Entropy and the additive combinatorics of probability densities on $\mathbb{R}^n$

Mokshay Madiman

University of Delaware and Yale University

Includes joint work with Sergey Bobkov, University of Minnesota
and Ioannis Kontoyiannis, Athens U. of E.&B.

Asymptotic Geometric Analysis II, St. Petersburg
20–24 June 2013

Outline

- Background: Functional/probabilistic setting for convex geometry
- Reverse Entropy Power Inequality for convex measures
- Additive combinatorics and sumset inequalities
- A unified setting: entropy inequalities in LCA groups
- Miscellanea:
  - Effective version of Reverse Entropy Power Inequality for i.i.d. summands
  - Plünnecke-Ruzsa inequality for convex sets
  - A Freiman-type observation

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 entropy power of X is $N(X) = e^{\frac{2h(X)}{n}}$

Remarks

- Usual abuse of notation: we write $h(X)$ even though the entropy is a functional depending only on the density of X
- $N(X) \in [0, \infty]$ can be thought of as a "measure of randomness"
- N is an (inexact) analogue of volume: if U_A is uniformly distributed on a bounded Borel set A,

$$h(U_A) = \log |A| \quad \text{or} \quad N(U_A) = |A|^{2/n}$$

Entropy power

The entropy power of X is $N(X) = e^{\frac{2h(X)}{n}}$

Remarks

- The reason we don't define entropy power by $e^{h(X)}$ (which would give $|A|$ for Unif(A)) is that the "correct" comparison is not to uniforms but to Gaussians
- Just as Euclidean balls are special among subsets of $\mathbb{R}^n$, Gaussians are special among distributions on $\mathbb{R}^n$
- If Z is $N(0, \sigma^2 I)$, the entropy power of Z is

  $$N(Z) = (2\pi e)\sigma^2$$

  Thus the entropy power of X is (up to a universal constant) the variance of the (isotropic) normal that has the same entropy as X:

  $$N(X) = N(Z) = (2\pi e)\sigma_Z^2$$

- entropy power: random variables :: volume$^{\frac{1}{n}}$: sets
  since $|A|^{\frac{1}{n}}$ is (up to a universal constant) the radius of the ball that has the same volume as A

Brunn-Minkowski inequality and entropy power inequality

The Inequalities

- Let A, B be any Borel-measurable sets in $\mathbb{R}^n$. Write $A + B = \{x + y : x \in A,\ y \in B\}$ for the Minkowski sum, and $|A|$ for the n-dimensional volume. The Brunn-Minkowski inequality says that

$$\left|A + B\right|^{1/n} \geq |A|^{1/n} + |B|^{1/n} \quad [BM]$$

- For a random vector X in $\mathbb{R}^n$, the entropy power is $N(X) = e^{2h(X)/n}$. For any two independent random vectors X and Y in $\mathbb{R}^n$,

$$N(X + Y) \geq N(X) + N(Y) \quad [EPI]$$

Remarks

- BM was proved by [Brunn 1887, Minkowski 1890s, Lusternik '35]
- EPI was proved by [Shannon '48, Stam '59]; equality holds iff X, Y are normal with proportional covariances

Sidenote: Two kinds of functional versions

For the goal of embedding the geometry of convex sets in a more analytic setting, several approaches are possible:

- Replace sets by functions, and convex sets by log-concave or s-concave functions. Replace volume by integral. E.g. [Klartag-Milman '05, Milman-Rotem '13]
- Replace sets by random variables, and convex sets by random variables with log-concave or s-concave distributions. Replace volume by entropy (actually entropy power). E.g. [Dembo-Cover-Thomas '91, Lutwak-Yang-Zhang '04-'13, Bobkov-Madiman '11-'13]

Another example: Blaschke-Santaló inequality

The Inequalities

- If K, L are compact sets in $\mathbb{R}^n$, then
$$|K| \cdot |L| \leq \omega_n^2 \max_{x \in K, y \in L} |\langle x, y \rangle|^n$$
- For any two independent random vectors X and Y in $\mathbb{R}^n$, there is an (explicit) universal constant c such that
$$N(X) \cdot N(Y) \leq c\mathbf{E}\left[|\langle X, Y \rangle|^2\right]$$
[Lutwak-Yang-Zhang '04]

Remarks

- The first inequality implies the Blaschke-Santaló inequality by taking K to be a symmetric convex body, and L to be the polar of K
- Functional versions of the other kind also exist [Artstein-Klartag-Milman '05, Fradelizi-Meyer '07, Lehec '09]

Reverse Brunn-Minkowski inequality

Given two convex bodies A and B in $\mathbb{R}^n$, one can find an affine volume-preserving map $u : \mathbb{R}^n \to \mathbb{R}^n$ such that with some absolute constant C,

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

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

Remarks

- The reverse Brunn-Minkowski inequality was proved by [V. Milman '86], with other proofs in [Milman '88, Pisier '89]
- Seminal result in convex geometry/asymptotic theory of Banach spaces; closely connected to the hyperplane conjecture
- Is there a reverse EPI under some "convexity" assumption?

Reverse entropy power inequality

If X and Y are independent and have log-concave densities, then for some linear entropy-preserving map $u : \mathbb{R}^n \to \mathbb{R}^n$,

$$N\big(\widetilde{X} + Y\big) \;\leq\; C\,(N(X) + N(Y)), \qquad \text{[Bobkov–M.'11, CRAS]}$$

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

Remarks

- Recall that a probability density function f on $\mathbb{R}^n$ is *log-concave* (or LC) if
$$f(\alpha x + (1-\alpha)y) \geq f(x)^\alpha f(y)^{1-\alpha},$$
for each $x, y \in \mathbb{R}^n$ and each $0 \leq \alpha \leq 1$
- Can recover reverse BM inequality as a special case, though this is not immediately obvious
- *Question*: Can we generalize to a larger class of measures?

Convex measures

Fix a parameter $\beta \geq n$. A density f on $\mathbb{R}^n$ is β*-concave* if

$$f(x) = V(x)^{-\beta}, \qquad x \in \mathbb{R}^n$$

where V is a positive convex function on $\mathbb{R}^n$

Remarks

- Probability measures μ on $\mathbb{R}^n$ with β-concave densities satisfy the geometric inequality

$$\mu\big(tA + (1-t)B\big) \geq \big[\, t\mu(A)^\kappa + (1-t)\mu(B)^\kappa \big]^{1/\kappa}$$

  for all $t \in (0,1)$ and for all Borel measurable sets $A, B \subset \mathbb{R}^n$, with negative power $\kappa = -\frac{1}{\beta - n}$
- For growing β, the families of β-concave densities shrink and converge in the limit as $\beta \to +\infty$ to the family of log-concave densities
- The largest class is thus the class of n-concave densities; the corresponding class of measures is said to be "convex"
- One main reason to consider β-concave densities is that they allow heavy tails, unlike log-concave densities (e.g., Cauchy density on $\mathbb{R}$ is 2-concave)

Reverse EPI for β-concave class

Let X and Y be independent random vectors in $\mathbb{R}^n$ with densities, for $\beta \geq \max\{2n+1, \beta_0 n\}$ with $\beta_0 > 2$. There exists a linear entropy-preserving map $u : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$N\big(\widetilde{X} + Y\big) \leq C_{\beta_0}\,(N(X) + N(Y)), \qquad \text{[Bobkov–M.'12, JFA]}$$

where $\widetilde{X} = u(X)$, and C_{β_0} is a constant depending on β_0 only

Remarks

- *Question*: Is it possible to relax the assumption on the range of β, or even to remove any convexity hypotheses?

No reverse EPI for convex measures

This is impossible already for the class of all one-dimensional convex probability distributions (note that for $n = 1$, there are only two admissible linear transformations, $\widetilde{X} = X$ and $\widetilde{X} = -X$)

Theorem: [Bobkov–M.'13] For any constant C, there is a convex probability distribution μ on the real line with a finite entropy, such that

$$\min\{N(X+Y), N(X-Y)\} \geq C\, N(X),$$

where X and Y are i.i.d. random variables drawn from μ

Intuition: A main reason for $N(X+Y)$ and $N(X-Y)$ to be much larger than $N(X)$ is that the distributions of the sum $X+Y$ and the difference $X-Y$ may lose convexity properties, when the distribution μ of X is not "sufficiently convex"

Question: Is it possible to say anything about the relationship between $N(X+Y)$, $N(X-Y)$ and $N(X), N(Y)$ in general (i.e., no convexity hypotheses at all)?

mile-marker

- Background: Functional/probabilistic setting for convex geometry
- Reverse Entropy Power Inequality for convex measures
- Additive combinatorics and sumset inequalities
- A unified setting: entropy inequalities in LCA groups
- Miscellanea:
  - Effective version of Reverse Entropy Power Inequality for i.i.d. summands
  - Plünnecke-Ruzsa inequality for convex sets
  - A Freiman-type observation

Motivation: 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

Reminder: Discrete Entropy

For a discrete random variable X with probability mass function p, i.e., $\mathbf{P}\{X = x\} = p(x)$,
entropy $\quad H(X) = H(p) = -\sum_x p(x) \log p(x)$

Key Properties

- If X is supported on a finite set A, then

$$0 \leq H(X) \leq \log |A|$$

  with the first being equality iff X is deterministic, and the second being equality iff $X \sim \text{Unif}(A)$
- The entropy is the "minimum number of bits needed to represent X", and so can be thought of as the amount of information in X

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

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

May be rewritten as

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

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

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

Upper bound proved by Ruzsa '09, Tao '10, lower bound by M.-Marcus-Tetali '09

mile-marker

- Background: Functional/probabilistic setting for convex geometry
- Reverse Entropy Power Inequality for convex measures
- Additive combinatorics and sumset inequalities
- A unified setting: entropy inequalities in LCA groups
- Miscellanea:
  - Effective version of Reverse Entropy Power Inequality for i.i.d. summands
  - Plünnecke-Ruzsa inequality for convex sets
  - A Freiman-type observation

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

Reminder: Entropy in General Setting

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

Key cases

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

A Question and an Answer

Setup: Let Y and Y' be i.i.d. 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 [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: Let $\mathcal{G} = \mathbb{R}$ or $\mathcal{G} = \mathbb{Z}$. Given any $M > 0$, there exist i.i.d. $\mathcal{G}$-valued random variables Y, Y' of finite 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. $\mathcal{G}$-valued 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. $\mathcal{G}$-valued 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 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

Proof outline

We obtain the desired inequality from two more general facts:

Fact 1: [*Entropy analogue of the Plünnecke-Ruzsa inequality*] If Y, Y', Z are independent random variables, then the Submodularity Lemma says

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

Fact 2: [*Entropy analogue of the Ruzsa triangle inequality*] If Y, Y', Z are independent random variables, then

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

Proof of Upper Bound Since $h(Y + Y') \leq h(Y + Y' + Z)$, Fact 1 implies

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

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

$$\begin{aligned} & h(Y + Y') + h(Y) \leq 2h(Y - Y') \\ \text{or} \quad & h(Y + Y') - h(Y) \leq 2[h(Y - Y') - h(Y)] \end{aligned}$$

Proof of Lower Bound The other half follows similarly from Fact 2

mile-marker

- Background: Functional/probabilistic setting for convex geometry
- Reverse Entropy Power Inequality for convex measures
- Additive combinatorics and sumset inequalities
- A unified setting: entropy inequalities in LCA groups
- Miscellanea:
  - Effective version of Reverse Entropy Power Inequality for i.i.d. summands
  - Plünnecke-Ruzsa inequality for convex sets
  - A Freiman-type observation

An effective reverse EPI

Let X and Y be i.i.d. random vectors in $\mathbb{R}^n$ with a LC density. Then

$$H(X - Y) \leq e^2 H(X)$$

and

$$H(X + Y) \leq 4H(X)$$

Remarks

- Proof of first part is an easy consequence of a Gaussian comparison inequality of

Continuous Plünnecke-Ruzsa inequality

Let A and $B_1, \ldots, B_m$ be convex bodies in $\mathbb{R}^n$, such that

$$\Big|A + B_i\Big|^{\frac{1}{n}} \le c_i |A|^{\frac{1}{n}}$$

for each $i \in [m]$. Then

$$\Big|A + \sum_{i\in[m]} B_i\Big|^{\frac{1}{n}} \le \Big[\prod_{i\in[m]} c_i\Big] |A|^{\frac{1}{n}}$$

Remarks

- Proved in [Bobkov-M.'12]; seem to be the first such upper bounds for volumes of Minkowski sums that do not use non-volumetric information and do not require invoking affine transformation
- This is the exact analogue of the discrete Plünnecke-Ruzsa inequality
- Unclear if such an inequality extends to larger classes of sets

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

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

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)

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

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

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

- Almost complete characterization of when a reverse EPI can hold
- Along the way, developed tools of independent interest:
  - Exponential concentration of information content for LC random vectors
  - A Gaussian comparison inequality for entropy of LC random vectors
  - Submodularity of entropy of convolutions
- Reverse EPI with explicit constants in IID case
- Beginnings of a probabilistic study of additive combinatorics on $\mathbb{R}^n$

Thank you for your attention!

○ — ○ — ○

Extras

Reminder: Three Useful Facts about Entropy

- Shannon's Chain Rule:

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

- 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

$$I(X;Y|Z) = h(X|Z) - h(X|Y,Z),$$

consequently $\quad h(X|Z) \geq h(X|Y,Z)$ ("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

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
- For general locally compact abelian groups, it is due to M.'08, Kontoyiannis-M.'13
- 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
- Differential entropy ($\mathcal{G} = \mathbb{R}$) 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:

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

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

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

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

By Lemma B, this is the same as

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

The entropy 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

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

mile-marker

$\surd$ Background: EPI and BMI

$\surd$ Reverse EPI for log-concave measures

$\surd$ More generally: When does a reverse EPI hold?

$\surd$ Key proof ideas

- Effective versions and additive combinatorics in $\mathbb{R}^n$

A Gaussian Comparison Inequality

If a random vector X in $\mathbb{R}^n$ has a log-concave density f, let Z in $\mathbb{R}^n$ be any normally distributed random vector with maximum density being the same as that of X. Then

$$\frac{1}{n}h(Z) - \tfrac{1}{2} \leq \frac{1}{n}h(X) \leq \frac{1}{n}h(Z) + \tfrac{1}{2}$$

Equality holds in LB iff $X \sim$ Unif(A), for a convex set A with non-empty interior. Equality holds in UB if X has coordinates that are i.i.d. exponentially distributed.

Remarks

- Suppose "amount of randomness" is measured by entropy per coordinate. Then *any LC random vector of any dimension* contains randomness that differs from that in the normal random variable with the same maximal density value by at most $1/2$

A Gaussian comparison inequality

Write $\|f\| = \operatorname{ess\,sup}_x f(x)$. If a random vector X in $\mathbb{R}^n$ has density f, then

$$\frac{1}{n} h(X) \geq \log \|f\|^{-1/n}.$$

If, in addition, f is log-concave, then

$$\frac{1}{n} h(X) \leq 1 + \log \|f\|^{-1/n},$$

with equality for the n-dimensional exponential distribution, concentrated on the positive orthant with density $f(x) = e^{-(x_1 + \cdots + x_n)}$, $x_i > 0$.

Remarks

- The lower bound is trivial and holds without any assumption on the density: $h(X) \geq \int_{\mathbb{R}^n} f(x) \log \frac{1}{\|f\|}\, dx = \log \frac{1}{\|f\|}$
- Observe that the maximum density of the $N(0, \sigma^2 I)$ distribution is $(2\pi\sigma^2)^{-n/2}$. Thus matching the maximum density of f and the isotropic normal Z leads to $(2\pi\sigma^2)^{1/2} = \|f\|^{-1/n}$, and $\frac{1}{n} h(Z) = \frac{1}{2}\log(2\pi e \sigma^2) = \frac{1}{2} + \log \|f\|^{-1/n}$. Thus the above inequality may be written as

$$\frac{\left|h(X) - h(Z)\right|}{n} \leq \tfrac{1}{2}$$

Proof of upper bound

By definition of log-concavity, for any $x, y \in \mathbb{R}^n$,

$$f(tx + sy) \geq f(x)^t f(y)^s, \; t, s > 0, t + s = 1.$$

Integrating with respect to x,

$$t^{-n} \int f(x)\, dx \geq f(y)^s \int f(x)^t\, dx.$$

Using the fact that $\int f = 1$ and maximizing over y, we obtain

$$t^{-n} \geq \|f\|^{1-t} \int f(x)^t\, dx$$

Observe that the left and right sides are equal for $t = 1$, and the left side dominates the right side for $0 < t \leq 1$. Thus we can compare derivatives in t of the two sides at $t = 1$. Specifically,

$$-n \leq -\log \|f\| + \int f(x) \log f(x)\, dx,$$

which yields the desired inequality. It is easy to check that a product of exponentials is an instance of equality.

Sidenote: EPI and Central Limit Theorem

For a random vector X in $\mathbb{R}^n$, the entropy power is $H(X) = e^{2h(X)/n}$. For any two independent random vectors X and Y in $\mathbb{R}^n$,

$$H(X+Y) \geq H(X) + H(Y) \quad [EPI]$$

Connection to Entropic CLT (say, on $\mathbb{R}$)

- $N(0, \sigma^2)$ has maximum entropy among all densities with variance σ^2
- If X_1 and X_2 are i.i.d., then $H(X_1 + X_2) \geq 2H(X_1)$ implies
  $$H\left(\frac{X_1 + X_2}{\sqrt{2}}\right) \geq H(X_1)$$
  using the scaling property $H(aX) = a^2 H(X)$
- Entropic CLT: Let X_i be i.i.d. with $EX_1 = 0$ and $EX_1^2 = \sigma^2$, and
  $$S_M = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} X_i$$
  Then under minimal conditions, as $M \to \infty$, $h(S_M) \uparrow h(N(0, \sigma^2))$
  [Barron '86, Artstein–Ball–Barthe–Naor '04, Barron–M.'07]

Entropy: reminder

When random vector $X \in \mathbb{R}^n$ has density $f(x)$, the entropy of X is

$$h(X) = h(f) := -\int f(x) \log f(x) dx = E[-\log f(X)]$$

Remarks

- 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 $f = g$
- For $X \sim f$ in $\mathbb{R}^n$, its relative entropy from Gaussianity is
$$D(f) := D(f \| f^G),$$
where f^G is the Gaussian with the same mean and covar. matrix as X
- *Fact*: For any f, $D(f) = h(f^G) - h(f)$
Implies: Under the variance constraint Var$(X) \leq \sigma^2$,
$$X \text{ has maximum entropy if } X \sim N(0, \sigma^2)$$

Log-concavity and Gaussianity

For $X \sim f$ in $\mathbb{R}^n$, let $h(X)$ or $h(f)$ denote its differential entropy, and let $D(f)$ denote its relative entropy from Gaussianity, i.e.,

$$D(f) = D(f\|g) = h(g) - h(f),$$

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

Theorem 2: [Log-concave densities are Gaussian-like]

Let f be any log-concave (LC) density on $\mathbb{R}^n$. Then

$$D(f) \leq \frac{1}{4} n \log n + O(n) =: C_n \quad \text{uniformly over all LC } f$$

Remarks

- Quantifies the intuition
- Based on a result of [Klartag '06] in convex geometry
- In fact, we conjecture that something much stronger is true

Entropic Form of Hyperplane Conjecture

Conjecture 1': For any LC density f on $\mathbb{R}^n$ and some universal constant c,

$$\frac{D(f)}{n} \leq c.$$

Remarks

- Theorem: Conjectures 1 and 1' are equivalent
- Pleasing formulation: The slicing problem is a statement about the (dimension-free) closeness of an arbitrary log-concave measure to a Gaussian measure

Another Entropic Form of Hyperplane Conjecture

For a random vector $X = (X_1, \ldots, X_n)$ in $\mathbb{R}^n$ with density $f(x)$, let $I(f)$ denote its *relative entropy from independence*, i.e.,

$$I(f) = D(f \| f_1 \otimes f_2 \otimes \ldots \otimes f_n)$$

where f_i denotes the i-th marginal of f

Conjecture 1": For any LC density f with identity covariance matrix on $\mathbb{R}^n$ and some universal constant c,

$$\frac{I(f)}{n} \leq c.$$

Remarks

- Theorem: Conjectures 1, 1' and 1" are equivalent
- Pleasing formulation: The slicing problem is a statement about the (dimension-free) closeness of an uncorrelated log-concave measure to a product measure

Conjecture 1' $\iff$ Conjecture 1”

The following identity is often used in information theory: if f is an arbitrary density on $\mathbb{R}^n$ and $f^{(0)}$ is the density of some product distribution (i.e., of a random vector with independent components), then

$$D(f\|f_0) = \sum_{i=1}^{n} D(f_i\|f_i^{(0)}) + I(f),$$

where f_i and $f_i^{(0)}$ denote the i-th marginals of f and $f^{(0)}$ respectively.

Now Conjecture 1' is equivalent to its restriction to those log-concave measures with zero mean and identity covariance (since $D(f)$ is an affine invariant). Applying the above identity to such measures,

$$D(f) = \sum_{i=1}^{n} D(f_i) + I(f),$$

since the standard normal is a product measure. By Theorem 2, each $D(f_i)$ is bounded from above by some universal constant since these are one-dimensional LC distributions. Thus $D(f)$ being uniformly $O(n)$ is equivalent to $I(f)$ being uniformly $O(n)$.

mile-marker

$\surd$ Background: EPI and BMI

$\surd$ Reverse EPI for log-concave measures

$\surd$ More generally: When does a reverse EPI hold?

- Key proof ideas

- Effective versions and additive combinatorics in $\mathbb{R}^n$

Entropy and Information Content

Let $X = (X_1, \ldots, X_n)$ be a random vector in $\mathbb{R}^n$, with (joint) density f. The random variable

$$\widetilde{h}(X) = -\log f(X)$$

may be thought of as the *information content* of X

Discrete case: $\widetilde{h}(X)$ is the number of bits needed to represent X by an optimal coding scheme [Shannon '48]

Continuous case: No coding interpretation, but may think of it as the log likelihood function in a nonparametric model

The entropy of X is defined by

$$h(X) = -\int f(x) \log f(x)\, dx = \mathbf{E}\widetilde{h}(X)$$

Remarks

- In general, $h(X)$ may or may not exist (in the Lebesgue sense); if it does, it takes values in the extended real line $[-\infty, +\infty]$
- h always exists and is finite for LC random vectors

Background: Shannon-McMillan-Breiman Theorem

Let $\mathbb{X}$ be a stationary, ergodic process, with $X^{(n)} = (X_1, \ldots, X_n) \in \mathbb{R}^n$ having joint density $f^{(n)}$ w.r.t Lebesgue measure on $\mathbb{R}^n$. Then

$$\frac{\widetilde{h}(X^{(n)})}{n} := -\frac{1}{n} \log f^{(n)}(X^{(n)}) \to h(\mathbb{X}) \quad \text{w.p. } 1$$

History

- If $\mathbb{X}$ is stationary, the limit $h(\mathbb{X}) = \lim_{n\to\infty} \frac{h(X^{(n)})}{n}$ typically exists, and is called the *entropy rate* of the process $\mathbb{X}$
- IID case is a simple instance of the Law of Large Numbers: if $X_i \sim f$,

$$-\frac{1}{n} \log f^{(n)}(X^{(n)}) = -\frac{1}{n} \sum_{i=1}^{n} \log f(X_i) \to h(X_1) \quad \text{w.p. } 1$$

- Has been called "the basic theorem of information theory"
- [Shannon '48, McMillan '53, Breiman '57] for discrete case; [Moy '61, Perez '64, Kieffer '74] partially for the continuous case; [Barron '85, Orey '85] for definitive version

A Motivation

The SMB theorem says

$$\frac{\widetilde{h}(X^{(n)})}{n} := -\frac{1}{n}\log f^{(n)}(X^{(n)}) \to h(\mathbb{X}) \quad \text{w.p. } 1$$

Asymptotic Equipartition Property: With high probability, the distribution of $X^{(n)}$ is effectively the uniform distribution on the class of typical observables, or the "typical set"

IID case: For some small fixed $\varepsilon > 0$, let

$$A = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : f(x_1, \ldots, x_n) \in [e^{-n[h(X_1)+\varepsilon]}, e^{-n[h(X_1)-\varepsilon]}]$$

Then $\Pr(X^{(n)} \in A) \to 1$, and distribution of $X^{(n)}$ on A is close to uniform

Applications

- *Likelihood*: Since the SMB Theorem describes the asymptotic behavior of the likelihood function, it and its relatives have strong implications for consistency of maximum likelihood estimators, etc.
- *Coding*: Just encode the typical set. . .

Concentration of Information Content

Given a random vector X in $\mathbb{R}^n$ with log-concave density f,

$$\mathbf{P}\left\{ \left| -\frac{1}{n}\log f(X)) - \frac{h(X)}{n} \right| \geq s \right\} \leq 4e^{-cns^2}, \qquad 0 \leq s \leq 2$$

where $c \geq 1/16$ is a universal constant

Remarks

- When X has i.i.d. components, the CLT suggests a Gaussian bound of this type. The theorem extends this for the special function $\log f$ to a large class with dependence, with a *universal constant*
- With high probability, the distribution of X itself is effectively the uniform distribution on the class of typical observables, or the "ε-typical set"

$$\{x \in \mathbb{R}^n : f(x) \in [e^{-h(X)-n\varepsilon}, e^{-h(X)+n\varepsilon}]\}$$

- Bound can be extended to all $s > 0$ at the cost of an $e^{-O(\sqrt{n}s)}$ bound

Proof of Reverse EPI

Let $Z \sim \text{Unif}(D)$, where D is the centered Euclidean ball with volume one. Since $H(Z) = 0$, Theorem 5 implies

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

for random vectors X and Y in $\mathbb{R}^n$ independent of each other and of Z.

Let X and Y have LC densities. Due to homogeneity of the reverse EPI, assume w.l.o.g. that $\|f\| \geq 1$ and $\|g\| \geq 1$. Then, our task reduces to showing that both $H(X+Z)$ and $H(Y+Z)$ can be bounded from above by universal constants.

For some affine volume preserving map $u : \mathbb{R}^n \to \mathbb{R}^n$, the distribution $\widetilde{\mu}$ of $\widetilde{X} = u(X)$ satisfies

$$\widetilde{\mu}(D)^{1/n} \geq c_0$$

with a universal constant $c_0 > 0$. Let $\tilde{f}$ denote the density of $\widetilde{X} = u(X)$. Then the density p of $S = \widetilde{X} + Z$, given by $p(x) = \int_D \tilde{f}(x-z)\,dz = \widetilde{\mu}(D - x)$, satisfies

$$\|p\| \geq p(0) \geq c_0^n$$

Applying Theorem 3' to the random vector S,

$$H(S) \leq C \, \|p\|^{-2/n} \leq C \cdot c_0^{-2}$$