Entropy and Information Inequalities Workshop
April 2013

Rearrangement and Entropy Inequalities

Liyao Wang

Yale University

Joint work with Mokshay Madiman (University of Delaware and Yale)

Differential Entropy and Entropy Power

Let X be a random vector in $\mathbb{R}^n$ with probability density f. Then the differential entropy of X, $h(X)$, is defined as

$$h(X) = -\int f(x)\log(f(x))dx$$

The entropy power of X, $N(X)$, is defined as

$$N(X) = e^{\frac{2}{n}h(X)}$$

Remarks

- The differential entropy differs in many ways from discrete entropy. For example, $h(AX) = h(X) + \log(|A|)$
- The differential entropy can take all values in $[-\infty, +\infty]$. Hence $N(X)$ can assume any values in $[0, +\infty]$
- If X is an isotropic Gaussian N$(0, \sigma^2\mathbb{I})$, then $N(X) = 2\pi e\sigma^2$

Original Entropy Power Inequality

The Entropy Power Inequality (EPI) of [Shannon '48, Stam '59] is as follows:

If X_1 and X_2 be two independent random vectors in $\mathbb{R}^n$:

$$e^{\frac{2h(X_1+X_2)}{n}} \geq e^{\frac{2h(X_1)}{n}} + e^{\frac{2h(X_2)}{n}}$$

Many motivations

- Useful in proving converses in coding theorems
- Implies, for i.i.d. X_i,

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

  and hence is a key step to the entropic Central Limit Theorem
- Implies the Heisenberg Uncertainty Principle

Equivalent formulation of EPI

2nd Formulation If X_1, X_2 are independent random vectors in $\mathbb{R}^n$ and Z_1, Z_2 are independent isotropic Gaussian random vectors in $\mathbb{R}^n$ such that $h(X_1) = h(Z_1), h(X_2) = h(Z_2)$,

$$h(X_1 + X_2) \geq h(Z_1 + Z_2)$$

with equality iff X_1 and X_2 are normal with proportional covariance matrices

Why? Let Z_1 and Z_2 be two independent, isotropic (i.e., with covariance matrix a multiple of identity) Gaussian random vectors in $\mathbb{R}^n$ such that

$$h(X_1) = h(Z_1), \qquad h(X_2) = h(Z_2)$$

Then, assuming the original formulation,

$$e^{\frac{2h(X_1+X_2)}{n}} \geq e^{2h(X_1)/n} + e^{2h(X_2)/n} = e^{2h(Z_1)/n} + e^{2h(Z_2)/n} = e^{2h(Z_1+Z_2)/n}$$

Conversely, assuming the second formulation,

$$e^{\frac{2h(X_1+X_2)}{n}} \geq e^{2h(Z_1+Z_2)/n} = e^{2h(Z_1)/n} + e^{2h(Z_2)/n} = e^{2h(X_1)/n} + e^{2h(X_2)/n}$$

Digression: An Open Problem

We define the Stam region to be the subset of $\mathbb{R}^{2^k-1}$ given by

$$\Gamma(k) = \{[N\Big(\sum_{i\in s} X_i\Big)]_{s\neq\emptyset\subset[k]} : X_1, X_2, \ldots, X_k$$

are independent random vectors (in some $\mathbb{R}^n$)}

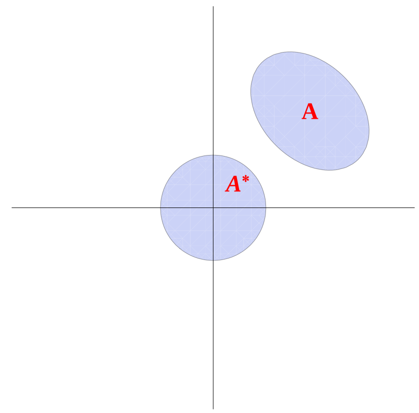
Question Can we characterize the Stam region?

Known Results

- It was shown in [Madiman and Barron '07] that the set function $v(s) = N\Big(\sum_{i\in s} X_i\Big)$ is fractionally superadditive
- We can show that the set function $v(s)$ is NOT always supermodular

A Brief Introduction to Rearrangements

The rearrangement of a set $A \subset \mathbb{R}^n$ is just the Euclidean ball with the same volume as A centered at 0 and is denoted A^*



For a function $f : \mathbb{R}^n \to [0, \infty)$, let $A_t = \{x : f(x) \geq t\}$. Define

$$f^*(x) = \int_0^\infty 1_{\{x \in {A_t}^*\}} dt$$

where $1_{\{x \in {A_t}^*\}} = 1$ if $x \in {A_t}^*$ and 0 otherwise. We call f^* the spherically symmetric decreasing rearrangement of f

Remarks

- Note that $f(x) = \int_0^{f(x)} dt = \int_0^\infty 1_{\{f(x) \geq t\}} dt = \int_0^\infty 1_{\{x \in A_t\}} dt$
- So the idea of the definition is to build up f^* from the rearranged super-level sets in the same way that we can build f from its super-level sets

Discrete Analogues of f^*

Consider a probability mass function $p(i)$ on $\mathbb{Z}$. Let us define p^+ to be a permutation of p

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

and ${}^+p$ to be a permutation of p

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

If $p^+ = {}^+p$, we define

$$p^* = p^+$$

Note that p^* is symmetric and decreasing on the set of non-negative integers. This is precisely the discrete analogue of f^*

Example

Let $f(x) = \frac{1}{\sqrt{(2\pi)^n|A|}} e^{-\frac{x^T A^{-1} x}{2}}$ be a non-degenerate Gaussian with covariance matrix A. Then

$$A_t = \{y : f(y) > t\} = \{y : y^T A^{-1} y < -2\log(t\sqrt{(2\pi)^n|A|})\}$$

if $t < \frac{1}{\sqrt{(2\pi)^n|A|}}$ and empty otherwise

$$A_t^* = \{y : y^T y < -2|A|^{\frac{1}{n}} \log(t\sqrt{(2\pi)^n|A|})\}$$

if $t < \frac{1}{\sqrt{(2\pi)^n|A|}}$ and empty otherwise. Recall that:

$$f^*(x) = \int_0^\infty 1_{\{x \in A_t{}^*\}} dt$$

But $1_{\{x \in A_t{}^*\}} = 1$ if and only if $t < \frac{1}{\sqrt{(2\pi)^n|A|}} e^{-\frac{x^T x}{2|A|^{\frac{1}{n}}}}$. Hence finally,

$$f^*(x) = \frac{1}{\sqrt{(2\pi)^n|A|}} e^{-\frac{x^T x}{2|A|^{\frac{1}{n}}}}$$

Some Lemmas

Lemma 1 For any function $f : \mathbb{R}^n \to [0, \infty)$, and all $t \geq 0$,

$$\{x : f(x) > t\}^* = \{x : f^*(x) > t\}$$

Lemma 2 (Rearrangement preserves densities) For any function $f : \mathbb{R}^n \to [0, \infty)$, and any $p \geq 1$,

$$\|f\|_p = \|f^*\|_p$$

In particular, the rearrangement of a density function is a density function

Lemma 3 (Rearrangement preserves entropy) For any random vector X with density f,

$$h(X) = h(X^*)$$

where X^* is distributed according to f^*

Remark

Lemma 1 implies that f^* is a spherically symmetric decreasing function (i.e., $f^*(x)$ only depends on $|x|$ and is non-increasing in it) such that, for any measurable subset $B \subset [0, \infty)$, the volumes of the sets $\{x : f(x) \in B\}$ and $\{x : f^*(x) \in B\}$ are the same

Main Result: An EPI with Rearrangements

Suppose X_1 and X_2 are two independent random vectors in $\mathbb{R}^n$ with densities f_1 and f_2. Then

$$h(X_1 + X_2) \geq h(X_1^* + X_2^*)$$

where X_1^* and X_2^* are independent with densities f_1^* and f_2^*

Remarks

- Since $h(X_i^*) = h(X_i)$, the 2nd formulation of EPI applied to X_1^* and X_2^* implies

  $$h(X_1^* + X_2^*) \geq h(Z_1 + Z_2)$$

  so that we can write

  $$h(X_1 + X_2) \geq h(X_1^* + X_2^*) \geq h(Z_1 + Z_2)$$

  Hence it can indeed be seen as a kind of strengthening

- However, note that this does not directly give a new proof of the EPI since we used the EPI to show that it was a strengthening!

Rényi Entropy of Order p

Definition

$$h_p(X) = \frac{1}{1-p} \log \int f^p(x)dx$$

where $p \in (0,1) \cup (1, +\infty)$

Remarks

- $h_1(X)$ is defined as the Shannon differential entropy

$$h(X) = h_1(X) = -\int_{\mathbb{R}^n} f(x) \log f(x) dx$$

- As $p \to 0$, $h_p(X)$ reduces to

$$h_0(X) = \log |\text{Supp}(f)|$$

where $\text{Supp}(f) = \{x : f(x) > 0\}$ is the support of the density f

- As $p \to +\infty$, $h_p(X)$ reduces to

$$h_\infty(X) = -\log(\text{esssup}_x f(x))$$

where $\text{esssup}_x f(x)$ is the essential supremum of f

A More General Result

Suppose X_1 and X_2 are two independent random vectors in $\mathbb{R}^n$ with densities f_1 and f_2. Then

$$h_p(X_1 + X_2) \geq h_p(X_1^* + X_2^*)$$

where X_1^* and X_2^* are independent with densities f_1^* and f_2^* and $p \in [0, +\infty]$

Remarks

- In fact, we proved the following, which is even stronger

$$\int \phi(f_1 \star f_2 \star \cdot \cdot \star f_k)(x)dx \leq \int \phi(f_1^* \star f_2^* \star \cdot \cdot f_k^*)(x)dx$$

  where ϕ is a continuous convex function on $[0, \infty)$ such that $\phi(0) = 0$

- By taking $\phi = x^p$ for $p > 1$, $\phi = x \log(x)$ for $p = 1$ and $\phi = -x^p$ for $p < 1$, we obtain

$$h_p(X_1 + X_2 + \cdot \cdot + X_k) \geq h_p(X_1^* + X_2^* + \cdot \cdot + X_k^*)$$

- The proof is based on the generalized Riesz Sobolev convolution inequality and ideas from majorization theory

Application 1

Question Can we get a new proof of the Original EPI from our Main Result?

Special Case: Assume X_1 and X_2 are IID

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

Proof

- It suffices to show

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

since then by our main result and Lemma 3 (Rearrangement preserves entropy)

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

- By independence, $h(X_1^*, X_2^*) = h(X_1^*) + h(X_2^*)$. By spherical symmetry and i.i.d. assumption

$$\frac{X_1^* + X_2^*}{\sqrt{2}} =^d \frac{X_1^* - X_2^*}{\sqrt{2}}$$

By the scaling property for entropy and subadditivity

$$h(X_1^*, X_2^*) = h\left(\frac{X_1^* + X_2^*}{\sqrt{2}}, \frac{X_1^* - X_2^*}{\sqrt{2}}\right) \leq h\left(\frac{X_1^* + X_2^*}{\sqrt{2}}\right) + h\left(\frac{X_1^* - X_2^*}{\sqrt{2}}\right)$$

Hence

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

Application 2

Brunn-Minkowski inequality

Let A, B be any Borel 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. Then

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

Proof

BM follows from $p = 0$ case of our more general result. The Brunn-Minkowski inequality can be rewritten as

$$\left|A+B\right|^{\frac{1}{n}} \geq \left|A^* + B^*\right|^{\frac{1}{n}}$$

But the $p = 0$ case of our strengthened result implies

$$\left|A+B\right| \geq \left|\text{supp}(X+Y)\right| \geq \left|\text{supp}(X^* + Y^*)\right| = \left|A^* + B^*\right|$$

where all random vectors are independent and X and Y are uniform distributions on A and B

Summary

- A New EPI refining the classical one
- May be useful in reducing information-theoretic problems to the special case involving spherically symmetric, decreasing densities, for which the symmetry can be used
- Recovers several classical inequalities as corollaries

Thank you!

○ — ○ — ○