X\} = 2\,\mathbf{H}\{X\}$. Thus we conclude that $\mathbf{H}\{X\} = 0$.

Now suppose that X takes more than one value; let A be the set of a such that $\mathbf{P}\{X = a\} > 0$ and let $\alpha = \max_{a \in A} \mathbf{P}\{X = a\}$. For all n let X^n denote the tuple of n independent copies of X; the maximum probability of any particular value (in A^n) that X^n takes is α^n. But $\alpha < 1$ since X takes more than one value, so for any $\epsilon > 0$ we can find n such that $\alpha^n < \epsilon$. This means that we can partition A^n into two disjoint sets E and F such that $\mathbf{P}\{X^n \in E\}$ and $\mathbf{P}\{X^n \in F\}$ are both in the range $[1/2 - \epsilon, 1/2 + \epsilon]$.

Let Y be the random variable taking the value 0 if $X^n \in E$ and 1 if $X^n \in F$. Then by Corollary 2, $\mathbf{H}\{X^n\} = n\,\mathbf{H}\{X\}$, and since X^n determines Y,

$$\mathbf{H}\{X^n\} = \mathbf{H}\{Y\} + \mathbf{H}\{X^n \mid Y\} \geq \mathbf{H}\{Y\}.$$

But $\mathbf{H}\{Y\} > 0$ for ϵ small enough, the normalisation and continuity axioms. So $\mathbf{H}\{X\} \geq \mathbf{H}\{Y\}/n > 0$ as well. ∎

Mutual information. For random variables X and Y, the *mutual information* $\mathbf{I}\{X : Y\}$ is defined by the equivalent formulas

$$\begin{aligned}\mathbf{I}\{X : Y\} &= \mathbf{H}\{X\} + \mathbf{H}\{Y\} - \mathbf{H}\{X, Y\} \\ &= \mathbf{H}\{X\} - \mathbf{H}\{X \mid Y\} \\ &= \mathbf{H}\{Y\} - \mathbf{H}\{Y \mid X\}.\end{aligned}$$

It measures, roughly speaking, how much information one can get from one variable by looking at the other one. From the formula it is clear that $\mathbf{I}\{X : Y\} = 0$ if and only if X and Y are independent, and we also have the inequality $\mathbf{H}\{X|Y\} \leq \mathbf{H}\{X\}$; that is, *conditioning cannot increase the entropy of a random variable*. Given a triple (X, Y, Z) of random variables, we can apply this with $(X \mid Z = z)$ to obtain the inequality

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

which is called *submodularity*. An equivalent statement is

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

If Z takes values in a set C, the *conditional mutual information* is defined by

$$\begin{aligned}\mathbf{I}\{X : Y \mid Z\} &= \sum_{z \in C} p_Z(z)\, \mathbf{I}\big\{(X \mid Z = z) : (Y \mid Z = z)\big\} \\ &= \mathbf{H}\{X \mid Z\} + \mathbf{H}\{Y \mid Z\} - \mathbf{H}\{X, Y \mid Z\} \\ &= \mathbf{H}\{X, Z\} - 2\,\mathbf{H}\{Z\} + \mathbf{H}\{Y, Z\} - \mathbf{H}\{X, Y, Z\} + \mathbf{H}\{Z\} \\ &= \mathbf{H}\{X, Z\} + \mathbf{H}\{Y, Z\} - \mathbf{H}\{X, Y, Z\} - \mathbf{H}\{Z\},\end{aligned}$$

so we see that submodularity of entropy is equivalent to the statement that $\mathbf{I}\{X : Y \,|\, Z\} \geq 0$. Analogously to the unconditional case, we have equality if and only if X and Y are independent when conditioned on Z.

2. Group-valued random variables

Now we will examine the case where the random variables in question take values in an abelian group G, meaning we can take sums $X + Y$ and differences $X - Y$ of them. Note that if we condition on Y, then the values taken $X + Y$ are in bijection with values taken by X. This leads to the following proposition.

Proposition 9. *Let X and Y be random variables each taking finitely many values in an abelian group G. We have*

$$\max\big(\mathbf{H}\{X\}, \mathbf{H}\{Y\}\big) - \mathbf{I}\{X : Y\} \leq \mathbf{H}\{X \pm Y\}.$$

Furthermore, for any random variable Z, we have the conditional version

$$\max\big(\mathbf{H}\{X \mid Z\}, \mathbf{H}\{Y \mid Z\}\big) - \mathbf{I}\{X : Y \mid Z\} \leq \mathbf{H}\{X \pm Y \mid Z\}$$

of the same statement.

Proof. Since conditioning does not increase entropy, we have

$$\mathbf{H}\{X \pm Y\} \geq \mathbf{H}\{X \pm Y \mid Y\},$$

and since the probabilities $\mathbf{P}\{X + Y = z \mid Y = y\} = \mathbf{P}\{X = z - y \mid Y = y\}$ for all $z \in G$, by invariance we have

$$\mathbf{H}\{X \pm Y\} \geq \mathbf{H}\{X \mid Y\} = \mathbf{H}\{X\} - \mathbf{I}\{X : Y\}.$$

Repeating the same argument but exchanging the roles of X and Y, we get

$$\mathbf{H}\{X \pm Y\} \geq \mathbf{H}\{Y \mid X\} = \mathbf{H}\{Y\} - \mathbf{I}\{X : Y\},$$

so

$$\mathbf{H}\{X \pm Y\} \geq \max\big(\mathbf{H}\{X\}, \mathbf{H}\{Y\}\big) - \mathbf{I}\{X : Y\}.$$

Now let Z be any random variable with finite support.

$$\begin{aligned}\mathbf{H}\{X \pm Y \mid Z\} &= \sum_{z \in G} \mathbf{P}\{Z = z\}\, \mathbf{H}\{X \pm Y \mid Z = z\} \\ &\geq \Big(\max\big(\mathbf{H}\{X \mid Z\}, \mathbf{H}\{Y \mid Z\}\big) - \mathbf{I}\{X : Y \mid Z\}\Big) \sum_{z \in G} \mathbf{P}\{Z = z\} \\ &= \max\big(\mathbf{H}\{X \mid Z\}, \mathbf{H}\{Y \mid Z\}\big) - \mathbf{I}\{X : Y \mid Z\},\end{aligned}$$

which completes the proof. ∎

Corollary 10. *If X and Y are independent, then*

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

Proof. The mutual information $\mathbf{I}\{X : Y\}$ is zero whenever X and Y are independent. ∎

Entropic Ruzsa distance. In additive combinatorics, whenever we have two finite subsets A and B of the same abelian group, we can compute the Ruzsa distance

$$d(A, B) = \log \frac{|A - B|}{\sqrt{|A| \cdot |B|}}$$

between them. (This satisfies all the axioms of a metric except the one requiring $d(A, A) = 0$ for all sets A.)

The entropic analogue of the Ruzsa distance is defined as follows. For finitely supported random variables X and Y taking values in the same abelian group, we let X' and Y' be independent copies of X and Y, respectively, and define the *entropic Ruzsa distance* by

$$\mathbf{d}\{X, Y\} = \mathbf{H}\{X' - Y'\} - \frac{\mathbf{H}\{X'\}}{2} - \frac{\mathbf{H}\{Y'\}}{2}.$$

This definition only depends on the individual distributions of X and Y and does not require them to have the same sample space. Once again, we don't necessarily have $\mathbf{d}\{X, X\} = 0$, but we do have the triangle inequality, which shall now prove.

Proposition 11. *Let X, Y, and Z be random variables with finite support in the same abelian group. Then*

$$\mathbf{d}\{X, Z\} \leq \mathbf{d}\{X, Y\} + \mathbf{d}\{Y, Z\},$$

which is equivalent to

$$\mathbf{H}\{X' - Z'\} \leq \mathbf{H}\{X' - Y'\} + \mathbf{H}\{Y' - Z'\} - \mathbf{H}\{Y'\}$$

for X', Y', and Z' independent and distributed as X, Y, and Z, respectively.

Proof. That the two statements are equivalent is easily obtained by expanding the definition of entropic Ruzsa distance and cancelling some terms. So without loss of generality, we may assume that that X, Y, and Z are independent and just prove the second statement.

By submodularity, we have $\mathbf{I}\big[(X-Y:Z) \mid X-Z\big] \geq 0$, so

$$\begin{aligned} 0 &\leq \mathbf{I}\big[(X-Y:Z) \mid X-Z\big] \\ &\leq \mathbf{H}\{X-Y \mid X-Z\} + \mathbf{H}\{Z \mid X-Z\} - \mathbf{H}\{X-Y, Z \mid X-Z\} \\ &\leq \mathbf{H}\{X-Y, X-Z\} + \mathbf{H}\{Z, X-Z\} - \mathbf{H}\{X-Y, Z, X-Z\} - \mathbf{H}\{X-Z\}. \end{aligned} \tag{1}$$

Now, since the values $(x-y, x-z)$ taken by $(X-Y, X-Z)$ are in bijection with values $(x-z, y-z)$ taken by $(X-Z, Y-Z)$ via the map $(v,w) \mapsto (w, w-v)$, by the invariance axiom we have

$$\mathbf{H}\{X-Y, X-Z\} = \mathbf{H}\{X-Z, Y-Z\},$$

and

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

follows by submodularity. Similar invocations of the invariance axiom give

$$\mathbf{H}\{Z, X-Z\} = \mathbf{H}\{X, Z\}$$

and

$$\mathbf{H}\{X-Y, Z, X-Z\} = \mathbf{H}\{X, Y, Z\} = \mathbf{H}\{X, Z\} + \mathbf{H}\{Y\},$$

where in the latter statement the second equality follows from the fact that (X, Y) and Z are independent. Substituting these three inequalities into (1), we have

$$0 \leq \mathbf{H}\{X-Y\} + \mathbf{H}\{Y-Z\} + \mathbf{H}\{X, Z\} - \mathbf{H}\{X, Z\} + \mathbf{H}\{Y\} - \mathbf{H}\{X-Z\},$$

whence

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

which completes the proof. ∎

We also define a conditional version of the entropic Ruzsa distance. If X and Y are G-valued random variables with finite support and Z and W are any random variables with finite supports A and B respectively, then we define

$$\mathbf{d}\{X \mid Z; Y \mid W\} = \sum_{z \in A} \sum_{w \in B} \mathbf{P}\{Z=z\}\, \mathbf{P}\{W=w\}\, \mathbf{d}\big[(X \mid Z=z); (Y \mid W=w)\big].$$

If (X', Z') and (Y', W') are independent copies of (X, Z) and (Y, W) respectively, then this distance is also given by the formula

$$\mathbf{d}\{X \mid Z; Y \mid W\} = \mathbf{H}\{X' - Y' \mid Z', W'\} - \frac{\mathbf{H}\{X' \mid Z'\}}{2} - \frac{\mathbf{H}\{Y' \mid W'\}}{2}.$$

3. The Plünnecke–Ruzsa inequality

In additive combinatorics, one of the most useful sumset inequalities is the following.

Theorem A (*Plünnecke–Ruzsa inequality*). *Let A and B be finite subsets of an abelian group and suppose that $|A + B| \leq K|A|$ for some constant K. Then for any integers $r, s \geq 0$, not both zero, we have $|rB - sB| \leq K^{r+s}|A|$.* ∎

In this section we will develop an entropic analogue of this statement, in which sets are replaced by random variables of finite support and cardinality is replaced with the exponential of entropy. First, a technical lemma.

Lemma 12. *Let X, Y, and Z be independent random variables taking values in a common abelian group. Then*

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

Proof. By submodularity, the quantity $\mathbf{I}\{X : Z \mid X + Y + Z\}$ is nonnegative, so we have

$$\begin{aligned} 0 &\leq \mathbf{I}\{X : Z \mid X + Y + Z\} \\ &= \mathbf{H}\{X, X + Y + Z\} + \mathbf{H}\{Z, X + Y + Z\} \\ &\qquad - \mathbf{H}\{X, Z, X + Y + Z\} - \mathbf{H}\{X + Y + Z\}. \end{aligned}$$

Since X, Y, and Z are independent, we have

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

where in the first equality we use invariance. By similar reasoning we have

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

and

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

Plugging these three identities into the inequality above yields

$$\begin{aligned} 0 &\leq \mathbf{H}\{X\} + \mathbf{H}\{Y + Z\} + \mathbf{H}\{Z\} + \mathbf{H}\{X + Y\} \\ &\qquad - \mathbf{H}\{X\} - \mathbf{H}\{Y\} - \mathbf{H}\{Z\} - \mathbf{H}\{X + Y + Z\} \\ &= \mathbf{H}\{Y + Z\} + \mathbf{H}\{X + Y\} - \mathbf{H}\{Z\} - \mathbf{H}\{X + Y + Z\}, \end{aligned}$$

whence the claim follows upon rearranging. ∎

From here we are not far from proving the entropic Plünnecke–Ruzsa inequality, a result of T. Tao.

Theorem 13. *Let $X, Y_1, \ldots, Y_m$ be independent random variables of finite entropy taking values in an abelian group G, such that*

$$\mathbf{H}\{X + Y_i\} \leq \mathbf{H}\{X\} + \log K_i$$

for all $1 \le i \le m$ *and some scalars* $K_1, \dots, K_m \ge 1$. *Then*

$$\mathbf{H}\{X + Y_1 + \cdots + Y_m\} \le \mathbf{H}\{X\} + \log(K_1 \cdots K_m).$$

Proof. We prove the claim by induction on m. If $m = 1$, then we are done by hypothesis. Now suppose that $\mathbf{H}\{X+Y_1+\cdots+Y_{m-1}\} \le \mathbf{H}\{X\}+\log(K_1 \cdots K_{m-1})$. Then by the previous lemma, the induction hypothesis, and the hypothesis on $\mathbf{H}\{X + Y_m\}$, we have

$$\begin{aligned}\mathbf{H}\{Y_1 + \cdots + Y_{m-1} + X + Y_m\} &\le \mathbf{H}\{Y_1 + \cdots + Y_{m-1} + X\} \\ &\qquad + \mathbf{H}\{X + Y_m\} - \mathbf{H}\{X\} \\ &\le \mathbf{H}\{X\} + \log(K_1 \cdots K_{m-1}) + \log K_m \\ &\le \mathbf{H}\{X\} + \log(K_1 \cdots K_m),\end{aligned}$$

which is what we sought to prove. ∎

We can make this look bit more like the version of the Plünnecke–Ruzsa inequality above by using the triangle inequality.

Corollary 14 (*Entropic Plunnecke–Ruzsa inequality*). *Let* X *and* Y *be random variables with* $\mathbf{H}\{X+Y\} \le \mathbf{H}\{X\}+\log K$. *Then for any* $r, s \ge 0$ *not both zero, we have*

$$\mathbf{H}\{Y_1 + \cdots + Y_r - Z_1 - \cdots - Z_s\} \le \mathbf{H}\{X\} + (r+s)\log K,$$

where $Y_1, \dots, Y_r, Z_1, \dots, Z_s$ *are independent copies of* Y.

Proof. By the entropic Ruzsa triangle inequality, we have

$$\begin{aligned}&\mathbf{H}\{Y_1 + \cdots + Y_r - Z_1 - \cdots - Z_s\} \le \\ &\qquad \mathbf{H}\{Y_1 + \cdots + Y_r + X\} + \mathbf{H}\{-X - Z_1 - \cdots - Z_s\} - \mathbf{H}\{-X\}.\end{aligned}$$

The values of $-X$ are in bijection with values of X, and the values of $-X - Z_1 - \cdots - Z_s$ are in bijection with the values of $X + Z_1 + \cdots + Z_s$ (with the same probabilities in both cases), so by the invariance axiom, we have

$$\begin{aligned}&\mathbf{H}\{Y_1 + \cdots + Y_r - Z_1 - \cdots - Z_s\} \le \\ &\qquad \mathbf{H}\{X + Y_1 + \cdots + Y_r\} + \mathbf{H}\{X + Z_1 + \cdots + Z_s\} - \mathbf{H}\{X\},\end{aligned}$$

and we can apply the the previous theorem twice to get

$$\begin{aligned}\mathbf{H}\{Y_1 + \cdots + Y_r - Z_1 - \cdots - Z_s\} &\le \mathbf{H}\{X\} + \log(K^r) + \log(K^s), \\ &= \mathbf{H}\{X\} + (r+s)\log K. \quad ∎\end{aligned}$$

References