Upper bounds for entropies of sums, and ramifications

Mokshay Madiman

University of Delaware and Yale University

Includes joint work with Ioannis Kontoyiannis, Athens U. of E.&B.
Adam Marcus, Yale
and Prasad Tetali, Georgia Tech

EII, CUHK

16 April 2013

Outline

- Motivation: Additive combinatorics, Network information theory
- Background: Entropy, submodularity, hypergraphs

PART I: *Inequalities for Discrete Entropy and Cardinalities*

- Entropy inequalities for sums and differences
- Entropy and sumset cardinality inequalities for P-D functions

PART II: *Inequalities for Continuous Entropy*

- Why is the continuous case different?
- (Continuous) Entropy inequalities for sums and differences

Motivation I: Network information theory

Inequalities for entropies of sums have long played an important role in information theory

The key example is the entropy power inequality (1948/59) and its many variants. Shannon used it to explore the capacity of additive noise channels; subsequently it has found particular use in network settings (e.g., broadcast channel)

Such inequalities also play a key role in understanding the central limit theorem using entropy, and are thus of significant interest in probability theory

The focus of our talk is upper bounds rather than lower bounds on entropies of sums; these are starting to find significant applications (e.g., the recent work of Wu-Shamai-Verdú '12 building on Etkin-Ordentlich '09 on the degrees of freedom of the interference channel)

Motivation II: The additive side of number theory

A lot of modern problems in number theory have to do with inherently "additive structure". E.g.:

- van der Corput's theorem (1939):
  The set of prime numbers contains infinitely many arithmetic progressions (AP's) of size 3

- Szemerédi's theorem (1975):
  Any set A of integers such that
  $$\limsup_{n\to\infty} \frac{|A \cap \{1, \ldots, n\}|}{n} > 0$$
  contains an AP of length k, for all $k \geq 2$

- Green-Tao theorem (2008):
  For each $k \geq 2$, the set of prime numbers contains an arithmetic progression of length k

Additive combinatorics

In all three results above, the problem is to count the number of occurrences of a certain $additive$ pattern in a given set

Classical "multiplicative" combinatorial results are insufficient for these purposes

The theory of additive combinatorics, and in particular the so-called $sumset$ $inequalities$, provides a set of very effective tools

Sumset inequalities

- "sumset" $A + B = \{a + b : a \in A, b \in B\}$, where A, B are finite sets in some group G
- "sumset inequality": inequalities for the cardinalities of sumsets under a variety of conditions

Simplest (trivial) example of a sumset inequality:
For any discrete subset A of an additive group $(G, +)$ [WLOG think of $G = \mathbb{Z}$],

$$|A| \leq |A + A| \leq |A|^2$$

Classical Sumset inequalities

Examples from the Plünnecke-Ruzsa (direct) theory

- Ruzsa triangle inequality

$$|A-C| \leq \frac{|A-B| \cdot |B-C|}{|B|}$$

- Plünnecke-Ruzsa inequality: Although it is not true in general that

$$|A+B+C| \cdot |B| \leq |A+B| \cdot |B+C|,$$

  it is true under appropriate conditions on the pair (A, B)

There is also the so-called Freiman or inverse theory, which deduces structural information about sets from the fact that their sumset is small. We will not discuss this much today

Combinatorics and Entropy

Natural connection: For a finite set A,

$$H(\mathsf{Unif}(A)) = \log |A|$$

is the maximum entropy of any distribution supported on A

Applications of entropy in combinatorics

- Intersection families [Chung-Graham-Frankl-Shearer '86]
- New proof of Bregman's theorem, etc. [Radhakrishnan '97-'03]
- Various counting problems [Kahn '01, Friedgut-Kahn '98, Brightwell-Tetali '03, Galvin-Tetali '04, M.-Tetali '07, Johnson-Kontoyiannis-M.'09]

Entropy in Additive Combinatorics?

Natural question: Can sumset inequalities be derived via entropy inequalities? Even more interestingly, are sumset inequalities special cases of entropy inequalities for sums of group-valued discrete random variables?

The answer to this question was developed by Ruzsa '09, M.-Marcus-Tetali '09, and Tao '10 in the discrete setting, and partially generalized to continuous settings by Kontoyiannis-M.'12, '13

Fact: Doubling and difference constants (sets)

Let A and B be arbitrary subsets of the integers (or discrete subsets of any commutative group).

A classical inequality in additive combinatorics

The difference set $A - B = \{a - b : a \in A, b \in B\}$

Define the *doubling constant* of A by

$$\sigma[A] = \frac{|A + A|}{|A|}$$

and the *difference constant* of A by

$$\delta[A] = \frac{|A - A|}{|A|}.$$

Then $$\delta[A]^{\frac{1}{2}} \leq \sigma[A] \leq \delta[A]^2$$

May be rewritten as

$$\frac{1}{2}\big[\log|A - A| - \log|A|\big] \leq \log|A + A| - \log|A| \leq 2\big[\log|A - A| - \log|A|\big]$$

Question: Doubling and difference constants (RV's)

Formal translation procedure

- Replace discrete sets by independent discrete random variables
- Replace the log-cardinality of a set by the discrete entropy function

Translation of the previous inequality

Let Y, Y' be i.i.d. discrete random variables. Define the *doubling constant* of Y by

$$\sigma_+(Y) = H(Y + Y') - H(Y)$$

and the *difference constant* of Y by

$$\sigma_-(Y) = H(Y - Y') - H(Y)$$

where $H(\cdot)$ denotes the discrete entropy function. Then the entropy analog of the doubling–difference sumset inequality is

$$\frac{1}{2}\sigma_-(Y) \leq \sigma_+(Y) \leq 2\sigma_-(Y)$$

Is this true?

mile-marker

- $\surd$ Motivation: Additive combinatorics, Network information theory
- Background: Entropy, submodularity, hypergraphs

PART I: *Inequalities for Discrete Entropy and Cardinalities*

- Entropy inequalities for sums and differences
- Entropy and sumset cardinality inequalities for P-D functions

PART II: *Inequalities for Continuous Entropy*

- Why is the continuous case different?
- (Continuous) Entropy inequalities for sums and differences

Entropy and Mutual Information

For random element X, entropy $\mathcal{H}(X) = \mathcal{H}(p) = E[-\log p(X)]$

Here if X is discrete, p is the p.m.f of X, and $\mathcal{H}$ is denoted H
and if X is continuous, p is the p.d.f of X, and $\mathcal{H}$ is denoted h

Conditional entropy of X given Y is

$$H(X|Y) = \sum_y p_Y(y) H(X|Y=y) \quad , \quad h(X|Y) = \int h(X|Y=y) p_Y(y) dy$$

where the term in the integrand is the entropy of $p(x|Y=y)$.

Mutual Information

- The *mutual information*

$$I(X;Y) = E\left[\log \frac{p_{X,Y}(X,Y)}{p_X(X)\, p_Y(Y)}\right]$$

represents the information shared between X and Y; it is non-negative, and symmetric in X and Y

- One has

$$I(X;Y) = \mathcal{H}(X) - \mathcal{H}(X|Y) = \mathcal{H}(Y) - \mathcal{H}(Y|X)$$

Three Useful Facts about Entropy

- Shannon's Chain Rule:

$$\mathcal{H}(X,Y) = \mathcal{H}(Y) + \mathcal{H}(X|Y)$$

- The *conditional mutual information* $I(X;Y|Z)$ represents the information shared between X and Y given that Z is already known; since it is non-negative and can be written as

$$I(X;Y|Z) = \mathcal{H}(X|Z) - \mathcal{H}(X|Y,Z),$$

consequently $\mathcal{H}(X|Z) \geq \mathcal{H}(X|Y,Z)$ ("conditioning reduces entropy")

- Things that we can rely on *only* in the discrete case:
  - $H(X|Y) \geq 0$ and $H(X) \geq 0$
  - $H(X|Y) = 0$ if and only if X is a function of Y

Consequences: A plethora of entropy inequalities

Hypergraphs and fractional subadditivity

- $[n]$ is the index set $\{1, 2, \ldots, n\}$
- A collection $\mathcal{C}$ of subsets of $[n]$ is a *hypergraph*, and the sets in $\mathcal{C}$ are hyperedges. E.g.: $\big\{\{1,2\},\{1,3\},\{2,3,4\}\big\}$
- For any index i in $[n]$, define the *degree* of i in $\mathcal{C}$ as

$$r(i) = |\{\mathbf{t} \in \mathcal{C} : i \in \mathbf{t}\}|$$

  The hypergraph $\mathcal{C}$ is *r-regular* if each i has the same degree r

Remarks

- A set function $g : 2^{[n]} \to \mathbb{R}$ is submodular if

$$g(\mathbf{s} \cup \mathbf{t}) + g(\mathbf{s} \cap \mathbf{t}) \leq g(\mathbf{s}) + g(\mathbf{t}) \text{ for any } \mathbf{s}, \mathbf{t}$$

- Lemma FSA: If a set function f is submodular, and $f(\phi) = 0$, then f is *fractionally subadditive*, i.e., for any r-regular hypergraph $\mathcal{C}$,

$$g([n]) \leq \frac{1}{r} \sum_{\mathbf{s} \in \mathcal{C}} g(\mathbf{s})$$

- Most general formulation with fractional coverings/packings and conditioning in M.-Tetali '10 (cf. Bondareva '63, Shapley '67-'71, Ollagnier-Pinchon '82)

Key example: Entropy of joint distributions

Submodularity of joint entropy:

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

Implications

We wish to consider various subsets of the random variables $X_1, \ldots, X_n$. For any $\mathbf{s} \in \mathcal{C}$, let $X_{\mathbf{s}}$ stand for the collection of random variables $(X_i : i \in \mathbf{s})$, with the indices taken in their increasing order.

- For any $\mathbf{s}, \mathbf{t} \subset [n]$,

$$H(X_{\mathbf{s}\cup\mathbf{t}}) + H(X_{\mathbf{s}\cap\mathbf{t}}) \leq H(X_{\mathbf{s}}) + H(X_{\mathbf{t}})$$

- (Shearer's inequality) For any r-regular hypergraph on $[n]$,

$$H(X_{[n]}) \leq \frac{1}{r} \sum_{\mathbf{s}\in\mathcal{C}} H(X_{\mathbf{s}})$$

- These are *nicely structured* Shannon-type inequalities that may be of use in the study of the entropic region

mile-marker

$\surd$ Motivation: Additive combinatorics, Network information theory

$\surd$ Background: Entropy, submodularity, hypergraphs

PART I: *Inequalities for Discrete Entropy and Cardinalities*

- Entropy inequalities for sums and differences
- Entropy and sumset cardinality inequalities for P-D functions

PART II: *Inequalities for Continuous Entropy*

- Why is the continuous case different?
- (Continuous) Entropy inequalities for sums and differences

Illustrative inequalities

The Inequalities

- Let $Z_1, \ldots, Z_n$ be independent discrete random variables taking values in the abelian group G. For any r–regular hypergraph $\mathcal{C}$ on $[n]$,

$$\text{ENT}: \qquad H(Z_1 + \cdots + Z_n) \leq \frac{1}{r} \sum_{\mathbf{S} \in \mathcal{C}} H\left(\sum_{i \in \mathbf{S}} Z_i \right)$$

- Let $X_1, \ldots, X_n$ be finite subsets of the abelian group G. For any r–regular hypergraph $\mathcal{C}$ on $[n]$,

$$\text{CARD}: \qquad |X_1 + \cdots + X_n| \leq \prod_{\mathbf{S} \in \mathcal{C}} \left| \sum_{i \in \mathbf{S}} X_i \right|^{\frac{1}{r}}$$

Remarks

- CARD implies trivially the known fact (e.g., Nathanson '96) that $|kA|^{1/k}$ is a non-increasing sequence
- CARD was conjectured by Gyarmati-Matolcsi-Ruzsa '10 and proved independently by Gyarmati-Matolcsi-Ruzsa '08, Balister-Bollobás '07, M.-Marcus-Tetali '10
- ENT was proved by M.-Marcus-Tetali '10

Relation between ENT and CARD

Let $X_1, \ldots, X_n \subset G$ be finite sets,
and $Z_1, \ldots, Z_n$ be RVs supported on $X_1, \ldots, X_n$ respy.

Note: For any $\mathbf{s} \subset [n]$, $Z^+_{\mathbf{s}} := \sum_{i \in \mathbf{s}} Z_i$ is supported on $X^+_{\mathbf{s}} := \sum_{i \in \mathbf{s}} X_i$

Remarks

- We can bound both the LHS and RHS of ENT:

$$H(Z^+_{[n]}) \overset{(a)}{\leq} \log \left| X^+_{[n]} \right|$$

and

$$\frac{1}{r} \sum_{\mathbf{s} \in \mathcal{C}} H\left(Z^+_{\mathbf{s}}\right) \leq \frac{1}{r} \sum_{\mathbf{s} \in \mathcal{C}} \log \left| X^+_{\mathbf{s}} \right|$$

  CARD says that these upper bounds themselves are ordered

- ENT would imply CARD immediately if:
  1. (a) was an equality [i.e., if the distribution $Z_{[n]}$ on $X_1 \times \ldots \times X_n$ makes sum $Z^+_{[n]}$ uniformly distributed on its range], AND
  2. ENT held [i.e., this distribution for $Z_{[n]}$ is a product distribution]

  In general, this is not possible

Message: no immediate implication between ENT and CARD

Basic notions for a general new framework

- *Compound sets*: For subsets $X_1, \ldots, X_k$ of some ambient space $\mathcal{X}$, consider

$$f(X_1, \ldots, X_k) = \{f(x_1, \ldots, x_k) : x_1 \in X_1, \ldots, x_k \in X_k\}.$$

  When the ambient space is a group, the only operation available is the sum, and all compound sets are sumsets. When the ambient space is a ring, one may consider compound sets built from polynomials.

- *Partition-determined function*: For what kind of functions f (with variable number of arguments) can one relate the entropy/cardinality of compound sets?

  Idea: Key fact for sums was that for fixed a, the sum $a+b$ depends only on b

Partition-determined functions

Let X_i be finite sets, and $X_{\mathbf{S}} = \prod_{i \in \mathbf{S}} X_i$ for nonempty $\mathbf{s} \subset [n]$ (assume that the indices i_j are labeled in increasing order for clarity). Let

$$Q(X_1, X_2, \ldots, X_n) = \bigsqcup_{\phi \neq \mathbf{S} \subset [n]} X_{\mathbf{S}}$$

For any $\mathbf{s} \subset [n]$, the projection function $\pi_{\mathbf{S}} : X_{[n]} \to X_{\mathbf{S}}$ is defined by $\pi_{\mathbf{S}}(x) = (x_{i_1}, \ldots, x_{i_{|\mathbf{S}|}})$ where $i_j \in \mathbf{s}$

Definitions

- $f : Q(X_1, \ldots, X_n) \to Y$ is partition-determined if for all $x, y \in X_{[n]}$, $f(\pi_{\mathbf{S}}(x)) = f(\pi_{\mathbf{S}}(y))$ and $f(\pi_{\mathbf{S}^c}(x)) = f(\pi_{\mathbf{S}^c}(y))$ for any $\mathbf{s} \subset [n]$ implies $f(x) = f(y)$
- $f : Q(X_1, \ldots, X_n) \to Y$ is strongly partition-determined if for any $x \in X_{[n]}$ and for any $\mathbf{s} \subset [n]$, any two of the objects $f(x)$, $f(\pi_{\mathbf{S}}(x))$ and $f(\pi_{\mathbf{S}^c}(x))$ determine the third

Partition-determined functions (II)

Let $n = 5$, and $\mathbf{s} = \{1, 3, 5\} \subset [5]$

If	$x = (x_1, x_2, x_3, x_4, x_5)$	$\in X_{[n]} \subset Q(X_1, \ldots, X_n),$
then	$x_{\mathbf{S}} = (x_1, \quad x_3, \quad x_5)$	$\in X_{\mathbf{S}} \subset Q(X_1, \ldots, X_n)$
and	$x_{\mathbf{S}^c} = (\quad x_2, \quad x_4 \quad)$	$\in X_{\mathbf{S}^c} \subset Q(X_1, \ldots, X_n)$

Reprising Definitions

- If $f : Q(X_1, \ldots, X_n) \to Y$ is P-D, then $f(x)$ is determined by $f(x_{\mathbf{S}})$ and $f(x_{\mathbf{S}^c})$
- If $f : Q(X_1, \ldots, X_n) \to Y$ is strongly P-D, then any two of the objects $f(x)$, $f(x_{\mathbf{S}})$ and $f(x_{\mathbf{S}^c})$ determine the third

Running Examples

- Identity function (i.e., $f(x_{\mathbf{S}}) = x_{\mathbf{S}}$) is strongly P-D
- Sum function in an abelian group (i.e., $f(x_{\mathbf{S}}) = \sum_{i \in \mathbf{S}} x_i$) is strongly P-D

Submodularity for Strongly P-D Functions

Theorem: [Submodularity of Entropy for Strongly P-D Functions]

Let X_i be finite sets, and $f : Q(X_1, \ldots, X_n) \to Y$ be strongly P-D. Let $Z_1, \ldots, Z_n$ (taking values in $X_1, \ldots, X_n$) be independent RVs. Then, writing $f_{\mathbf{S}} = f(\pi_{\mathbf{S}}(Z_1, \ldots, Z_n))$,

$$\text{GEN-ENT}: \qquad H(f_{\mathbf{S}\cup\mathbf{T}}) + H(f_{\mathbf{S}\cap\mathbf{T}}) \leq H(f_{\mathbf{S}}) + H(f_{\mathbf{T}})$$

for any nonempty subsets $\mathbf{S}$ and $\mathbf{T}$ of $[n]$

Corollary: For any r-regular hypergraph $\mathcal{C}$ on $[n]$,

$$H(f_{[n]}) \leq \frac{1}{r} \sum_{\mathbf{S}\in\mathcal{C}} H\big(f_{\mathbf{S}}\big)$$

Submodularity for Sums

Corollary: [Submodularity of Entropy for Sums]

If Z_i are independent discrete RVs taking values in an abelian group, and $Z_{\mathbf{S}}^+ = \sum_{i \in \mathbf{S}} Z_i$, then $g(\mathbf{s}) = H(Z_{\mathbf{S}}^+)$ is submodular. That is:

$$H(Z_1 + Z_2 + Z_3) + H(Z_1) \leq H(Z_1 + Z_2) + H(Z_1 + Z_3)$$

Corollary: For any r-regular hypergraph $\mathcal{C}$ on $[n]$,

$$H(Z_{[n]}^+) \leq \frac{1}{r} \sum_{\mathbf{S} \in \mathcal{C}} H(Z_{\mathbf{S}}^+)$$

Remarks

- The second statement follows from the first by Lemma FSA
- The first statement is implicit in Kaĭmanovich-Vershik '83 as pointed out by Tao '09; both corollaries explicit and generalized to continuous settings in M.'08

On the proof of GEN-ENT (I)

Lemma A: ("Data processing inequality") The mutual information cannot increase when one looks at functions of the random variables:

$$I(g(Z);Y) \leq I(Z;Y).$$

Lemma B: If f is strongly partition-determined and $Z_1, \ldots, Z_n$ are independent, then for disjoint sets $\mathbf{s}, \mathbf{t} \subset [n]$,

$$I(f_{\mathbf{s}\cup\mathbf{t}}; f_{\mathbf{t}}) = H(f_{\mathbf{s}\cup\mathbf{t}}) - H(f_{\mathbf{s}}). \tag{1}$$

Proof

Since conditioning reduces entropy,

$$\begin{aligned} H(f_{\mathbf{s}\cup\mathbf{t}}) - H(f_{\mathbf{s}}) &= H(f_{\mathbf{s}\cup\mathbf{t}}) - H(f_{\mathbf{s}}|f_{\mathbf{t}}) && \text{[independence of } Z_i\text{]} \\ &= H(f_{\mathbf{s}\cup\mathbf{t}}) - H(f_{\mathbf{s}\cup\mathbf{t}}|f_{\mathbf{t}}) && \text{[}f\text{ strongly P-D]} \\ &= I(f_{\mathbf{s}\cup\mathbf{t}}; f_{\mathbf{t}}) \end{aligned}$$

On the proof of GEN-ENT (II)

Suffices to prove the result for $n = 3$, i.e., we want to show

$$H(f_{\{1,2\}}) + H(f_{\{2,3\}}) \geq H(f_{\{1,2,3\}}) - H(f_{\{2\}})$$

Now, using Lemma B,

$$\begin{aligned} & H(f_{\{1,2\}}) + H(f_{\{2,3\}}) - H(f_{\{1,2,3\}}) - H(f_{\{2\}}) \\ & = H(f_{\{1,2\}}) - H(f_{\{2\}}) - \left[H(f_{\{1,2,3\}}) - H(f_{\{2,3\}})\right] \\ & = I(f_{\{1,2\}}; f_{\{1\}}) - I(f_{\{1,2,3\}}; f_{\{1\}}) \end{aligned}$$

But

$$\begin{aligned} I(f_{\{1,2,3\}}; f_{\{1\}}) & \leq I(f_{\{1,2\}}, f_{\{3\}}\,; f_{\{1\}}) && \text{[Lemma A]} \\ & = I(f_{\{1,2\}}; f_{\{1\}}) + I(f_{\{3\}}; f_{\{1\}} | f_{\{1,2\}}) && \text{[“chain rule”]} \\ & = I(f_{\{1,2\}}; f_{\{1\}}) && \text{[independence]} \end{aligned}$$

Thus

$$I(f_{\{1,2\}}; f_{\{1\}}) \geq I(f_{\{1,2,3\}}; f_{\{1\}})$$

and the proof is complete

Half of the doubling–difference inequality

Goal: If X, Z are i.i.d.,

$$H(X+Z) - H(X) \leq 2[H(X-Z) - H(X)]$$

Proof

If X, Y, Z are independent random variables, then

$$H(X+Y+Z) + H(Y) \leq H(X+Y) + H(Z+Y)$$

Since $H(X+Z) \leq H(X+Y+Z)$,

$$H(X+Z) + H(Y) \leq H(X+Y) + H(Z+Y)$$

Taking $X, -Y$ and Z i.i.d.,

$$H(X+Z) + H(X) \leq 2H(X-Z)$$

which is the required upper bound

Functional Submodularity

If $X_0 = F(X_1) = G(X_2)$ and $X_{12} = R(X_1, X_2)$, then

$$H(X_{12}) + H(X_0) \leq H(X_1) + H(X_2)$$

Proof

By data processing for mutual information and entropy,

$$\begin{aligned} H(X_1) + H(X_2) - H(X_{12}) &\geq H(X_1) + H(X_2) - H(X_1, X_2) \\ &= I(X_1; X_2) \\ &\geq I(X_0; X_0) \\ &= H(X_0) \end{aligned}$$

Note: Does **not** hold for differential entropy h

The other half: discrete case

Goal: [Implicit in Ruzsa '09, Tao–Vu '06] If X, Y are i.i.d.,

$$H(X-Y) - H(X) \leq 2[H(X+Y) - H(X)]$$

Proof

By functional submodularity,

$$H(X,Y,Z) + H(X-Z) \leq H(X-Y, Y-Z) + H(X,Z).$$

Rearranging and using independence,

$$H(X-Z) \leq H(X-Y, Y-Z) - H(Y)$$

Thus one obtains the Ruzsa triangle inequality:

$H(X-Z) \leq H(X-Y) + H(Y-Z) - H(Y)$ [Ruzsa '09, Tao–Vu '06, Tao '09]

Replacing Y by $-Y$ and noting that $H(W) = H(-W)$ for any W,

$$H(X-Z) + H(Y) \leq H(X+Y) + H(Y+Z)$$

Letting X, Z have the same distribution as Y gives

$$H(X-Y) + H(X) \leq 2H(X+Y),$$

which is the desired result

mile-marker

$\surd$ Motivation: Additive combinatorics, Network information theory

$\surd$ Background: Entropy, submodularity, hypergraphs

PART I: *Inequalities for Discrete Entropy and Cardinalities*

$\surd$ Entropy inequalities for sums and differences

$\surd$ Entropy and sumset cardinality inequalities for P-D functions

PART II: *Inequalities for Continuous Entropy*

- Why is the continuous case different?
- (Continuous) Entropy inequalities for sums and differences

A Unified Setting

Let $\mathcal{G}$ be a Hausdorff topological group that is abelian and locally compact, and λ be a Haar measure on $\mathcal{G}$. If $\mu \ll \lambda$ is a probability measure on $\mathcal{G}$, the entropy of $X \sim \mu$ is defined by

$$h(X) = -\int \frac{d\mu}{d\lambda}(x) \log \frac{d\mu}{d\lambda}(x)\, \lambda(dx)$$

Remarks

- In general, $h(X)$ may or may not exist; if it does, it takes values in the extended real line $[-\infty, +\infty]$
- If $\mathcal{G}$ is compact and λ is the Haar ("uniform") probability measure on $\mathcal{G}$, then $h(X) = -D(\mu\|\lambda) \leq 0$ for every RV X
- Covers both the classical cases: $\mathcal{G}$ discrete with counting measure, and $\mathcal{G} = \mathbb{R}^n$ with Lebesgue measure

A Question and an Answer

Setup: Let Y and Y' be i.i.d. random variables (continuous, with density f). As usual, the differential entropy is $h(Y) = E[-\log f(Y)]$

Question

How different can $h(Y+Y')$ and $h(Y-Y')$ be?

First answer [Lapidoth–Pete '08]

The entropies of the sum and difference of two i.i.d. random variables *can differ by an arbitrarily large amount*

Precise formulation: Given any $M > 0$, there exist i.i.d. random variables Y, Y' of finite differential entropy, such that

$$h(Y-Y') - h(Y+Y') > M \qquad \text{(Ans. 1)}$$

A Question and another Answer

Question

If Y and Y' are i.i.d. continuous random variables, how different can $h(Y+Y')$ and $h(Y-Y')$ be?

Our answer [Kontoyiannis–M.'12]

The entropies of the sum and difference of two i.i.d. random variables *are not too different*

Precise formulation: For any two i.i.d. random variables Y, Y' with finite differential entropy:

$$\frac{1}{2} \leq \frac{h(Y+Y') - h(Y)}{h(Y-Y') - h(Y)} \leq 2 \qquad \text{(Ans. 2)}$$

What do the two Answers tell us?

Together, they suggests that the natural quantities to consider are the differences

$$\Delta_+ = h(Y + Y') - h(Y) \quad \text{and} \quad \Delta_- = h(Y - Y') - h(Y)$$

Then (Ans. 1) states that the *difference* $\Delta_+ - \Delta_-$ can be arbitrarily large, while (Ans. 2) asserts that the *ratio* Δ_+/Δ_- must always lie between $\frac{1}{2}$ and 2

Why is this interesting?

- Seems rather intriguing in its own right
- Observe that Δ_+ and Δ_- are affine-invariant; so these facts are related to the *shape* of the density
- This statement for *discrete* random variables (one half of which follows from [Ruzsa '09, Tao '10], and the other half of which follows from [M.-Marcus-Tetali '12]) is the exact analogue of the inequality relating doubling and difference constants of sets in additive combinatorics
- This and possible extensions may be relevant for studies of "polarization" phenomena and/or interference alignment in information theory

Half the proof

Want to show: If Y, Y' are i.i.d.,

$$h(Y + Y') - h(Y) \leq 2[h(Y - Y') - h(Y)]$$

Proof: If Y, Y', Z are independent random variables, then the Submodularity Lemma says

$$h(Y + Y' + Z) + h(Z) \leq h(Y + Z) + h(Y' + Z) \quad \text{[M. '08]}$$

Since $h(Y + Y') \leq h(Y + Y' + Z)$,

$$h(Y + Y') + h(Z) \leq h(Y + Z) + h(Y' + Z) \qquad (2)$$

Taking now Y, Y' to be i.i.d. and Z to be an independent copy of $-Y$,

$$h(Y + Y') + h(Y) \leq 2h(Y - Y')$$

which is the required upper bound

Remark: The other half would follow similarly if we could prove the following slight variant of (2):

$$h(Y - Y') + h(Z) \leq h(Y + Z) + h(Y' + Z)$$

This is the entropy analogue of the Ruzsa triangle inequality and is a bit more intricate to prove

The Submodularity Lemma

Given independent $\mathcal{G}$-valued RVs X_1, X_2, X_3 with finite entropies,

$$h(X_1 + X_2 + X_3) + h(X_2) \leq h(X_1 + X_2) + h(X_3 + X_2)$$

Remarks

- For discrete groups, the Lemma is implicit in Kaĭmanovich-Vershik '83, but was rediscovered and significantly generalized by M.-Marcus-Tetali '12 en route to proving some conjectures of Ruzsa
- Discrete entropy is subadditive; trivially,

  $$H(X_1 + X_2) \leq H(X_1, X_2) \leq H(X_1) + H(X_2)$$

  This corresponds to putting $X_2 = 0$ in discrete form of the Lemma
- Continuous entropy is not subadditive; it is easy to construct examples with

  $$h(X_1 + X_2) > h(X_1) + h(X_2)$$

  Note that putting $X_2 = 0$ in the Lemma is no help since $h(\text{const.}) = -\infty$

Proof of Submodularity Lemma

Lemma A: ("Data processing inequality") The mutual information cannot increase when one looks at functions of the random variables:

$$I(g(Z); Y) \leq I(Z; Y).$$

Lemma B: If X_i are independent RVs, then

$$I(X_1 + X_2; X_1) = H(X_1 + X_2) - H(X_2).$$

Proof of Lemma B

Since conditioning reduces entropy,

$$\begin{aligned} h(X_1 + X_2) - h(X_2) &= h(X_1 + X_2) - h(X_2|X_1) && \text{[independence of } X_i\text{]} \\ &= h(X_1 + X_2) - h(X_1 + X_2|X_1) && \text{[translation-invariance]} \\ &= I(X_1 + X_2; X_1) \end{aligned}$$

Proof of Submodularity Lemma

$$I(X_1 + X_2 + X_3; X_1) \overset{(a)}{\leq} I(X_1 + X_2, X_3; X_1) \overset{(b)}{=} I(X_1 + X_2; X_1)$$

where (a) follows from Lemma A and (b) follows from independence

By Lemma B, this is the same as

$$h(X_1 + X_2 + X_3) + h(X_2) \leq h(X_1 + X_2) + h(X_2 + X_3)$$

Aside: Applications in Convex Geometry

Continuous Plünnecke-Ruzsa inequality: Let A and $B_1, \ldots, B_n$ be convex bodies in $\mathbf{R}^d$, such that for each i,

$$\left|A + B_i\right|^{\frac{1}{d}} \leq c_i |A|^{\frac{1}{d}}.$$

Then

$$\left|A + \sum_{i \in [n]} B_i\right|^{\frac{1}{d}} \leq \left[\prod_{i=1}^{n} c_i\right] |A|^{\frac{1}{d}}$$

The proof combines the Submodularity Lemma with certain reverse Hölder-type inequalities developed in [Bobkov-M.'12]

Reverse Entropy Power Inequality: The Submodularity Lemma is one ingredient (along with a deep theorem of V. Milman on the existence of "M-ellipsoids") used in Bobkov-M.'11, '12 to prove a reverse entropy power inequality for convex measures (generalizing the reverse Brunn-Minkowski inequality)

mile-marker

$\surd$ Motivation: Additive combinatorics, Network information theory

$\surd$ Background: Entropy, submodularity, hypergraphs

PART I: *Inequalities for Discrete Entropy and Cardinalities*

- Entropy inequalities for sums and differences
- Entropy and sumset cardinality inequalities for P-D functions

PART II: *Inequalities for Continuous Entropy*

- Why is the continuous case different?
- (Continuous) Entropy inequalities for sums and differences

Continuous analogue of Ruzsa triangle inequality

Goal: If X, Y, Z are independent,

$$h(X - Z) \leq h(X - Y) + h(Y - Z) - h(Y)$$

Proof

Note $$\text{RHS} \geq h(X - Y, Y - Z) + h(X, Z) - h(X, Y, Z)$$

But $$\begin{aligned} h(X, Y, Z) &= h(X - Y, Y - Z, X) \\ &= h(X - Y, Y - Z) + h(X | X - Y, Y - Z). \end{aligned}$$

so

$$\begin{aligned} \text{RHS} &\geq h(X, Z) - h(X | X - Y, Y - Z) \\ &= h(X) - h(X | X - Y, Y - Z) + h(Z) \\ &= I(X; X - Y, Y - Z) + h(Z) \\ &\geq I(X; X - Z) + h(Z) \\ &= h(X - Z) - h(X - Z | X) + h(Z) \\ &= h(X - Z) - h(-Z | X) + h(Z) \\ &= h(X - Z) \end{aligned}$$

Non-Gaussianity

For $X \sim f$ in $\mathbf{R}^n$, its relative entropy from Gaussianity is

$$D(X) = D(f) := D(f \| f^G),$$

where f^G is the Gaussian with the same mean and covar. matrix as X

Observe:

- For any density f, its non-Gaussianity $D(f) = h(f^G) - h(f)$

  Proof: Gaussian density is exponential in first two moments

- Thus Gaussian is MaxEnt: $N(0, \sigma^2)$ has maximum entropy among all densities on $\mathbb{R}$ with variance $\leq \sigma^2$

  Proof: $D(f) \geq 0$

Towards the Entropic CLT

Two observations . . .

- Gaussian is MaxEnt: $N(0, \sigma^2)$ has maximum entropy among all densities on $\mathbb{R}$ with variance $\leq \sigma^2$
- Let X_i be i.i.d. with $EX_1 = 0$ and $EX_1^2 = \sigma^2$.

  For the CLT, we are interested in $S_M := \frac{1}{\sqrt{M}} \sum_{i=1}^{M} X_i$

  The CLT scaling preserves variance

suggest . . .

Question: Is it possible that the CLT may be interpreted like the 2nd law of thermodynamics, in the sense that $h(S_M)$ monotonically increases in M until it hits the maximum entropy possible (namely, the entropy of the Gaussian)?

Entropic Central Limit Theorem

If $D(S_M) < \infty$ for some M, then as $M \to \infty$,

$$D(S_M) \downarrow 0 \qquad \text{or equivalently,} \qquad h(S_M) \uparrow h(N(0, \sigma^2))$$

Convergence shown by Barron '86; monotonicity shown by Artstein-Ball-Barthe-Naor '04 with simple proof by Barron–M.'07

Remarks

- The proof in Barron–M.'07 of a general inequality that implies monotonicity is a direct consequence of 3 ingredients:
  - An (almost) standard reduction to statements about Fisher information of sums
  - An integration-by-parts trick to reduce the desired Fisher information inequality to a variance inequality
  - A proof of the variance inequality, which generalizes Hoeffding's variance bounds for U-statistics
- *Question*: Can such a "2nd law" interpretation be given to other limit theorems in probability?
  Answer: Yes, but it is harder to do so, and the theory is incomplete

  E.g.: Partial results in the Compound Poisson case by [Johnson-Kontoyiannis-M.'09, Barbour-Johnson-Kontoyiannis-M.'10]

Original Entropy Power Inequality

If X_1 and X_2 are independent RVs,

$$e^{2h(X_1+X_2)} \geq e^{2h(X_1)} + e^{2h(X_2)} \quad \text{[Shannon '48, Stam '59]}$$

with equality if and only if both X_1 and X_2 are Gaussian

Remarks

- Implies the Gaussian logarithmic Sobolev inequality in 3 lines
- Implies Heisenberg's uncertainty principle (stated using Fourier transforms for unit vectors in $L_2(\mathbb{R}^n)$)
- Since $h(aX) = h(X) + \log|a|$, implies for i.i.d. X_i,

$$h\left(\frac{X_1 + X_2}{\sqrt{2}}\right) \geq h(X_1)$$

Thus we have monotonicity for doubling sample size: $h(S_{2n}) \geq h(S_n)$

An elementary observation

If X_i are independent,

$$\begin{aligned} h(X_1) + h(X_2) &= h(X_1, X_2) \\ &= h\left(\frac{X_1 + X_2}{\sqrt{2}}, \frac{X_1 - X_2}{\sqrt{2}}\right) \\ &\leq h\left(\frac{X_1 + X_2}{\sqrt{2}}\right) + h\left(\frac{X_1 - X_2}{\sqrt{2}}\right) \end{aligned}$$

When X_1 and X_2 are IID...

- If X_1 has a symmetric (even) density, this immediately yields $h(S_2) \geq h(S_1)$ in the CLT
- If $h(X_1 - X_2) < h(X_1 + X_2) - C$, then

$$h(Z) \geq h\left(\frac{X_1 + X_2}{\sqrt{2}}\right) > h(X_1) + \frac{C}{2}$$

  so that $D(X_1) > \frac{C}{2}$
- Thus any distribution of X for which $|h(X_1 - X_2) - h(X_1 + X_2)|$ is large must be far from Gaussianity

What does small doubling mean?

Let X be a $\mathbb{R}$-valued RV with finite (continuous) entropy and variance σ^2. The EPI implies $h(X+X')-h(X) \geq \frac{1}{2}\log 2$, with equality iff X is Gaussian

A (Conditional) Freiman theorem in $\mathbb{R}^n$

If X has finite Poincaré constant $R = R(X)$, and

$$h(X + X') - h(X) \leq \frac{1}{2}\log 2 + C, \tag{3}$$

then X is approximately Gaussian in the sense that

$$D(X) \leq \left(\frac{2R}{\sigma^2} + 1\right) C$$

Remarks

- Follows from a convergence rate result in the entropic CLT obtained independently by [Johnson-Barron '04] and [Artstein-Ball-Barthe-Naor '04]
- A construction of [Bobkov-Chistyakov-Götze '11] implies that in general such a result does not hold
- A *sufficient* condition for small doubling is log-concavity: in this case, $h(X + X') \leq h(X) + \log 2$ and $h(X - X') \leq h(X) + 1$

- There are still structural conclusions to be drawn just from (3)...

Summary

- Entropy of strongly partition-determined functions is submodular
- Recover and generalize numerous sumset and entropy inequalities in abelian and nonabelian groups (and resolve several conjectures of Ruzsa '09, Gyarmati-Matolcsi-Ruzsa '10)
- Continuous analogues have non-trivial consequences in convex geometry and probability
- *Broader message*: Entropy inequalities are promising tools:
  – in information theory, for communications applications – in additive combinatorics, as an efficient way of proving sumset inequalities
  – as a source of additional intuition (e.g., how discrete and continuous cases differ)

Thank you for your attention!

○ — ○ — ○

Extras

What changes for non-abelian groups?

- If A, B, C are finite subsets of a non-abelian group (with multiplication as the group operation), then:

  ABC need not have any relation with AC

  (OR)

  the (naive) product function $f(x_{\mathbf{s}}) = x_{i_1}x_{i_2}\ldots x_{i_k}$, where $i_1 < i_2 < \ldots < i_k$ are the elements of $\mathbf{s}$, is not P-D

- However, the additive combinatorics of non-abelian groups generally aims to use the fact that even though ABC need not have any relation with AC, ABC may be related to the sets AbC that one gets by running over $b \in B$

- $Question$: Can any of the preceding technology be applied to non-abelian groups?

CARD for non-abelian groups

Theorem 3: Let X_i be subsets of a non-abelian group, and define

$$A(i,j) = \max\{|X_i x_{i+1} \dots x_{j-1} X_j| : \quad x_{i+1} \in X_{i+1}, \dots, x_{j-1} \in X_{j-1}\}$$

for all $1 \le i < j \le n$. Then, for $n \ge 2$,

$$|X_1 X_2 \dots X_n|^{n-1} \le \prod_{1 \le i < j \le n} A(i,j)$$

Proof ideas

- *Non-trivial choice of P-D function*: If $\mathbf{s}$ is contiguous, of form $\{k, k+1, \dots, l\}$, then $f_{\mathbf{s}}(x) = x_k x_{k+1} \dots x_l$; otherwise $f_{\mathbf{s}}(x) = \pi_{\mathbf{s}}(x)$
- *New inequality for joint entropy*: Let Z_i be random variables, and define for all $i < j$, $\quad Z_{(i,j)} = \{Z_t : i < t < j\}$. Then

$$(n-1) H(Z_1, \dots, Z_n) \le \sum_{i=1}^{n} \sum_{j>i} H(Z_i, Z_j \mid Z_{(i,j)})$$

- Resolves partially a conjecture of Ruzsa '09, and gives hints about the general conjecture

Cardinality inequalities of Plünnecke-Ruzsa type

Let $A, B_1, \ldots, B_n$ be finite subsets of an abelian group, and assume $|A + B_{\mathbf{s}}^+| \le K_{\mathbf{s}}|A|$. Let $\mathcal{C}$ be any r-regular hypergraph on $[n]$.

1. Conjecture: There exists a nonempty set $A' \subset A$ such that
$$|A' + B_{[n]}^+| \le (\textstyle\prod_{\mathbf{s}\in\mathcal{C}} K_{\mathbf{s}})^{1/r}|A'|$$
2. Theorem 4: Set $c := |\mathcal{C}|/r$. For any $D \subseteq B_{[n]}^+$,
$$|A + D| \le (\textstyle\prod_{\mathbf{s}\in\mathcal{C}} K_{\mathbf{s}})^{1/|\mathcal{C}|}|D|^{1-1/c} \cdot |A|$$

Remarks

- (1) was proved for $\mathcal{C}_m$ (all m-sets) by Gyarmati, Matolcsi and Ruzsa '08. Specializing to $\mathcal{C}_1$ (singletons) and $n = 2$ gives the original Plünnecke-Ruzsa inequality for different summands:
$$|A' + B_1 + B_2| \le \left(\frac{|A + B_1|}{|A|}\right)\left(\frac{|A + B_2|}{|A|}\right)|A'| \qquad \text{[Ruzsa '89]}$$
- (2) was proved in M.-Marcus-Tetali '10, with special cases proved independently [$\mathcal{C}_1$ (singletons) by Gyarmati-Matolcsi-Ruzsa '10, Balister-Bollobás '07 and $\mathcal{C}_{n-1}$ (leave-one-out sets) by Gyarmati-Matolcsi-Ruzsa '08]

Entropic inequalities of Plünnecke-Ruzsa type

Theorem 5: Let $Z_0, Z_1, \ldots, Z_n$ be *independent* discrete RVs taking values in the abelian group G. Define the nonnegative constants $\beta(\mathbf{s})$ for each $\mathbf{s} \in \mathcal{C}$ by $H(Z_0 + Z_{\mathbf{s}}^+) = H(Z_0) + \beta(\mathbf{s})$. Then, with $c := |\mathcal{C}|/r$,

$$H(Z_0 + Z_{[n]}^+) \le H(Z_0) + \big(1 - \tfrac{1}{c}\big) H(Z_{[n]}^+) + \tfrac{1}{|\mathcal{C}|} \textstyle\sum_{\mathbf{s} \in \mathcal{C}} \beta(\mathbf{s})$$

$$H(Z_0 + Z_{[n]}^+) \le H(Z_0) + \tfrac{1}{r} \textstyle\sum_{\mathbf{s} \in \mathcal{C}} \beta(\mathbf{s})$$

Remarks

- These are the entropy analogues of the Conjecture and Theorem 4, but they are of the same form. On the other hand, there is an essential difference between the Conjecture and Theorem 4: the Conjecture can only be true for *some* $A' \subset A$ (one hopes not too small)
- Neither of these bounds seems *a priori* better than the other

Some reflections

- Q: Why can A' not be taken to be A in the original Plünnecke-Ruzsa inequality (and hence in the Conjecture)?

  Ans: Putting $A' = A$ yields

  $$|A + B_1 + B_2| \cdot |A| \leq |A + B_1| \cdot |A + B_2| \qquad (4)$$

  This is FALSE since log cardinality of sumsets is not submodular [Ruzsa '09], although from CARD, we know it is fractionally subadditive

- Q: Why can something be said with A' a nonempty subset of A?

  Ans: Recent work of Petridis '11 makes it clear that the heart of the Plünnecke-Ruzsa inequality is that although (4) is not true,

  $$|A' + B_1 + B_2| \cdot |A'| \leq |A' + B_1| \cdot |A' + B_2|$$

  holds for some subset A' of A

- The entropy of sums is submodular; so we have nicer entropic analogues

Entropic Central Limit Theorem?

Two observations . . .

- Gaussian is MaxEnt: $N(0, \sigma^2)$ has maximum entropy among all densities on $\mathbb{R}$ with variance $\leq \sigma^2$
- Let X_i be i.i.d. with $EX_1 = 0$ and $EX_1^2 = \sigma^2$.

  For the CLT, we are interested in $S_M := \dfrac{1}{\sqrt{M}} \displaystyle\sum_{i=1}^{M} X_i$

  The CLT scaling preserves variance

suggest . . .

Question: Is it possible that the CLT may be interpreted like the 2nd law of thermodynamics?

Specifically: Is it true that $h(S_M)$ monotonically increases in M until it hits the maximum entropy possible (namely, the entropy of the Gaussian)?

The Entropic Central Limit Theorem

If $D(S_M) < \infty$ for some M, then as $M \to \infty$,

$$D(S_M) \downarrow 0 \qquad \text{or equivalently,} \qquad h(S_M) \uparrow h(N(0, \sigma^2))$$

Remarks

- Convergence shown by Barron '86
- Monotonicity shown by Artstein-Ball-Barthe-Naor '04 with simple proofs by Barron–M. '06-'07, Tulino–Verdú '06
- Monotonicity in n indicates that the entropy is a *natural measure* for CLT convergence (cf. second law of thermodynamics)

Reverse Brunn-Minkowski inequality

Given two convex bodies A and B in $\mathbf{R}^d$, one can find an affine volume-preserving map $u : \mathbf{R}^d \to \mathbf{R}^d$ (i.e., $u \in SL_d(\mathbb{R})$) such that with some absolute constant C,

$$\left|\widetilde{A} + B\right|^{1/d} \leq C \left(|A|^{1/d} + |B|^{1/d}\right)$$

where $\widetilde{A} = u(A)$

Remarks

- Note that by the Brunn-Minkowski (BM) inequality, we always have

$$\left|\widetilde{A} + B\right|^{1/d} \geq |A|^{1/d} + |B|^{1/d}$$

- The Reverse BM Inequality was proved by [Milman '86], with other proofs in [Milman '88, Pisier '89]; all proofs use deep techniques from convex geometry and functional analysis
- Is there an entropic form of the Reverse BM Inequality, for random vectors, under some "convexity" assumption?

Reverse entropy power inequality

Theorem 8: If X and Y are independent $\mathbf{R}^d$-valued RVs, and have log-concave densities, then for some affine entropy-preserving map $u : \mathbf{R}^d \to \mathbf{R}^d$ (i.e., $u \in SL_d(\mathbb{R})$),

$$\mathcal{N}\left(\widetilde{X} + Y\right) \leq C\left(\mathcal{N}(X) + \mathcal{N}(Y)\right) \qquad \text{[Bobkov-M.'10]}$$

where $\widetilde{X} = u(X)$ and C is an absolute constant

Remarks

- *Entropy power inequality* (EPI): For any two independent random vectors X and Y in $\mathbf{R}^n$,

  $$\mathcal{N}(X+Y) \geq \mathcal{N}(X) + \mathcal{N}(Y) \quad \text{[Shannon '48, Stam '59]}$$

  If $X \sim$Unif(A) for a convex body $A \subset \mathbf{R}^n$, $\mathcal{N}(X) = |A|^{2/n}$. Even though this is not an exact correspondence, BM and EPI are closely related
- Reverse BM inequality can be recovered as a special case of Theorem 8
- Theorem 8 can be generalized to the larger class of "convex measures"