Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 July 2018 doi:10.20944/preprints201807.0145.v1

Article

Rényi Entropy Power Inequalities *via* Normal Transport and Rotation

Olivier Rioul

LTCI, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France; olivier.rioul@telecom-paristech.fr
École Polytechnique, Université Paris-Saclay, 91128 Palaiseau, France; olivier.rioul@polytechnique.edu

Abstract: Following a recent proof of Shannon's entropy power inequality (EPI), a comprehensive framework for deriving various EPIs for the Rényi entropy is presented, that uses transport arguments from normal densities and a change of variable by rotation.

Keywords: Rényi entropy; Entropy power inequalities; Transportation Arguments; Normal distributions; Escort distributions; Log-concave distributions

1. Introduction

The entropy power inequality (EPI) dates back to Shannon's seminal paper [1] and has a long history [2]. The link with the Rényi entropy was first made by Dembo, Cover and Thomas [3] in connection with Young's convolutional inequality with sharp constants, where Shannon's EPI is obtained by letting the Rényi entropy orders tend to one [4, Theorem 17.8.3].

Recently there has been significant interest in Rényi EPIs for several independent variables. Bobkov and Chistyakov [5] extended the classical Shannon's EPI to the Rényi entropy by incorporating a multiplicative constant that depends on the order of the Rényi entropy. Ram and Sason [6] improved the constant by making it depend also on the number of variables. Even more recently, Bobkov and Marsiglietti [7] proved another modification of the EPI for the Rényi entropy for two independent variables, with a power exponent parameter α which was improved by Li [8]. All these EPIs were found for Rényi entropies of orders > 1. The α-modification of the Rényi EPI was extended to orders < 1 for two independent variables having log-concave densities by Marsiglietti and Melbourne [9]. The starting point of all the above works was Young's strengthened convolutional inequality.

Also recently, Shannon's original EPI was proved anew [10] using a simple transport argument from normal variables and a change of variable by rotation. In this paper, we exploit these ingredients, described in the following lemmas, to establish all previously known Rényi EPIs and derive new ones.

Notation 1. *Throughout this article the considered zero-mean random variables $X \in \mathbb{R}^n$ admit a density which is implicitly assumed continuous inside its support. Write $X \sim f$ if X has density f.*

Lemma 1 (Normal Transport [10])**.** *Let $X^* \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. There exists a diffeomorphism $T : \mathbb{R}^n \to \mathbb{R}^n$ such that $X = T(X^*) \sim f$. Moreover, T can be chosen such that its Jacobian matrix T' is (lower) triangular with positive diagonal elements.*

Two different proofs of Lemma 1 are given in [10]. The proof is very simple for one-dimensional variables [11], where T is just an increasing function with continuous derivative $T' > 0$.

Lemma 2 (Normal Rotation [10])**.** *If X^*, Y^* are i.i.d. normal, then for any $0 < \lambda < 1$, the rotation*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 July 2018 doi:10.20944/preprints201807.0145.v1

$$\begin{pmatrix}\widetilde{X}\\ \widetilde{Y}\end{pmatrix} = \begin{pmatrix}\sqrt{\lambda} & \sqrt{1-\lambda}\\ -\sqrt{1-\lambda} & \sqrt{\lambda}\end{pmatrix}\begin{pmatrix}X^*\\ Y^*\end{pmatrix} \tag{1}$$

yields i.i.d. normal variables $\widetilde{X}, \widetilde{Y}$.

Notice that the the starred variables can be expressed in terms of the tilde variables by the inverse rotation

$$\begin{pmatrix}X^*\\ Y^*\end{pmatrix} = \begin{pmatrix}\sqrt{\lambda} & -\sqrt{1-\lambda}\\ \sqrt{1-\lambda} & \sqrt{\lambda}\end{pmatrix}\begin{pmatrix}\widetilde{X}\\ \widetilde{Y}\end{pmatrix}. \tag{2}$$

The proof of Lemma 2 is trivial considering covariance matrices. A deeper result states that this property of remaining i.i.d. by rotation characterizes the normal distribution—this is known as Bernstein's lemma (see e.g., [12, Chap. 5], [11, Lemma 4]). This explains why one obtains equality in the EPI only for normal variables.

This article is a revised, full version of what was presented in part in a previous conference communication [13]. It is organized as follows. Preliminary definitions and known properties are presented in Section 2. Section 3 derives a crucial "information inequality" for Rényi entropies that enjoys a transformational invariance. The central result is in Section 4, where the first version of the Rényi EPI by Dembo, Cover and Thomas is proved using the ingredients of Lemmas 1 and 2. All previously known Rényi EPIs—and new ones—are then derived using a simple method in Section 5. Section 6 concludes.

2. Preliminary Definitions and Properties

Throughout this article we consider exponents $p > 0$ with $p \neq 1$. The following definition is well known and used e.g., in Hölder's inequality.

Definition 1 (Conjugate Exponent)**.** *The* conjugate exponent *of p is*

$$p' = \frac{p}{p-1}, \tag{3}$$

that is, the number p' such that

$$\frac{1}{p} + \frac{1}{p'} = 1. \tag{4}$$

Remark 1. *There are two situations depending on whether p' is positive or negative, as summarized in the following table.*

$p > 1$	$0 < p < 1$
$p' > 1$	$p' < 0$

Definition 2 (Rényi Entropy)**.** *If X has density $f \in L^p(\mathbb{R}^n)$, its Rényi entropy of order p is defined by*

$$h_p(X) = \frac{1}{1-p}\log\int_{\mathbb{R}^n} f^p(x)\,\mathrm{d}x \tag{5}$$

$$= -p'\log\|f\|_p \tag{6}$$

It is known that the limit as $p \to 1$ is the Shannon entropy

$$h_1(X) = -\int_{\mathbb{R}^n} f(x)\log f(x)\,\mathrm{d}x. \tag{7}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 July 2018 doi:10.20944/preprints201807.0145.v1

The Rényi entropy enjoys well-known properties similar to those of the Shannon entropy, which are recalled here for completeness.

Lemma 3 (Scaling Property)**.** *For any $a \in \mathbb{R}$,*

$$h_p(aX) = h_p(X) + n \log |a|. \tag{8}$$

Proof. Making a change of variables, $h_p(aX) = \frac{1}{1-p} \log \int \big(\frac{1}{|a|^n} f(\frac{x}{a}) \big)^p \,\mathrm{d}x = \frac{1}{1-p} \log \int f^p(\frac{x}{a}) \frac{\mathrm{d}x}{|a|^n} + \frac{1}{1-p} \log |a|^{n(1-p)} = h_p(X) + n \log |a|$. □

One recovers the usual scaling property for the Shannon entropy by letting $p \to 1$.

Lemma 4 (Rényi Entropy of the Normal)**.** *If $X^* \sim \mathcal{N}(0, \mathbf{K})$ for some nonsingular covariance matrix* $\mathbf{K}$*, then*

$$h_p(X^*) = \frac{1}{2} \log((2\pi)^n |\mathbf{K}|) + \frac{n}{2} p' \frac{\log p}{p} \tag{9}$$

where $|\cdot|$ denotes the determinant. In particular for $X^* \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$,

$$h_p(X^*) = \frac{n}{2} \log(2\pi\sigma^2) + \frac{n}{2} p' \frac{\log p}{p}. \tag{10}$$

Proof. By direct calculation, $h_p(X^*) = \frac{1}{1-p} \log \int \Big(\frac{\exp(-\frac{1}{2} x^t \mathbf{K}^{-1} x)}{\sqrt{(2\pi)^n |\mathbf{K}|}} \Big)^p \,\mathrm{d}x = \frac{1}{1-p} \log \frac{\sqrt{(2\pi)^n |\mathbf{K}| p^{-n}}}{\sqrt{(2\pi)^n |\mathbf{K}|}^p} = \frac{1}{2} \log((2\pi)^n |\mathbf{K}|) - \frac{n}{2} \frac{\log p}{1-p}$. □

Again one recovers the Shannon entropy of a normal variable by letting $p \to 1$ (then $p' \frac{\log p}{p} \to \log e$).

The following notion of escort distribution [14,15] is useful in the sequel.

Definition 3 (Escort Density [15, § 2.2])**.** *If $f \in L^p(\mathbb{R}^n)$, its* escort density *of exponent p is the density defined by*

$$f_p(x) = \frac{f^p(x)}{\int_{\mathbb{R}^n} f^p(x) \,\mathrm{d}x}. \tag{11}$$

In other words $f_p = \frac{f^p}{\|f\|_p^p}$ where $\|f\|_p$ denotes the L^p norm of f. We also use the notation X_p to denote the corresponding escort random variable with density f_p.

Lemma 5 (Monotonicity Property)**.** *If $p < q$ then $h_p(X) \geq h_q(X)$ with equality if and only if X is uniformly distributed.*

Proof. Let $p \neq 1$ and assume that $f \in L^q(\mathbb{R}^n)$ for all q in a neighborhood of p so that one can freely differentiate under the integral sign:

$$\frac{\partial}{\partial p} h_p(X) = \frac{1}{(1-p)^2} \log \int f^p + \frac{1}{1-p} \frac{\int f^p \log f}{\int f^p} \tag{12}$$

$$= \frac{1}{(1-p)^2} \big(\log \int f^p + \int f_p \log f^{1-p}\big) \tag{13}$$

$$= \frac{1}{(1-p)^2} \int f_p \log \frac{f}{f_p} \tag{14}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 July 2018 doi:10.20944/preprints201807.0145.v1

$$= -\frac{D(f_p\|f)}{(1-p)^2} \leq 0 \tag{15}$$

where $D(\cdot\|\cdot)$ denotes the Kullback-Leibler divergence. Equality $D(f_p\|f) = 0$ can hold only if $f = f_p$ a.e., which since $p \neq 1$ implies that f is constant over some measurable subset $A \subset \mathbb{R}^n$, and is zero elsewhere. It follows that $h_p(X) > h_q(X)$ for any $p < q$ if X is not uniformly distributed. Conversely, if X is uniformly distributed over some measurable subset $A \subset \mathbb{R}^n$, its density can be written as $f(x) = 1/\mathrm{vol}(A)$ for all $x \in A$ and $= 0$ elsewhere. Then $h_p(X) = \frac{1}{1-p}\log\frac{\mathrm{vol}(A)}{\mathrm{vol}(A)^p} = \log\mathrm{vol}(A)$ is independent of p. □

Remark 2. *The identity established in the proof:*

$$\frac{\partial}{\partial p}h_p(X) = -\frac{D(f_p\|f)}{(1-p)^2} \tag{16}$$

seems new, although a similar formula for discrete variables can be found in [16, § 5.3].

3. An Information Inequality

The Shannon entropy satisfies a fundamental "information inequality" [4, Theorem 2.6.3] from which many classical information-theoretic inequalities can be derived. This can be written as

$$h_1(X) \leq -\mathbb{E}\log\varphi(X) \tag{17}$$

for any density φ, with equality if and only if $\varphi = f$ a.e. The following Theorem can be seen as the natural extension of the information inequality to Rényi entropies and is central in the following derivations of this paper. J.F. Bercher has pointed out to the author that it is similar to an inequality for discrete distributions established by Campbell [17] in the context of source coding (see also [14]).

Theorem 1 (Information Inequality)**.** *For any density φ,*

$$h_p(X) \leq -p'\log\mathbb{E}\big(\varphi^{1/p'}(X)\big) \tag{18}$$

with equality if and only if $\varphi = f_p$ a.e.

By letting $p \to 1$ one recovers the classical information inequality (17) for the Shannon entropy.

Proof. By definition (6),

$$h_p(X) = -p'\log\|f\|_p \tag{19}$$

$$= -p'\log\Big(\int (f^p\varphi^{-1})\varphi\Big)^{1/p} \tag{20}$$

$$\leq -p'\log\int (f^p\varphi^{-1})^{1/p}\varphi \tag{21}$$

$$= -p'\log\int f\varphi^{1/p'} \tag{22}$$

where the inequality follows from Jensen's inequality applied to the function $x \mapsto x^{1/p}$, which is strictly concave if $p > 1$ (that is, $p' > 0$) and strictly convex if $p < 1$ (that is, $p' < 0$). Equality holds if and only if $f^p\varphi^{-1}$ is constant a.e., which means that φ and f^p are proportional a.e. Normalizing gives the announced condition $\varphi = f_p$ a.e. □

Remark 3. *An alternate proof is obtained using Hölder's inequality or its reverse applied to f and $\varphi^{1/p'}$. Notice that the equality case for $\varphi = f_p$ gives*

 doi:10.20944/preprints201807.0145.v1

$$h_p(X) = -p' \log \mathbb{E}(f_p^{1/p'}(X)) \tag{23}$$

as can be easily checked directly.

The following conditional version of Theorem 1 involves a more complicated relation for dependent variables.

Corollary 1 (Conditional Information Inequality). *For any two random variables $X, Y \in \mathbb{R}^n$,*

$$-p' \log \mathbb{E}_Y \exp(-h_p(X|Y)/p') \leq -p' \log \mathbb{E}(\varphi^{1/p'}(X|Y)) \tag{24}$$

where $h_p(X|y)$ denotes the Rényi entropy of X knowing $Y = y$ and the expectation on the l.h.s. is taken over Y (the expectation in the r.h.s. is taken over (X, Y)).

In particular, when X and Y are independent,

$$h_p(X) \leq -p' \log \mathbb{E}(\varphi^{1/p'}(X|Y)). \tag{25}$$

with equality if and only if $\varphi(x|y)$ does not depend on y and equals $f_p(x)$ a.e.

Proof. From (18) for fixed y, one has $\mathbb{E}(\varphi^{1/p'}(X|y)) \leq \exp(-h_p(X|y)/p')$ for $p' > 0$ ($p > 1$) and the opposite inequality for $p' < 0$ ($p < 1$), with equality if and only if $\varphi(x|y) = f_p(x|y)$ a.e. Taking the expectation over Y yields $\mathbb{E}(\varphi^{1/p'}(X|Y)) \leq \mathbb{E}_Y \exp(-h_p(X|Y)/p')$ for $p' > 0$ ($p > 1$) and the opposite inequality for $p' < 0$ ($p < 1$). The result follows by taking the logarithm and multiplying by $-p'$. When X and Y are independent, equality holds if and only if $\varphi(x|y) = f_p(x)$ a.e. for all y. □

For the Shannon entropy, the difference between the two sides of the information inequality (17) is the Kullback-Leibler divergence:

$$D(f\|\varphi) = \mathbb{E} \log \frac{f(X)}{\varphi(X)} \geq 0$$

which can also be noted $D(X\|Z)$ where $X \sim f$ and $Z \sim \varphi$. It is known (and easy to check) that the divergence is invariant by reversible transformations T. This means that when $X = T(X^*)$ and $Z = T(Z^*)$, one has $D(X\|Z) = D(X^*\|Z^*)$. A natural extension to Rényi entropies can be obtained on the difference

$$-p' \log \mathbb{E}(\varphi^{1/p'}(X)) - h_p(X) \geq 0 \tag{26}$$

between the two sides of the information inequality (18).

Theorem 2 (Transformational Invariance). *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a diffeomorphism and suppose that*

$$X_p = T(X_p^*) \tag{27}$$
$$Z = T(Z^*) \tag{28}$$

where $Z \sim \varphi$ and $Z^* \sim \varphi^*$. Then

$$-p' \log \mathbb{E}(\varphi^{1/p'}(X)) - h_p(X) = -p' \log \mathbb{E}(\varphi^{*1/p'}(X^*)) - h_p(X^*). \tag{29}$$

Note that from (6), this identity can be rewritten as

$$\frac{\mathbb{E}(\varphi^{*1/p'}(X^*))}{\|f^*\|_p} = \frac{\mathbb{E}(\varphi^{1/p'}(X))}{\|f\|_p} \tag{30}$$

Proof. Proceed to prove (30). Let f, f^* be the respective densities of X, X^* and recall that $X_p \sim f_p$ and $X_p^* \sim f_p^*$. By the transformation T the densities are related by

$$f_p^*(x^*) = f_p(T(x^*))|T'(x^*)| \tag{31}$$
$$\varphi^*(x^*) = \varphi(T(x^*))|T'(x^*)|. \tag{32}$$

where $|T'|$ denotes the Jacobian determinant of T. Using these relations and Definition 3,

$$\mathbb{E}(\varphi^{*1/p'}(X^*))/\|f^*\|_p = \mathbb{E}(\varphi^{1/p'}(T(X^*))|T'(X^*)|^{1/p'})/\|f^*\|_p \tag{33}$$
$$= \int \varphi^{1/p'}(T(x^*))|T'(x^*)|^{1/p'} f^*(x^*)\,\mathrm{d}x^*/\|f^*\|_p \tag{34}$$
$$= \int \varphi^{1/p'}(T(x^*))|T'(x^*)|^{1/p'} f_p^*(x^*)^{1/p}\,\mathrm{d}x^* \tag{35}$$
$$= \int \varphi^{1/p'}(T(x^*)) f_p(T(x^*))^{1/p}|T'(x^*)|\,\mathrm{d}x^* \tag{36}$$
$$= \int \varphi^{1/p'}(x) f_p(x)^{1/p}\,\mathrm{d}x \tag{37}$$
$$= \mathbb{E}(\varphi^{1/p'}(X))/\|f\|_p \quad \square \tag{38}$$

Remark 4. *The fact that φ is a density was not used in the proof of Theorem 2. Therefore (29) holds more generally for any function φ satisfying (32).*

4. First Version of the Rényi EPI

For two independent random variables X and Y, the Shannon entropy power inequality can be expressed as follows [2,3]: For any $0 < \lambda < 1$,

$$h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \geq \lambda h(X) + (1-\lambda)h(Y) \tag{39}$$

with equality if and only if X, Y are i.i.d. normal. That is, the difference $h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h(X) - (1-\lambda)h(Y)$ is minimum (zero) for i.i.d. normal X, Y. In this section, we study the natural generalization for Rényi entropies [3, Theorem 12], namely that the quantity

$$h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h_p(X) - (1-\lambda)h_q(Y) \tag{40}$$

is minimum for i.i.d. normal X, Y. Here the triple (p, q, r) and its associated λ satisfy the following condition, which is used e.g., in Young's convolutional inequality.

Definition 4 (Exponent Triple)**.** *An* exponent triple $(p, q, r)_\lambda$ *has conjugates p', q', r' of the same sign and such that*

$$\frac{1}{p'} + \frac{1}{q'} = \frac{1}{r'}. \tag{41}$$

The corresponding coefficient $\lambda \in (0, 1)$ is defined by

$$\lambda = \frac{r'}{p'} = 1 - \frac{r'}{q'} \tag{42}$$

In other words, the exponents p, q, r are such that

 doi:10.20944/preprints201807.0145.v1

$$\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r} \tag{43}$$

and fulfill one the following two conditons:

$p, q, r > 1$ $p', q', r' > 1$	$0 < p, q, r < 1$ $p', q', r' < 0$
$r' < p', r' < q'$ $r > p, r > q$	$\lvert r'\rvert < \lvert p'\rvert, \lvert r'\rvert < \lvert q'\rvert$ $r < p, r < q$

The key argument used in this section is the following. If $X \sim f$ and $Y \sim g$, then for the escort variables, $X_p \sim f_p$ and $Y_q \sim g_q$. By normal transport (Lemma 1), one can write

$$X_p = T(X_p^*) \tag{44}$$

$$Y_q = U(Y_q^*) \tag{45}$$

for two diffeomorphims T and U, where X^*, Y^* are, say, i.i.d. standard normal $\mathcal{N}(0, \mathbf{I})$. (It follows that $X_p^* \sim \mathcal{N}(0, \mathbf{I}/p)$ and $Y_q^* \sim \mathcal{N}(0, \mathbf{I}/q)$.) We then have the following straightforward extension of Theorem 2:

Lemma 6 (Transformational Invariance for Two Independent Variables)**.** *For a two-dimensional* $\varphi(x, y)$*,*

$$\begin{aligned} &-r' \log \mathbb{E}\big(\varphi^{1/r'}(X, Y)\big) - \lambda h_p(X) - (1-\lambda) h_q(Y) \\ &\qquad = -r' \log \mathbb{E}\big(\varphi^{*1/r'}(X^*, Y^*)\big) - \lambda h_p(X^*) - (1-\lambda) h_q(Y^*) \end{aligned} \tag{46}$$

where

$$\varphi^*(x^*, y^*) = \varphi(T(x^*), U(y^*)) |T'(x^*)|^{\lambda} |U'(y^*)|^{1-\lambda}. \tag{47}$$

Proof. From (6) and the definition of λ, (57) can be rewritten as

$$\frac{\mathbb{E}\big(\varphi^{*1/r'}(X^*, Y^*)\big)}{\|f^*\|_p \|g^*\|_q} = \frac{\mathbb{E}\big(\varphi^{1/r'}(X, Y)\big)}{\|f\|_p \|g\|_q} \tag{48}$$

By the transformations T and U the densities of the escort variables are related by $f_p^*(x^*) = f_p(T(x^*))|T'(x^*)|$ and $g_q^*(y^*) = g_q(U(y^*))|U'(y^*)|$. Now by the same calculation as in the proof of Theorem 2,

$$\begin{aligned} \frac{\mathbb{E}\big(\varphi^{*1/r'}(X^*, Y^*)\big)}{\|f^*\|_p \|g^*\|_q} &= \frac{\mathbb{E}\big(\varphi^{1/r'}(T(X^*), U(Y^*)) |T'(X^*)|^{1/p'} |U'(Y^*)|^{1/q'}\big)}{\|f^*\|_p \|g^*\|_q} && (49) \\ &= \frac{\int \varphi^{1/r'}(T(x^*), U(y^*)) |T'(x^*)|^{1/p'} |U'(y^*)|^{1/q'} f^*(x^*) g^*(y^*) \,\mathrm{d}x^* \,\mathrm{d}y^*}{\|f^*\|_p \|g^*\|_q} && (50) \\ &= \int \varphi^{1/r'}(T(x^*), U(y^*)) |T'(x^*)|^{1/p'} |U'(y^*)|^{1/q'} f_p^*(x^*)^{1/p} g_q^*(y^*)^{1/q} \,\mathrm{d}x^* \,\mathrm{d}y^* && (51) \\ &= \int \varphi^{1/r'}(T(x^*), U(y^*)) |T'(x^*)| |U'(y^*)| f_p(T(x^*))^{1/p} g_q(U(y^*))^{1/q} \,\mathrm{d}x^* \,\mathrm{d}y^* && (52) \\ &= \int \varphi^{1/r'}(x, y) f_p(x)^{1/p} g_q(y)^{1/q} \,\mathrm{d}x \,\mathrm{d}y && (53) \\ &= \frac{\mathbb{E}\big(\varphi^{1/r'}(X, Y)\big)}{\|f\|_p \|g\|_q} \quad \square && (54) \end{aligned}$$

Lemma 7. *Let φ be the density of $\sqrt{\lambda}X + \sqrt{1-\lambda}Y$. Then*

$$\begin{aligned} &h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h_p(X) - (1-\lambda)h_q(Y) \\ &\quad \geq -r' \log \mathbb{E}\Big\{ \big(\varphi_r(\sqrt{\lambda}T(X^*) + \sqrt{1-\lambda}U(Y^*)) \cdot |\lambda T'(X^*) + (1-\lambda)U'(Y^*)|\big)^{1/r'} \Big\} \\ &\quad - \lambda h_p(X^*) - (1-\lambda)h_q(Y^*). \end{aligned} \tag{55}$$

Proof. By the equality case of Theorem 1 (see (23)), one has

$$h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) = -r' \log \mathbb{E}\varphi_r^{1/r'}(\sqrt{\lambda}X + \sqrt{1-\lambda}Y). \tag{56}$$

Now by Lemma 6 applied to $\varphi(x,y) = \varphi_r(\sqrt{\lambda}x + \sqrt{1-\lambda}y)$, we have $\varphi^*(x,y) = \varphi_r(\sqrt{\lambda}T(x^*) + \sqrt{1-\lambda}U(y^*))|T'(x^*)|^{\lambda}|U'(y^*)|^{1-\lambda}$, and, therefore,

$$\begin{aligned} &h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h_p(X) - (1-\lambda)h_q(Y) \\ &\quad = -r' \log \mathbb{E}\Big\{ \big(\varphi_r(\sqrt{\lambda}T(X^*) + \sqrt{1-\lambda}U(Y^*)) \cdot |T'(X^*)|^{\lambda}|U'(Y^*)|^{1-\lambda}\big)^{1/r'} \Big\} \\ &\quad - \lambda h_p(X^*) - (1-\lambda)h_q(Y^*). \end{aligned} \tag{57}$$

Since from Lemma 1, T and U can be chosen such that T' and U' are (lower) triangular with positive diagonal elements, it follows easily from the arithmetic-geometric mean inequality that

$$|T'(X^*)|^{\lambda}|U'(Y^*)|^{1-\lambda} \leq |\lambda T'(X^*) + (1-\lambda)U'(Y^*)|. \tag{58}$$

The result follows at once (for either positive or negative r'). □

We can now use the normal rotation Lemma 2 to conclude by proving the following

Theorem 3 (Rényi EPI [3])**.** *For independent X, Y and exponent triple $(p,q,r)_\lambda$,*

$$h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h_p(X) - (1-\lambda)h_q(Y) \geq \frac{n}{2} r' \Big(\frac{\log r}{r} - \frac{\log p}{p} - \frac{\log q}{q} \Big) \tag{59}$$

with equality if and only if X, Y are i.i.d. normal.

Proof. If X, Y are i.i.d. normal, then $\sqrt{\lambda}X + \sqrt{1-\lambda}Y$ is also identically distributed as X and Y, and from Lemma 4, it is immediate to check that equality holds (irrespective of their covariances). Therefore, inequality (59) is equivalent to

$$h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h_p(X) - (1-\lambda)h_q(Y) \geq h_r(\sqrt{\lambda}X^* + \sqrt{1-\lambda}Y^*) - \lambda h_p(X^*) - (1-\lambda)h_q(Y^*) \tag{60}$$

where X^*, Y^* are, say, i.i.d. standard normal $\mathcal{N}(0, \mathbf{I})$.

To prove (60), consider the normal rotation of Lemma 2 and write X^*, Y^* in terms of $\widetilde{X}, \widetilde{Y}$ using (2) in the first term of the r.h.s. of (55) (Lemma 7). One obtains:

$$\begin{aligned} &h_r(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h_p(X) - (1-\lambda)h_q(Y) \\ &\quad \geq -r' \log \mathbb{E}\Big\{ \big(\psi(\widetilde{X}|\widetilde{Y})\big)^{1/r'} \Big\} - \lambda h_p(X^*) - (1-\lambda)h_q(Y^*). \end{aligned} \tag{61}$$

where

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 July 2018 doi:10.20944/preprints201807.0145.v1

$$\begin{aligned}\psi(\widetilde{x}|\widetilde{y}) =&\varphi_r(\sqrt{\lambda}T(\sqrt{\lambda}\widetilde{x}-\sqrt{1-\lambda}\widetilde{y})+\sqrt{1-\lambda}U(\sqrt{1-\lambda}\widetilde{x}+\sqrt{\lambda}\widetilde{y}))\\ &\times|\lambda T'(\sqrt{\lambda}\widetilde{x}-\sqrt{1-\lambda}\widetilde{y})+(1-\lambda)U'(\sqrt{1-\lambda}\widetilde{x}+\sqrt{\lambda}\widetilde{y})|\end{aligned} \tag{62}$$

Making the change of variable $z=\sqrt{\lambda}T(\sqrt{\lambda}\widetilde{x}-\sqrt{1-\lambda}\widetilde{y})+\sqrt{1-\lambda}U(\sqrt{1-\lambda}\widetilde{x}+\sqrt{\lambda}\widetilde{y})$, one obtains

$$\int \psi(\widetilde{x}|\widetilde{y})\,\mathrm{d}\widetilde{x}=\int \varphi_r(z)\,\mathrm{d}z=1, \tag{63}$$

since φ_r is a density. Hence $\psi(\widetilde{x}|\widetilde{y})$ is also a density in $\widetilde{x}$ for fixed $\widetilde{y}$. Now since by Lemma 2, $\widetilde{X}$ and $\widetilde{Y}$ are independent, by the conditional information inequality (25) of Corollary 1, one has

$$-r'\log\mathbb{E}\Big\{\big(\psi(\widetilde{X}|\widetilde{Y})\big)^{1/r'}\Big\}\geq h_r(\widetilde{X})=h_r(\sqrt{\lambda}X^*+\sqrt{1-\lambda}Y^*). \tag{64}$$

Combining with (61) yields the announced inequality (60).

It remains to settle the equality case in (60). From the above proof, equality holds in (60) if and only if both (58) and (64) are equalities. Equality in (58) holds if and only if for all $i=1,2,\ldots,n$,

$$\frac{\partial T_i}{\partial x_i}(X^*)=\frac{\partial U_i}{\partial y_i}(Y^*)\text{ a.s.} \tag{65}$$

Since X^* and Y^* are independent normal variables, this implies that $\frac{\partial T}{\partial x_i}$ and $\frac{\partial U}{\partial y_i}$ are constant and equal. In particular the Jacobian $|\lambda T'(\sqrt{\lambda}\widetilde{x}-\sqrt{1-\lambda}\widetilde{y})+(1-\lambda)U'(\sqrt{1-\lambda}\widetilde{x}+\sqrt{\lambda}\widetilde{y})|$ in (62) is constant.

From Corollary 1 equality in (64) holds if and only if $\psi(\widetilde{x}|\widetilde{y})$ does not depend on $\widetilde{y}$, which implies that $\sqrt{\lambda}T(\sqrt{\lambda}\widetilde{x}-\sqrt{1-\lambda}\widetilde{y})+\sqrt{1-\lambda}U(\sqrt{1-\lambda}\widetilde{x}+\sqrt{\lambda}\widetilde{y})$ does not depend on the value of $\widetilde{y}$. Taking derivatives with respect to y_j for all $j=1,2,\ldots,n$,

$$-\sqrt{\lambda}\sqrt{1-\lambda}\frac{\partial T_i}{\partial x_j}(\sqrt{\lambda}\widetilde{X}-\sqrt{1-\lambda}\widetilde{Y})+\sqrt{\lambda}\sqrt{1-\lambda}\frac{\partial U_i}{\partial x_j}(\sqrt{1-\lambda}\widetilde{X}+\sqrt{\lambda}\widetilde{Y})=0 \tag{66}$$

which implies

$$\frac{\partial T_i}{\partial x_j}(X^*)=\frac{\partial U_i}{\partial y_j}(Y^*)\text{ a.s.} \tag{67}$$

for all $i,j=1,2,\ldots,n$. Therefore, T and U are linear transformations, equal up to an additive constant (equal to 0 since all variables are assumed of zero mean). It follows that $X_p=T(X_p^*)$ and $Y_q=U(Y_q^*)$ are normal with respective distributions $X_p\sim\mathcal{N}(0,\mathbf{K}/p)$ and $Y_q\sim\mathcal{N}(0,\mathbf{K}/q)$. Hence X and Y are i.i.d. normal $\mathcal{N}(0,\mathbf{K})$. □

A straightforward generalization to several independent variables is the following

Corollary 2 (Rényi EPI for Several Variables). *Let $r_1,r_2,\ldots,r_m,r$ be exponents those conjugates $r_1',r_2',\ldots,r_m',r'$ are of the same sign and satisfy*

$$\sum_{i=1}^m\frac{1}{r_i'}=\frac{1}{r'} \tag{68}$$

and let $\lambda_1,\lambda_2,\ldots,\lambda_m$ be defined by

$$\lambda_i=\frac{r'}{r_i'}\qquad(i=1,2,\ldots,m). \tag{69}$$

Then for independent $X_1,X_2,\ldots,X_m$,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 July 2018 doi:10.20944/preprints201807.0145.v1

$$h_r\Big(\sum_{i=1}^m \sqrt{\lambda_i}X_i\Big) - \sum_{i=1}^m \lambda_i h_{r_i}(X_i) \geq \frac{n}{2}r'\Big(\frac{\log r}{r} - \sum_{i=1}^m \frac{\log r_i}{r_i}\Big) \tag{70}$$

with equality if and only if the X_i are i.i.d. normal.

Proof. By induction on m: The result for $m = 2$ is Theorem 3. Suppose the result satisfied at order $m-1$ and let $Y_m = \sum_{i=1}^{m-1} \sqrt{\lambda_i}X_i/\sqrt{1-\lambda_m}$ and s_m be such that $\frac{1}{s'_m} = \sum_{i=1}^{m-1} \frac{1}{r'_i}$. Notice that $\frac{1}{r'} = \frac{1}{r'_m} + \frac{1}{s'_m} = \frac{\lambda_m}{r'} + \frac{1}{s'_m}$, hence $r' = (1-\lambda_m)s'_m$. By Theorem 3, $h_r\big(\sum_{i=1}^m \sqrt{\lambda_i}X_i\big) = h_r(\sqrt{\lambda_m}X_m + \sqrt{1-\lambda_m}Y_m) \geq \lambda_m h_{r_m}(X_m) + (1-\lambda_m)h_{s_m}(Y_m) + \frac{n}{2}r'\big(\frac{\log r}{r} - \frac{\log r_m}{r_m} - \frac{\log s_m}{s_m}\big)$ with equality if and only if X_m, Y_m are i.i.d. normal. Now by the induction hypothesis, $h_{s_m}(Y_m) \geq \sum_{i=1}^{m-1} \frac{\lambda_i}{1-\lambda_m} h_{r_i}(X_i) + \frac{n}{2}s'_m\big(\frac{\log s_m}{s_m} - \sum_{i=1}^{m-1}\frac{\log r_i}{r_i}\big)$ with equality if and only if the X_i $(i = 1, 2, \ldots, m-1)$—and hence Y_m—are i.i.d. normal. The result at order m follows by combining the two inequalities since $(1-\lambda_m)s'_m = r'$. □

5. Recent Versions of the Rényi EPI

Definition 5 (Rényi Entropy Power [5]). *The* Rényi entropy power *of order r is defined by*

$$N_r(X) = e^{2h_r(X)/n}. \tag{71}$$

Up to a multiplicative constant, $N_r(X)$ is the (average) power of a white normal variable having the same Rényi entropy as X—hence the name "entropy power". In fact, if $X^* \sim \mathcal{N}(0, \sigma^2)$ has the same Rényi entropy $h_r(X^*) = h_r(X)$, then by Lemma 4,

$$\sigma^2 = \frac{e^{2h_r(X)/n}}{2\pi r^{r'/r}}. \tag{72}$$

The Renyi entropy power enjoys the same scaling property as for the usual power: By Lemma 3, for any $a \in \mathbb{R}$,

$$N_r(aX) = a^2 N_r(X). \tag{73}$$

For independent $X_1, X_2, \ldots, X_m$, Rényi entropy power inequalities take the either the form [5,6]

$$N_r\Big(\sum_{i=1}^m X_i\Big) \geq c\sum_{i=1}^m N_r(X_i) \tag{74}$$

for some positive constant c, or the form [7–9]

$$N_r^{\,\alpha}\Big(\sum_{i=1}^m X_i\Big) \geq \sum_{i=1}^m N_r^{\,\alpha}(X_i) \tag{75}$$

for some positive exponent α. The constants c and α may depend on the order r, the number m of variables and the dimension n. What is desired is:

- a maximum possible value of c in (74) since the inequality is automatically satisfied for all positive constants $c' < c$.
- a minimum possible value of α in (75) since the inequality is automatically satisfied for all positive exponents $\alpha' > \alpha$; in fact, since (75) is homogeneous by scaling the variables $X_i \mapsto aX_i$ as in (73), one may suppose without loss of generality that the r.h.s. of (75) is $= 1$; then $N_p(X_i) < 1$ for all i and $\sum_{i=1}^m N_r^{\,\alpha'}(X_i) \leq \sum_{i=1}^m N_r^{\,\alpha}(X_i)$.

The following useful characterization, which generalizes [8, Lemma 2.1], makes the link between the various versions (70), (74), (75) of the Rényi entropy power inequality.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 July 2018 doi:10.20944/preprints201807.0145.v1

Lemma 8. *For independent $X_1, X_2, \dots, X_m$, the Rényi EPI in the general form*

$$N_r^{\alpha}\Big(\sum_{i=1}^m X_i\Big) \geq c \sum_{i=1}^m N_r^{\alpha}(X_i) \tag{76}$$

for some constant $c > 0$ and exponent $\alpha > 0$ is equivalent to the following inequality

$$h_r\Big(\sum_{i=1}^m \sqrt{\lambda_i} X_i\Big) - \sum_{i=1}^m \lambda_i h_r(X_i) \geq \frac{n}{2}\Big(\frac{\log c}{\alpha} + (\frac{1}{\alpha} - 1) H(\lambda)\Big) \tag{77}$$

for any positive $\lambda_1, \lambda_2, \dots, \lambda_m$ such that $\sum_{i=1}^m \lambda_i = 1$, where $H(\lambda) > 0$ denotes the discrete entropy

$$H(\lambda) = \sum_{i=1}^m \lambda_i \log \frac{1}{\lambda_i} > 0. \tag{78}$$

Proof. Suppose (76) holds. Then

$$\begin{aligned} h_r\Big(\sum_{i=1}^m \sqrt{\lambda_i} X_i\Big) &= \frac{n}{2\alpha} \log N_r^{\alpha}\Big(\sum_{i=1}^m \sqrt{\lambda_i} X_i\Big) && (79)\\ &\geq \frac{n}{2\alpha} \log \sum_{i=1}^m N_r^{\alpha}(\sqrt{\lambda_i} X_i) + \frac{n}{2\alpha} \log c && (80)\\ &= \frac{n}{2\alpha} \log \sum_{i=1}^m \lambda_i^{\alpha} N_r^{\alpha}(X_i) + \frac{n}{2\alpha} \log c && (81)\\ &\geq \frac{n}{2\alpha} \sum_{i=1}^m \lambda_i \log(\lambda_i^{\alpha-1} N_r^{\alpha}(X_i)) + \frac{n}{2\alpha} \log c && (82)\\ &= \sum_{i=1}^m \lambda_i h_r(X_i) + \frac{n(\alpha-1)}{2\alpha} \sum_{i=1}^m \lambda_i \log \lambda_i + \frac{n}{2\alpha} \log c && (83) \end{aligned}$$

where the scaling property (73) is used in (81) and the concavity of the logarithm is used in (82). Conversely, suppose that (77) is satisfied for all $\lambda_i > 0$ such that $\sum_{i=1}^m \lambda_i = 1$. Set $\lambda_i = N_r^{\alpha}(X_i) / \sum_{i=1}^m N_r^{\alpha}(X_i)$. Then

$$\begin{aligned} N_r^{\alpha}\Big(\sum_{i=1}^m X_i\Big) &= \exp \frac{2\alpha}{n} h_r\Big(\sum_{i=1}^m \sqrt{\lambda_i} \frac{X_i}{\sqrt{\lambda_i}}\Big) && (84)\\ &\geq \exp \frac{2\alpha}{n} \sum_{i=1}^m \lambda_i h_r\big(\frac{X_i}{\sqrt{\lambda_i}}\big) \cdot c \cdot \exp(1-\alpha) \sum_{i=1}^m \lambda_i \log \frac{1}{\lambda_i} && (85)\\ &= c \prod_{i=1}^m \Big(N_r^{\alpha}\big(\frac{X_i}{\sqrt{\lambda_i}}\big) \lambda_i^{\alpha-1}\Big)^{\lambda_i} && (86)\\ &= c \prod_{i=1}^m \Big(N_r^{\alpha}(X_i) \lambda_i^{-1}\Big)^{\lambda_i} && (87)\\ &= c \Big(\sum_{i=1}^m N_r^{\alpha}(X_i)\Big)^{\sum_{i=1}^m \lambda_i} && (88)\\ &= c \sum_{i=1}^m N_r^{\alpha}(X_i). \quad \square && (89) \end{aligned}$$

5.1. Rényi Entropy Power Inequalities for Orders >1

From Lemma 8 and Corollary 2 it is easy to recover known Rényi EPIs and obtain new ones for orders $r > 1$. In fact, if $r > 1$ then $r' > 0$ and all r_i' are positive and $> r'$. Therefore all r_i are $< r$ and by monotonicity (Lemma 5),

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 July 2018 doi:10.20944/preprints201807.0145.v1

$$h_{r_i}(X_i) \geq h_r(X_i) \qquad (i = 1, 2, \ldots, m). \tag{90}$$

Plugging this into (70) one obtains

$$h_r\Big(\sum_{i=1}^{m} \sqrt{\lambda_i} X_i\Big) - \sum_{i=1}^{m} \lambda_i h_r(X_i) \geq \frac{n}{2} r' \Big(\frac{\log r}{r} - \sum_{i=1}^{m} \frac{\log r_i}{r_i}\Big) \tag{91}$$

where $\lambda_i = r'/r'_i$ for $i = 1, 2, \ldots, m$. For future reference define[1]

$$A(\lambda) = |r'| \Big(\frac{\log r}{r} - \sum_{i=1}^{m} \frac{\log r_i}{r_i}\Big) \tag{92}$$

$$= |r'| \Big(\frac{\log r}{r} + \sum_{i=1}^{m} (1 - \frac{\lambda_i}{r'}) \log(1 - \frac{\lambda_i}{r'})\Big) \tag{93}$$

This function is strictly convex in $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ because $x \mapsto (1 - x/r') \log(1 - x/r')$ is strictly convex. Note that $A(\lambda)$ vanishes in the limiting cases where λ tends to one of the vectors $(1, 0, \ldots, 0)$, $(0, 1, 0, \ldots, 0)$, …, $(0, 0, \ldots, 0, 1)$ and since every λ is a convex combination of these vectors and $A(\lambda)$ is strictly convex, one has $A(\lambda) < 0$.

Theorem 4 (Ram and Sason [6])**.** *The Rényi EPI (74) holds for $r > 1$ and $c = r^{r'/r}(1 - \frac{1}{mr'})^{mr'-1}$.*

Proof. By Lemma 8 for $\alpha = 1$ we only need to check that the r.h.s. of (91) is greater than $\frac{n}{2} \log c$ for any choice of the λ_i's, that is, for any choice of exponents r_i such that $\sum_{i=1}^{m} \frac{1}{r'_i} = \frac{1}{r'}$. Thus (74) will hold for $\log c = \min_\lambda A(\lambda)$. Now by the log-sum inequality [4, Theorem 2.7.1],

$$\sum_{i=1}^{m} \frac{1}{r_i} \log \frac{1}{r_i} \geq \Big(\sum_{i=1}^{m} \frac{1}{r_i}\Big) \log \frac{\sum_{i=1}^{m} \frac{1}{r_i}}{m} = (m - 1/r') \log \frac{m - 1/r'}{m} \tag{94}$$

with equality if and only if all r_i are equal, that is, the λ_i are equal to $1/m$. Thus $\min_\lambda A(\lambda) = r'\Big(\frac{\log r}{r} + (m - 1/r') \log \frac{m-1/r'}{m}\Big)$ which yields $c = r^{r'/r}(1 - \frac{1}{mr'})^{mr'-1}$. □

An alternate proof is to argue that $A(\lambda)$ is convex and symmetrical in $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ and is, therefore, minimized when all λ_i are equal.

Remark 5. *The above constant c is certainly not optimal since equality in (90) holds if and only if the X_i are uniformly distributed (Lemma 5) while equality in (70) holds if and only if the X_i are identically normally distributed (Corollary 2). Ram and Sason [6] tightened (74) further using optimization techniques, resulting in a constant that depends on the relative values of the entropy powers themselves.*

Remark 6. *It can be noted that $\log c = r' \frac{\log r}{r} + (mr' - 1) \log(1 - \frac{1}{mr'}) < 0$ decreases as m increases; in fact $\frac{\partial \log c}{\partial m} = r' \log(1 - \frac{1}{mr'}) + \frac{mr'}{r'm^2} < r'(-\frac{1}{mr'}) + \frac{1}{m} = 0$. Thus, a universal constant independent of m is obtained by taking*

$$c = \inf_m r^{r'/r}(1 - \frac{1}{mr'})^{mr'-1} = r^{r'/r} \lim_{m\to\infty} (1 - \frac{1}{mr'})^{mr'-1} = \frac{r^{r'/r}}{e} \tag{95}$$

This was the constant established by Bobkov and Chistyakov [5].

Theorem 5. *The Rényi EPI (75) holds for $r > 1$ and $\alpha = \Big(1 + \frac{r' \frac{\log r}{r} + (mr'-1) \log\big(1 - \frac{1}{mr'}\big)}{\log m}\Big)^{-1}$.*

---

[1] The absolute value $|r'|$ is needed in the next subsection where r' will ne negative.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 July 2018 doi:10.20944/preprints201807.0145.v1

Proof. By Lemma 8 for $c = 1$ we only need to check that the r.h.s. of (91) is greater than $\frac{n}{2}(1/\alpha - 1)H(\lambda)$ for any choice of the λ_i's, that is, for any choice of exponents r_i such that $\sum_{i=1}^m \frac{1}{r'_i} = \frac{1}{r'}$. Thus (75) will hold for $\frac{1}{\alpha} - 1 = \min_\lambda \frac{A(\lambda)}{H(\lambda)}$. By the proof of the preceding theorem, the numerator is minimized when all λ_i are equal and this also maximizes the entropy $= \log m$ in the denominator. However on cannot conclude yet since the minimum in the numerator is negative.

Suppose that the minimum is *not* attained when all λ_i are equal. Since $A(\lambda)/H(\lambda)$ is symmetric in λ, if the minimum is attained in λ, then it is also attained at any permutation λ' of λ. Without loss of generality we may assume that $\lambda_1 \neq \lambda_2$ and take $\lambda' = (\lambda_2, \lambda_1, \lambda_3, \ldots, \lambda_m)$. Now for every $\mu \in [0, 1]$ consider

$$\varphi(\mu) = \frac{A(\mu\lambda + (1-\mu)\lambda')}{H(\mu\lambda + (1-\mu)\lambda')} \tag{96}$$

so that $\min_\lambda \frac{A(\lambda)}{H(\lambda)} = \varphi(0) = \varphi(1)$. Differentiating one finds

$$\varphi'(1) = \frac{B(\lambda)}{H(\lambda)} - \frac{A(\lambda)C(\lambda)}{H(\lambda)^2} \tag{97}$$

where $B(\lambda) = \frac{\partial A(\mu\lambda+(1-\mu)\lambda')}{\partial\mu}\big|_{\mu=1} = \sum_{i=1}^m (\lambda'_i - \lambda_i)\log\big(e(1-\lambda_i/r')\big) = (\lambda_2 - \lambda_1)\log\frac{1-\lambda_1/r'}{1-\lambda_2/r'} > 0$ and $C(\lambda) = \frac{\partial H(\mu\lambda+(1-\mu)\lambda')}{\partial\mu}\big|_{\mu=1} = \sum_{i=1}^m (\lambda'_i - \lambda_i)\log(e\lambda_i) = (\lambda_2 - \lambda_1)\log\frac{\lambda_1}{\lambda_2} < 0$. Therefore, $\varphi'(1) > 0$ and similarly one finds $\varphi'(0) < 0$. It follows that there exists $0 < \mu < 1$ such that $\varphi(\mu) < \varphi(0) = \varphi(1)$, a contradiction.

Therefore, the minimum of $A(\lambda)/H(\lambda)$ is attained when all λ_i are equal to $1/m$. This gives $\frac{1}{\alpha} - 1 = \Big(r'\frac{\log r}{r} + (mr'-1)\log(1 - \frac{1}{mr'})\Big)/\log m$. □

Remark 7. *The case $m = 2$ yields $\alpha = \frac{r-1}{(r+1)\log_2(r+1) - r\log_2 r - 2}$ which was found by Li [8] who remarked that this value of α is strictly smaller (better) than the value $\alpha = \frac{r+1}{2}$ obtained by Bobkov and Marsiglietti [7].*

For $m > 2$ the exponent of Theorem 5 is even smaller (better). In fact, by Remark 6, it is immediate to see that $\Big(r'\frac{\log r}{r} + (mr'-1)\log(1 - \frac{1}{mr'})\Big)/\log m$ is negative and decreases toward 0 as m increases. Therefore, the exponent α decreases (is improved) as m increases.

In all cases $\alpha > 1$ and tends to 1 as $m \to +\infty$. However, using the same method it is easy to obtain Rényi EPIs with exponent values $\alpha < 1$. This is given by the following Theorem.

Theorem 6. *The Rényi EPI (76) holds for $r > 1$, $0 < \alpha < 1$ with $c = \Big(m\, r'^{r'/r}(1 - \frac{1}{mr'})^{mr'-1}\Big)^\alpha/m$.*

Proof. By Lemma 8 we only need to check that the r.h.s. of (91) is greater than $\frac{n}{2}\big((\log c)/\alpha + (1/\alpha - 1)H(\lambda)\big)$, that is, $A(\lambda) \geq (\log c)/\alpha + (1/\alpha - 1)H(\lambda)$ for any choice of the λ_i's, that is, for any choice of exponents r_i such that $\sum_{i=1}^m \frac{1}{r'_i} = \frac{1}{r'}$. Thus for a given $0 < \alpha < 1$, (76) will hold for $\log c = \min_\lambda \alpha A(\lambda) - (1-\alpha)H(\lambda)$. From the preceding proofs the minimum attained when all λ_i are equal. This gives $\log c = \alpha\Big(r'\frac{\log r}{r} + (mr'-1)\log(1 - \frac{1}{mr'})\Big) - (1-\alpha)\log m$. □

5.2. Rényi Entropy Power Inequalities for Orders <1 and Log-Concave Densities

If $r < 1$ then $r' < 0$ and all r'_i are negative and $< r'$. Therefore all r_i are $> r$ and by monotonicity (Lemma 5), the opposite inequality of (90) holds and the method of the preceding subsection fails. For log-concave densities, however, (90) can be replaced by a similar inequality in the right direction.

Definition 6 (Log-Concave Density)**.** *A density f is log-concave if $\log f$ is concave in it support, i.e., for all $0 < \mu < 1$,*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 July 2018 doi:10.20944/preprints201807.0145.v1

$$f(x)^{\mu} f(y)^{1-\mu} \leq f(\mu x + (1-\mu)y). \tag{98}$$

Lemma 9. *If X has a log-concave density, then $h_p(pX) - ph_p(X) = n\log p + (1-p)h_p(X)$ is concave in p.*

As noted below in Remark 8, this is essentially a result obtained by Fradelizi, Madiman and Wang [18]. The following alternate proof uses the transport properties seen in Section 4.

Proof. Define $r = \lambda p + (1-\lambda)q$ where $0 < \lambda < 1$. By Lemma 1 there exists two diffeomorphisms T, U such that one can write $pX_p = T(X^*)$ and $qX_q = U(X^*)$. Then X^* has density

$$\tfrac{1}{p^n} f_p\big(\tfrac{T(x^*)}{p}\big)|T'(x^*)| = \tfrac{1}{q^n} f_q\big(\tfrac{U(x^*)}{q}\big)|U'(x^*)| = \tfrac{1}{p^{\lambda n} q^{(1-\lambda)n}} f_p\big(\tfrac{T(x^*)}{p}\big)^{\lambda} f_q\big(\tfrac{U(x^*)}{q}\big)^{1-\lambda} |T'(x^*)|^{\lambda} |U'(x^*)|^{1-\lambda} \tag{99}$$

Now, by log-concavity (98) with $\mu = \lambda p / r$,

$$f_p\big(\tfrac{T(x^*)}{p}\big)^{\lambda} f_q\big(\tfrac{U(x^*)}{q}\big)^{1-\lambda} = \frac{1}{\|f\|_p^{\lambda p} \|f\|_q^{(1-\lambda)q)}} f\big(\tfrac{T(x^*)}{p}\big)^{\lambda p} f\big(\tfrac{U(x^*)}{q}\big)^{(1-\lambda)q} \tag{100}$$

$$\leq \frac{1}{\|f\|_p^{\lambda p} \|f\|_q^{(1-\lambda)q)}} f\big(\tfrac{\lambda T(x^*) + (1-\lambda)U(x^*)}{r}\big)^{r} \tag{101}$$

$$= \frac{\|f\|_r^r}{\|f\|_p^{\lambda p} \|f\|_q^{(1-\lambda)q)}} f_r\big(\tfrac{\lambda T(x^*) + (1-\lambda)U(x^*)}{r}\big) \tag{102}$$

Using the arithmetic-geometric mean inequality (58) and integrating the density (99) over $x^* \in \mathbb{R}^n$, one obtains

$$(p^{\lambda} q^{1-\lambda})^n \|f\|_p^{\lambda p} \|f\|_q^{(1-\lambda)q)} \leq r^n \|f\|_r^r. \tag{103}$$

Taking the logarithm yields the announced concavity. □

Remark 8. *Since $n\log p + (1-p)h_p(X)$ is concave, one has $\frac{\partial^2}{\partial p^2}(n\log p + (1-p)h_p(X)) \leq 0$. Differentiating (16) with respect to p it is easy to show that this is equivalent to the following "varentropy bound" $\operatorname{Var}\log f(X_p) \leq n/p^2$, that is, $\operatorname{Var}\log f_p(X_p) \leq n$ which was obtained in [18].*

Since $n\log p + (1-p)h_p(X)$ is concave and vanishes for $p = 1$, the slopes $\frac{n\log p + (1-p)h_p(X) - 0}{p-1}$ are nonincreasing in p. In other words $h_p(X) + n\frac{\log p}{1-p}$ is nondecreasing. Therefore:

Corollary 3 (Marsiglietti and Melbourne [9]). *If $p < q$ then for any X with log-concave density, $h_p(X) + n\frac{\log p}{1-p} \leq h_q(X) + n\frac{\log q}{1-q}$.*

We can now use Lemma 8 and Corollary 2 to obtain Rényi EPIs for orders $r < 1$. Since all r_i are $> r$, by Corollary 3,

$$h_{r_i}(X) + n\frac{\log r_i}{1-r_i} \geq h_r(X) + n\frac{\log r}{1-r} \qquad (i = 1, 2, \ldots, m). \tag{104}$$

Plugging this into (70) one obtains

$$h_r\Big(\sum_{i=1}^m \sqrt{\lambda_i} X_i\Big) - \sum_{i=1}^m \lambda_i h_r(X_i) \geq n\Big(\frac{\log r}{1-r} - \sum_{i=1}^m \lambda_i \frac{\log r_i}{1-r_i}\Big) + \frac{n}{2} r' \Big(\frac{\log r}{r} - \sum_{i=1}^m \frac{\log r_i}{r_i}\Big) \tag{105}$$

$$= \frac{n}{2} r' \Big(\sum_{i=1}^m \frac{\log r_i}{r_i} - \frac{\log r}{r}\Big) \tag{106}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 July 2018 doi:10.20944/preprints201807.0145.v1

where we have used that $\lambda_i = r'/r'_i$ for $i = 1, 2, \ldots, m$. Notice that the r.h.s. of (106) for $r < 1$ ($r' < 0$) is the opposite of that of (91) for $r > 1$ ($r' > 0$). Note, however, that since r' is now negative, the r.h.s. is exactly $= \frac{n}{2} A(\lambda)$ which is still convex and negative. For this reason, the proofs of the following theorems for $r < 1$ are such repeats of the theorems obtained previously for $r > 1$.

Theorem 7. *The Rényi EPI (74) for log-concave densities holds for* $c = r^{-r'/r}\big(1 - \frac{1}{mr'}\big)^{1-mr'}$ *and* $r < 1$.

Proof. Identical to that of Theorem 4 except for the change $|r'| = -r'$ in the expression of $A(\lambda)$. □

Theorem 8. *The Rényi EPI (75) for log-concave densities holds for* $\alpha = \Big(1 - \frac{r' \frac{\log r}{r} + (mr'-1)\log\big(1-\frac{1}{mr'}\big)}{\log m}\Big)^{-1}$ *and* $r < 1$.

Proof. Identical to that of Theorem 5 except for the change $|r'| = -r'$ in the expression of $A(\lambda)$. □

Remark 9. *The case* $m = 2$ *yields* $\alpha = \frac{1-r}{(r+1)\log_2(r+1) - r\log_2 r - 2r}$ *which was found by Marsiglietti and Melbourne [9]. Again the exponent of the Theorem is improved (strictly smaller) for* $m > 2$. *In fact* α *decreases and tends to one as* $m \to +\infty$.

Theorem 9. *The Rényi EPI (76) for log-concave densities holds for* $c = \Big(mr^{-r'/r}\big(1 - \frac{1}{mr'}\big)^{1-mr'}\Big)^{\alpha}/m$ *where* $r < 1$ *and* $0 < \alpha < 1$.

Proof. Identical to that of Theorem 6 except for the change $|r'| = -r'$ in the expression of $A(\lambda)$. □

6. Conclusion

This article provides a comprehensive framework to derive all known Rényi entropy power inequalities (with much shorter proofs), and prove new ones. The framework is based on a transport argument from normal densities and a change of variable by rotation. Only basic properties of Rényi entropies are used in the proofs.

In particular, the α-modification of the EPI is generalized to more than two independent variables for Rényi entropy orders > 1 as well as for orders < 1. Also, the Rényi EPI with multiplicative constant c is extended to Rényi entropy orders < 1, and a more general formulation with both exponent α and constant c is obtained for all orders.

As a perspective, the methods developed in this paper can perhaps be generalized to obtain reverse Rényi entropy power inequalities (see e.g., the discussion in [8]).

1. Shannon, C.E. A mathematical theory of communication. *Bell System Technical Journal* **1948**, *27*, 623–656.
2. Rioul, O. Information Theoretic Proofs of Entropy Power Inequalities. *IEEE Transactions on Information Theory* **2011**, *57*, 33–55.
3. Dembo, A.; Cover, T.M.; Thomas, J.A. Information Theoretic Inequalities. *IEEE Transactions on Information Theory* **1991**, *37*, 1501–1518.
4. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley, 2006.
5. Bobkov, S.G.; Chistyakov, G.P. Entropy Power Inequality for the Rényi Entropy. *IEEE Transactions on Information Theory* **2015**, *61*, 708–714.
6. Ram, E.; Sason, I. On Rényi Entropy Power Inequalities. *IEEE Transactions on Information Theory* **2016**, *62*, 6800–6815.
7. Bobkov, S.G.; Marsiglietti, A. Variants of the Entropy Power Inequality. *IEEE Transactions on Information Theory* **2017**, *63*, 7747–7752.
8. Li, J. Rényi entropy power inequality and a reverse. *Studia Mathematica* **2018**, *242*, 303–319.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 July 2018 doi:10.20944/preprints201807.0145.v1

9. Marsiglietti, A.; Melbourne, J. On the entropy power inequality for the Rényi entropy of order $[0, 1]$. *Preprint available online: https://arxiv.org/abs/1710.00800* **2018**.
10. Rioul, O. Yet Another Proof of the Entropy Power Inequality. *IEEE Transactions on Information Theory* **2017**, *63*, 3595–3599.
11. Rioul, O. Optimal transportation to the entropy-power inequality. IEEE Information Theory and Applications Workshop (ITA 2017), 2017.
12. Bryc, W. *The Normal Distribution - Characterizations with Applications*; Vol. 100, *Lecture Notes in Statistics*, Springer, 1995.
13. Rioul, O. Optimal Transport to Rényi Entropies. Proc. 3rd Conference on Geometric Science of Information (GSI 2017). Springer, 2017, Vol. 10589, *Lecture Notes in Computer Science*.
14. Bercher, J.F. Source Coding with Escort Distributions and Rényi Entropy Bounds. *Physics Letters A* **2009**, *373*, 3235–3238.
15. Bercher, J.F. On Generalized Cramér–Rao Inequalities, Generalized Fisher Information and Characterizations of Generalized q-Gaussian Distributions. *Journal of Physics A: Mathematical and Theoretical* **2012**, *45*, 255303 (15pp).
16. Beck, C.; Schlögl, F. *Thermodynamics of Chaotic Systems: An Introduction*; Cambridge University Press, 1993.
17. Campbell, L.L. A Coding Theorem and Rényi's Entropy. *Information and Control* **1965**, *8*, 423–429.
18. Fradelizi, M.; Madiman, M.; Wang, L. Optimal Concentration of Information Content for Log-Concave Densities. High Dimensional Probability VII: The Cargèse Volume; Houdré, C.; Mason, D.M.; Reynaud-Bouret, P.; Rosiński, J., Eds.; Birkhäuser: Basel, 2016.