High Dimensional Probability VII Conference
Cargèse, 26–30 May 2014

Rearrangements and the Entropy Power Inequality

Mokshay Madiman

University of Delaware

Joint work with Liyao Wang (Yale University)

Original Entropy Power Inequality

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

EPI: Suppose X_1 and X_2 are two independent random vectors in $\mathbb{R}^n$ with finite differential entropies. 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

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

Many motivations

- 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
- Useful in proving coding theorems in information theory
- Implies the Heisenberg Uncertainty Principle, and the Log Sobolev Inequality for Gaussian measures

A Comment and a Question

Comment: The functional $N(X) = e^{\frac{2h(X)}{n}}$ is called the "entropy power" of X. The EPI stated above is equivalent to the usual formulation: If X_1, X_2 are independent random vectors in $\mathbb{R}^n$,

$$N(X_1 + X_2) \geq N(X_1) + N(X_2)$$

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 Gaussian random vectors in $\mathbb{R}^n$ such that

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

Then

$$N(X_1) + N(X_2) = N(Z_1) + N(Z_2) = N(Z_1 + Z_2)$$

using the fact that for an isotropic Gaussian Z, $N(Z)$ is (up to an absolute constant) the variance of each component

Question: Can we refine the EPI, in the sense that we insert a meaningful quantity in between the two sides?

Reminder: What are Rearrangements?

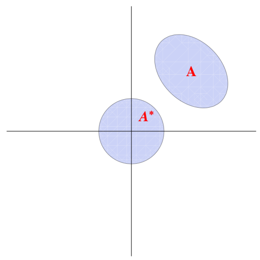
For a function $f : \mathbb{R}^n \to [0, \infty)$, let $A_t = \{x : f(x) \geq t\}$. Define

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

where $1_{\{x \in A_t{}^*\}} = 1$ if $x \in A_t{}^*$ and 0 otherwise.

Remarks

- Note that $f(x) = \int_0^{f(x)} dt = \int_0^\infty 1_{\{f(x) \geq t\}} dt = \int_0^\infty 1_{\{x \in A_t\}} dt$
- 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^*
- Rearrangements for sets $\subset \mathbb{R}^n$:



- 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

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 density f,

$$h(f) = h(f^*)$$

Remarks

- Lemma 1 says 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
- Lemma 1 and Lemma 2 are classical; Lemma 3 appears to be a new observation but is not very difficult
- Basic idea of proofs: Tonelli's theorem

An entropy inequality with rearrangements

Theorem: Suppose X_1 and X_2 are two independent random vectors in $\mathbb{R}^n$ with densities f_1 and f_2, such that . Then, provided they exist,

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

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

Remarks

- We will show that $h(X_i^*) = h(X_i)$. Hence, the 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!

Beyond Shannon entropy

Rényi entropy of order p: For $p > 1$,

$$h_p(X) = \frac{p}{p-1} \log \frac{1}{\|f\|_p}$$

where $\|f\|_p = \left(\int_{\mathbb{R}^n} f^p \, dx \right)^{1/p}$ is the usual L^p-norm on $\mathbb{R}^n$

The definition of $h_p(X)$ continues to make sense for $p \in (0,1)$ even though $\|f\|_p$ is then not a norm.

Special values of p: For $p = 0, 1, \infty$, $h_p(X)$ is defined "by continuity". In particular, as $p \to 1$, $h_p(X)$ reduces to the Shannon entropy

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

Also

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

where $\text{Supp}(f) =$ is the support of the density f, and

$$h_\infty(X) = -\log \|f\|_\infty$$

The General Results

Main Theorem: If X_i are independent $\mathbb{R}^n$-valued random vectors with densities f_i, and X_i^* are independent random vectors with densities f_i^*,

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

for any $p \in [0, \infty]$, provided both sides are well defined

Remarks

- Suppose $f_i, 1 \leq i \leq k$ are probability densities. Let $\phi(x)$ be a convex function defined on the non-negative real line such that $\phi(0) = 0$ and ϕ is continuous at 0. Then we actually show the even stronger result:

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

provided that both sides are well defined

Proof ideas

For densities f and g on $\mathbb{R}^n$, we say that f is *majorized* by g if

$$\int_{\{x:\|x\|<r\}} f^*(x)dx \leq \int_{\{x:\|x\|<r\}} g^*(x)dx$$

for all $r > 0$. In this case, we write $f \prec g$

Fact: Let $\phi(x)$ be a convex function defined on $[0, \infty)$ such that $\phi(0) = 0$ and ϕ is continuous at 0. If f and g are densities with $f \prec g$, then

$$\int \phi(f(x))dx \leq \int \phi(g(x))dx,$$

provided that both sides are well defined

Remarks

- For non-negative ϕ, the Fact was proved by [Burchard '94]
- By taking $\phi(x) = x^p$ for $p > 1$, $\phi(x) = -x^p$ for $0 < p < 1$ and $\phi(x) = x\log(x)$ for $p = 1$, we recover the Main Theorem
- The fact that $f = f_1 \star f_2 \star \cdots \star f_k \prec g = f_1^* \star f_2^* \star \cdots \star f_k^*$ can be obtained as a consequence of the classical Rogers-Brascamp-Lieb-Luttinger inequalities for rearrangements [Rogers '57, Brascamp–Lieb–Luttinger '74]

The Unifying Character of The General Result

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

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

Remarks

- BM follows from $p = 0$ case
- It has long been observed that BM resembles the EPI, and the two can be given a common proof via Young's inequality for convolution with a sharp constant [Lieb '78, Dembo–Cover–Thomas '91]
- $p = \infty$ case yields as a corollary an inequality due to Riesz and Sobolev

Implication for Fisher information inequalities

Recall the definition of Fisher information $I(f)$ for a density f:

$$I(f) = \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f}$$

Corollary:

$$I(f) \geq I(f^*).$$

Remarks

- This turns out to be equivalent to the classical Pólya-Szegő inequality $\|\nabla f\|_p \geq \|\nabla f^*\|_p$ ($p \in [1, \infty]$) for $p = 2$, and gives a different proof of it
- Let g be an isotropic Gaussian density such that $h(g) = h(f)$. Then

$$I(f) \geq I(g) = \frac{1}{N(f)} \quad \text{[Stam '59]}$$

  is the "isoperimetric inequality for entropy". Just as our Main Theorem strengthens the EPI, the above strengthens the isoperimetric inequality for entropy since it inserts $I(f^*)$ in between $I(f)$ and $I(g)$

mile-marker

$\surd$ Background: EPI; Rearrangements

$\surd$ An entropy inequality involving rearrangements

$\surd$ General formulations and proof ideas

$\surd$ The unifying character of the general result

- A conjecture involving generalized Gaussians
- A new approach to the classical EPI
- Discrete analogues

Towards a Rényi EPI

The *Rényi entropy power* of order p is $N_p(X) = \exp\{\frac{2h_p(X)}{n}\}$

Theorem:[Bobkov–Chistyakov '13] If $X_1, \ldots, X_k$ are independent random vectors taking values in $\mathbb{R}^n$, then for any $p \geq 1$,

$$N_p(X_1 + \ldots + X_k) \geq c_p \sum_{i=1}^{k} N_p(X_i),$$

where c_p is a constant depending only on p

Remarks

- For $p \in (1, \infty)$, [Bobkov–Chistyakov '13] showed that $c_p = \frac{1}{e} p^{\frac{1}{p-1}}$ works
- The $p = 1$ case is simply the EPI, which is sharp for Gaussians with proportional covariance matrices
- For $p = 1$, [Rogozin '87] showed that $c_\infty = \frac{1}{2}$ is sharp when $n = 1$, while [Bobkov–Chistyakov '12] showed that $c_\infty = \frac{1}{2}$ is sharp when $k = 2$
- In general, the theorem is not sharp

Generalized Gaussians

For $-\infty < \beta \le \frac{2}{n+2}$, the *standard generalized Gaussian* of order β has density

$$g_\beta(x) = A_\beta \left(1 - \frac{\beta}{2}\|x\|^2\right)_+^{\frac{1}{\beta}-\frac{n}{2}-1},$$

where A_β is a normalizing constant

Remarks

- g_0 is the standard Gaussian density in $\mathbb{R}^n$
- For $\beta < 0$, g_β is proportional to a negative power of $(1 + b\|x\|^2)$ for a positive constant b, and therefore correspond to heavy-tailed measures with full support on $\mathbb{R}^n$ (Student-t distributions)
- For $0 < \beta \le \frac{2}{n+2}$, g_β is a concave function supported on a centered Euclidean ball of finite radius (Student-r distributions)
- Although the first class includes many distributions from what one might call the "Cauchy family", it excludes the standard Cauchy distribution. Indeed, the form of g_β has been chosen so that, for $Z \sim g_\beta$, $E[\|Z\|^2] = n$ for any β

A conjecture involving generalized Gaussians

$$\text{For } p > \frac{n}{n+2}, \quad \text{define} \quad \frac{1}{\beta_p} = \frac{1}{p-1} + \frac{n+2}{2}$$

Fact:[Costa–Hero–Vignat '03, Lutwak–Yang–Zhang '07] If X is a random vector taking values in $\mathbb{R}^n$, then for any $p > \frac{n}{n+2}$,

$$\frac{E[\|X\|^2]}{N_p(X)} \geq \frac{E[\|Z^{(p)}\|^2]}{N_p(Z^{(p)})},$$

where $Z^{(p)} \sim g_{\beta_p}$

Conjecture: Let $X_1, \ldots, X_k$ be independent random vectors taking values in $\mathbb{R}^n$, and $p > \frac{n}{n+2}$. Suppose Z_i are independent random vectors, each a scaled version of $Z^{(p)}$. such that $h_p(X_i) = h_p(Z_i)$. Then

$$h_p(X_1 + \ldots + X_k) \geq h_p(Z_1 + \ldots + Z_k).$$

Remarks

- The conjecture suggests optimal constants for the Rényi EPI in general
- It is true for the three special cases where sharp constant is known
- Our main result "refines" this conjecture

A New Proof of the EPI: IID case

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

Folklore proof for symmetric densities

By the scaling property for entropy, and by subadditivity,

$$h(Y_1, Y_2) \ = \ h\left(\frac{Y_1 + Y_2}{\sqrt{2}}, \frac{Y_1 - Y_2}{\sqrt{2}}\right) \ \leq \ h\left(\frac{Y_1 + Y_2}{\sqrt{2}}\right) + h\left(\frac{Y_1 - Y_2}{\sqrt{2}}\right)$$

But, by independence, $h(Y_1, Y_2) = h(Y_1) + h(Y_2)$, and by spherical symmetry (in fact, we only need central symmetry) and i.i.d. assumption, we have

$$Y_1 + Y_2 \ =^{\mathcal{D}} \ Y_1 - Y_2.$$

Hence

$$h\left(\frac{Y_1 + Y_2}{\sqrt{2}}\right) \geq h(Y_1) \tag{1}$$

Remarks

- The Main Result implies $h(Y_1 + Y_2) \geq h(Y_1^* + Y_2^*)$, which combined with the folklore observation, yields (1) for all i.i.d. Y_1, Y_2
- There are many proofs of EPI [Stam '59, Blachman '65, Lieb '78, Szarek–Voiculescu '00, Verdú–Guo '06, Rioul '11]

A New Proof of the EPI: general case

Our goal is to recover, from our Main Theorem, the following formulation of EPI: for any $0 < \lambda < 1$,

$$h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \geq \lambda h(X) + (1-\lambda)h(Y)$$

where, for notational simplicity, we will only consider $n = 1$

Three Reductions of the Problem

1. By the Main Theorem, we can assume that X and Y are symmetric, unimodal random variables
2. By the tensorization trick, it is sufficient to show

$$h(\sqrt{\lambda}\mathbf{X} + \sqrt{1-\lambda}\mathbf{Y}) \geq \lambda h(\mathbf{X}) + (1-\lambda)h(\mathbf{Y}) + o(M)$$

   where $\mathbf{X}$ and $\mathbf{Y}$ consist of M independent copies of X and Y respectively
3. Suppose $X \sim f$ and $Y \sim g$. Standard approximation arguments allow us to take f and g to be simple functions, i.e., mixtures of uniform distributions on symmetric intervals. Then:
   - densities of $\mathbf{X}$ and $\mathbf{Y}$ are simple functions
   - densities of $\mathbf{X}^*$ and $\mathbf{Y}^*$ are mixtures of uniforms on balls

Concavity Properties of Entropy

If $p = \sum_{i=1}^{r} c_i f_i$ is a mixture of densities, then

$$\sum_{i=1}^{r} c_i h(f_i) \le h(p) \le \sum_{i=1}^{r} c_i h(f_i) + \log(r)$$

Remarks

- The right inequality– a "reverse concavity" of the entropy functional– is very easy but seems to be a new observation
- Since a convolution of mixtures is a mixture of convolutions, the above concavity properties can be used to reduce the EPI involving $\mathbf{X}$ and $\mathbf{Y}$ (from previous slide) to just an EPI involving uniforms on balls
- By comparison with Gaussians and a use of Stirling's formula, we can show that

$$h(\sqrt{\lambda}\mathbf{Z_1} + \sqrt{1-\lambda}\mathbf{Z_2}) \ge \lambda h(\mathbf{Z_1}) + (1-\lambda)h(\mathbf{Z_2}) - C\log(M)$$

  where C is a universal constant and $\mathbf{Z_1}, \mathbf{Z_2}$ are uniforms on balls of possibly different radii

Summary

- We can refine the EPI using rearrangements. In fact, such an inequality (the Main Theorem) extends to Rényi entropies of any order $p \in [0, \infty]$
- EPI can be obtained as a consequence of the Main Theorem
- The Main Theorem gives a clean unification of EPI and Brunn-Minkowski inequality (and also of some other inequalities)
- There is an analogue on the integers, which has connections to additive combinatorics