Entropy and the Additive Combinatorics of Probability Densities on LCA groups

Mokshay Madiman

University of Delaware

Based on joint work with Ioannis Kontoyiannis, Athens Univ. of Economics
Jiange Li, University of Delaware
Jae Oh Woo, University of Texas, Austin
Liyao Wang, J. P. Morgan

Additive Combinatorics at Bordeaux

11–15 April 2016

Outline

- Background and Motivations

- Entropy inequalities and additive combinatorics
  - Basic question: Entropy of sum vs. entropy of difference
  - The Ruzsa divergence and its properties

- Phenomena that arise for particular groups
  - $\mathbb{Z}$
  - $\mathbb{R}^n$

Information theory: then to now

Early days...

- Q.1: What are the fundamental limits of data compression?
- Q.2: What are the fundamental limits of communication across a noisy channel?
- Answered for simple settings by Shannon '48, using a probabilistic formulation

Over the decades...

- Recognition of the fundamental role of notions like entropy in probability (e.g., in large deviations theory)
- Use of information theory to explore fundamental limits of statistical inference
- Growing recognition of the usefulness of information theoretic ideas in "pure mathematics"

Entropy of a Discrete Random Variable

- When random variable X has probability mass function p on a finite or countable set A, the entropy of X is

$$H(X) = H(p) := -\sum_{x \in A} p(x) \log p(x) = \mathbf{E}[-\log p(X)]$$

- If the "alphabet" A is finite, $H(X) \leq \log |A|$ with equality iff X is uniformly distributed on A
- If X and Y are discrete random variables (on the same probability space), then so is (X, Y); so one can talk about the "joint entropy" $H(X, Y)$. It is easy to see that $H(X, Y) \leq H(X) + H(Y)$, with equality iff X and Y are independent

Why is it relevant?

- Entropy is a very meaningful measure of randomness: $H(X)$ is the number of bits needed on average to encode X [Shannon '48]

Motivation for this talk

The entropy of sums of random variables is ubiquitous in information theory. What is the most general setting in which studying these makes sense?

Why should we care?

- Information theory: Our work has led to recent advances in the understanding of the interference channel [Wu-Shamai-Verdú '15, Stotz-Boelcskei '15], and carries much promise for other problems
- Probability: Related to basic questions, even when the group is plain old $\mathbb{R}^n$. E.g.: rate of convergence in the (entropic) Central Limit Theorem [Stam '59, Johnson-Barron '04, Artstein-Ball-Barthe-Naor '04, M.-Barron '07, etc.]
- Additive combinatorics: Sumset inequalities (inequalities for cardinalities of sums of sets) play a key role in this fast-developing area of mathematics, and entropy allows one to adopt a more general probabilistic approach to additive combinatorics [Ruzsa '09, M.-Marcus-Tetali '10–'12, Tao '10, etc.]
- Convex geometry: Related to the "geometrization of probability" program popularized by V. Milman (also C. Borell, K. Ball, etc.) [Lutwak-Yang-Zhang '04-'15, Bobkov-M.'11-'15, Fradelizi-M.-Wang '15]

Classical Sumset inequalities

Despite the origins of additive combinatorics in number theory, the canvas of the field has expanded in recent years to locally compact groups (we only consider locally compact abelian– or LCA– groups). Sumset inequalities are a very useful tool...

Examples of "direct" inequalities

- Ruzsa triangle inequality
$$|A - C| \leq \frac{|A - B| \cdot |B - C|}{|B|}$$
- Sum-difference inequality
$$|A + B| \leq \frac{|A - B|^3}{|A| \cdot |B|}$$
- The "Cauchy-Davenport inequality on $\mathbb{Z}$" says that
$$|A + B| \geq |A| + |B| - 1$$
with equality iff A and B are AP's with the same "step"

Examples from the Freiman (inverse) theory

- 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, $H(\mathrm{Unif}(A)) = \log |A|$ 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, Jog-Anantharam '12, M.-Wang-Woo '14 (entropic "direct" theory)
- Tao '10 (entropic "inverse" theory, including Freiman's theorem)

Differential Entropy on $\mathbb{R}^n$

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

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

where dx represents Lebesgue measure on $\mathbb{R}^n$

Key properties

- Translation-invariance:
  $h(X + b) = h(X)$ for any constant $b \in \mathbb{R}^n$
- $SL(n, \mathbb{R})$-invariance:
  Since $h(AX) = h(X) + \log \det(A)$ for any $n \times n$ matrix A of real entries,
  $h(AX) = h(X)$ when $\det(A) = 1$

Key non-properties

- Unlike discrete entropy, differential entropy is NOT always non-negative
- Unlike discrete entropy, differential entropy is NOT invariant with respect to bijections

Non-Gaussianity

- 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 $f = g$

- Relative Entropy is a very useful notion of "distance" between probability measures (e.g., dominates total variation distance)

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

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

  where X^G is Gaussian with the same mean and covariance matrix as X

Observe:

- For any random vector X, its non-Gaussianity $D(X) = h(X^G) - h(X)$

  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: $D(X) \geq 0$

Entropic Central Limit Theorem– I

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– II

Entropic CLT

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

- Monotonicity in n indicates that the entropy is a *natural measure* for CLT convergence (cf. second law of thermodynamics)
- A key step towards the Entropic CLT is the Entropy Power Inequality ("EPI")

mile-marker

- Background and Motivations
- Entropy inequalities and additive combinatorics
  - Basic question: Entropy of sum vs. entropy of difference
  - The Ruzsa divergence and its properties
- Phenomena that arise for particular groups
  - $\mathbb{Z}$
  - $\mathbb{R}^n$

Haar measure on LCA groups

Haar measure: Under some topological assumptions, an abelian group G admits a measure λ that is translation-invariant, i.e., such that

$$\lambda(A + x) = \lambda(A) \qquad \forall\, A \subset G, \forall x \in G$$

where $A + x = \{a + x : a \in A\}$. Such a measure is called a Haar measure, and is unique up to scaling by a positive constant

What assumptions? "LCA"

- The set G is a topological space such that $x + y$ is a continuous function of (x, y)
- The topology on G is Hausdorff and locally compact
- The Haar measure is then a countably additive measure defined on the Borel σ-field on G

Entropy on Groups

Let G be an LCA group, and λ be a Haar measure on G. If $\mu \ll \lambda$ is a probability measure on G, the entropy of $X \sim \mu$ is defined by

$$h(X) = -\int_G f(x) \log f(x)\, \lambda(dx)$$

where $f(x) = \frac{d\mu}{d\lambda}(x)$ is the density of μ with respect to the Haar measure λ

Remarks

- Recall that $X \sim \mu$ is said to have density f when
$$\mathbf{P}(X \in A) = \int_A f(x)\lambda(dx), \qquad A \in \mathcal{G}$$
- Usual abuse of notation: we write $h(X)$ though h depends only on f
- 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, the Haar measure λ is finite, and so we can normalize it to get the "uniform" probability measure on $\mathcal{G}$. Then, for every RV X,
$$h(X) = -D(\mu \| \lambda) \leq 0$$

Entropy on Groups: Examples

Classical examples

- G discrete with counting measure: discrete entropy
- $G = \mathbb{R}^n$ with Lebesgue measure: differential entropy

Non-classical examples

- $G = \mathbb{T}^n$, torus with Lebesgue measure: differential entropy on the torus
- $G = (0, \infty)$ with the Haar measure $\lambda(dx) = x^{-1}dx$:
  if f is the density (w.r.t. Lebesgue measure) of a positive random variable X, then $\mathbf{P}(X \in A) = \int_A f(x)dx = \int_A xf(x)\frac{dx}{x}$, so

$$h_G(X) = -\int_0^\infty [xf(x)] \log[xf(x)]\lambda(dx) = -\int_0^\infty f(x)[\log x + \log f(x)]dx$$
$$= h_{\mathbb{R}}(X) - \mathbf{E}[\log X]$$

Key Properties of Entropy on Groups

We cannot even talk about things like linear transformations on general groups because they do not have a linear structure. Yet one has...

Lemma 1 (Translation-invariance)
Let X be a random variable taking values in G. If $b \in G$, then

$$h(X + b) = h(X)$$

Lemma 2 ($SL(n, \mathbb{Z})$-invariance) [M.-Singla '15]
Let X be a random variable taking values in G^n, and denote by $SL_n(\mathbb{Z})$ the set of $n \times n$ matrices A with integer entries and determinant 1. If $A \in SL_n(\mathbb{Z})$, then

$$h(AX) = h(X)$$

Remark

- Integer linear combinations of group elements always makes sense in an abelian group, e.g., $2x - 3y$ represents $x + x + (-y) + (-y) + (-y)$

A Question and an Answer

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

Question

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

First answer

For $\mathcal{G} = \mathbb{Z}$ or $\mathcal{G} = \mathbb{R}$, the entropies of the sum and difference of two i.i.d. random variables *can differ by an arbitrarily real number*

Precise formulation: [Abbe–Li–M.'16] Let $\mathcal{G} = \mathbb{Z}$ or $\mathcal{G} = \mathbb{R}$. Given any $M \in \mathbb{R}$, there exist i.i.d. random variables Y, Y' of finite entropy such that

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

Note: [Lapidoth–Pete '08] showed that for every $M > 0$, there exist i.i.d. random variables Y, Y' such that $h(Y-Y') - h(Y+Y') > M$; our answer also implies the opposite (which is not symmetrical)

Achievable differences: examples

Fact: There exist i.i.d. $\mathbb{Z}$-valued (or $\mathbb{R}$-valued) random variables Y and Y' with finite entropy, such that $h(Y-Y')$ and $h(Y+Y')$ differ by an arbitrary real number

Remarks:

- The analogous question for sets has a long history: Conway's 1967 "conjecture"; MSTD set $\{1,2,3,5,8,9,13,15,16\}$ of [Marica '69]; Conway's MSTD set $\{0,2,3,4,7,11,12,14\}$
- Non-MSTD sets can give rise to examples relevant to us: $\{0,1,3,4,5,6,7,10\}$
- [Hegarty '07] proved that there is no MSTD set in $\mathbb{Z}$ of size 7 and, up to linear transformations, Conway's set is the unique MSTD set of size 8.
- On $\mathbb{Z}/3\mathbb{Z}$, $h(X-X') \geq h(X+X')$ [Abbe-Li-M.'16]
- Just to prove that the difference can be arbitrarily large in either direction, one can use a construction based on a dilation observation of [S. K. Stein '73] (for more differences than sums) or [Ruzsa '92] (for both)

Achievable differences: proof

Fact: There exist i.i.d. $\mathbb{Z}$-valued (or $\mathbb{R}$-valued) random variables Y and Y' with finite entropy, such that $h(Y-Y')$ and $h(Y+Y')$ differ by an arbitrary real number

Proof: If $X \in \{0, 1, \cdots, n-1\}$, $H(X+Y) - H(X-Y)$ is a continuous function of n variables. By assuming n large enough, we know that this function can take both positive and negative values. Since the function is continuous, the intermediate value theorem implies that its range must contain an open interval (a, b) with $a < 0 < b$. Let $X' = (X_1, \cdots, X_k)$, where X_i are independent copies of X, and let Y' be an independent copy of X'. Then

$$H(X'+Y') - H(X'-Y') = k[H(X+Y) - H(X-Y)]$$

The range of $H(X'+Y') - H(X'-Y')$ thus contains (ka, kb), which includes any real number by taking k large enough. While $X', Y' \in \mathbb{Z}^k$, the linear transformation $(x_1, \cdots, x_k) \to x_1 + dx_2 + \cdots + d^{k-1}x_k$ maps X, Y to $\mathbb{Z}$-valued random variables, and moreover, preserves entropy by taking d large enough.

A Question and another Answer

Question

If Y and Y' are i.i.d. random variables taking values in the LCA group G, how different can $h(Y+Y')$ and $h(Y-Y')$ be?

Another answer [Kontoyiannis–M.'15]

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

Precise formulation: Let G be any LCA group. For any two G-valued i.i.d. random variables Y, Y' with finite 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 suggest 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

Remarks

- Observe that if $\mathcal{G} = \mathbb{R}^n$, Δ_+ and Δ_- are affine-invariant; so these facts are related to the *shape* of the density
- This statement for *discrete* random variables (half due to [Ruzsa '09, Tao '10], half due to [M.-Marcus-Tetali '12]) is the exact analogue of the doubling–difference inequality for sets in additive combinatorics
- Only for discrete setting, [Abbe–Li–M.'16] observe that the analog of Freiman-Pigarev inequality follows:

$$\frac{3}{4} < \frac{H(X+Y)}{H(X-Y)} < \frac{4}{3}$$

mile-marker

- Entropy inequalities and additive combinatorics
  - Background and Motivations
  - Basic question: Entropy of sum vs. entropy of difference
  - The Ruzsa divergence and its properties

- Phenomena that arise for particular groups
  - $\mathbb{Z}$
  - $\mathbb{R}^n$

Conditional entropy and mutual information

Conditional entropy of X given Y is

$$h(X|Y) = \int h(X|Y=y) P_Y(dy)$$

where $h(X|Y=y)$ is the entropy of the (regular) conditional distribution $P_X(\cdot|Y=y)$.

Two useful facts

- Shannon's Chain Rule:

$$h(X,Y) = h(Y) + h(X|Y)$$

- Conditioning reduces entropy (or) mutual information is non-negative:

$$h(X) - h(X|Y) = D(p_{X,Y} \| p_X \times p_Y) := I(X;Y) \geq 0$$

The Ruzsa divergence

Suppose X and Y are G-valued random variables with finite entropy. The quantity

$$d_R(X\|Y) := h(X - Y') - h(X),$$

where X and Y' are taken to be independent random vectors with Y' having the same distribution as Y, will be called the Ruzsa divergence

Lemma: If X and Y are independent RVs, then

$$d_R(X\|Y) = I(X - Y; Y)$$

Proof:

$$\begin{aligned} d_R(X\|Y) &= h(X - Y) - h(X) \\ &= h(X - Y) - h(X|Y) \qquad \text{[independence]} \\ &= h(X - Y) - h(X - Y|Y) \quad \text{[translation-invariance]} \\ &= I(X - Y; Y) \end{aligned}$$

Note: In particular, $d_R(X, Y) \geq 0$ (for some groups like $\mathbb{R}^n$, it is never 0 in non-degenerate situations)

Key properties of the Ruzsa divergence

Theorem 1: If X_i are independent, then

$$d_R(X_1\|X_3) \le d_R(X_1\|X_2) + d_R(X_2\|X_3)$$

Theorem 2: If X and Y_i are all mutually independent, then

$$d_R\left(X\middle\|\sum_{i=1}^{k} Y_i\right) \le \sum_{i=1}^{k} d_R(X\|Y_i)$$

Remarks

- Theorem 1 is the analog of Ruzsa's triangle inequality for sumsets
- Theorem 2 is the analog of Plünnecke-Ruzsa inequality for sumsets
- Theorem 2 is equivalent to the Submodularity Lemma:

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

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
- Entropy is not subadditive in continuous settings; it is easy to construct examples on $\mathbb{R}$ 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 $h(\text{const.}) = -\infty$
- This Lemma has many other applications, e.g., to convex geometry [Bobkov-M.'12]

Completing the proof of (Ans.2)

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

$$\frac{h(Y+Y')-h(Y)}{h(Y-Y')-h(Y)} \in [\tfrac{1}{2}, 2]$$

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

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

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

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

Also the Ruzsa triangle inequality can be rewritten:

$$h(Y-Y')+h(Z) \leq h(Y-Z)+h(Y'-Z) \tag{2}$$

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

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

which are the desired bounds

mile-marker

- Background and Motivations
- Entropy inequalities and additive combinatorics
  - Basic question: Entropy of sum vs. entropy of difference
  - The Ruzsa divergence and its properties
- Phenomena that arise for particular groups
  - $\mathbb{Z}$
  - $\mathbb{R}^n$

Notation

Entropies

- Differential Entropy

$$h(f) = -\int_{\mathbb{R}} f \log f$$

- Shannon Entropy

$$H(p) = H_1(p) = -\sum_{i \in \mathbb{Z}} p(i) \log p(i)$$

- Rényi Entropy of order $r \in (0,1) \cup (1,\infty)$, $r = 0$, and $r = \infty$

$$H_r(p) = \frac{1}{1-r} \log \left[\sum_{i \in \mathbb{Z}} p(i)^r \right]$$

$$H_0(p) = \log |\mathsf{supp}(p)|$$

$$H_\infty(p) = -\log \left[\sup_{i \in \mathbb{Z}} p(i) \right]$$

Given a random variable $X \sim f$ on $\mathbb{R}$, or $Y \sim p$ on $\mathbb{Z}$, we write

$$h_r(X) = h_r(f) \quad \text{and} \quad H_r(Y) = H_r(p)$$

Entropy Power Inequality on $\mathbb{R}$

If X and Y independent $\mathbb{R}$-valued random variables

$$e^{2h(X+Y)} \geq e^{2h(X)} + e^{2h(Y)}$$

Remark

- Equivalent Formulation:

  $$h(X+Y) \geq h(X^{\#} + Y^{\#})$$

  where $X^{\#}$ and $Y^{\#}$ independent Gaussians with $h(X^{\#}) = h(X)$ and $h(Y^{\#}) = h(Y)$
- Sharp lower bounds on entropies of convolutions
- Weak Formulation: If X and X' are IID,

  $$h(X+X') - h(X) \geq \frac{1}{2}\log 2$$
- Closely related to the central limit theorem, log-Sobolev inequality, Heisenberg uncertainty principle etc.

Question: Is there a discrete analogue of the EPI?

Earlier Work

Unconditional results for $\mathbb{Z}$-valued random variables

- [Tao '10] proved the asymptotically sharp result: If X and X' are IID,

$$H(X+X') - H(X) \geq \frac{1}{2}\log 2 - o(1)$$

  where $o(1)$ disappears as $H(X)$ tends to infinity

- [Haghighatshoar–Abbe–Telatar '13] If X and X' are IID,

$$H(X+X') - H(X) \geq g(H(X))$$

  for an increasing function $g : \mathbb{R}_+ \to \mathbb{R}_+$ with $g(0) = 0$

Conditional results

- [Johnson–Yu '10] have results involving thinning for ultra-log-concave random variables (not directly related to our goals)
- Results mimicking EPI on $\mathbb{R}$ available for special subclasses of distributions on $\mathbb{Z}$

Naive EPI for integers

The first (naive) conjecture that one might make for distributions on the integers is

$$N(X+Y) \geq N(X) + N(Y) \tag{3}$$

where $N(X) = \exp\{2H(X)\}$ and $H(\cdot)$ denotes the Shannon entropy of the independent $\mathbb{Z}$-valued random variables X, Y

Remarks

- Counterexample: take both X and Y to be constants. More generally, if X and Y are i.i.d. Bernoulli(p) with $p \neq 0.5$, then $N(X+Y) < N(X) + N(Y)$ [Sharma–Das–Muthukrishnan '11]
- However, it is known that the inequality (3) holds true for binomial random variables with parameter $1/2$ [Harremöes–Vignat '03]
- [Sharma–Das–Muthukrishnan '11] showed that

  $$N(Bin(m+n,p)) \geq N(Bin(n,p)) + N(Bin(m,p))$$

  if $m, n \geq n_0(p)$
- Not obvious that $N(X)$ is the right functional to look at for $\mathbb{Z}$-valued random variables (in the real-valued case, the motivation comes from Gaussian comparison, which is no longer directly relevant)

Suggestive set analogues

- Brunn-Minkowski inequality over $\mathbb{R}$: For $A, B \subset \mathbb{R}$,

$$|A+B| \geq |A| + |B|$$

where m is Lebesgue measure

Proof: Since $m(A+a) = m(A)$ for any $a \in \mathbb{R}$, we may assume that $\sup(A) = \inf(B) = \{0\}$, and also that $0 \in A \cap B$ without changing the Lebesgue measure of any of the sets A, B and $A+B$. Then $A \cup B \subseteq A+B$. This implies

$$\begin{aligned} m(A+B) &\geq m(A \cup B) \\ &= m(A) + m(B) - m(\{0\}) = m(A) + m(B). \end{aligned}$$

- Cauchy-Davenport inequality over $\mathbb{Z}$: For $A, B \subset \mathbb{Z}$,

$$|A+B| \geq |A| + |B| - 1 \tag{4}$$

Proof: If c is counting measure on $\mathbb{Z}$,

$$\begin{aligned} c(A+B) &\geq c(A \cup B) \\ &= c(A) + c(B) - c(\{0\}) = c(A) + c(B) - 1 \end{aligned}$$

- It is clear that the additional term -1 appears to remove the redundancy: With a counting measure, the volume of a singleton set should be considered, whereas it does not matter for Lebesgue measure.

A Natural Conjecture

By analogy with the inequalities for sets, one may wonder if

$$N(X+Y) \geq N(X) + N(Y) - 1$$

for independent $\mathbb{Z}$-valued random variables X, Y

Remarks

- The old counterexample (constants) does not falsify this: the -1 is precisely the term needed to counter the zero entropy of the second constant random variable.
- [Woo–M.'15] show Theorem: If X and Y are uniformly distributed over finite sets $A \subset \mathbb{Z}$ and $B \subset \mathbb{Z}$ respectively

$$N(X+Y) \geq N(X) + N(Y) - 1$$

- Nonetheless the conjecture fails to hold in general: if X, Y are IID Bernoulli(p) with $0 < p < 0.08$ or $0.92 < p < 1$, $N(X+Y) < N(X) + N(Y) - 1$ (a "small p effect")

Other conjectures

More generally, is a Kneser-type inequality true for entropy? If X, Y, and Z are independent, is it true that

$$N(X+Y+Z)+N(Z) \geq N(X+Z)+N(Y+Z)?$$

Remarks

- False as asked; X and Y uniform on $\{0,1\}$ and Z uniform on $\{0,1,2\}$ gives a counterexample
- Heuristic reason for believing this may hold when Z is log-concave: If A, B, C are finite subsets of $\mathbb{Z}$, with C being a contiguous subset, then

  $$|A+B+C|+|C| \geq |A+C|+|B+C|$$

  Proof is a simple extension of the reasoning for Cauchy-Davenport. Such a result also holds in $\mathbb{R}^n$ when A, B, C convex and conjectured to hold just under convexity of C [Fradelizi–M.–Marsiglietti–Zvavitch '16]

A General EPI

Discrete Analogue of EPI [Wang–Woo–M.'14]

$$H(X+Y) \geq H(X^+ + Y^*)$$

where X and Y independent random variables, X^+ and Y^* defined later

Remarks

- Sharp lower bound on entropy of convolutions
- Based on discrete rearrangements
- Unifies results for sets and random variables since it generalizes to Rényi entropy
- Similar continuous version proved by [Wang–M.'14]

Rearrangements

A non-negative function p indexed by $\mathbb{Z}$, $p \in c_0(\mathbb{Z})$, the space of functions on $\mathbb{Z}$ vanishing at infinity

- p^+ to be a permutation of p such that

$$p^+(0) \geq p^+(1) \geq p^+(-1) \geq p^+(2) \geq p^+(-2) \geq \cdots$$

- ${}^+p$ to be a permutation of p such that

$${}^+p(0) \geq^+ p(-1) \geq^+ p(1) \geq^+ p(-2) \geq^+ p(2) \geq \cdots$$

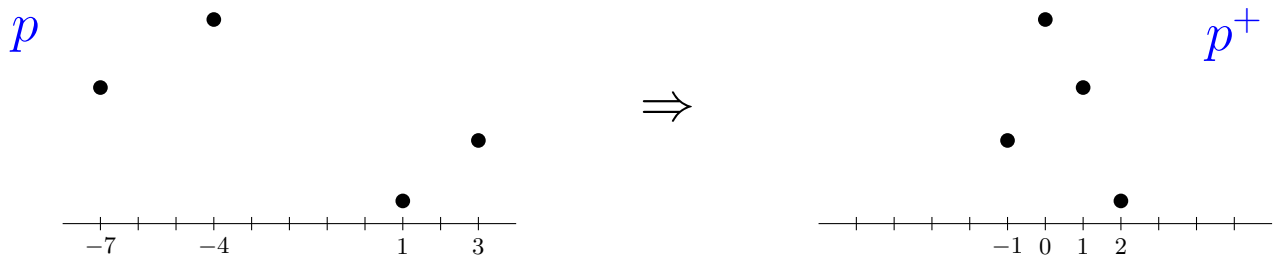
- If $p^+ =^+ p$, define $p^* = p^+$(call p regular)

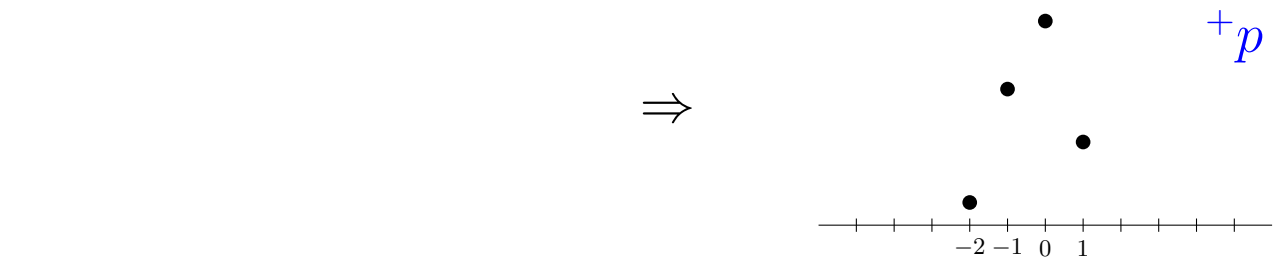
Remarks

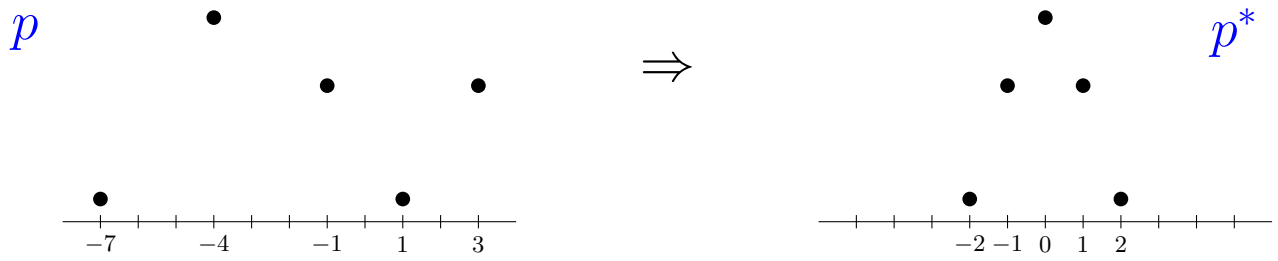
- p^+ and ${}^+p$ mirror images each other
- If p^* exists, p largest value an odd number of times and each of the other values an even number of times
- Rearrangements do not change Rényi entropy

$$H_r(p) = H_r(p^+) = H_r({}^+p)$$

Examples of Rearrangement







Majorization

For two probability mass functions p and q on $\mathbb{Z}$, we say that p is majorized by q (and write $p \prec q$) if and only if for all non-negative integer k, we have

$$\sum_{|i| \leq k} p^+(i) \leq \sum_{|i| \leq k} q^+(i)$$

$$\sum_{i=-k}^{k+1} p^+(i) \leq \sum_{i=-k}^{k+1} q^+(i)$$

Remarks

- The largest sum of k points in p is less than or equal to that in q
- Entire sums are the same

Rearrangement Inequality

Lemma 1

Let $p_j \in c_0(\mathbb{Z}), j = 1, 2, \cdot\cdot, n$ be non-negative functions on $\mathbb{Z}$. Suppose $p_j, j \geq 2$ are regular. Let a non-negative function $g \in c_0(\mathbb{Z})$. Then

$$\sum_{i\in\mathbb{Z}} g(i)p_1 \star p_2 \star \cdot\cdot \star p_n(i) \leq \sum_{i\in\mathbb{Z}} g^+(i)p_1^+ \star p_2^* \star p_3^* \cdot\cdot \star p_n^*(i)$$

Remarks

- Based on Hardy-Littlewood-Pólya's rearrangement inequality

$$q \star p_1 \star p_2 \star \cdot\cdot \star p_n(0) \leq {}^+q \star p_1^+ \star p_2^* \star p_3^* \cdot\cdot \star p_n^*(0)$$

- Let $g(i) = q(-i)$ on $i \in \mathbb{Z}$
- Restriction that $p_j, j \geq 2$ are regular is essential
- Implies a powerful majorization

Lower Bounds on Rényi Entropy of Convolutions

Theorem 1

Let $p_j, j = 1, 2, \cdot\cdot, n$ be probability mass functions on $\mathbb{Z}$. Suppose $p_j, j \geq 2$ are regular.

$$p_1 \star p_2 \star \cdot\cdot \star p_n \prec p_1^+ \star p_2^* \star p_3^* \star \cdot\cdot \star p_n^*$$

Theorem 2

Let $\phi(x)$ be a convex function defined on the non-negative real line such that $\phi(0) = 0$ and it is continuous at 0. Suppose p and q are probability mass functions on $\mathbb{Z}$. If $p \prec q$

$$\sum_i \phi(p(i)) \leq \sum_i \phi(q(i))$$

provided that both sides are well defined

Theorem 3

Given same assumptions of Theorem 1 and 2

$$\sum_i \phi(p_1 \star p_2 \star \cdot\cdot \star p_n(i)) \leq \sum_i \phi(p_1^+ \star p_2^* \star p_3^* \star \cdot\cdot \star p_n^*(i))$$

provided that both sides are well defined

Lower Bounds on Rényi Entropy of Convolutions

Key Corollary

By using the continuous convexity and the nullity at zero of x^r for $1 < r < +\infty$, $x\log(x)$, and $-x^r$ for $0 < r < 1$, we may conclude for any $r \in [0, \infty]$,

$$H_r(p \star q) \geq H_r(p^* \star q^+)$$

where p is regular

Implication

Consider two finite sets A and B on $\mathbb{Z}$. Let p and q be uniformly distributed on A and B with $|A|$ odd.

$$\text{supp}(p \star q) = A + B \quad \text{supp}(p^* \star q^+) = A^\# + B^\#$$

where $A^\# := \text{supp}(p^*)$ and $B^\# := \text{supp}(q^+)$. Then $|A^\#| = |A|$ and $|B^\#| = |B|$. Since p is regular, $r = 0$ case implies

$$|A + B| \geq |A^\# + B^\#| = |A| + |B| - 1$$

which is Cauchy-Davenport Inequality on $\mathbb{Z}$

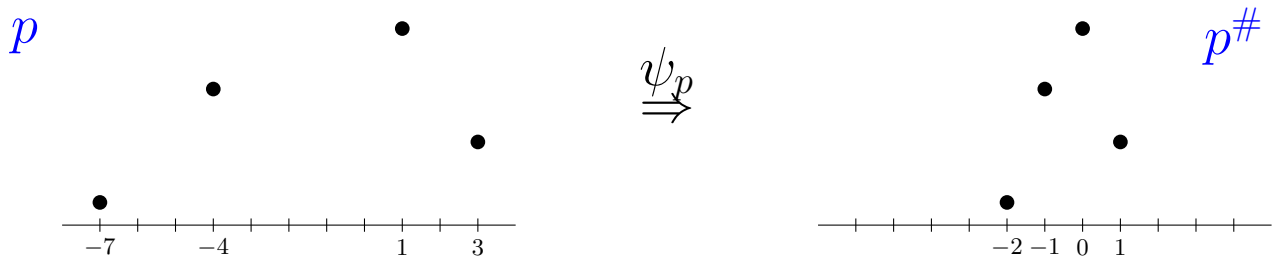
More Entropy Inequalities

More entropy inequalities are available if we use Sperner Theory

#-log-concave

Let $p \in c_0(\mathbb{Z})$ be a non-negative function on $\mathbb{Z}$. Define $p^{\#}$ to be a re-arrangement of p such that $\text{supp}(p^{\#})$ is both order-preserving (i.e., $i < j$ implies $p^{\#}(i) < p^{\#}(j)$) and is supported on a set of consecutive integers

We call p #-log-concave if the distribution of $p^{\#}$ is log-concave which is equivalent to $p^{\#}(i)^2 \geq p^{\#}(i-1)p^{\#}(i+1)$ for all $i \in \mathbb{Z}$



More Entropy Inequalities

One of Our Results [Wang–Woo–M.'14]

For any $0 \leq r \leq +\infty$,

$$H_r(p \star q) \geq H_r(p^{\#} \star q^{\#})$$

where p and q are $\#$-log-concave

Implication

Let us consider p and q, uniformly distributed on finite sets A and B respectively. Both p and q are $\#$-log-concave. Let $A^{\#} := \text{supp}(p^{\#})$ and $B^{\#} := \text{supp}(q^{\#})$, then $|A| = |A^{\#}|$ and $|B| = |B^{\#}|$. We may conclude Cauchy-Davenport Inequality on $\mathbb{Z}$

$$|A + B| \geq |A^{\#} + B^{\#}| = |A| + |B| - 1$$

Note that this removes the odd cardinality constraint of p in the previous result

Probabilistic statement of Littlewood-Offord lemma

Setup

Let X_i be i.i.d. Bernoulli random variables, taking values 0 and 1, each with probability $\frac{1}{2}$. Suppose v_i are real numbers with $v_i \neq 0$. Define the random variable

$$S_v = \sum_{i=1}^{n} X_i v_i,$$

and let $Q(v) = \max_x \mathbf{P}(S_v = x)$

Classical results

- [Littlewood–Offord '43] proved that

$$Q(v) = O(n^{-1/2})$$

- [Erdős '45] refined this by identifying an extremal set of v_i; specifically he showed that if $v_i^* = 1$ for each i, then

$$Q(v) \leq Q(v^*)$$

Application to Littlewood-Offord-type Inequalities

General Littlewood-Offord-Erdős-type question:

Let X_i be independent $\mathbb{Z}$-valued random variables, and $V = (0, \infty)^n$. For given $v = (v_1, \ldots, v_n) \in V$, form the sum $S_v = \sum_{i=1}^n X_i v_i$ and define

$$Q_r(v) = e^{-H_r(S_v)}$$

We wish to study the problem

$$\text{Maximize } Q_r(v) \quad \text{(or)} \quad \text{Minimize } H_r(S_v)$$
$$\text{subject to } v \in V$$

Theorem: Let X_i be independent $\mathbb{Z}$-valued random variables, whose probability mass functions are symmetric and unimodal. For any $v \in V$,

$$Q_r(v) \leq Q_r(v^*)$$

or equivalently, $H_r(S_v) \geq H_r(S_v^*)$

Remarks

- Observe that $Q(v) = \max_x \mathbf{P}(S_v = x)$ defined earlier is just $Q_\infty(v)$
- The Theorem generalizes the Littlewood-Offord-Erdős lemma in 2 ways: much larger class of distributions, and large class of entropies, *but* the extremizers stay the same!

Summary

- Steps towards an entropy theory for additive combinatorics of probability densities in the general abelian setting
- Special phenomena of combinatorial interest in the finite abelian setting
- For $\mathbb{R}^n$, the theory has close connections to convex geometry/geometric functional analysis, as well as probability

Thank you!

○ — ○ — ○