Entropy power inequality for the Rényi entropy

Sergey G. Bobkov and Gennadiy P. Chistyakov

***Abstract*—The classical entropy power inequality is extended to the Rényi entropy. We also discuss the question of the existence of the entropy for sums of independent random variables.**

***Index Terms*—Renyi entropy; entropy power inequality.**

I. Introduction

Given a random vector X in the Euclidean space $\mathbb{R}^d$ with density p, the (differential) Rényi entropy of order $\alpha > 1$ is defined by

$$h_\alpha(X) = -\frac{1}{\alpha - 1}\,\log \int_{\mathbb{R}^d} p(x)^\alpha\,dx.$$

Introduce the corresponding Rényi entropy power

$$N_\alpha(X) = \exp\Big[\,\frac{2}{d}\,h_\alpha(X)\Big] = \bigg(\int_{\mathbb{R}^d} p(x)^\alpha\,dx\bigg)^{-\frac{2}{d}\,\frac{1}{\alpha-1}}.$$

Here the ratio $\frac{2}{d}$ in the exponent makes this functional to be homogeneous of order 2 (although sometimes a different constant is used, cf. e.g. [16], or it is just omitted).

Both quantities are well-defined and may take the values $-\infty \leq h_\alpha(X) < \infty$ and $0 \leq N_\alpha(X) < \infty$. If the distribution of X is not absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$, put $h_\alpha(X) = -\infty$ and $N_\alpha(X) = 0$.

As α increases, the entropies $\alpha \rightarrow h_\alpha(X)$ do not increase, so the limits

$$h_1(X) = \lim_{\alpha \rightarrow 1} h_\alpha(X), \qquad N_1(X) = \lim_{\alpha \rightarrow 1} N_\alpha(X)$$

exist. Moreover, if $\int_{\mathbb{R}^d} p(x)^\alpha\,dx < \infty$, for some $\alpha > 1$, we arrive at the usual Shannon entropy $h_1 = h$ and the entropy power $N_1 = N$, where

$$h(X) = -\int_{\mathbb{R}^d} p(x)\log p(x)\,dx, \quad N(X) = \exp\Big[\,\frac{2}{d}\,h(X)\Big].$$

On the other side, as $\alpha \rightarrow \infty$, we deal with the functions of the maximum of the density,

$$h_\infty(X) = \lim_{\alpha \rightarrow \infty} h_\alpha(X) = \log \frac{1}{M(X)},$$

$$N_\infty(X) = \lim_{\alpha \rightarrow \infty} N_\alpha(X) = M^{-\frac{2}{d}}(X),$$

where

$$M(X) = \operatorname{ess\,sup}_x p(x).$$

School of Mathematics, University of Minnesota, 127 Vincent Hall, 206 Church St. S.E., Minneapolis, MN 55455 USA. Email: bobkov@math.umn.edu

Falultät für Mathematik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Deutschland. Email: chistyak@math.uni-bielefeld.de

S. G. Bobkov was supported by NSF grant DMS-1106530 and Simons Fellowship, and G. P. Chistyakov was supported by SFB 701.

These functionals contain an important information about the distributions and possess a number of remarkable properties. For example, all N_α are affine invariant, i.e.,

$$N_\alpha(TX) = N_\alpha(X),$$

where $T : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is an arbitrary linear operator preserving the Lebesgue measure ($|\det(T)| = 1$). In addition, they are translation invariant and, as was mentioned, homogeneous of order 2, i.e., for all $h \in \mathbb{R}^d$, $\lambda \in \mathbb{R}$,

$$N_\alpha(X + h) = N_\alpha(X), \qquad N_\alpha(\lambda X) = \lambda^2 N_\alpha(X).$$

These two identities resemble a similar property of the variance functional.

As another general property, one should mention monotonicity under convolutions ([5]): We have

$$N_\alpha(X + Y) \geq N_\alpha(X),$$

whenever X and Y are independent. (Since we assume that $\alpha > 1$, this property follows from the convexity of the power function $t \rightarrow t^\alpha$). In other words, the Rényi entropy may only increase when adding an independent summand.

If one restricts to the usual case $\alpha = 1$, the monotonicity property may considerably be sharpened. Namely, the classical entropy power inequality indicates that

$$N(X + Y) \geq N(X) + N(Y), \tag{1}$$

cf. [17], [6], [10], [11], [14]. More generally, if $X_1, \dots, X_n$ are independent random vectors in $\mathbb{R}^d$,

$$N(X_1 + \dots + X_n) \geq \sum_{k=1}^n N(X_k). \tag{2}$$

One may therefore wonder how to extend this inequality to other values of α. The point is that with the current definition of N_α, the last two relations do not hold in general for the Rényi powers.

Example. Let $d = 1$ and let X and Y be independent random variables uniformly distributed in the interval $(0, 1)$. They have density $p(x) = 1_{\{0<x<1\}}$, so

$$N_\alpha(X) = 1, \quad \text{for all } \alpha \geq 1.$$

The sum $X + Y$ has the triangle distribution with density $1 - |1 - x|$ $(0 < x < 2)$, so,

$$\begin{aligned} N_\alpha(X + Y) &= \bigg(\int_0^2 (1 - |1 - x|)^\alpha\,dx\bigg)^{-\frac{2}{\alpha-1}} \\ &= \Big(\frac{\alpha+1}{2}\Big)^{\frac{2}{\alpha-1}}. \end{aligned}$$

In particular, $N_3(X+Y) = 2$. Since $\alpha \to N_\alpha(X+Y)$ is decreasing, we conclude that

$$N_\alpha(X+Y) \geq N_\alpha(X) + N_\alpha(Y), \quad \text{if and only if } 1 \leq \alpha \leq 3.$$

A similar conclusion can be made, when X and Y are independent random vectors uniformly distributed in the cube $(0,1)^d$.

Nevertheless, one can show that, at the expense of a universal factor in front of the sum, the inequality (2) does hold for all N_α regardless of the number of the summands. More precisely, we prove:

Theorem I.1. *If $X_1, \dots, X_n$ are independent random vectors in $\mathbb{R}^d$, then*

$$N_\alpha(X_1 + \dots + X_n) \geq c_\alpha \sum_{k=1}^n N_\alpha(X_k), \tag{3}$$

where $c_\alpha > 0$ depends on α, only. Moreover, one may take

$$c_\alpha = \frac{1}{e}\, \alpha^{\frac{1}{\alpha-1}} \quad (\alpha > 1).$$

Hence, we have an effect which is similar to the usual entropy power: The Rényi entropy power increases at least linearly with respect to the summands. Note, however, that in contrast with (1)-(2), the inequality (3) cannot be reduced to the case of two summands, since an application of the induction argument would lead to a logarithmically decaying constant (with respect to n).

As a function of α, the constants c_α decrease from 1 to $\frac{1}{e}$. This can already be seen from the identity

$$\frac{1}{2\pi}\, N_\alpha(Z) = \alpha^{\frac{1}{\alpha-1}},$$

where Z is a standard normal random variable. Hence, in the limit as $\alpha \to 1$, (3) contains the entropy power inequality (2). In the other limit, as $\alpha \to \infty$, we get a similar inequality for the maximum of the density,

$$M^{-\frac{2}{d}}(X_1 + \dots + X_n) \geq \frac{1}{e} \sum_{k=1}^n M^{-\frac{2}{d}}(X_k).$$

This particular case has recently been considered in [7]. Moreover, combining some delicate results due to Rogozin [15] and Ball [1], in dimension $d = 1$ one may replace $\frac{1}{e}$ with constant $\frac{1}{2}$. It is best possible and is attained for $n = 2$ in the above mentioned example of the uniform distribution on $(0,1)$.

Anyhow, in all cases,

$$N_\alpha(X_1 + \dots + X_n) \geq \frac{1}{e} \sum_{k=1}^n N_\alpha(X_k).$$

Let us also state another immediate consequence of Theorem I.1, which makes use of the homogeneity of the Rényi power.

Corollary I.2. *If the independent random vectors $X_1, \dots, X_n$ satisfy $N_\alpha(X_k) \geq N$, for all $k \leq n$, then*

$$N_\alpha(\theta_1 X_1 + \dots + \theta_n X_n) \geq \frac{1}{e}\, N,$$

whenever $\theta_1^2 + \dots + \theta_n^2 = 1$.

Note that the Rényi entropy and the corresponding entropy power are also defined and treated for the parameter values $0 < \alpha < 1$ (with the same formula). However, we do not know, whether or not an analog of Theorem I.1 remains to hold in this case. Theorem I.1 can be proved with the argument which is similar to the one of Lieb [12], who applied the Young inequality with sharp constants when deriving the entropy power inequality (1). The general case $\alpha > 1$ is however a bit more delicate and requires to work with convolutions of an increasing number of densities.

In the next section we remind a basic general result, and then specialize it to convolutions of probability densities (Section III). Final steps of the proof of Theorem I.1 are made in Section IV. Then we turn to the classical case $\alpha = 1$ (Sections V-VI) to address the following question which seems to be ignored in the literature: Under what most general assumptions should one formulate the entropy power inequality (1)?

II. Sharp Young's inequality

For $1 \leq \nu \leq \infty$, we denote by L^ν the space of all measurable functions u on $\mathbb{R}^d$ with finite norm

$$\|u\|_\nu = \bigg(\int_{\mathbb{R}^d} |u(x)|^\nu\, dx\bigg)^{1/\nu}.$$

In particular, $\|u\|_\infty = \operatorname{ess\,sup}_x |u(x)|$.

Define the conjugate power

$$\nu' = \frac{\nu}{\nu - 1},$$

so that $\frac{1}{\nu} + \frac{1}{\nu'} = 1$.

The Young inequality with sharp constants was discovered by Beckner [4], cf. also [8], [2]; it is stated below for several functions following [4], Theorem 4. Put

$$A_\nu = \nu^{1/\nu}\, (\nu')^{-1/\nu'}, \quad A_1 = A_\infty = 1,$$

and

$$A = \Big(\frac{A_{\nu_1} \dots A_{\nu_n}}{A_\nu}\Big)^{d/2}$$

for given numbers $1 \leq \nu_k \leq \infty$ $(1 \leq k \leq n)$ and $1 \leq \nu \leq \infty$.

Theorem II.1. *Assume that $u_k \in L^{\nu_k}$ $(k = 1, \dots, n)$, and let $\sum_{k=1}^n \frac{1}{\nu_k'} = \frac{1}{\nu'}$. If $\nu_k, \nu \geq 1$, then the convolution $u = u_1 * \dots * u_n$ belongs to L^ν and has the norm*

$$\|u\|_\nu \leq A\, \|u_1\|_{\nu_1} \dots \|u_n\|_{\nu_n}.$$

For example, when $\nu = \nu' = 2$, the basic condition on the parameters becomes

$$\sum_{k=1}^n \frac{1}{\nu_k'} = \frac{1}{2}.$$

If $\nu_1 = 2$, then all remaining values must be $\nu_k = 1$, so that $\nu_k' = \infty$ $(k = 2, \dots, n)$. Hence, in general, the L^2-norm of u cannot be estimated from above in terms of the L^2-norms of u_k. Nevertheless, this is possible, if we additionally know the L^1-norms of u_k. Such a conclusion can be made, in particular, when dealing with probability densities on $\mathbb{R}^d$.

III. The case of densities

Here we develop one application of Theorem II.1 involving (probability) densities on $\mathbb{R}^d$.

First note that, given a measurable function $u \geq 0$ on a measure space (Ω, μ), by Hölder's inequality, there is a family of relations

$$\int u^{(1-\theta)p+\theta q}\, d\mu \leq \Big(\int u^p\, d\mu\Big)^{1-\theta} \Big(\int u^q\, d\mu\Big)^{\theta},$$

holding whenever $p, q \geq 0$, $0 \leq \theta \leq 1$. In particular, taking $p = 1$ and assuming that $\int u\, d\mu = 1$, we have

$$\int u^{(1-\theta)+\theta q}\, d\mu \leq \Big(\int u^q\, d\mu\Big)^{\theta}, \quad q \geq 0,\ 0 \leq \theta \leq 1. \tag{4}$$

Let $q = \alpha > 1$ and $\nu \geq 1$. Then $(1-\theta)+\theta\alpha = \nu \Longleftrightarrow \theta = \frac{\nu-1}{\alpha-1}$. Hence, if $1 \leq \nu \leq \alpha$,

$$\int u^{\nu}\, d\mu \leq \Big(\int u^{\alpha}\, d\mu\Big)^{\frac{\nu-1}{\alpha-1}}.$$

Let us state this particular case of (4) separately by raising the above inequality to the power $1/\nu$ and assuming that μ is the Lebesgue measure on $\Omega = \mathbb{R}^d$. As before, $\nu' = \frac{\nu}{\nu-1}$.

Lemma III.1. *For any density p on $\mathbb{R}^d$,*

$$\|p\|_{\nu} \leq \bigg[\Big(\int_{\mathbb{R}^d} p(x)^{\alpha}\, dx\Big)^{\frac{1}{\alpha-1}}\bigg]^{1/\nu'} \quad (1 \leq \nu \leq \alpha,\ \alpha > 1).$$

Next, given densities p_k on $\mathbb{R}^d$ $(k = 1, \dots, n)$, consider the convolution $p = p_1 * \cdots * p_n$. By Theorem II.1, for $1 \leq \nu_k \leq \infty$, $\alpha > 1$, we have

$$\|p\|_{\alpha} \leq\ A\, \|p_1\|_{\nu_1} \dots \|p_n\|_{\nu_n} \tag{5}$$

as long as

$$\sum_{k=1}^{n} \frac{1}{\nu_k'} = \frac{1}{\alpha'} = 1 - \frac{1}{\alpha}, \tag{6}$$

and with constant

$$A = \Big(\frac{A_{\nu_1} \dots A_{\nu_n}}{A_{\alpha}}\Big)^{d/2}. \tag{7}$$

The condition (6) requires that $\frac{1}{\nu_k'} = 1 - \frac{1}{\nu_k} \leq 1 - \frac{1}{\alpha}$ which is equivalent to $1 \leq \nu_k \leq \alpha$. Hence, one may apply Lemma III.1 to p_k, ν_k and α, and get

$$\|p_k\|_{\nu_k} \leq \bigg[\Big(\int_{\mathbb{R}^d} p(x)^{\alpha}\, dx\Big)^{\frac{1}{\alpha-1}}\bigg]^{1/\nu_k'}. \tag{8}$$

Combining (5) with (8), we obtain that

$$\begin{aligned} \|p\|_{\alpha} &\leq A \prod_{k=1}^{n} \bigg[\Big(\int_{\mathbb{R}^d} p_k(x)^{\alpha}\, dx\Big)^{\frac{1}{\alpha-1}}\bigg]^{1/\nu_k'} \\ &= A \prod_{k=1}^{n} \|p_k\|_{\alpha}^{\alpha'/\nu_k'} \end{aligned} \tag{9}$$

with constant described in (7).

Now we see that in contrast with Theorem II.1, for any $\alpha > 1$, the L^{α}-norm of p can indeed be estimated in terms of the L^{α}-norms of p_k. Moreover, the equality (6) provides some freedom in choosing ν_k which may be used to make (9) as sharp, as possible.

But, first let us restate the inequality (9) in terms of the Rényi power. If a random vector X in $\mathbb{R}^d$ has density p,

$$\|p\|_{\alpha} = \bigg[\Big(\int_{\mathbb{R}^d} p(x)^{\alpha}\, dx\Big)^{\frac{1}{\alpha-1}}\bigg]^{1/\alpha'} = N_{\alpha}(X)^{-\frac{d}{2\alpha'}}.$$

Hence, raising (9) to the power $-\frac{2\alpha'}{d}$, we arrive at the inequality with a constant which does not depend on d.

Corollary III.2. *If $X_1, \dots, X_n$ are independent random vectors in $\mathbb{R}^d$, then for the sum $S_n = X_1 + \dots + X_n$ we have*

$$N_{\alpha}(S_n) \geq B \prod_{k=1}^{n} \big(N_{\alpha}(X_k)\big)^{\frac{\alpha'}{\nu_k'}}, \tag{10}$$

whenever (6) holds with $\nu_k \geq 1$, $\alpha > 1$, and with constant

$$B = (A_{\nu_1} \dots A_{\nu_n} A_{\alpha'})^{-\alpha'}. \tag{11}$$

Strictly speaking, the application of Lemma III.1 still requires that X_k have densities. This assumption may, however, be removed. Indeed, in the case when the distribution of any of X_k is not absolutely continuous, then $N_{\alpha}(X_k) = 0$, and (10) is fulfilled automatically.

IV. Proof of Theorem I.1

One can try to optimize the inequality (10) over ν_k, but an appropriate (in general non-optimal) choice may considerably simplify this inequality. Put $t_k = \frac{\alpha'}{\nu_k'}$, so that, by Corollary III.2,

$$N_{\alpha}(S_n) \geq B \prod_{k=1}^{n} N_{\alpha}^{t_k}(X_k), \tag{12}$$

or equivalently,

$$\log N_{\alpha}(S_n) \geq \sum_{k=1}^{n} t_k \log N_{\alpha}(X_k) + \log B. \tag{13}$$

Here, according to the condition (6), t_k's may be arbitrary positive numbers such that $t_1 + \dots + t_n = 1$.

To work with (12)-(13), in view of the homogeneity of the functional N_{α}, we may and do assume that

$$\sum_{k=1}^{n} N_{\alpha}(X_k) = 1. \tag{14}$$

To further specify (13), write down the coefficients A_{ν} for $\nu = \nu_k$ as

$$A_{\nu_k} = \frac{(\frac{1}{\nu_k'})^{\frac{1}{\nu_k'}}}{(\frac{1}{\nu_k})^{\frac{1}{\nu_k}}},$$

so that

$$\log\big(A_{\nu_k}^{\alpha'}\big) = \frac{\alpha'}{\nu_k'} \log\frac{1}{\nu_k'} - \frac{\alpha'}{\nu_k} \log\frac{1}{\nu_k}.$$

Using $\frac{\alpha'}{\nu_k'} = t_k$, one may also write

$$-\log\big(A_{\nu_k}^{\alpha'}\big) = \frac{\alpha'}{\nu_k} \log\frac{1}{\nu_k} - t_k \log t_k + t_k \log\alpha',$$

which gives

$$-\sum_{k=1}^{n} \log\left(A_{\nu_k}^{\alpha'}\right) = \alpha' \sum_{k=1}^{n} \frac{1}{\nu_k} \log \frac{1}{\nu_k} - \sum_{k=1}^{n} t_k \log t_k + \log \alpha'.$$

To simplify the sum on the right-hand side, first note that

$$\frac{1}{\nu_k} = 1 - \frac{1}{\nu_k'} = 1 - \frac{t_k}{\alpha'}.$$

Hence,

$$\begin{aligned} \log B &= -\log\left(A_{\alpha'}^{\alpha'}\right) - \sum_{k=1}^{n} \log\left(A_{\nu_k}^{\alpha'}\right) \\ &= -\log\left(A_{\alpha'}^{\alpha'}\right) + \log \alpha' - \sum_{k=1}^{n} t_k \log t_k \\ &\quad + \alpha' \sum_{k=1}^{n} \Big(1 - \frac{t_k}{\alpha'}\Big) \log \Big(1 - \frac{t_k}{\alpha'}\Big). \end{aligned}$$

Choosing in (12) $t_k = N_\alpha(X_k)$, we thus get

$$\begin{aligned} \log N_\alpha(S_n) &\geq -\log\left(A_{\alpha'}^{\alpha'}\right) + \log \alpha' \\ &\quad + \alpha' \sum_{k=1}^{n} \Big(1 - \frac{t_k}{\alpha'}\Big) \log \Big(1 - \frac{t_k}{\alpha'}\Big). \end{aligned}$$

To bound further the right-hand side, one may apply the elementary inequality $(1-\varepsilon)\log(1-\varepsilon) \geq -\varepsilon$ $(0 \leq \varepsilon \leq 1)$, to get

$$\sum_{k=1}^{n} \Big(1 - \frac{1}{\alpha'}\, t_k\Big) \log \Big(1 - \frac{t_k}{\alpha'}\Big) \geq -\sum_{k=1}^{n} \frac{t_k}{\alpha'} = -\frac{1}{\alpha'},$$

thus,

$$\log N_\alpha(S_n) \geq -\log\left(A_{\alpha'}^{\alpha'}\right) + \log \alpha' - 1.$$

Finally,

$$\begin{aligned} -\log\left(A_{\alpha'}^{\alpha'}\right) + \log \alpha' &= \alpha' \left[\frac{1}{\alpha'} \log \frac{1}{\alpha'} - \frac{1}{\alpha} \log \frac{1}{\alpha}\right] + \log \alpha' \\ &= -\frac{\alpha'}{\alpha} \log \frac{1}{\alpha} = \frac{1}{\alpha - 1} \log \alpha, \end{aligned}$$

so that

$$\log N_\alpha(S_n) \geq \frac{1}{\alpha - 1} \log \alpha - 1.$$

Equivalently,

$$N_\alpha(S_n) \geq c_\alpha = \frac{1}{e}\, \alpha^{\frac{1}{\alpha - 1}}.$$

In view of (14), Theorem I.1 is proved.

V. The case of the Shannon entropy

Let X and Y be independent random vectors in $\mathbb{R}^d$ with absolutely continuous distributions. The following natural question has to be still clarified: Under what assumptions does the entropy power inequality

$$N(X+Y) \geq N(X) + N(Y) \tag{15}$$

hold true? In general, the entropy

$$h(X) = h(p) = -\int_{\mathbb{R}^d} p(x) \log p(x)\, dx,$$

where p is density of X, may or may not exist as the Lebesgue integral, and so is the entropy power

$$N(X) = \exp\Big[\,\frac{2}{d}\, h(X)\Big].$$

Hence, at least the existence of $h(X)$ and $h(Y)$ has to be postulated in (15), as is done e.g. in [10]. In that case, what can one say about $h(X+Y)$?

Proposition V.1. *For some i.i.d. random vectors X and Y in $\mathbb{R}^d$ the entropy exists, while the entropy of $X+Y$ does not exist.*

Some examples illustrating such a "bad" behaviour with necessarily $N(X) = N(Y) = 0$ are described in the next section.

Let us however make the convention that $N(X) = 0$, whenever the entropy of X does not exist, including the cases when the distribution of X is not absolutely continuous with respect to Lebesgue measure. Then $N(X)$ is always defined, and the entropy power inequality may indeed be formulated as a universal principle:

Proposition V.2. *With this convention, the inequality* (15) *is true for all independent random vectors X and Y in $\mathbb{R}^d$.*

For the proof, one can start with Theorem I.1 in the limit case $\alpha = 1$, when it becomes

$$N_1(X+Y) \geq N_1(X) + N_1(Y), \tag{16}$$

where $N_1 = \lim_{\alpha \to 1^+} N_\alpha$. To obtain this inequality for N in place of N_1, first we recall an elementary lemma whose proof we include for completeness.

Lemma V.3. *Let X be a random vector in $\mathbb{R}^d$ with density p. If*

$$\int_{\mathbb{R}^d} p^\alpha(x)\, dx < \infty, \quad \textit{for some } \alpha > 1, \tag{17}$$

then the entropy of X exists, and $N_1(X) = N(X)$.

Proof. The lemma is a particular case of the following more general assertion. If a random variable $\xi \geq 0$ satisfies $\|\xi\|_r = (\mathbf{E}\xi^r)^{1/r} < \infty$, for some $r > 0$, then $\mathbf{E} \log \xi$ exists as the Lebesgue integral, and

$$\|\xi\|_{0^+} = \lim_{r \to 0^+} \|\xi\|_r = \exp\{\mathbf{E} \log \xi\}. \tag{18}$$

To prove the latter, one may assume that $\xi > 0$ a.s. First note that the log-moment function $u(r) = \log \mathbf{E}\xi^r$ is continuously differentiable and convex in some interval $0 < r < r_0$. In particular, the derivative $u'(r)$ is non-decreasing. Since also $\lim_{r \to 0^+} u(r) = 0$, we have

$$\lim_{r \to 0^+} \frac{u(r)}{r} = \lim_{r \to 0^+} \int_0^1 u'(rs)\, ds = \lim_{r \to 0^+} u'(r),$$

which is exactly the relation (18). It remains to apply it to $\xi = p(X)$ in which case $N_\alpha(X) = \|\xi\|_{\alpha-1}^{-2/d}$ $(\alpha > 1)$. $\square$

If both random vectors X and Y satisfy the integrability condition (17) on their densities, then so does $X+Y$ (by Young's inequality). Hence, by Lemma V.3, $h(X+Y)$ exists, as well, and $N_1(X+Y) = N(X+Y)$. As a result, we obtain from (16) the entropy power inequality (15) under the condition (17) posed on the summands X and Y (as was stated by Lieb [12]).

It is however desirable to remove the requirement (17), and to this aim a truncation-of-density argument can be used. Given a density p and $n \geq 1$ large enough, define

$$p_n(x) = \frac{1}{c_n}\, p(x)\, 1_{\{p(x) \leq n\}}, \quad x \in \mathbb{R}^d,$$

where

$$c_n = \int_{\{p(x) \leq n\}} p(x)\,dx$$

is a normalizing constant (assuming that $c_n > 0$). If a random vector X_n has density p_n, let us say that X_n has density p, truncated at level n. By the construction, p_n is bounded, so $h(X_n)$ exists regardless of whether $h(X)$ exists, and we have $-\infty < h(X_n) \leq \infty$.

Lemma V.4. *Given a random vector X in $\mathbb{R}^d$ with density p, let X_n have density p, truncated at level n. Then, the limit $\lim_{n\to\infty} h(X_n)$ always exists and is equal to $h(X)$, when $h(X)$ exists.*

Introduce the positive and negative parts of the entropy of X,

$$h^+(X) = \int_{p(x) \leq 1} p(x) \log \frac{1}{p(x)}\,dx$$

and

$$h^-(X) = \int_{p(x) \geq 1} p(x) \log p(x)\,dx.$$

Both integrals are well-defined and non-negative. Moreover, $h(X)$ exists, if and only if $h^+(X) < \infty$ or $h^-(X) < \infty$, in which case

$$h(X) = h^+(X) - h^-(X).$$

An application of the monotone convergence theorem together with the Lebesgue dominated convergence theorem (like in the proof of the next lemma) easily yields

$$\lim_{n\to\infty} h^+(X_n) = h^+(X), \qquad \lim_{n\to\infty} h^-(X_n) = h^-(X).$$

To prove Lemma V.4, first note that $c_n \uparrow 1$. Hence, by the very definition,

$$\begin{aligned} h(X_n) &= \frac{\log c_n}{c_n} \int_{p(x) \leq n} p(x)\,dx \\ &\quad - \frac{1}{c_n} \int_{p(x) \leq n} p(x) \log p(x)\,dx \\ &= o(1) - (1 + o(1)) \int_{p(x) \leq n} p(x) \log p(x)\,dx. \end{aligned}$$

The last integral is finite, if and only if $h^+(X)$ is finite. In case $h^+(X) = \infty$, we have $h(X_n) = \infty$, for all n. If $h^+(X) < \infty$, then $h(X_n) < \infty$, and moreover,

$$\begin{aligned} \lim_{n\to\infty} h(X_n) &= h^+(X) - \lim_{n\to\infty} \int_{1 \leq p(x) \leq n} p(x) \log p(x)\,dx \\ &= h^+(X) - h^-(X) \\ &= h(X). \end{aligned}$$

This proves the lemma.

Lemma V.5. *Given two independent random vectors X and Y in $\mathbb{R}^d$ with densities p and q, let X_n and Y_n be independent random vectors in $\mathbb{R}^d$ with densities p and q, truncated at levels n, respectively. If $h(X+Y)$ exists, then*

$$\lim_{n\to\infty} h(X_n + Y_n) = h(X+Y).$$

Proof. Denote by p_n and q_n the corresponding truncated densities for X and Y with normalizing constants

$$c_n = \int_{p(x) \leq n} p(x)\,dx, \quad c_n' = \int_{q(x) \leq n} q(x)\,dx.$$

Then $X_n + Y_n$ has a bounded density

$$r_n(x) = (p_n * q_n)(x) = \int_{\mathbb{R}^d} p_n(x-y) q_n(y)\,dy.$$

In particular, the entropy $h(X_n + Y_n)$ exists.

Since $c_n p_n(x) \uparrow p(x)$ and $c_n' q_n(x) \uparrow q(x)$, as $n \to \infty$, we get

$$r_n(x) \to r(x) = (p * q)(x) = \int_{\mathbb{R}^d} p(x-y) q(y)\,dy,$$

which is the density of $Z = X+Y$. Moreover, $c_n c_n' \cdot r_n(x) \uparrow r(x)$, so, by the monotone convergence theorem,

$$\begin{aligned} &\int_{r(x) \geq 1} c_n c_n' \cdot r_n(x)\, \log\big(c_n c_n' \cdot r_n(x)\big)\,dx \quad \uparrow \\ &\qquad\qquad \int_{r(x) \geq 1} r(x) \log r(x)\,dx \;=\; h^-(Z). \end{aligned}$$

But $c_n c_n' \log(c_n c_n') \int_{r(x) \geq 1} r_n(x)\,dx \to 0$, and we obtain that

$$\int_{r(x) \geq 1} r_n(x)\, \log r_n(x)\,dx \;\to\; h^-(Z).$$

By a similar argument, using the property that the function $t \log \frac{1}{t}$ is increasing in $0 < t \leq 1/e$, we get that

$$\int_{r(x) \leq 1/e} r_n(x)\, \log \frac{1}{r_n(x)}\,dx \;\to\; \int_{r(x) \leq 1/e} r(x)\, \log \frac{1}{r(x)}\,dx.$$

Finally, to cover the region $A = \{1/e < r(x) < 1\}$, first note that $\mathrm{mes}(A) \leq e$, in view of $\int r(x)\,dx = 1$. Since the sequence $\tilde{r}_n = c_n c_n' r_n$ is non-decreasing, we have $0 \leq \tilde{r}_n \leq r \leq 1$ on the set A, and hence,

$$0 \leq \tilde{r}_n(x) \log \frac{1}{\tilde{r}_n(x)} \leq \frac{1}{e}.$$

Consequently, we are in position to apply on A (which has a finite measure) the Lebesgue dominated convergence theorem, which gives

$$\int_A \tilde{r}_n(x)\, \log \frac{1}{\tilde{r}_n(x)}\,dx \;\to\; \int_A r(x)\, \log \frac{1}{r(x)}\,dx.$$

But the integrals on the left-hand side (as a function of n) are convergent or divergent together with $\int_A r_n(x) \log \frac{1}{r_n(x)}\,dx$, while the integral on the right-hand side is finite.

As a result,

$$\int_{r(x)<1} r_n(x) \log \frac{1}{r_n(x)}\ dx \quad \to \quad \int_{r(x)<1} r(x) \log \frac{1}{r(x)}\ dx \ = \ h^+(Z).$$

The convergences over the two regions $\{r(x) \geq 1\}$ and $\{r(x) < 1\}$ imply that $h(X_n + Y_n) \to h(Z)$, as long as $h(Z)$ exists. Lemma V.5 is proved. □

Applying (15) to X_n and Y_n and using the two lemmas, we arrive at the entropy power inequality, where the condition (17) does not appear anymore.

Corollary V.6. *The entropy power inequality* (15) *holds true, whenever the entropies of X, Y, and $X+Y$ exist.*

Note that the existence of the entropies is guaranteed, for example, when both X and Y have finite second moments. This is a very popular condition, under which the entropy power inequality is stated. Moreover, it sufficies to require that $\mathbf{E}\log(1+|X|) < \infty$ and $\mathbf{E}\log(1+|Y|) < \infty$.

Now, to make the last step towards the more general Proposition V.2, it remains to combine Corollary V.6 with the monotonicity property of the entropy, which we state separately as the following proposition.

Proposition V.7. *Let X and Y be independent random vectors in $\mathbb{R}^d$. If the entropy of X exists and $N(X) > 0$, then the entropy of $X+Y$ exists as well, and $N(X+Y) \geq N(X)$.*

This statement is easily obtained by applying Jensen's inequality, and we omit the details. Here, it does not matter whether Y has density or not. However, the assumption that $N(X) > 0$ is important, since otherwise the entropy of $X+Y$ may not exist (even if Y has density with well-defined entropy).

What also seems interesting, the monotonicity property does not exclude the cases of a strong "discontinuity" of the entropy under convolutions.

Proposition V.8. *There exists a random vector X in $\mathbb{R}^d$ with a finite entropy, such that $h(X+Y) = \infty$ for every random vector Y in $\mathbb{R}^d$, which is independent of X and which has a finite entropy.*

VI. EXAMPLES

To illustarte a possible behaviour of the entropy on convolutions, let us start with two examples. For simplicity, we only consider the one-dimensional case.

Example 1. For $\varepsilon > 0$, consider probability distributions supported on the interval $(0, \frac{1}{e})$ with densities

$$p_\varepsilon(x) = \frac{\varepsilon}{x \log^{1+\varepsilon}(1/x)}, \qquad 0 < x < \frac{1}{e}.$$

The entropy for each of these densities exists, since all p_ε are lower-bounded. Moreover, $h(p_\varepsilon) > -\infty$, if and only if $\varepsilon > 1$. The convolution $p_\varepsilon * p_\varepsilon$ behaves near zero like $p_{2\varepsilon}$, so,

$$h(p_\varepsilon * p_\varepsilon) > -\infty, \quad \text{if and only if} \quad \varepsilon > \frac{1}{2}$$

(cf. e.g. [3]).

Example 2. Let a random variable X be uniformly distributed in the set A of Lebesgue measure 1, namely, of the form

$$A = \bigcup_{n=1}^{\infty} (2^n, 2^n + a_n),$$

where $a_n \geq 0$, $\sum_{n=1}^\infty a_n = 1$. Then $h(X) = 0$, while, as easy to check,

$$h(X+U) = \frac{1}{2}\sum_{n=1}^\infty a_n^2 + \sum_{n=1}^\infty a_n \log \frac{1}{a_n},$$

where the random variable U is independent of X and has a uniform distribution in $(0,1)$. Here, both X and $X+U$ have densities bounded by 1, so there is no need to speak about the existence of the entropies.

This example already shows that in general $h(X+Y)$ cannot be bounded from above in terms of $h(X)$ and $h(Y)$. This is so also in the i.i.d. case, as the next example shows. Moreover, it may occur that $h(X+Y) = \infty$, while the entropy of the summands is finite.

Example 3. Let the random variables X and X' be independent with the same uniform density $q = 1_A$ as in the previous example, based on a decreasing sequence a_n. Write $q = \sum_{n=1}^\infty a_n q_n$, where q_n is the density of a random variable uniformly distributed in $(2^n, 2^n + a_n)$. Hence,

$$q * q \ = \sum_{k>n\geq 1} 2\, a_k a_n\, q_k * q_n + \sum_{n\geq 1} a_n^2\, q_n * q_n.$$

The entropy of this convolution exists and is non-negative, since $q * q \leq 1$. To compute it, first note that all densities $q_k * q_n$ $(k \geq n)$ are supported on non-overlapping intervals. On the other hand, if U_1, U_2 are independent and uniformly distributed in the intervals of lengths $c_1 \geq c_2 > 0$, then

$$h(U_1 + U_2) = \log c_1 + \frac{c_2}{2c_1}.$$

Using this formula, we find that

$$h(q*q) \ = \ -2\log 2 + 2\log 2 \sum_{n=1}^\infty s_n a_n + \sum_{n=1}^\infty \Big(n - \frac{1}{2}\Big)\, a_n^2 + \ \sum_{n=1}^\infty a_n^2 \log \frac{1}{a_n} + 2\sum_{n=2}^\infty s_{n-1}\, a_n \log \frac{1}{a_n},$$

where $s_n = a_1 + \dots + a_n$. Here, since a_n is decreasing and summing to 1, it readily follows that $\sum_{n=1}^\infty n a_n^2 \leq 1$. Indeed, necessarily $a_n \leq \frac{1}{n}$, since otherwise, by the monotonicity, we would have $a_k > \frac{1}{n}$ for all $k \leq n$, and then $\sum_{k=1}^\infty a_k > 1$ which contradicts to the assumption. Hence, $\sum_{n=1}^\infty (n a_n)\, a_n \leq \sum_{n=1}^\infty a_n = 1$. Next, using $t \log \frac{1}{t} \leq \frac{1}{e}$

for $0 \le t \le 1$, we also have the bound $\sum_{n=1}^\infty a_n^2 \log \frac{1}{a_n} \le \frac{1}{e}$. Therefore,

$$h(q*q) < \infty, \quad \text{if and only if} \quad \sum_{n=1}^\infty a_n \log \frac{1}{a_n} < \infty. \quad (19)$$

Obviously, in this characterization the assumption on the monotonicity of a_n may be removed.

The cases where $\sum_n a_n \log \frac{1}{a_n} = \infty$ provide examples proving Proposition V.8. Indeed, using the submodularity property of the entropy (due to Madiman, cf. [13]), for any random variable Y, which is independent of X and has finite entropy,

$$\begin{aligned} h(X+X'+Y) + h(Y) &\le h(X+Y) + h(X'+Y) \\ &= 2h(X+Y). \end{aligned}$$

But $h(X+X'+Y) \ge h(X+X') = \infty$, according to (19) and Proposition V.7. So it is necessary that $h(X+Y) = \infty$ as well.

Example 4. Given $0 < \varepsilon \le \frac{1}{4}$, let X have the density

$$p = \frac{1}{2}\, p_\varepsilon + \frac{1}{2}\, q,$$

where $q = 1_A$ is the uniform density from Examples 2-3 with $\sum_n a_n \log \frac{1}{a_n} = \infty$.

Obviously, the entropy of X exists. More precisely, since p_ε and q are supported on disjoint sets, and since $h(q) = 0$,

$$\begin{aligned} h(X) &= \Big(\frac{\log 2}{2} + \frac{1}{2}\, h(p_\varepsilon)\Big) + \Big(\frac{\log 2}{2} + \frac{1}{2}\, h(q)\Big) \\ &= \log 2 + \frac{1}{2}\, h(p_\varepsilon) \ = \ -\infty. \end{aligned}$$

Now, let Y be an independent copy of X, and write the density of $X+Y$ in the form

$$p*p = \frac{1}{4}\, p_\varepsilon * p_\varepsilon + \frac{3}{4}\, r, \quad \text{where} \quad r = \frac{2}{3}\, p_\varepsilon * q + \frac{1}{3}\, q*q.$$

Here, $p_\varepsilon * p_\varepsilon$ is supported on $(0, \frac{2}{e})$, while the density r is supported on $(2, \infty)$. Let us show that $h(r) = \infty$, which together with $h(p_\varepsilon * p_\varepsilon) = -\infty$ (as in Example 1) would imply that $h(X+Y)$ is not defined, thus proving Proposition V.1.

Note that $p_\varepsilon * q \le 1$, which implies $h(p_\varepsilon * q) \ge 0$. Using the concavity of the function $t \to t \log \frac{1}{t}$, Jensen's inequality yields

$$h(r) \ \ge \ \frac{2}{3}\, h(p_\varepsilon * q) + \frac{1}{3}\, h(q*q) \ \ge \ \frac{1}{3}\, h(q*q) \ = \ \infty,$$

according to (19). One may take $\varepsilon = \frac{1}{4}$, for example.

Example 5. One can also show that $h(p_\varepsilon * q) = \infty$, provided that

$$\sum_{n=1}^\infty a_n \log \frac{1}{a_n} = \infty, \qquad \sum_{n=1}^\infty a_n \log^{1-\varepsilon} \frac{1}{a_n} < \infty,$$

where $0 < \varepsilon \le \frac{1}{4}$. For instance, the choice $a_n \sim \text{const} \cdot \frac{1}{n \log^2 n}$ for large n meets these reqirements.

Hence, taking independent random variables X and Y with densities p and p_ε as above, we obtain that $h(X) = h(Y) = -\infty$, but $h(X+Y)$ is not defined. In this example, the random variables are not identically distributed, but one of the distributions is compactly supported.

The argument is based on the following estimate: If a random variable Y has the density p_ε $(0 < \varepsilon \le 1)$, and U is independent of Y and has a uniform distribution in $(0,1)$, then for each $0 < a \le 1/(2e)$,

$$h(Y + aU) \ge -C \log^{1-\varepsilon}(1/a),$$

where C is an absolute constant. We leave it for the reader as an excercise.

Acknowledgement. The authors are grateful to Igal Sason for valuable comments.

References

[1] K. Ball. Cube slicing in R^n. Proc. Amer. Math. Soc. 97 (1986), no. 3, 465-473.

[2] F. Barthe. Optimal Young's inequality and its converse: a simple proof. Geom. Funct. Anal. 8 (1998), no. 2, 234–242.

[3] A. R. Barron. Entropy and the central limit theorem. Ann. Probab. 14 (1986), no. 1, 336-342.

[4] W. Beckner. Inequalities in Fourier analysis. Ann. of Math. (2) 102 (1975), no. 1, 159–182.

[5] J.-F. Bercher; C. Vignat. A Renyi entropy convolution inequality with application. In: EUSIPCO 2002, Toulouse, France.

[6] N. M. Blachman. The convolution inequality for entropy powers. IEEE Trans. Inform. Theory, IT-11, 1965, 267–271.

[7] S. G. Bobkov; G. P. Chistyakov. Bounds for the maximum of the density of the sum of independent random variables. (Russian) Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 408 (2012), Veroyatnost i Statistika. 18, 62–73, 324; translation in J. Math. Sci. (N.Y.) 199 (2014), no. 2, 100-106.

[8] H. J. Brascamp; E. H. Lieb. Best constants in Young's inequality, its converse, and its generalization to more than three functions. Advances in Math. 20 (1976), no. 2, 151–173.

[9] T. M. Cover; J. A. Thomas. Elements of information theory. Wiley Series in Telecommunications. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1991. xxiv+542 pp.

[10] A. Dembo; T. M. Cover; J. A. Thomas. Information-theoretic inequalities. IEEE Trans. Inform. Theory, 37(6), 1501–1518, 1991.

[11] O. Johnson. Information theory and the central limit theorem. Imperial College Press, London, 2004, xiv+209 pp.

[12] E. H. Lieb. Proof of an entropy conjecture of Wehrl. Comm. Math. Phys. 62 (1978), no. 1, 35–41.

[13] M. Madiman. On the entropy of sums. Proceedings of the 2008 IEEE Information Theory Workshop, Porto, Portugal, May 2008.

[14] O. Rioul. Information theoretic proofs of entropy power inequalities. IEEE Trans. Inform. Theory 57 (2011), no. 1, 33-55.

[15] B. A. Rogozin. An estimate for the maximum of the convolution of bounded densities. (Russian) Teor. Veroyatnost. i Primenen. 32 (1987), no. 1, 53-61.

[16] G. Savaré; G. Toscani. The concavity of Rényi entropy power. IEEE Trans. Inform. Theory 60 (2014), no. 5, 2687-2693.

[17] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. Information and Control, 2:101–112, 1959.