Rényi entropy dimension of the mixture of measures

Marek Śmieja
Institute of Computer Science
Department of Mathematics and Computer Science
Jagiellonian University
Lojasiewicza 6, 30-348, Krakow, Poland
marek.smieja@ii.uj.edu.pl

Jacek Tabor
Institute of Computer Science
Department of Mathematics and Computer Science
Jagiellonian University
Lojasiewicza 6, 30-348, Krakow, Poland
jacek.tabor@ii.uj.edu.pl
*

(Received 2012)

Abstract. Rényi entropy of order α is a general measure of entropy. It provides the limit for optimal coding assuming that the cost of encoding a given symbol varies exponentially with its code length. In this paper we provide the bounds for the Rényi entropy dimension of the mixture of measures in terms of the dimensions of the single measures. Our proof relies on the equivalent weighted definition of the Rényi entropy which allows to deal well with the calculation of the entropy of the mixture of measures.

1. Introduction

The Rényi entropy of order α is a quantitative measure of information which extends classical notion of Shannon entropy. It has been successfully applied in information theory, statistics, and, more recently, in thermodynamics and quantum mechanics [3,4,6,9,12,15]. P. Grasberger and I. Proccacia [5] proposed an efficient method for computing Rényi entropy for some values of parameter α.

The definition of Rényi entropy is based on an axiomatic approach to information and was introduced as the one parametric family of generalized entropy functions [10]. Given a measurable partition $\mathcal{P}$ of measurable space

*This research is supported by National Centre of Science of Poland - Grants ID: 2011/01/B/ST6/01887).

(X, Σ), the Rényi entropy is defined by:

$$h_\alpha(\mu; \mathcal{P}) := \frac{1}{1-\alpha} \log_2 \big(\sum_{P \in \mathcal{P}} \mu(P)^\alpha \big), \text{ for } \alpha \in (0, \infty) \setminus \{1\}. \tag{1}$$

When modifying the parameter α it is possible to emphasize or weaken the relevance of some probability events.

It is well-known that the Shannon entropy provides a limit for best possible compression assuming that the cost of encoding symbol depends linearly on its code length [13]. L. L. Campbell [1] considered the case when the cost grows exponentially with the code length. Under this assumption he stated that it is possible to encode so that the measure of length is arbitrarily close to the Rényi entropy of the input.

In case of metric space, it is very important to investigate the asymptotic behavior of the entropy function when the partition becomes finer. This dependence is described by the Rényi entropy dimension of order α (see DEFINITION 12). I. Csiszár obtained the bounds for the Rényi entropy of the mixture of measures of arbitrary order in the case of real probability space [2, Lemma 1]. Those inequalities lead directly to the estimation of the Rényi entropy dimension:

CSISZÁR ESTIMATION OF THE RÉNYI ENTROPY DIMENSION OF THE MIXTURE. *Let* $(\mathbb{R}, \Sigma)$ *be a Borel measurable space and let* $a_1, a_2 \in [0, 1]$ *satisfy* $a_1 + a_2 = 1$. *If* μ_1, μ_2 *are probability measures which both have Rényi entropy dimension then* $a_1\mu_1 + a_2\mu_2$ *also has Rényi entropy dimension and*

$$\dim_\alpha(a_1\mu_1 + a_2\mu_2) = \begin{cases} \max\{\dim(\mu_1), \dim(\mu_2)\} & \text{for } \alpha \in (0, 1), \\ \min\{\dim(\mu_1), \dim(\mu_2)\} & \text{for } \alpha \in (1, \infty). \end{cases} \tag{2}$$

In this paper we show that similar result can be obtained for any probability measures in general metric space. For this purpose we use a different technique which relies on the adaptation of the reasoning used in [14].

More precisely, we have recently provided an equivalent weighted approach to the Shannon entropy which is based on sub-probability measures instead of partitions. Such a formulation of the entropy describes the idea of random lossy coding which, from a practical point of view, can be more important than the classical deterministic coding. Moreover, weighted entropy allows to deal well with the calculation of the entropy of the mixture of measures.

In the next section we adapt the idea of weighted entropy for the case of Rényi entropy. Making use of the techniques presented in [14] and some simple properties of the Rényi entropy function, we justify that the weighted Rényi entropy is equivalent to the classical one (see Corollary 7) and provide bounds for Rényi entropy of the mixture of measures for arbitrary parameter $\alpha \in (0,\infty) \setminus \{1\}$ (see Corollary 10). Our bounds are sharp and with $\alpha \to 1$ they converge to the corresponding bounds obtained in [14, Theorem III.1] for the Shannon entropy. Finally, third section contains our main result concerning the Rényi entropy dimension of the mixture of measures (see Theorem 13).

2. Weighted Rényi Entropy.

Paper [14] introduced the notion of weighted Shannon entropy. This equivalent definition allowed to prove some important properties concerning the Shannon entropy of the mixture of measures. We will show that similar approach can be applied for the Rényi entropy. As a consequence, we will also give the expressions for the Rényi entropy of the mixture of measures in terms of the entropies of single measures. If not stated otherwise, we always assume that (X, Σ, μ) is a probability space.

We say that a family $\mathcal{P} \subset \Sigma$ is a μ-partition of X if $\mathcal{P}$ is countable family of pairwise disjoint sets and

$$\mu(X \setminus \bigcup_{P \in \mathcal{P}} P) = 0. \tag{3}$$

Rényi entropy was introduced as one parametric family of generalized entropy functions and represents an extension to the traditional Shannon entropy:

DEFINITION 1. Let $\alpha \in (0,\infty) \setminus \{1\}$ and let $\mathcal{P}$ be a μ-partition of X. *The Rényi μ-entropy of order α of $\mathcal{P}$* is defined by [11]

$$h_\alpha(\mu; \mathcal{P}) := \frac{1}{1-\alpha} \log_2 \big(\sum_{P \in \mathcal{P}} \mu(P)^\alpha \big). \tag{4}$$

When we consider the problem of lossy data compression then the partition plays a role of a coding alphabet. We code a given $x \in X$ by a binary representation of $P \in \mathcal{P}$ iff $x \in P$.

Shannon coding theorem [13] states that it is not possible to find a uniquely decipherable code of μ-partition $\mathcal{P}$ whose average length is less than Shannon μ-entropy of $\mathcal{P}$. Implicit assumption of this fact is that the cost of encoding a given symbol varies linearly with its code length. As it was explained by L. L. Campbell [1], this is not always the case. Assuming that the cost is exponential function of code length, we get that it is not possible to encode so that the statistical amount of memory used per one symbol is less than Rényi μ-entropy of $\mathcal{P}$ [1]. This justifies the great importance of Rényi entropy in the theory of coding. Through the paper we consider the case of exponential dependence between the codes lengths and the cost of encoding.

The coding with use of a given partition causes specific level of error. If we are interested in controlling the maximal error of lossy coding then we fix the error-control family $\mathcal{Q}$ which is simply the family of measurable subsets of X. Then as the coding partitions we consider only $\mathcal{Q}$-acceptable μ-partitions $\mathcal{P}$ (which we write $\mathcal{P} \prec \mathcal{Q}$) i.e. we desire that for every $P \in \mathcal{P}$ there exists $Q \in \mathcal{Q}$ such that $P \subset Q$.

Consequently, to describe the statistical memory used by optimal lossy coding determined by $\mathcal{Q}$-acceptable μ-partitions we define the Rényi μ-entropy of an error-control family $\mathcal{Q}$.

DEFINITION 2. Let $\alpha \in (0, \infty) \setminus \{1\}$ and let $\mathcal{Q}$ be an error-control family. We define *Rényi μ-entropy of order α of $\mathcal{Q}$* by

$$H_\alpha(\mu; \mathcal{Q}) := \inf\{h_\alpha(\mu; \mathcal{P}) \in [0, \infty] : \mathcal{P} \text{ is a } \mu\text{-partition and } \mathcal{P} \prec \mathcal{Q}\}. \quad (5)$$

Observe that if there is no $\mathcal{Q}$-acceptable μ-partition then directly from the definition[1] $H_\alpha(\mu; \mathcal{Q}) = \infty$. Moreover, if $\mathcal{Q}$ itself is a μ-partition of X then trivially $H_\alpha(\mu; \mathcal{Q}) = h_\alpha(\mu; \mathcal{Q})$.

REMARK 3. Similar notion of the entropy was studied by A. Rényi [10,11]. However, he considered error-control families in metric spaces consisted of balls with given radius or cubes with specific edge length. The simplest example of the error-control family is given by the set of all intervals in $\mathbb{R}$ with length at most δ. Then to find the entropy we have to consider the entropies of all partitions consisting of intervals with length at most δ.

To see that $H_\alpha(\mu; \mathcal{Q})$ does not have to be attained, it is sufficient to use the trivial example from [14, Example II.1]:

EXAMPLE 4. Let $X = (0, 1)$, Σ be a sigma algebra generated by all Borel subsets of $(0, 1)$, μ be a Lebesgue measure and $\mathcal{Q}$ be an error-control family defined by

$$\mathcal{Q} = \{[a, b] : 0 < a < b < 1\}. \quad (6)$$

Clearly $H_\alpha(\mu; \mathcal{Q}) = 0$ but for every μ-partition $\mathcal{P} \prec \mathcal{Q}$, we have $H_\alpha(\mu; \mathcal{P}) > 0$ when $\alpha \in (0, \infty) \setminus \{1\}$.

[1] We put $\inf(\emptyset) = \infty$.

The inspiration of weighted entropy, lies in the substitution of the division of space X into partition by the division of measure μ into sub-probability measures[2]. It enables us to use of the operations on functions rather than on plain sets. Roughly speaking, this approach provides the computation and interpretation of the entropy with respect to "formal" convex combination $a_1\mathcal{P}_1 + a_2\mathcal{P}_2$, where $\mathcal{P}_1, \mathcal{P}_2$ are partitions (which clearly does not make sense in the classical approach).

Let us denote the space of divisions of measure μ with respect to $\mathcal{Q} \subset \Sigma$ by:

$$\begin{aligned}&W(\mu;\mathcal{Q}) := \{\mathfrak{m} : \mathcal{Q} \ni Q \to \mathfrak{m}_Q \in M(X,\Sigma) :\\ &\mathfrak{m}_Q(X \setminus Q) = 0 \text{ for every } Q \in \mathcal{Q} \text{ and } \textstyle\sum_{Q\in\mathcal{Q}} \mathfrak{m}_Q = \mu\},\end{aligned} \tag{7}$$

where $M(X,\Sigma)$ is the family of all sub-probability measures on (X,Σ) (when we consider the set of all probability measures on (X,Σ) then we write $M_1(X,\Sigma)$). Observe that every function $\mathfrak{m} \in W(\mu;\mathcal{Q})$ is non-zero on at most countable number of sets from $\mathcal{Q}$. Hence the sum in (7) is countable.

We define the weighted Rényi μ-entropy of $\mathfrak{m} \in W(\mu;\mathcal{Q})$.

DEFINITION 5. Let $\alpha \in (0,\infty) \setminus \{1\}$ and let $\mathcal{Q}$ be an error-control family. We define the *weighted Rényi μ-entropy* of order α of $\mathfrak{m} \in W(\mu;\mathcal{Q})$ by

$$h_\alpha^W(\mu;\mathfrak{m}) := \frac{1}{1-\alpha}\log_2(\sum_{Q\in\mathcal{Q}} \mathfrak{m}_Q(X)^\alpha) \tag{8}$$

and the weighted Rényi μ-entropy of order α of $\mathcal{Q}$ by

$$H_\alpha^W(\mu;\mathcal{Q}) := \inf\{h_\alpha^W(\mu;\mathfrak{m}) \in [0,\infty] : \mathfrak{m} \in W(\mu,\mathcal{Q})\}. \tag{9}$$

Weighted entropy has very practical interpretation. It describes the idea of random (nondeterministic) lossy coding. Let us consider the following example.

EXAMPLE 6. Let $X = \{a,b\}$, $\Sigma = \mathcal{Q}$ be families of all subsets of X and μ be uniformly distributed measure on X. We define $\mathfrak{m} \in W(\mu;\mathcal{Q})$ by:

$$\mathfrak{m}(\{a,b\}) = \nu_1,$$

$$\mathfrak{m}(\{b\}) = \nu_2,$$

$$\mathfrak{m}(\{a\}) \equiv \mathfrak{m}(\emptyset) \equiv 0,$$

where

$$\nu_1(\{a\}) = 1/2,\ \nu_1(\{b\}) = 1/6$$

[2]The idea of weighted entropy is indebted to the notion of weighted Hausdorff measures considered by J. Howroyd [7,8].

and

$$\nu_2(\{a\}) = 0,\ \nu_2(\{b\}) = 1/3.$$

We consider the coding as follows: ν_1 determines code 1 and ν_2 determines code 01. It implies that symbol b is coded by 1 with probability

$$\frac{\nu_1(\{b\})}{\nu_1(\{b\}) + \nu_2(\{b\})} = \frac{1/6}{1/6 + 1/3} = 1/3$$

and by 01 with probability

$$\frac{\nu_2(\{b\})}{\nu_1(\{b\}) + \nu_2(\{b\})} = \frac{1/3}{1/6 + 1/3} = 2/3$$

while symbol a is always by code 1. The coding is nondeterministic since symbol b can be coded by two different codes with positive probabilities. One of the advantages of this approach is that we do not have to control strictly the way of coding since it is determined by a nondeterministic rule.

As in the case of the Shannon entropy [14, Theorem II.1], weighted Rényi μ-entropy coincides with classical Rényi μ-entropy of error-control family $\mathcal{Q}$:

COROLLARY 7. *Let $\mathcal{Q}$ be an error-control family. Then, weighted Rényi μ-entropy equals the classical Rényi μ-entropy, i.e.*

$$H_\alpha^W(\mu; \mathcal{Q}) = H_\alpha(\mu; \mathcal{Q}),\ \textit{for } \alpha \in (0, \infty) \setminus \{1\}. \tag{10}$$

The proof of the aforementioned theorem relies on the technique used in [14] and simple facts concerning the Rényi entropy function (see Observation 8 below).

Let us denote by g_α and its inverse g_α^{-1} the following functions:

$$g_\alpha(x) = 2^{(1-\alpha)x},\ g_\alpha^{-1}(x) = \frac{1}{1-\alpha} \log_2(x). \tag{11}$$

Then the observation is valid:

OBSERVATION 8.

1. *If $\alpha \in (0, 1)$ then:*
   (a) *g_α^{-1} and g_α are increasing,*
   (b) *$x \to x^\alpha$ is subadditive,*
   (c) *$x \to x^\alpha$ is concave*
2. *If $\alpha \in (1, \infty)$ then:*

(a) g_α^{-1} *and* g_α *are decreasing,*

(b) $x \to x^\alpha$ *is superadditive,*

(c) $x \to x^\alpha$ *is convex.*

The definition of weighted Shannon μ-entropy [14] allowed to approximate the entropy of the mixture of measures in terms of the entropies of single measures:

SHANON ENTROPY OF THE MIXTURE [14, Theorem III.1]: *Let* $a_1, a_2 \in [0,1]$ *be such that* $a_1 + a_2 = 1$. *If* μ_1, μ_2 *are probability measures and* $\mathcal{Q}$ *is an error-control family then:*

$$H(a_1\mu_1 + a_2\mu_2; \mathcal{Q}) \geq a_1\, H(\mu_1; \mathcal{Q}) + a_2\, H(\mu_2; \mathcal{Q}) \tag{12}$$

and

$$H(a_1\mu_1 + a_2\mu_2; \mathcal{Q}) \leq a_1\, H(\mu_1; \mathcal{Q}) + a_2\, H(\mu_2; \mathcal{Q}) - a_1 \log_2(a_1) - a_2 \log_2(a_2), \tag{13}$$

where $H(\mu; \mathcal{Q})$ denotes the Shannon μ-entropy of $\mathcal{Q}$.

REMARK 9. We can interpret μ_1 and μ_2 as two separate sources which send us information with probability a_1 and a_2, respectively. The aforementioned formula states that it is possible to encode combined information from these sources with average memory not higher than the convex combination of individual entropies and the entropy of probability distribution (a_1, a_2).

Using similar reasoning to the one used in [14] we can find the bounds of the Rényi μ-entropy.

COROLLARY 10. *Let* $\alpha \in (0,\infty) \setminus \{1\}$ *and* $n \in \mathbb{N}$. *We assume that* $a_k \in [0,1]$ *for* $k \in \{1,\ldots,n\}$ *be such that* $\sum\limits_{k=1}^{n} a_k = 1$. *Let* $\{\mu_k\}_{k=1}^n \subset M_1(X,\Sigma)$ *and* $\mu := \sum\limits_{k=1}^{n} a_k\mu_k \in M_1(X,\Sigma)$. *If* $\mathcal{Q}$ *is an error-control family then*

$$H_\alpha(\mu; \mathcal{Q}) \geq g^{-1}\big[\sum_{k=1}^{n} a_k g(H_\alpha(\mu_k; \mathcal{Q}))\big] \tag{14}$$

and

$$H_\alpha(\mu; \mathcal{Q}) \leq g^{-1}\big[\sum_{k=1}^{n} a_k^\alpha g(H_\alpha(\mu_k; \mathcal{Q}))\big]. \tag{15}$$

The proof is based on applying the weighted entropy and using Observation 8 – for the convenience of the reader we place it in Appendix A.

The following example confirms that the above estimation (14) and (15) cannot be improved.

EXAMPLE 11. Let us assume that $\alpha \in (0, \infty) \setminus \{1\}$, $X = \{0, 1\}$, $\Sigma = \mathcal{Q}$ be the families of all subsets of X and μ_1, μ_2 be discrete measures such that:

$$\mu_1(\{0\}) = 1 \text{ and } \mu_2(\{1\}) = 1. \tag{16}$$

Then, we have

$$H_\alpha(a_1\mu_1 + a_2\mu_2; \mathcal{Q}) = \frac{1}{1-\alpha} \log_2(a_1^\alpha + a_2^\alpha). \tag{17}$$

It is exactly the right side of the inequality (15).

On the other hand, if we consider two measures which satisfy $\mu_1 = \mu_2$, then

$$H_\alpha(a_1\mu_1 + a_2\mu_2; \mathcal{Q}) = H_\alpha(\mu_1; \mathcal{Q}) = H_\alpha(\mu_2; \mathcal{Q}) \tag{18}$$

and it equals the right side of (14).

Let us observe the similarity between bounds obtained for both, Shannon μ-entropy [14, Theorem III.1] and Rényi μ-entropy from Corollary 10. Let us consider the functions:

$$l_\alpha(x, y) = g_\alpha^{-1}(a_1 g_\alpha(x) + a_2 g_\alpha(y)), \tag{19}$$

$$u_\alpha(x, y) = g_\alpha^{-1}(a_1^\alpha g_\alpha(x) + a_2^\alpha g_\alpha(y)), \tag{20}$$

which describe the lower and upper bounds for the Rényi μ-entropy of order α. If x, y are non negative real numbers then these functions converge to the corresponding bounds calculated for Shannon μ-entropy as $\alpha \to 1$, i.e.:

$$\begin{cases} l_\alpha(x, y) \to a_1 x + a_2 y \\ u_\alpha(x, y) \to a_1 x + a_2 y - a_1 \log_2(a_1) - a_2 \log_2(a_2). \end{cases} \tag{21}$$

3. Bounds for Rényi Entropy Dimension of the Mixture of Measures.

I. Csiszár [2] studied the properties of Rényi entropy dimension of the mixture of measures. He gave the bounds for the entropy of the mixture of measures in $\mathbb{R}$ [2, Lemma 1] which lead to the approximation of Rényi entropy dimension of the mixture. We will consider the case of arbitrary probability metric space and show that the similar approximation holds. Our proof relies on applying Corollary 10.

Before proceeding with it let us recall the definition of Rényi entropy dimension of order α. In this section we additionally assume that X is a metric space and (X, Σ, μ) is a probability space, where Σ contains all Borel subsets of X.

Given $\delta > 0$ let us denote a family of all balls in X with radius δ by

$$\mathcal{B}_\delta := \{B(x,\delta) : x \in X\}, \tag{22}$$

where $B(x,\delta)$ is a closed ball centred at x with radius δ.

DEFINITION 12. The *upper and lower Rényi entropy dimension of order* $\alpha \in (0,\infty) \setminus \{1\}$ of measure $\mu \in M_1(X,\Sigma)$ are defined by

$$\overline{\dim}_\alpha(\mu) := \limsup_{\delta \to 0} \frac{H_\alpha(\mu; \mathcal{B}_\delta)}{-\log_2(\delta)}, \tag{23}$$

$$\underline{\dim}_\alpha(\mu) := \liminf_{\delta \to 0} \frac{H_\alpha(\mu; \mathcal{B}_\delta)}{-\log_2(\delta)}. \tag{24}$$

If the above are equal we say that μ has the *Rényi entropy dimension of order* α and denote it by $\dim_\alpha(\mu)$.

The following theorem gives the estimation of the Rényi entropy dimension of the mixture of measures.

THEOREM 13. *Let $a_1, a_2 \in (0,1)$ be such that $a_1 + a_2 = 1$ and let $\mu_1, \mu_2 \in M_1(X,\Sigma)$. If $\overline{\dim}_\alpha(\mu_1) < \infty$ and $\overline{\dim}_\alpha(\mu_2) < \infty$ then*

$$\overline{\dim}_\alpha(a_1\mu_1 + a_2\mu_2) \leq \begin{cases} \max\{\overline{\dim}_\alpha(\mu_1), \overline{\dim}_\alpha(\mu_2)\}, \textit{for } \alpha \in (0,1), \\ \min\{\overline{\dim}_\alpha(\mu_1), \overline{\dim}_\alpha(\mu_2)\}, \textit{for } \alpha \in (1,\infty) \end{cases} \tag{25}$$

and

$$\underline{\dim}_\alpha(a_1\mu_1 + a_2\mu_2) \geq \begin{cases} \max\{\underline{\dim}_\alpha(\mu_1), \underline{\dim}_\alpha(\mu_2)\}, \textit{for } \alpha \in (0,1), \\ \min\{\underline{\dim}_\alpha(\mu_1), \underline{\dim}_\alpha(\mu_2)\}, \textit{for } \alpha \in (1,\infty). \end{cases} \tag{26}$$

Proof. Let us show first inequality from formula (26). The rest of them can be proven in similar manner.

Directly from the definition of Rényi entropy dimension of order $\alpha \in (0,\infty) \setminus \{1\}$, we have:

$$\liminf_{\delta \to 0} \frac{H_\alpha(\mu_k; \mathcal{B}_\delta)}{-\log_2(\delta)} = \underline{\dim}_\alpha(\mu_k), \text{ for } k = 1,2. \tag{27}$$

Then for arbitrary $\varepsilon_1, \varepsilon_2 > 0$, there exists $\delta_1, \delta_2 > 0$, such that:

$$\frac{H_\alpha(\mu_k; \mathcal{B}_{\delta_k})}{-\log_2(\delta_k)} \geq \underline{\dim}_\alpha(\mu_k) - \varepsilon_k \tag{28}$$

and consequently

$$H_\alpha(\mu_k;\mathcal{B}_{\delta_k}) \geq -\log_2(\delta_k)(\underline{\dim}_\alpha(\mu_k)-\varepsilon_k), \tag{29}$$

for $k=1,2$.

We put $\delta := \min\{\delta_1,\delta_2\}$. Making use of Observation 8: 1a, we get:

$$g_\alpha^{-1}\{a_1 g_\alpha[\,H_\alpha(\mu_1;\mathcal{B}_\delta)] + a_2 g_\alpha[\,H_\alpha(\mu_2;\mathcal{B}_\delta)]\} \tag{30}$$

$$\geq g_\alpha^{-1}\{a_1 g_\alpha[-\log_2(\delta)(\underline{\dim}_\alpha(\mu_1)-\varepsilon_1)] \tag{31}$$

$$+a_2 g_\alpha[-\log_2(\delta)(\underline{\dim}_\alpha(\mu_2)-\varepsilon_2)]\}. \tag{32}$$

By Corollary 10, we have:

$$H_\alpha(a_1\mu_1+a_2\mu_2;\mathcal{B}_\delta) \tag{33}$$

$$\geq g_\alpha^{-1}\{a_1 g_\alpha[\,H_\alpha(\mu_1;\mathcal{B}_\delta)] + a_2 g_\alpha[\,H_\alpha(\mu_2;\mathcal{B}_\delta)]\} \tag{34}$$

$$\geq g_\alpha^{-1}\{a_1 g_\alpha[-\log_2(\delta)(\underline{\dim}_\alpha(\mu_1)-\varepsilon_1)] \tag{35}$$

$$+a_2 g_\alpha[-\log_2(\delta)(\underline{\dim}_\alpha(\mu_2)-\varepsilon_2)]\} \tag{36}$$

$$= \frac{1}{1-\alpha}\log_2\left[a_1\delta^{-(1-\alpha)(\underline{\dim}_\alpha(\mu_1)-\varepsilon_1)}\right. \tag{37}$$

$$\left.+a_2\delta^{-(1-\alpha)(\underline{\dim}_\alpha(\mu_2)-\varepsilon_2)}\right] \tag{38}$$

$$= \frac{1}{1-\alpha}\log_2\left\{\delta^{-(1-\alpha)(\underline{\dim}_\alpha(\mu_1)-\varepsilon_1)}\right. \tag{39}$$

$$\left.\cdot\left[a_1+a_2\delta^{(1-\alpha)(\underline{\dim}_\alpha(\mu_1)-\underline{\dim}_\alpha(\mu_2)-\varepsilon_1+\varepsilon_2)}\right]\right\}. \tag{40}$$

Dividing the above inequality by $(-\log_2(\delta))$ and taking the limit as $\delta \to 0$, we conclude:

$$\liminf_{\delta\to 0}\frac{H_\alpha(a_1\mu_1+a_2\mu_2;\mathcal{B}_\delta)}{-\log_2(\delta)} \geq \underline{\dim}_\alpha(\mu_1)-\varepsilon_1. \tag{41}$$

Since $\varepsilon_1,\varepsilon_2$ was the arbitrary numbers, then desired inequality holds. □

Clearly, the above theorem can be generalized for any finite number of measures. When all measures have Rényi entropy dimension of order α then the entropy dimension of the convex combination of measures is determined precisely.

COROLLARY 14. *Let* $a_k \in (0,1)$ *for* $k = 1,\dots,n$ *be such that* $\sum\limits_{k=1}^{n} a_k = 1$ *where* $n \in \mathbb{N}$ *and let* $\{\mu_k\}_{k=1}^{n} \subset M_1(X,\Sigma)$. *If every* μ_k *has finite Rényi entropy dimension for* $k \in \{1,\dots,n\}$ *then* $\sum\limits_{k=1}^{n} a_k\mu_k$ *also have Rényi entropy dimension. Moreover,*

$$\dim_\alpha(\sum_{k=1}^{n} a_k\mu_k) = \begin{cases} \max\limits_{k=1,\dots,n} \dim(\mu_k) & \text{for } \alpha \in (0,1), \\ \min\limits_{k=1,\dots,n} \dim(\mu_k) & \text{for } \alpha \in (1,\infty). \end{cases} \tag{42}$$

4. Conclusion

This paper provides a successive application of the weighted entropy. The definition of weighted Rényi entropy allows to decompose the Rényi entropy of the mixture of measures in terms of the entropies of the single measures. Moreover, the obtained bounds converge with $\alpha \to 1$ to the corresponding bounds calculated for the Shannon entropy.

Our main result gives an estimation of the Rényi entropy dimension of the mixture of measures in the case of arbitrary metric space. It generalizes the estimation of I. Csiszár.

5. Appendix A.

We prove Corollary 10. Let us start with the proposition:

PROPOSITION 15. *We assume that* $\alpha \in (0,\infty)\setminus\{1\}$ *and* $n \in \mathbb{N}$. *Let* $a_k \in (0,1)$ *for* $k \in \{1,\dots,n\}$ *be such that* $\sum\limits_{k=1}^{n} a_k = 1$ *and let* $\{\mu_k\}_{k=1}^{n} \subset M_1(X,\Sigma)$. *We define* $\mu := \sum\limits_{k=1}^{n} a_k\mu_k \in M_1(X,\Sigma)$.

- *If* $\mathcal{P}$ *is a* μ*-partition of* X *then* $\mathcal{P}$ *is a* μ_k*-partition of* X *for* $k \in \{1,\dots,n\}$ *and*

$$h_\alpha(\mu;\mathcal{P}) \geq g_\alpha^{-1}\big[\sum_{k=1}^{n} a_k g_\alpha(h_\alpha(\mu_k;\mathcal{P}))\big]. \tag{43}$$

- *If* $\mathcal{Q} \subset \Sigma$ *and* $\mathfrak{m}^k \in W(\mu_k;\mathcal{Q})$ *for* $k \in \{1,\dots,n\}$ *then* $\mathfrak{m} := \sum\limits_{k=1}^{n} a_k\mathfrak{m}^k \in W(\mu;\mathcal{Q})$ *and*

$$h_\alpha^W(\mu;\mathfrak{m}) \leq g_\alpha^{-1}\big[\sum_{k=1}^{n} a_k^\alpha g_\alpha(h_\alpha^W(\mu_k;\mathfrak{m}^k))\big]. \tag{44}$$

Proof. It is easy to see that $\mathcal{P}$ is a μ_k-partition of X for every $k \in \{1,\dots,n\}$.

Making use of Observation 8: 1a and 1c for $\alpha \in (0,1)$ or 2a and 2c for $\alpha \in (1,\infty)$, we have

$$h_\alpha(\mu;\mathcal{P}) = \frac{1}{1-\alpha}\log_2\big[\sum_{P\in\mathcal{P}}\big(\sum_{k=1}^n a_k\mu_k(P)\big)^\alpha\big] \tag{45}$$

$$\geq \frac{1}{1-\alpha}\log_2\big[\sum_{k=1}^n\big(a_k\sum_{P\in\mathcal{P}}\mu_k(P)^\alpha\big)\big] \tag{46}$$

$$= \frac{1}{1-\alpha}\log_2\big[\sum_{k=1}^n a_k 2^{(1-\alpha)\,h_\alpha(\mu_k;\mathcal{P})}\big] \tag{47}$$

$$= g_\alpha^{-1}\big[\sum_{k=1}^n a_k g_\alpha(h_\alpha(\mu_k;\mathcal{P}))\big], \tag{48}$$

which proves (43).

We derive the second part of Proposition. Clearly, $\mathfrak{m} \in W(\mu;\mathcal{Q})$. To verify (44) we use Observation 8: 1a and 1b for $\alpha \in (0,1)$ or 2a and 2b for $\alpha \in (1,\infty)$:

$$h_\alpha^W(\mu;\mathfrak{m}) = \frac{1}{1-\alpha}\log_2\big[\sum_{Q\in\mathcal{Q}}\big(\sum_{k=1}^n a_k\mathfrak{m}_Q^k(X)\big)^\alpha\big] \tag{49}$$

$$\leq \frac{1}{1-\alpha}\log_2\big[\sum_{k=1}^n\big(a_k^\alpha\sum_{Q\in\mathcal{Q}}\mathfrak{m}_Q^k(X)^\alpha\big)\big] \tag{50}$$

$$= \frac{1}{1-\alpha}\log_2\big[\sum_{k=1}^n a_k^\alpha 2^{(1-\alpha)\,h_\alpha^W(\mu_k;\mathfrak{m}^k)}\big] \tag{51}$$

$$= g_\alpha^{-1}\big[\sum_{k=1}^n a_k^\alpha g_\alpha(h_\alpha^W(\mu_k;\mathfrak{m}^k))\big]. \tag{52}$$

$\square$

Below we present the complete proof of Corollary 10

Proof. (of Corollary 10) Let us first consider the case when $H_\alpha(\mu_k;\mathcal{Q}) = \infty$ for a certain $k \in \{1,\dots,n\}$. Then also $H_\alpha(\mu;\mathcal{Q}) = \infty$ and the inequalities hold trivially.

Thus let us assume that for every $k \in \{1,\dots,n\}$, $H_\alpha(\mu_k;\mathcal{Q}) < \infty$. Without loss of generality, we may assume also that $a_k \neq 0$ for every $k \in \{1,\dots,n\}$. Let $\varepsilon > 0$ be arbitrary.

To prove the first inequality, we find a $\mathcal{Q}$-acceptable μ-partition $\mathcal{P}$ such that

$$H_\alpha(\mu;\mathcal{Q}) \geq h_\alpha(\mu;\mathcal{P}) - \varepsilon. \tag{53}$$

Consequently, by Proposition 15 and the definition of Rényi entropy, we have

$$h_\alpha(\mu;\mathcal{P}) = h_\alpha(\sum_{k=1}^{n} a_k\mu_k;\mathcal{P}) \tag{54}$$

$$\geq g_\alpha^{-1}\big[\sum_{k=1}^{n} a_k g_\alpha(h_\alpha(\mu_k;\mathcal{P}))\big] \geq g_\alpha^{-1}\big[\sum_{k=1}^{n} a_k g_\alpha(H_\alpha(\mu_k;\mathcal{Q}))\big]. \tag{55}$$

Finally by (53), we obtain

$$H_\alpha(\mu;\mathcal{Q}) \geq h_\alpha(\mu;\mathcal{P}) - \varepsilon \geq g_\alpha^{-1}\big[\sum_{k=1}^{n} a_k g_\alpha(H_\alpha(\mu_k;\mathcal{P}))\big] - \varepsilon, \tag{56}$$

which proves (14).

We prove the inequality (15). For each $k \in \{1,\dots n\}$ we find $\mathfrak{m}^k \in W(\mu_k;\mathcal{Q})$ satisfying

$$h_\alpha^W(\mu_k;\mathfrak{m}^k) \leq H_\alpha(\mu_k;\mathcal{Q}) + \frac{\varepsilon}{n}. \tag{57}$$

Making use of Proposition 15 and (57), we have

$$H_\alpha(\mu;\mathcal{Q}) \leq g_\alpha^{-1}\big[\sum_{k=1}^{n} a_k^\alpha g_\alpha(h_\alpha^W(\mu_k;\mathfrak{m}^k))\big] \tag{58}$$

$$\leq g_\alpha^{-1}\big[\sum_{k=1}^{n} a_k^\alpha g_\alpha(H_\alpha(\mu_k;\mathcal{Q}))\big] + \varepsilon. \tag{59}$$

This completes the proof as $\varepsilon > 0$ was an arbitrary number. □

Bibliography

1. **Campbell, L. L.** 1965. Inform. Control **8**:423–429.
2. **Csiszár, I.** 1962. Acta Math. Hungarica **13** (3–4):245–255.
3. **Ellis, R. S.** 1985. Springer, New York.
4. **Franchini, F., A. R. Its, and V. E. Korepin.** 2008. J. Phys. A-Math. Theor. **41** (2).
5. **Grassberger, P., and I. Procaccia.** 1983. Phys. rev. **28**:2591–2593.
6. **Gray, R. M.** 1990. *Entropy and Information Theory.* Springer, New York.

7. **Howroyd, J. D.** 1994. *On the Theory of Hausdorff Measure in Metric Space.* Ph.D. Dissertation, University Collage London.

8. **Howroyd, J. D.** 1995. Proc. London Math. Soc. **70** (3):581–604.

9. **Li, X., and S. Zhang.** 2011. Probab. Eng. Inform. Sc. **25**:237–250.

10. **Rényi, A.** 1959. Acta Math. Hungarica **10** (1–2):193–215.

11. **Rényi, A.** 1961. Proc. Fourth Berkeley Symp. on Math. Statist.and Prob. **1**:647–561.

12. **Seibt, P.** 2006. *Algorithmic Information Theory.* Springer-Verlag, Berlin.

13. **Shannon, C. E.** 1948. Bell Syst. Tech. J. **27**:379–423, 623–656.

14. **Śmieja, M., and J. Tabor.** 2012. IEEE T. Infor. Theory **58** (5):2719–2728.

15. **Wu, Y., and S. Verdú.** 2010. IEEE T. Inform. Theory **56** (8):3721–3748.