Entropy inequalities for sums and differences, and their relationship to limit theorems

Mokshay Madiman

University of Delaware / Yale University

Based on joint work with Andrew Barron, Yale University
Sergey Bobkov, University of Minnesota
Ioannis Kontoyiannis, AUEB
Adam Marcus, Yale University
Prasad Tetali, Georgia Tech

IMI Conference on Limit Theorems, 9–11 January 2013

Outline

- Entropy inequalities and additive combinatorics
  - Background and Motivations
  - Basic question: Entropy of sum vs. entropy of difference

- Entropic Limit Theorems
  - Entropic CLT
  - Role of entropy power inequalities

- Towards some structural results

Some results from number theory

Many 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

Classical Sumset inequalities

Examples from the Plünnecke-Ruzsa (direct) theory

- Ruzsa triangle inequality

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

- Sum-difference inequality

$$|A + B| \leq \frac{|A - B|^3}{|A| \cdot |B|}$$

These are special cases of the Plünnecke-Ruzsa inequalities

Examples from the Freiman (inverse) theory

- The *Cauchy-Davenport inequality* says that

$$|A + B| \geq |A| + |B| - 1$$

  with equality iff A and B are AP's

- The *Freiman theory* provides structural (inverse sumset) results
  E.g.: if $|A + A|$ is not too large relative to $|A|$, then A is "close" to a "generalized AP"

Combinatorics and Entropy

Discrete entropy: For probability mass function $p(\cdot)$ on a countable set A, entropy $H(p) = -\sum_{x \in A} p(x) \log p(x)$

Natural connection: For a finite set A, $H(\text{Unif}(A)) = \log |A|$ is the maximum entropy of any distribution supported on A

Entropy in Classical 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

- Ruzsa '09 (pioneered this approach, formulated basic questions)
- M.-Marcus-Tetali '10, '12 (entropic "direct" theory, including Plünnecke-Ruzsa inequalities)
- Tao '10 (entropic "inverse" theory, including Freiman's theorem)

Our Goal

So far, "entropy theory" in additive combinatorics has been focused on discrete abelian groups. Can we develop a theory that makes sense also in continuous settings, e.g., $\mathbb{R}^n$?

Why should we care?

- Probability: Related to basic questions. E.g.: rate of convergence in the (entropic) CLT
- Additive combinatorics: A thriving field in which discrete abelian groups have been well studied, but entropy techniques may be useful in more general settings that are under active investigation
- Convex geometry: Has fascinating unsolved problems that connect to high-dimensional probability and functional analysis. Understanding the entropy of sums of continuous RV's is useful in the context of the "geometrization of probability" program popularized by V. Milman
- Information theory: Studies fundamental limits of communication systems. Additive combinatorics has led to recent advances [Etkin-Ordentlich '09, Wu-Shamai-Verdú '12]

Continuous Entropy

- When random variable $X = (X_1, \ldots, X_n)$ has density $f(x)$ on $\mathbb{R}^n$, the entropy of X is

$$h(X) = h(f) := -\int_{\mathbb{R}^n} f(x) \log f(x) dx = E[-\log f(X)]$$

- The relative entropy between the distributions of $X \sim f$ and $Y \sim g$ is

$$D(f \| g) = \int f(x) \log \frac{f(x)}{g(x)} dx$$

For any f, g, $D(f\|g) \geq 0$ with equality iff $f = g$

Why are they relevant?

- Entropy is a measure of randomness
- Relative Entropy is a very useful notion of "distance" between probability measures (non-negative, and dominates several of the usual distances, although non-symmetric)

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

$$h(Y+Y') + h(Z) \leq h(Y+Z) + h(Y'+Z) \tag{1}$$

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 (1):

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

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,

$$h(X_1 + X_2 + X_3) + h(X_2) \leq h(X_1 + X_2) + h(X_3 + X_2) \qquad \text{[M. '08]}$$

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 $\mathbb{R}^d$, 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

- Entropy inequalities and additive combinatorics
  - Background and Motivations
  - Basic question: Entropy of sum vs. entropy of difference
- Entropic Limit Theorems
  - Entropic CLT
  - Role of entropy power inequalities
- Towards some structural results

Non-Gaussianity

For $X \sim f$ in $\mathbb{R}^n$, 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 $S_M := \dfrac{1}{\sqrt{M}} \displaystyle\sum_{i=1}^{M} X_i$

  The CLT scaling preserves variance

suggest . . .

Question: Is it possible that the CLT may be interpreted like the 2nd law of thermodynamics, 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: $h(S_{2n}) \geq h(S_n)$

mile-marker

- Entropy inequalities and additive combinatorics
  - Background and Motivations
  - Basic question: Entropy of sum vs. entropy of difference
- Entropic Limit Theorems
  - Entropic CLT
  - Role of entropy power inequalities
- Towards some structural results [from Kontoyiannis-M.'13]

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

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

then X is approximately Gaussian in the sense that

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

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 (2)...

Summary

- Took some initial steps towards developing an entropy theory for additive combinatorics in the general abelian setting
- Inequalities from this theory have applications in convex geometry/geometric functional analysis
- Looking at limit theorems using entropy is very natural and intuitive, and this study is also related to "continuous additive combinatorics"

Thank you!

○ — ○ — ○