On Rényi Entropy Power Inequalities

Eshed Ram

Andrew and Erna Viterbi Faculty of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel

M.Sc. thesis supervised by Igal Sason

January 19, 2017.

Outline

Entropy Power

Definition 1 (Entropy Power)

Let X be a d-dimensional random vector (r.v.) with differential entropy $h(X)$. The entropy power of X is

$$N(X) = \exp\left(\tfrac{2}{d}\, h(X)\right).$$

Entropy Power

Definition 1 (Entropy Power)

Let X be a d-dimensional random vector (r.v.) with differential entropy $h(X)$. The entropy power of X is

$$N(X) = \exp\left(\tfrac{2}{d}\, h(X)\right).$$

- $\frac{2}{d}$ in the exponent implies homogeneity of order 2:

$$N(\lambda X) = \lambda^2 N(X),\ \forall \lambda \in \mathbb{R}.$$

- If $X \sim N(0, \sigma^2 I_d)$, then $h(X) = \frac{d}{2}\log(2\pi e \sigma^2)$, and

$$N(X) = 2\pi e \sigma^2.$$

In some definitions the entropy power is normalized by $2\pi e$.

The Entropy Power Inequality

Introduced by Shannon in his 1948 fundamental paper: "A mathematical theory of communication."

The Entropy Power Inequality (EPI)

Let $\{X_k\}_{k=1}^n$ be independent r.v.'s. Then,

$$N\left(\sum_{k=1}^n X_k\right) \geq \sum_{k=1}^n N(X_k)$$

and equality holds if and only if $\{X_k\}_{k=1}^n$ are Gaussians with proportional covariances.

Applications of the EPI

Converse theorems for...

- The capacity region of the Gaussian broadcast channel - Bergmans, 1974
- The rate-equivocation region of the Gaussian wire-tap channel - Leung-Yan-Cheong & Hellman, 1978.
- The capacity region of the Gaussian interference channel - Costa, 1985.
- Multi-terminal rate-distortion theory (the quadratic Gaussian CEO problem) - Oohama, 1998.
- The capacity region of the Gaussian broadcast MIMO channel - Weingarten, Steinberg & Shamai, 2006.

Rényi's Entropy

Definition 2

Let X be a d-dimensional r.v. with density f_X, and $\alpha \in (0,1) \cup (1,\infty)$. The order-$\alpha$ Rényi entropy of X is

$$h_\alpha(X) = \frac{\alpha}{1-\alpha} \log \|f_X\|_\alpha = \frac{1}{1-\alpha} \log \left(\int_{\mathbb{R}^d} f_X^\alpha(x)\,\mathrm{d}x \right).$$

Rényi's Entropy

Definition 2

Let X be a d-dimensional r.v. with density f_X, and $\alpha \in (0,1) \cup (1,\infty)$. The order-$\alpha$ Rényi entropy of X is

$$h_\alpha(X) = \frac{\alpha}{1-\alpha} \log \|f_X\|_\alpha = \frac{1}{1-\alpha} \log \left(\int_{\mathbb{R}^d} f_X^\alpha(x)\,\mathrm{d}x \right).$$

By continuous extension in α, we have

- $h_0(X) = \log \mu\big(\mathrm{supp}(f_X)\big)$.
- $h_1(X) = h(X) = -\int_{\mathbb{R}^d} f_X(x) \log f_X(x)\,\mathrm{d}x$.
- $h_\infty(X) = -\log\big(\operatorname{ess\,sup}(f_X)\big)$.

where μ is the Lebesgue measure in $\mathbb{R}^d$.

Properties of Rényi's Entropy

Let X be a d-dimensional r.v. with density.

- $h_\alpha(X)$ is continuous in $\alpha \in [0, \infty]$
- $h_\alpha(X)$ is monotonically non-increasing in $\alpha \in [0, \infty]$,

$$0 \leq \beta \leq \alpha \implies h_\beta(X) \geq h_\alpha(X).$$

- If $X = (X_1, \ldots, X_d)$ has independent elements, then

$$h_\alpha(X) = \sum_{k=1}^{d} h_\alpha(X_k), \quad \forall \alpha \in [0, \infty].$$

  (similar to Shannon's entropy)
- Unlike Shannon's entropy,

$$h_\alpha(X) \nleq \sum_{k=1}^{d} h_\alpha(X_k).$$

Rényi's Entropy Power

Definition 3

- *Let X be a d-dimensional r.v. with density.*
- *Let $\alpha \in [0, \infty]$.*

The Rényi entropy power of X is

$$N_\alpha(X) = \exp\left(\tfrac{2}{d}\, h_\alpha(X)\right).$$

Rényi's Entropy Power

Definition 3

- *Let X be a d-dimensional r.v. with density.*
- *Let $\alpha \in [0, \infty]$.*

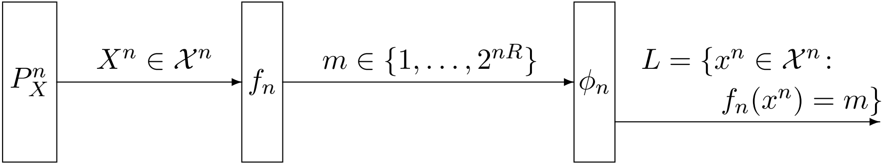
The Rényi entropy power of X is

$$N_\alpha(X) = \exp\left(\tfrac{2}{d}\, h_\alpha(X)\right).$$

Homogeneity of order 2:

$$N_\alpha(\lambda X) = \lambda^2 N_\alpha(X),\ \forall \lambda \in \mathbb{R},\ \alpha \in [0, \infty].$$

An Application of The Rényi Entropy - Example



- Fixed $\rho > 0$: rate R is called achievable if there exist encoders $\{f_n\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty} \mathbb{E}\left[|L|^{\rho}\right] = 1$.
- Direct and converse results: [1]

$$R > H_{\frac{1}{1+\rho}}(X) \Rightarrow R \text{ is achievable}$$
$$R < H_{\frac{1}{1+\rho}}(X) \Rightarrow R \text{ is not achievable}$$

[1] Bunte and Lapidoth, "Encoding Tasks and Rényi Entropy", *IEEE Trans. on Information Theory*, Sept. 2014."

A Rényi EPI (R-EPI)?

- Formally,
  1. Let $\{X_k\}_{k=1}^{n}$ be d–dimensional independent r.v.'s with densities.
  2. Let $\alpha \in [0, \infty]$, $n \in \mathbb{N}$.

  Does there exist a positive constant $c_\alpha^{(n,d)}$ such that

$$N_\alpha\left(\sum_{k=1}^{n} X_k\right) \geq c_\alpha^{(n,d)} \sum_{k=1}^{n} N_\alpha(X_k) \, ?$$

A Rényi EPI (R-EPI)?

- Formally,
  1. Let $\{X_k\}_{k=1}^n$ be d–dimensional independent r.v.'s with densities.
  2. Let $\alpha \in [0, \infty]$, $n \in \mathbb{N}$.

  Does there exist a positive constant $c_\alpha^{(n,d)}$ such that

$$N_\alpha\left(\sum_{k=1}^n X_k\right) \geq c_\alpha^{(n,d)} \sum_{k=1}^n N_\alpha(X_k) \ ?$$

- For independent Gaussian random vectors with proportional covariances, $N_\alpha\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n N_\alpha(X_k)$, for every $\alpha \in [0, \infty]$.

$$\Longrightarrow \ c_\alpha^{(n,d)} \leq 1, \quad \forall \alpha \in [0, \infty].$$

Related Work - EPI

1. Shannon, 1948 - the entropy power inequality (EPI)
   - Many information-theoretic proofs have been suggested (e.g., Stam - 1959, Gou-Shamai-Verdú - 2006, Rioul - 2011).
2. Zamir and Feder, 1993: a vector generalization of the EPI.
3. Baron and Madiman, 2007: Some generalizations of the EPI, and connection to the CLT.
4. EPI for discrete random variables:
   - Harremöes and Vignat, 2003.
   - Jog and Anantharam, 2014.
   - Telatar *et al.*, 2014.
5. Costa (1985), Toscani (2015) and Courtade (ISIT 2016): strengthening the EPI by restriction to some families of distributions.

Related Work - Rényi EPI

1. Bercher and Vignat (BV), 2002: for every $\alpha \in [0, \infty]$,
$$N_\alpha \left(\sum_{k=1}^n X_k\right) \geq \max_{1 \leq k \leq n} N_\alpha(X_k).$$
2. Wang, Woo & Madiman, 2014: lower bound on the Rényi entropy of convolutions in the integers.
3. Bobkov and Chistyakov (BC), 2015: for every $\alpha > 1$,
$$c_\alpha = \frac{1}{e}\alpha^{\frac{1}{\alpha-1}} \quad \text{(independently of } d \text{ and } n\text{).}$$
4. Wang and Madiman, 2014: conjectures on the optimal R-EPI.
5. Xu, Melboune & Madiman, ISIT 2016: reverse Rényi EPIs for s-concave densities ($s = 1 \Rightarrow$ log-concavity).

Our work provides the tightest R-EPIs known so far, for $\alpha > 1$.

Theorem 1

Let

- $\{X_k\}_{k=1}^n$ *be* d*–dimensional independent r.v's with densities.*
- $\alpha > 1$, $\alpha' = \frac{\alpha}{\alpha-1}$.
- $n \in \mathbb{N}$.

Then, the following R-EPI holds:

$$N_\alpha\left(\sum_{k=1}^{n} X_k\right) \geq c_\alpha^{(n)} \sum_{k=1}^{n} N_\alpha(X_k),$$

with

$$c_\alpha^{(n)} = \alpha^{\frac{1}{\alpha-1}} \left(1 - \frac{1}{n\alpha'}\right)^{n\alpha'-1}.$$

Theorem 1 Implications

Theorem 1 $\Rightarrow$ BC bound

Theorem 1 improves the R-EPI by Bobkov and Chistyakov ($c_\alpha = \frac{1}{e}\alpha^{\frac{1}{\alpha-1}}$) for every $\alpha > 1$ and $n \in \mathbb{N}$; for every $\alpha > 1$, it asymptotically coincides with the R-EPI by Bobkov and Chistyakov as $n \to \infty$.

Theorem 1 Implications

Theorem 1 $\Rightarrow$ BC bound

Theorem 1 improves the R-EPI by Bobkov and Chistyakov ($c_\alpha = \frac{1}{e}\alpha^{\frac{1}{\alpha-1}}$) for every $\alpha > 1$ and $n \in \mathbb{N}$; for every $\alpha > 1$, it asymptotically coincides with the R-EPI by Bobkov and Chistyakov as $n \to \infty$.

Theorem 1 $\Rightarrow$ EPI

If $\alpha \downarrow 1$, Theorem 1 coincides with the EPI.

Theorem 1 Implications

Theorem 1 ⇒ BC bound

Theorem 1 improves the R-EPI by Bobkov and Chistyakov ($c_\alpha = \frac{1}{e}\alpha^{\frac{1}{\alpha-1}}$) for every $\alpha > 1$ and $n \in \mathbb{N}$; for every $\alpha > 1$, it asymptotically coincides with the R-EPI by Bobkov and Chistyakov as $n \to \infty$.

Theorem 1 ⇒ EPI

If $\alpha \downarrow 1$, Theorem 1 coincides with the EPI.

Asymptotic Tightness of the Result in Theorem 1

If $n = 2$ and $\alpha \to \infty$, $c_\alpha^{(n)}$ tends to $\frac{1}{2}$ which is optimal; achieved when X_1 and X_2 are uniformly distributed in the cube $[0,1]^d$.

$c_\alpha^{(n)}$ as a function of α

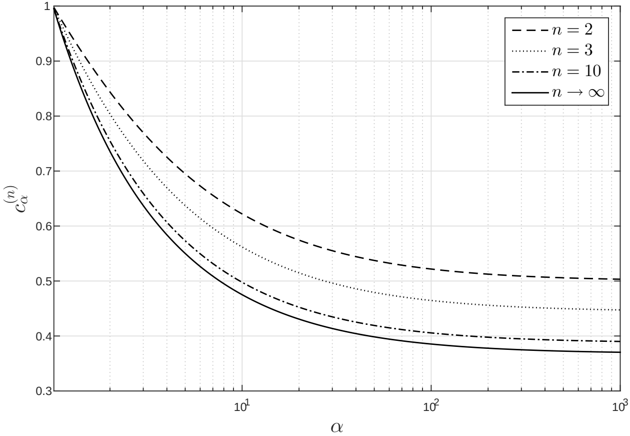


Figure: $c_\alpha^{(n)}$ as a function of α for $n = 2, 3, 10$ and $n \to \infty$

Outline of the Proof of Theorem 1

Main Tool: The Sharpened Young's Inequality

Let $p, q, r \geq 1$ satisfy $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ and let $f \in L^p(\mathbb{R}^d)$ and $g \in L^q(\mathbb{R}^d)$ be non-negative functions. Then

$$\|f * g\|_r \leq \left(\frac{A_p A_q}{A_r} \right)^{\frac{d}{2}} \|f\|_p \, \|g\|_q,$$

where $A_t = t^{\frac{1}{t}} \, t'^{-\frac{1}{t'}}$ and $t' = \frac{t}{t-1}$. Equality holds if and only if f and g are Gaussians or $r = p = q = 1$.

Outline of the Proof of Theorem 1

Main Tool: The Sharpened Young's Inequality

Let $p, q, r \geq 1$ satisfy $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$ and let $f \in L^p(\mathbb{R}^d)$ and $g \in L^q(\mathbb{R}^d)$ be non-negative functions. Then

$$\|f * g\|_r \leq \left(\frac{A_p A_q}{A_r}\right)^{\frac{d}{2}} \|f\|_p \, \|g\|_q,$$

where $A_t = t^{\frac{1}{t}} \, t'^{-\frac{1}{t'}}$ and $t' = \frac{t}{t-1}$. Equality holds if and only if f and g are Gaussians or $r = p = q = 1$.

- Reversed for $p, q, r \in (0, 1]$.
- Using mathematical induction:

$$\|f_1 * \ldots * f_n\|_\nu \leq A \prod_{k=1}^{n} \|f_k\|_{\nu_k}, \quad A = \left(\frac{1}{A_\nu} \prod_{k=1}^{n} A_{\nu_k}\right)^{\frac{d}{2}}.$$

Outline of the Proof of Theorem 1

Young's sharpened inequality and the monotonicity property of the Rényi entropy yield the following observation.

Let $\mathcal{P}^n = \{\underline{t} \in \mathbb{R}^n \colon t_k \geq 0,\ \sum_{k=1}^n t_k = 1\}$ be the probability simplex and let $\alpha > 1$. If $\sum_{k=1}^n N_\alpha(X_k) = 1$, then

$$\log N_\alpha\left(\sum_{k=1}^n X_k\right) \geq f_0(\underline{t}),\ \forall\, \underline{t} \in \mathcal{P}^n,$$

where

- $f_0(\underline{t}) = \frac{\log \alpha}{\alpha-1} - D(\underline{t} \| \underline{N}_\alpha) + \alpha' \sum_{k=1}^n \left(1 - \frac{t_k}{\alpha'}\right) \log\left(1 - \frac{t_k}{\alpha'}\right).$
- $\underline{N}_\alpha = (N_\alpha(X_1), \ldots, N_\alpha(X_n)).$
- $D(\underline{t} \| \underline{N}_\alpha) = \sum_{k=1}^n t_k \log\left(\frac{t_k}{N_\alpha(X_k)}\right).$

Outline of the Proof of Theorem 1

Young's sharpened inequality and the monotonicity property of the Rényi entropy yield the following observation.

Let $\mathcal{P}^n = \{\underline{t} \in \mathbb{R}^n \colon t_k \geq 0, \, \sum_{k=1}^n t_k = 1\}$ be the probability simplex and let $\alpha > 1$. If $\sum_{k=1}^n N_\alpha(X_k) = 1$, then

$$\log N_\alpha \left(\sum_{k=1}^n X_k \right) \geq f_0(\underline{t}), \; \forall \, \underline{t} \in \mathcal{P}^n,$$

$\implies$ The R-EPI can be tightened by maximizing $f_0(\underline{t})$.

Outline of the Proof of Theorem 1

Young's sharpened inequality and the monotonicity property of the Rényi entropy yield the following observation.
Let $\mathcal{P}^n = \{\underline{t} \in \mathbb{R}^n \colon t_k \geq 0, \, \sum_{k=1}^n t_k = 1\}$ be the probability simplex and let $\alpha > 1$. If $\sum_{k=1}^n N_\alpha(X_k) = 1$, then

$$\log N_\alpha \left(\sum_{k=1}^n X_k \right) \geq f_0(\underline{t}), \; \forall \, \underline{t} \in \mathcal{P}^n,$$

$\Longrightarrow$ The R-EPI can be tightened by maximizing $f_0(\underline{t})$.

- The solution of the optimization problem leads to an implicit bound in most cases
- Instead, we take a sub-optimal choice $t_k = N_\alpha(X_k)$ (it can be verified to be optimal if $N_\alpha(X_k)$ is independent of k).
- Some more steps yield Theorem 1.

Sub Optimality

- BV bound for $n = 2$:

$$N_\alpha(X_1 + X_2) \geq \max\{N_\alpha(X_1), N_\alpha(X_2)\}$$
$$\geq \tfrac{1}{2}\left(N_\alpha(X_1) + N_\alpha(X_2)\right), \ \alpha \in [0, \infty].$$

- Theorem 1 for $n = 2$ and $\alpha \to \infty$ yields

$$N_\infty(X_1 + X_2) \geq \tfrac{1}{2}\left(N_\infty(X_1) + N_\infty(X_2)\right).$$

- Since the maximal value of two numbers is larger than or equal to their average, the BV bound is tighter than our bound in Theorem 1 for $n = 2$ and large enough α's (unless $N_\infty(X_1) = N_\infty(X_2)$).

The Optimization Problem

Recall that $\log N_\alpha\left(\sum_{k=1}^n X_k\right) \geq f_0(\underline{t}),\ \forall\, \underline{t} \in \mathcal{P}^n$.

- The optimization problem is not convex

$$\begin{array}{ll} \text{maximize} & f_0(t_1, t_2, \ldots, t_{n-1}, t_n) \\ \text{subject to} & t_k \geq 0, \quad k \in \{1, \ldots, n\}, \\ & \sum_{k=1}^n t_k = 1 \end{array}$$

The Optimization Problem

Recall that $\log N_\alpha \left(\sum_{k=1}^{n} X_k\right) \geq f_0(\underline{t}), \ \forall\, \underline{t} \in \mathcal{P}^n$.

- The optimization problem is not convex

$$\begin{array}{ll} \text{maximize} & f_0(t_1, t_2, \ldots, t_{n-1}, t_n) \\ \text{subject to} & t_k \geq 0, \quad k \in \{1, \ldots, n\}, \\ & \sum_{k=1}^{n} t_k = 1 \end{array}$$

- An equivalent problem

$$\begin{array}{ll} \text{maximize} & f_0(t_1, t_2, \ldots, t_{n-1}, 1 - \sum_{k=1}^{n-1} t_k) \\ \text{subject to} & t_k \geq 0, \quad k \in \{1, \ldots, n-1\}, \\ & \sum_{k=1}^{n-1} t_k \leq 1 \end{array}$$

The Optimization Problem

Recall that $\log N_\alpha \left(\sum_{k=1}^n X_k\right) \geq f_0(\underline{t}),\ \forall\, \underline{t} \in \mathcal{P}^n$.

- The optimization problem is not convex

$$\begin{array}{ll} \text{maximize} & f_0(t_1, t_2, \ldots, t_{n-1}, t_n) \\ \text{subject to} & t_k \geq 0, \quad k \in \{1, \ldots, n\}, \\ & \sum_{k=1}^n t_k = 1 \end{array}$$

- An equivalent problem

$$\begin{array}{ll} \text{maximize} & f_0(t_1, t_2, \ldots, t_{n-1}, 1 - \sum_{k=1}^{n-1} t_k) \\ \text{subject to} & t_k \geq 0, \quad k \in \{1, \ldots, n-1\}, \\ & \sum_{k=1}^{n-1} t_k \leq 1 \end{array}$$

- This problem can be shown to be convex by a non trivial use of the next result from matrix theory (Bunch *et al.* 1978).

Rank–One Modification Theorem (Bunch *et al.* 1978)

Let

- $D \in \mathbb{R}^{n \times n}$ be a diagonal matrix with the eigenvalues $d_1 \leq d_2 \leq \ldots \leq d_n$.
- C be a rank-one modification of D i.e., $C = D + \rho z z^T$, where $z \in \mathbb{R}^n$, $\rho \in \mathbb{R}$, and let $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ be its eigenvalues.

Then,

Rank–One Modification Theorem (Bunch *et al.* 1978)

Let

- $D \in \mathbb{R}^{n \times n}$ be a diagonal matrix with the eigenvalues $d_1 \leq d_2 \leq \ldots \leq d_n$.
- C be a rank-one modification of D i.e., $C = D + \rho z z^T$, where $z \in \mathbb{R}^n$, $\rho \in \mathbb{R}$, and let $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ be its eigenvalues.

Then,

1. If $\rho > 0$, then $d_1 \leq \lambda_1 \leq d_2 \leq \lambda_2 \leq \ldots \leq d_n \leq \lambda_n$.
   If $\rho < 0$, then $\lambda_1 \leq d_1 \leq \lambda_2 \leq d_2 \leq \ldots \leq \lambda_n \leq d_n$.
2. If $d_j \neq d_i$ and $z_i, \rho \neq 0$, then the inequalities are strict, and for every $i \in \{1, \ldots, n\}$, λ_i is a zero of $W(x) = 1 + \rho \sum_{j=1}^{n} \frac{z_i^2}{d_j - x}$.

Applying The Rank–One Modification Theorem

1. The Hessian matrix of $f_0(t_1, t_2, \ldots, t_{n-1}, 1 - \sum_{k=1}^{n-1} t_k)$:

$$\nabla^2 f_0 = D + \rho \, \underline{1} \, \underline{1}^T$$

2. The Rank–One Modification Theorem is used to prove that $\nabla^2 f_0$ is negative semi-definite, hence f_0 is concave.

Applying The Rank–One Modification Theorem

1. The Hessian matrix of $f_0(t_1, t_2, \ldots, t_{n-1}, 1 - \sum_{k=1}^{n-1} t_k)$:

$$\nabla^2 f_0 = D + \rho \, \underline{1} \, \underline{1}^T$$

2. The Rank–One Modification Theorem is used to prove that $\nabla^2 f_0$ is negative semi-definite, hence f_0 is concave.
3. The optimization problem

$$\begin{array}{ll} \text{maximize} & f_0(t_1, t_2, \ldots, t_{n-1}, 1 - \sum_{k=1}^{n-1} t_k) \\ \text{subject to} & t_k \geq 0, \quad k \in \{1, \ldots, n-1\}, \\ & \sum_{k=1}^{n-1} t_k \leq 1 \end{array}$$

   is convex.
4. The solution can be found by solving the KKT conditions.

The KKT Conditions

- The optimization problem

$$\begin{array}{ll} \text{maximize} & f_0(t_1, t_2, \dots, t_{n-1}, 1 - \sum_{k=1}^{n-1} t_k) \\ \text{subject to} & t_k \geq 0, \quad k \in \{1, \dots, n-1\}, \\ & \sum_{k=1}^{n-1} t_k \leq 1 \end{array}$$

- Assume w.l.o.g that $N_\alpha(X_k) \leq N_\alpha(X_n), \quad k \in \{1, \dots, n-1\}$.
- Set $c_k = \frac{N_\alpha(X_k)}{N_\alpha(X_n)}$, $k \in \{1, \dots, n-1\}$.

The KKT Conditions

- The optimization problem

$$\begin{array}{ll} \text{maximize} & f_0(t_1, t_2, \ldots, t_{n-1}, 1 - \sum_{k=1}^{n-1} t_k) \\ \text{subject to} & t_k \geq 0, \quad k \in \{1, \ldots, n-1\}, \\ & \sum_{k=1}^{n-1} t_k \leq 1 \end{array}$$

- Assume w.l.o.g that $N_\alpha(X_k) \leq N_\alpha(X_n), \quad k \in \{1, \ldots, n-1\}$.
- Set $c_k = \frac{N_\alpha(X_k)}{N_\alpha(X_n)}$, $k \in \{1, \ldots, n-1\}$.
- After some simplifications, the KKT conditions are:
  1. $t_k(\alpha' - t_k) = c_k t_n(\alpha' - t_n), \quad k \in \{1, \ldots, n-1\}$
  2. $\sum_{k=1}^{n} t_k = 1$
  3. $t_k \geq 0, \quad k \in \{1, \ldots, n\}$

Theorem 2

- *Let* $X_1, \ldots, X_n$ *be* d*–dimensional independent r.v's with densities and assume, w.l.o.g, that* $N_\alpha(X_k) \leq N_\alpha(X_n), \quad k \in \{1, \ldots, n-1\}$.

Theorem 2

- *Let* $X_1, \ldots, X_n$ *be* d*–dimensional independent r.v's with densities and assume, w.l.o.g, that* $N_\alpha(X_k) \leq N_\alpha(X_n), \quad k \in \{1, \ldots, n-1\}$.
- *Let* $c_k = \frac{N_\alpha(X_k)}{N_\alpha(X_n)}, \qquad k \in \{1, \ldots, n-1\}$.

Theorem 2

- *Let* $X_1, \ldots, X_n$ *be* d*–dimensional independent r.v's with densities and assume, w.l.o.g, that* $N_\alpha(X_k) \le N_\alpha(X_n), \quad k \in \{1, \ldots, n-1\}$.
- *Let* $c_k = \frac{N_\alpha(X_k)}{N_\alpha(X_n)}, \qquad k \in \{1, \ldots, n-1\}$.
- *let* $t_n \in [0,1]$ *be the unique solution of* $t_n + \sum_{k=1}^{n-1} \psi_k(t_n) = 1$ *with* $\psi_k(x) = \frac{\alpha' - \sqrt{\alpha'^2 - 4c_k\, x(\alpha' - x)}}{2}, \quad x \in [0,1]$.

Theorem 2

- *Let* $X_1, \ldots, X_n$ *be* d*–dimensional independent r.v's with densities and assume, w.l.o.g, that* $N_\alpha(X_k) \leq N_\alpha(X_n), \quad k \in \{1, \ldots, n-1\}$.
- *Let* $c_k = \frac{N_\alpha(X_k)}{N_\alpha(X_n)}, \qquad k \in \{1, \ldots, n-1\}$.
- *let* $t_n \in [0,1]$ *be the unique solution of* $t_n + \sum_{k=1}^{n-1} \psi_k(t_n) = 1$ *with* $\psi_k(x) = \frac{\alpha' - \sqrt{\alpha'^2 - 4c_k\, x(\alpha' - x)}}{2}, \quad x \in [0,1]$.
- *Define* $t_k = \psi_k(t_n), \quad k \in \{1, \ldots, n-1\}$.

Theorem 2

- *Let* $X_1, \ldots, X_n$ *be* d*–dimensional independent r.v's with densities and assume, w.l.o.g, that* $N_\alpha(X_k) \leq N_\alpha(X_n), \quad k \in \{1, \ldots, n-1\}$.
- *Let* $c_k = \frac{N_\alpha(X_k)}{N_\alpha(X_n)}, \qquad k \in \{1, \ldots, n-1\}$.
- *let* $t_n \in [0,1]$ *be the unique solution of* $t_n + \sum_{k=1}^{n-1} \psi_k(t_n) = 1$ *with* $\psi_k(x) = \frac{\alpha' - \sqrt{\alpha'^2 - 4c_k\, x(\alpha' - x)}}{2}, \quad x \in [0,1]$.
- *Define* $t_k = \psi_k(t_n), \quad k \in \{1, \ldots, n-1\}$.

Then, the following R-EPI holds:

$$N_\alpha\left(\sum_{k=1}^{n} X_k\right) \geq e^{f_0(t_1,\ldots,t_n)} \sum_{k=1}^{n} N_\alpha(X_k),$$

Theorem 2 Implications

Theorem 2 ⇒ Theorem 1

- Theorem 2 improves the R-EPI in Theorem 1 unless $N_\alpha(X_k)$ is independent of k; in the latter case, the two R-EPIs coincide.

Theorem 2 Implications

Theorem 2 $\Rightarrow$ Theorem 1

- Theorem 2 improves the R-EPI in Theorem 1 unless $N_\alpha(X_k)$ is independent of k; in the latter case, the two R-EPIs coincide.
- However, Theorem 1 gives a closed-form bound.

Theorem 2 Implications

Theorem 2 ⇒ Theorem 1

- Theorem 2 improves the R-EPI in Theorem 1 unless $N_\alpha(X_k)$ is independent of k; in the latter case, the two R-EPIs coincide.
- However, Theorem 1 gives a closed-form bound.
- Recall that Theorem 1 ⇒ BC bound & the EPI.

Theorem 2 Implications

Theorem 2 ⇒ Theorem 1

- Theorem 2 improves the R-EPI in Theorem 1 unless $N_\alpha(X_k)$ is independent of k; in the latter case, the two R-EPIs coincide.
- However, Theorem 1 gives a closed-form bound.
- Recall that Theorem 1 ⇒ BC bound & the EPI.

Theorem 2 ⇒ BV Bound

- Improves the BV bound ($N_\alpha\left(\sum_{k=1}^n X_k\right) \geq \max_{1\leq k\leq n} N_\alpha(X_k)$).
- Both bounds asymptotically coincide as $\alpha \to \infty$ if and only if

$$\sum_{k=1}^{n-1} N_\infty(X_k) \leq N_\infty(X_n)$$

Closed-Form Expression of Theorem 2 for $n = 2$

Corollary 1

Let

- X_1 *and* X_2 *be* d*–dimensional independent r.v's with densities.*
- $\alpha > 1$, $\alpha' = \frac{\alpha}{\alpha-1}$.
- $\beta_\alpha = \frac{N_\alpha(X_1)}{N_\alpha(X_2)}$ *(Recall that w.l.o.g* $N_\alpha(X_1) \leq N_\alpha(X_2)$*).*
- $t_\alpha = \begin{cases} \frac{\alpha'(\beta_\alpha+1)-2\beta_\alpha-\sqrt{(\alpha'(\beta_\alpha+1))^2-8\alpha'\beta_\alpha+4\beta_\alpha}}{2(1-\beta_\alpha)} & \text{if } \beta_\alpha < 1 \\ \frac{1}{2} & \text{if } \beta_\alpha = 1 \end{cases}$

Closed-Form Expression of Theorem 2 for $n = 2$

Corollary 1

The following R-EPI holds:

$$N_\alpha(X_1 + X_2) \geq c_\alpha \left(N_\alpha(X_1) + N_\alpha(X_2)\right),$$

where

$$c_\alpha = \alpha^{\frac{1}{\alpha-1}} \exp\left\{-d\left(t_\alpha \,\Big\|\, \frac{\beta_\alpha}{\beta_\alpha + 1}\right)\right\} \left(1 - \frac{t_\alpha}{\alpha'}\right)^{\alpha' - t_\alpha} \left(1 - \frac{1 - t_\alpha}{\alpha'}\right)^{\alpha' - 1 + t_\alpha}$$

and $d(x\|y)$ is the binary relative entropy

$$d(x\|y) = x \log\left(\frac{x}{y}\right) + (1 - x) \log\left(\frac{1-x}{1-y}\right), \quad 0 \leq x, y \leq 1.$$

Closed-Form Expression of Theorem 2 for $n = 2$

For $n = 2$ (two summands), our tightest bound in Theorem 2 is asymptotically tight when $\alpha \to \infty$ and is achieved by two independent d-dimensional random vectors uniformly distributed in the cubes $[0, \sqrt{N_1}]^d$ and $[0, \sqrt{N_2}]^d$.

Comparing the R-EPIs ($n = 3$)

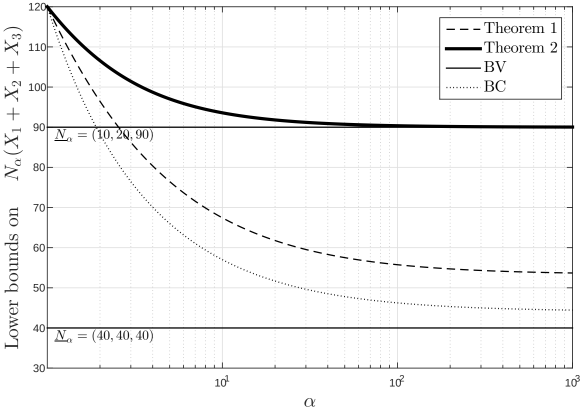


Figure: A comparison of the R-EPIs from Bobkov&Chistyakov (BC), Bercher&Vignat (BV), Theorem 1 and Theorem 2 for $n = 3$

Summary - Analytical Tools

- Theorem 1:
  1. The sharpened Young's inequality
  2. Monotonicity of the Rényi entropy power in its order

Summary - Analytical Tools

- Theorem 1:
  1. The sharpened Young's inequality
  2. Monotonicity of the Rényi entropy power in its order
- Theorem 2 - a further improvment:
  1. The rank-one modification theorem - proving convexity
  2. Convex optimization and solution of the KKT conditions

Publications

1. E. Ram and I. Sason, “On Rényi entropy power inequalities,” *IEEE Trans. on Information Theory*, vol. 62, no. 12, pp. 6800–6815, **December 2016**.
2. E. Ram and I. Sason, “On Rényi entropy power inequalities,” *Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT 2016)*, pp. 2289–2293, Barcelona, Spain, July 10–15, 2016.

Further Research: R-EPI For $\alpha \in [0, 1)$

- Are our bounding techniques extendible to $\alpha < 1$?

Further Research: R-EPI For $\alpha \in [0, 1)$

- Are our bounding techniques extendible to $\alpha < 1$?
- Unfortunately, not. In this case, Young's inequality and the monotonicity property of the Rényi entropy power yield inequalities in opposite directions.

Further Research: R-EPI For $\alpha \in [0, 1)$

- For $\alpha = 0$, one can use the Brunn-Minkowski (BM) inequality:

$$\mu^{\frac{1}{d}}(A + B) \geq \mu^{\frac{1}{d}}(A) + \mu^{\frac{1}{d}}(B).$$

Further Research: R-EPI For $\alpha \in [0, 1)$

- For $\alpha = 0$, one can use the Brunn-Minkowski (BM) inequality:

$$\mu^{\frac{1}{d}}(A + B) \geq \mu^{\frac{1}{d}}(A) + \mu^{\frac{1}{d}}(B).$$

  - By definition $N_0(X) = \mu^{\frac{2}{d}}\big(\mathrm{supp}(f_X)\big)$

Further Research: R-EPI For $\alpha \in [0, 1)$

- For $\alpha = 0$, one can use the Brunn-Minkowski (BM) inequality:

$$\mu^{\frac{1}{d}}(A + B) \geq \mu^{\frac{1}{d}}(A) + \mu^{\frac{1}{d}}(B).$$

  - By definition $N_0(X) = \mu^{\frac{2}{d}}\big(\mathrm{supp}(f_X)\big)$
  - BM yields $\sqrt{N_0(X+Y)} \geq \sqrt{N_0(X)} + \sqrt{N_0(Y)}$

Further Research: R-EPI For $\alpha \in [0, 1)$

- For $\alpha = 0$, one can use the Brunn-Minkowski (BM) inequality:

$$\mu^{\frac{1}{d}}(A + B) \geq \mu^{\frac{1}{d}}(A) + \mu^{\frac{1}{d}}(B).$$

  - By definition $N_0(X) = \mu^{\frac{2}{d}}\big(\mathrm{supp}(f_X)\big)$
  - BM yields $\sqrt{N_0(X+Y)} \geq \sqrt{N_0(X)} + \sqrt{N_0(Y)}$
  - By squaring both sides, $N_0(X+Y) \geq N_0(X) + N_0(Y)$, So $c_0 = 1$.

Further Research: R-EPI For $\alpha \in [0, 1)$

- For $\alpha = 0$, one can use the Brunn-Minkowski (BM) inequality:

$$\mu^{\frac{1}{d}}(A + B) \geq \mu^{\frac{1}{d}}(A) + \mu^{\frac{1}{d}}(B).$$

  - By definition $N_0(X) = \mu^{\frac{2}{d}}\big(\text{supp}(f_X)\big)$
  - BM yields $\sqrt{N_0(X+Y)} \geq \sqrt{N_0(X)} + \sqrt{N_0(Y)}$
  - By squaring both sides, $N_0(X+Y) \geq N_0(X) + N_0(Y)$, So $c_0 = 1$.
- From the EPI, $c_1 = 1$.

Further Research: R-EPI For $\alpha \in [0, 1)$

- For $\alpha = 0$, one can use the Brunn-Minkowski (BM) inequality:

$$\mu^{\frac{1}{d}}(A + B) \geq \mu^{\frac{1}{d}}(A) + \mu^{\frac{1}{d}}(B).$$

  - By definition $N_0(X) = \mu^{\frac{2}{d}}\big(\mathrm{supp}(f_X)\big)$
  - BM yields $\sqrt{N_0(X+Y)} \geq \sqrt{N_0(X)} + \sqrt{N_0(Y)}$
  - By squaring both sides, $N_0(X+Y) \geq N_0(X) + N_0(Y)$, So $c_0 = 1$.
- From the EPI, $c_1 = 1$.

It is conjectured that $c_\alpha = 1$ for all $\alpha \in (0, 1)$. This needs to be proved.

R-EPI For $\alpha \in [0, 1)$

Proposition 1 (R-EPI for $\alpha \in [0, 1)$)

Let $\{X_k\}_{k=1}^{n}$ be independent uniformly distributed random vectors and let $\alpha \in [0, 1)$. Then the following R-EPI holds,

$$N_\alpha\left(\sum_{k=1}^{n} X_k\right) \geq \sum_{k=1}^{n} N_\alpha(X_k)$$

R-EPI For $\alpha \in [0, 1)$

Proposition 1 (R-EPI for $\alpha \in [0, 1)$)

Let $\{X_k\}_{k=1}^{n}$ *be independent uniformly distributed random vectors and let* $\alpha \in [0, 1)$. *Then the following R-EPI holds,*

$$N_\alpha \left(\sum_{k=1}^{n} X_k \right) \geq \sum_{k=1}^{n} N_\alpha(X_k)$$

Proof.

1. Monotonicity of the Rényi entropy power in its order:
$$N_\alpha \left(\sum_{k=1}^{n} X_k\right) \geq N_1 \left(\sum_{k=1}^{n} X_k\right)$$
2. EPI: $N_1 \left(\sum_{k=1}^{n} X_k\right) \geq \sum_{k=1}^{n} N_1 \left(X_k\right)$.
3. For uniformly distributed random vectors, $N_1 \left(X_k\right) = N_\alpha \left(X_k\right)$

□

R-EPI For $\alpha \in [0,1)$

Proposition 1 (R-EPI for $\alpha \in [0,1)$)

Let $\{X_k\}_{k=1}^n$ be independent uniformly distributed random vectors and let $\alpha \in [0,1)$. Then the following R-EPI holds,

$$N_\alpha\left(\sum_{k=1}^n X_k\right) \geq \sum_{k=1}^n N_\alpha(X_k)$$

- Also holds (with equality) for Gaussian distributions with proportional covariances, for every $\alpha \in [0,\infty]$.

R-EPI For $\alpha \in [0, 1)$

Proposition 1 (R-EPI for $\alpha \in [0, 1)$)

Let $\{X_k\}_{k=1}^n$ be independent uniformly distributed random vectors and let $\alpha \in [0, 1)$. Then the following R-EPI holds,

$$N_\alpha \left(\sum_{k=1}^n X_k \right) \geq \sum_{k=1}^n N_\alpha(X_k)$$

- Also holds (with equality) for Gaussian distributions with proportional covariances, for every $\alpha \in [0, \infty]$.
- The uniform and Gaussian cases satisfy the conjecture

Further Research: More Topics

1. Rényi EPIs for discrete random vectors
2. Possible generalizations of the Rényi EPI in a way which generalizes the result by Feder and Zamir (IEEE Trans. on IT, 1993)
3. Possible generalizations with Rényi measures of the extended EPIs by Barron and Madiman (IEEE Trans. on IT, 2007)
4. Possible strengthening of the Rényi EPI by restriction to some families of distributions, e.g.,
   - extension of EPIs by Toscani (2015) for log-concave distributions;
   - extension of EPIs by Courtade (ISIT 2016).

Backup

Example: Data Filtering (FIR)

- Let $Y_k = 2X_k - X_{k-1} - X_{k-2}$ be an output of a FIR filter, where $\{X_k\}$ are i.i.d. random variables.
- Using the homogeneity of $N_\alpha(\cdot)$, we can consider the difference $h_2(Y) - h_2(X)$:

$$\begin{aligned} N_2(Y_k) &\geq c_2 \,\left(4N_2(X_k) + N_2(X_{k-1}) + N_2(X_{k-2})\right) \\ &= c_2\, 6\, N_2(X_k) \end{aligned}$$

Example: Data Filtering (FIR)

- Let $Y_k = 2X_k - X_{k-1} - X_{k-2}$ be an output of a FIR filter, where $\{X_k\}$ are i.i.d. random variables.
- Using the homogeneity of $N_\alpha(\cdot)$, we can consider the difference $h_2(Y) - h_2(X)$:

$$\begin{aligned} N_2(Y_k) &\geq c_2 \left(4N_2(X_k) + N_2(X_{k-1}) + N_2(X_{k-2})\right) \\ &= c_2\, 6\, N_2(X_k) \end{aligned}$$

1. Theorem 2: $h_2(Y) - h_2(X) \geq 0.8195$.
2. Theorem 1: $h_2(Y) - h_2(X) \geq 0.7866$.
3. Bobkov and Chistyakov: $h_2(Y) - h_2(X) \geq 0.7425$.
4. Bercher and Vignat: $h_2(Y) - h_2(X) \geq 0.6931$.

Example: Data Filtering (FIR)

- Let $Y_k = 2X_k - X_{k-1} - X_{k-2}$ be an output of a FIR filter, where $\{X_k\}$ are i.i.d. random variables.
- Using the homogeneity of $N_\alpha(\cdot)$, we can consider the difference $h_2(Y) - h_2(X)$:

$$\begin{aligned} N_2(Y_k) &\geq c_2\ (4N_2(X_k) + N_2(X_{k-1}) + N_2(X_{k-2})) \\ &= c_2\, 6\, N_2(X_k) \end{aligned}$$

1. Theorem 2: $h_2(Y) - h_2(X) \geq 0.8195.$
2. Theorem 1: $h_2(Y) - h_2(X) \geq 0.7866.$
3. Bobkov and Chistyakov: $h_2(Y) - h_2(X) \geq 0.7425.$
4. Bercher and Vignat: $h_2(Y) - h_2(X) \geq 0.6931.$
5. If X_k is a Gaussian: $h_2(Y) - h_2(X) = 0.8959.$

An Application of The Rényi Entropy - Example

Bunte and Lapidoth, 2014, “Encoding Tasks and Rényi Entropy”

- A task is drawn from a finite set $\mathcal{X}$ with probability P.
- The task should be described with a fixed number of bits.
- **No task should be neglected. Not even the atypical ones** (classic source coding cannot be used).

An Application of The Rényi Entropy - Example

Bunte and Lapidoth, 2014, "Encoding Tasks and Rényi Entropy"

- A task is drawn from a finite set $\mathcal{X}$ with probability P.
- The task should be described with a fixed number of bits.
- **No task should be neglected. Not even the atypical ones** (classic source coding cannot be used).
- The encoder partitions $\mathcal{X}$ in to M subsets,

$$f\colon \mathcal{X} \to \{1, \ldots, M\},$$

such that for every $x \in \mathcal{X}$, $f^{-1}(f(x))$ is the subset that contains x.

An Application of The Rényi Entropy - Example

Encoding Tasks Theorem

Let $\{X_i\}_{i=1}^{\infty}$ be a source over $\mathcal{X}$. Let $\rho > 0$.

1. **Direct:** If $R > \limsup_{n\to\infty} \frac{1}{n} H_{\frac{1}{1+\rho}}(X^n)$, then there exist encoders $f_n \colon \mathcal{X}^n \to \{1, \ldots, 2^{nR}\}$ such that
$$\lim_{n\to\infty} E\left[\left|f_n^{-1}(f_n(X^n))\right|^{\rho}\right] = 1.$$
2. **Converse:** If $R < \liminf_{n\to\infty} \frac{1}{n} H_{\frac{1}{1+\rho}}(X^n)$, then for any choice of encoders $f_n \colon \mathcal{X}^n \to \{1, \ldots, 2^{nR}\}$,
$$\lim_{n\to\infty} E\left[\left|f_n^{-1}(f_n(X^n))\right|^{\rho}\right] = \infty.$$

Proof of Theorem1 - Outline

1. Assume w.l.o.g that $\sum_{k=1}^{n} N_\alpha(X_k) = 1$ (homogeneity of the Rényi entropy power)
2. $\log N_\alpha\left(\sum_{k=1}^{n} X_k\right) \geq f(\underline{t})$
$$= \frac{\log \alpha}{\alpha-1} - D(\underline{t}\|\underline{N}_\alpha) + \alpha' \sum_{k=1}^{n} \left(1 - \frac{t_k}{\alpha'}\right) \log\left(1 - \frac{t_k}{\alpha'}\right)$$
3. Choose $t_k = N_\alpha(X_k)$ such that $D(\underline{t}\|\underline{N}_\alpha) = 0$
4. From the convexity of $f(x) = (1-x)\log(1-x),\ x \in (0,1)$,
$$\left(1 - \frac{t_k}{\alpha'}\right) \log\left(1 - \frac{t_k}{\alpha'}\right) \geq \log\left(1 - \frac{1}{n\alpha'}\right) + \frac{\log e}{n\alpha'} - \frac{t_k}{\alpha'}\left[\log e + \log\left(1 - \frac{1}{n\alpha'}\right)\right]$$
5. Combining 2., 3. and 4., yields the desired result (since $\sum_{k=1}^{n} t_k = 1$)

Discussion - Tightness

- The R-EPI in Theorem 2 provides the tightest R-EPI known to date for $\alpha \in (1, \infty)$.

Discussion - Tightness

- The R-EPI in Theorem 2 provides the tightest R-EPI known to date for $\alpha \in (1, \infty)$.
- Nevertheless, one of the inequalities involved in its derivation is loose:
  - The sharpened Young's inequality: equality only for Gaussians.
  - Monotonicity of the Rényi entropy power in its order: Equality only for uniformly distributed random vectors.

Discussion - Tightness

- The R-EPI in Theorem 2 provides the tightest R-EPI known to date for $\alpha \in (1, \infty)$.
- Nevertheless, one of the inequalities involved in its derivation is loose:
  - The sharpened Young's inequality: equality only for Gaussians.
  - Monotonicity of the Rényi entropy power in its order: Equality only for uniformly distributed random vectors.
- For $\alpha = \infty$ and $n = 2$, the sharpened Young's inequality reduces to

$$\|f * g\|_\infty \leq \|f\|_p \, \|g\|_{p'}.$$

  - Equality holds if f and g are scaled versions of a uniform distribution on the same convex set.
  - This is consistent with the fact that the R-EPIs in Theorems 1 and 2 are asymptotically tight for $n = 2$ by letting $\alpha \to \infty$.