Science and Information Conference 2014
August 27-29, 2014 | London, UK

Rényi entropy dimension of the mixture of measures

Marek Śmieja*, Jacek Tabor*
*Faculty of Mathematics and Computer Science
Jagiellonian University, Lojasiewicza 6, 30-348 Kraków, Poland
Email: marek.smieja@uj.edu.pl, jacek.tabor@uj.edu.pl

***Abstract*—Rényi entropy dimension describes the rate of growth of coding cost in the process of lossy data compression in the case of exponential dependence between the code length and the cost of coding. In this paper we generalize the Csiszár estimation of the Rényi entropy dimension of the mixture of measures for the case of general probability metric space. This result determines the cost of encoding of the information which comes from the combined sources assuming its exponential growth. Our proof relies on an equivalent definition of the Rényi entropy in weighted form which allows to deal well with a calculation of the entropy of the mixture of measures.**

I. Introduction

The fundamental limit of compression is determined by the Shannon entropy [1]. It is obtained by minimizing a function of average code length with a constraint that a code is uniquely decipherable. This applies well for the case when the coding cost depends linearly on the code length. However, in many practical situations the cost might be rather an exponential function of length. This may occur when the cost of encoding and decoding equipment is an important factor. Campbell considered such a measure of coding cost and showed then that the Rényi entropy is a limit of compression [2]. The idea was later continued and developed in [3]–[6].

An asymptotic property of the entropy is described by the entropy dimension [7]–[9] which is widely used in information theory [10]–[13]. In the process of lossy coding the entropy dimension (with specific entropy function as Shannon or Rényi entropy) stands for the rate of growth of the coding cost given by the successive increasing of the coding precision.

Csiszár obtained the bounds for the Rényi entropy of the mixture of measures in the case of real probability space [14, Lemma 1]. These 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

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

One can say that the rate of growth of the coding cost of information which comes from the combined sources is determined by the corresponding rate of dominant single source.

In this paper we investigate the problem of an approximation of Rényi entropy dimension of the mixture of measures in the more general case. Our main result presents how to obtain a similar estimation to the aforementioned one in a probability metric space (which is not necessarily a real probability space). For this reason we follow a different approach which relies on the idea introduced in [15] – the key for determining such an approximation is the definition of weighted entropy which provides an easy calculation of the entropy of the mixture of measures.

The paper is organized as follows. The next section contains a summary of the results concerning the weighted entropy and its application in the case of the entropy of the mixture of measures. In particular, in Theorem II.1 we show how to estimate the Rényi entropy of combined sources. In the third section we present our main result, Theorem III.1, which establishes the approximation of the Rényi entropy dimension of the mixture of measures in the case of general probability metric space.

II. Weighted Rényi entropy

The idea of weighted entropy, introduced in [15], allows the establishment of the estimation of the entropy of combined sources relatively easily. In this section we give a quick review of this approach and use it for the case of Rényi entropy. Let us first recall all necessary definitions.

If not stated otherwise, we always assume that (X, Σ, μ) is a probability space. As a μ-partition we consider a countable family $\mathcal{P} \subset \Sigma$ of pairwise disjoint sets of X which satisfies

$$\mu(X \setminus \bigcup_{P \in \mathcal{P}} P) = 0. \tag{2}$$

In the case of coding it plays a role of the coding alphabet.

The definition of the Rényi entropy [16] is based on an axiomatic approach to information and was introduced as the one parametric family of generalized entropy functions.

Definition II.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

$$h_\alpha(\mu; \mathcal{P}) := \frac{1}{1-\alpha} \log_2 \big(\sum_{P \in \mathcal{P}} \mu(P)^\alpha \big). \tag{3}$$

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

The Shannon Coding Theorem [1] 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 $\mathcal{P}$. Campbell considered another measure of length where the cost of encoding a symbol depends exponentially on the code length [2]. As a result he found that it is not possible to encode so that the introduced cost is less than the Rényi μ-entropy of $\mathcal{P}$. This justifies the importance of the Rényi entropy in the theory of coding. Through the paper we consider the case of the exponential dependence between the code length and the cost of encoding.

The lossy compression might use various coding partitions which differ in a coding precision. To define a maximal coding error we fix a family $\mathcal{Q}$ of measurable subsets of X which we call an error-control family. Consequently, as the coding partitions we consider only these which satisfy the condition: for every $P \in \mathcal{P}$ there exists $Q \in \mathcal{Q}$ such that $P \subset Q$ – partitions with this property are called $\mathcal{Q}$-acceptable which we write $\mathcal{P} \prec \mathcal{Q}$. In the case of $\mathbb{R}^N$ the error-control families usually consist of balls with a given radius or cubes with a specific edge length as explained in [7], [15], [17].

The statistical coding cost of optimal lossy compression is determined by the notion of the Rényi μ-entropy of an error-control family:

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

$$H_\alpha(\mu; \mathcal{Q}) := \inf\{h_\alpha(\mu; \mathcal{P}) \in [0, \infty] : \mathcal{P} \text{ is a } \mu\text{-partition and } \mathcal{P} \prec \mathcal{Q}\}. \tag{4}$$

The use of partition makes the coding deterministic – a given element $x \in X$ is always encoded by the unique $P \in \mathcal{P}$ such that $x \in P$. In many practical situations, when the coding cannot be controlled precisely, we have to allow some random factor in the coding procedure. We realize it by dividing the measure μ into sub-probability measures instead of dividing the data space into partition.

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

$$W(\mu; \mathcal{Q}) := \{\mathfrak{m} : \mathcal{Q} \ni Q \to \mathfrak{m}_Q \in M(X, \Sigma) : \mathfrak{m}_Q(X \setminus Q) = 0 \text{ for every } Q \in \mathcal{Q} \text{ and } \textstyle\sum_{Q \in \mathcal{Q}} \mathfrak{m}_Q = \mu\}, \tag{5}$$

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 (5) is countable.

The following example shows how to perform the random coding:

Example II.1. Let $X = [0, 1]$, Σ be a σ-algebra generated by all Borel subsets of X and μ be a measure on (X, Σ) presented in Figure 1a. The coding is controlled by a family $\mathcal{Q} = \{[0, \frac{1}{2}], [\frac{1}{4}, \frac{3}{4}], [\frac{1}{2}, 1]\}$.

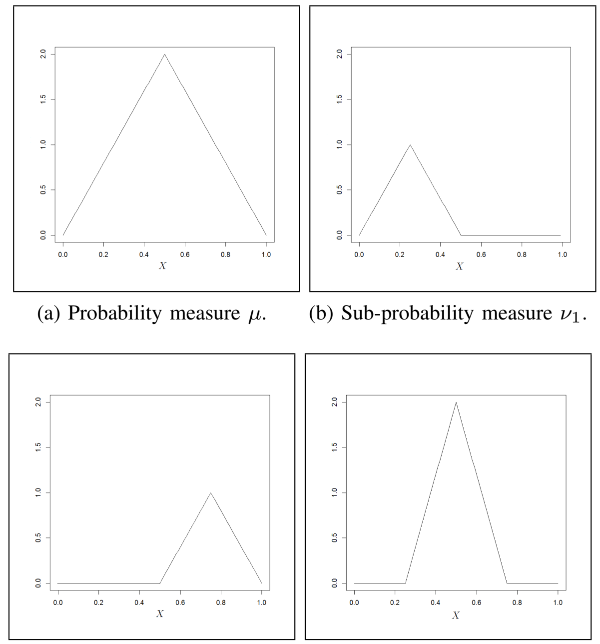
The random way of coding is defined by a function $\mathfrak{m} \in W(\mu; \mathcal{Q})$:

$$\mathfrak{m}([0, \frac{1}{2}]) = \nu_1, \tag{6}$$

$$\mathfrak{m}([\frac{1}{4}, \frac{3}{4}]) = \nu_2, \tag{7}$$

$$\mathfrak{m}([\frac{1}{2}, 1]) = \nu_3, \tag{8}$$

where ν_1, ν_2 and ν_3 are the measures shown in Figures 1b, 1c, 1d.



(a) Probability measure μ. (b) Sub-probability measure ν_1. (c) Sub-probability measure ν_2. (d) Sub-probability measure ν_3.

Fig. 1: The division $\mathfrak{m}$ of a probability measure μ which defines a random way of coding on X controlled by $\mathcal{Q}$.

Clearly, $\mu = \nu_1 + \nu_2 + \nu_3$.

This definition implies for instance that points contained in $[\frac{1}{4}, \frac{3}{4}]$ are coded by one of the measures ν_1, ν_2 or ν_3 with probabilities:

$$\begin{aligned} code(\nu_1) &= \frac{\int_{\frac{1}{4}}^{\frac{3}{4}} \nu_1(x)dx}{\int_{\frac{1}{4}}^{\frac{3}{4}} (\nu_1(x)+\nu_2(x)+\nu_3(x))dx} \\ &= \frac{1/8}{1/8+1/2+1/8} = 1/6, \end{aligned} \tag{9}$$

$$\begin{aligned} code(\nu_2) &= \frac{\int_{\frac{1}{4}}^{\frac{3}{4}} \nu_2(x)dx}{\int_{\frac{1}{4}}^{\frac{3}{4}} (\nu_1(x)+\nu_2(x)+\nu_3(x))dx} \\ &= \frac{1/2}{1/8+1/2+1/8} = 2/3, \end{aligned} \tag{10}$$

$$\begin{aligned} code(\nu_3) &= \frac{\int_{\frac{1}{4}}^{\frac{3}{4}} \nu_3(x)dx}{\int_{\frac{1}{4}}^{\frac{3}{4}} (\nu_1(x)+\nu_2(x)+\nu_3(x))dx} \\ &= \frac{1/8}{1/8+1/2+1/8} = 1/6. \end{aligned} \tag{11}$$

On the other hand, elements included in $[0, \frac{1}{4}] \cup [\frac{3}{4}, 1]$ are always coded with use of one measure – ν_1 and ν_3 respectively. The coding is nondeterministic, since points from $[\frac{1}{4}, \frac{3}{4}]$ can be written by three different codes with positive probabilities.

The amount of information associated with random coding is described by the weighted Rényi μ-entropy of $\mathfrak{m} \in W(\mu; \mathcal{Q})$.

Definition II.3. 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

$$h_\alpha^W(\mu;\mathfrak{m}) := \frac{1}{1-\alpha}\log_2\Big(\sum_{Q\in\mathcal{Q}} \mathfrak{m}_Q(X)^\alpha\Big) \tag{12}$$

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

$$H_\alpha^W(\mu;\mathcal{Q}) := \inf\{h_\alpha^W(\mu;\mathfrak{m}) \in [0,\infty] : \mathfrak{m} \in W(\mu,\mathcal{Q})\}. \tag{13}$$

The Rényi μ-entropy from Definition II.1 satisfies the condition of general entropy function [18, Definition 2.5 and Observation 2.1] i.e.:

$$h_\alpha(\mu;\mathcal{P}) = f\Big(\sum_{P\in\mathcal{P}} g(\mu(P))\Big), \tag{14}$$

where f is increasing, g is subadditive and concave (for $\alpha \in (0,1)$) or f is decreasing, g is superadditive and convex (for $\alpha \in (1,\infty)$). Thus the weighted Rényi entropy (13) is equivalent to the classical Rényi entropy of an error-control family [18, Theorem 3.1] which allows to use these terms interchangeably.

As mentioned in the Introduction the weighted entropy enables us to use of the operations on functions instead of the operations on sets when computing the entropy. In consequence, we can obtain the estimation of the entropy of the mixture of measures relatively easily. The formula for the Shannon entropy is established in [15, Theorem 3.1] while the result concerning Tsallis entropy in calculated in [18, Theorem 4.1]. Making use of similar technique, we get the bounds for the Rényi entropy.

For the sake of transparency let us denote by g_α and its inverse g_α^{-1} the following functions:

$$g_\alpha(x) = 2^{(1-\alpha)x},\; g_\alpha^{-1}(x) = \frac{1}{1-\alpha}\log_2(x). \tag{15}$$

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

$$H_\alpha(\mu;\mathcal{Q}) \geq g_\alpha^{-1}\Big[\sum_{k=1}^n a_k g_\alpha(H_\alpha(\mu_k;\mathcal{Q}))\Big] \tag{16}$$

and

$$H_\alpha(\mu;\mathcal{Q}) \leq g_\alpha^{-1}\Big[\sum_{k=1}^n a_k^\alpha g_\alpha(H_\alpha(\mu_k;\mathcal{Q}))\Big]. \tag{17}$$

Before presenting the sketch of the proof of Theorem II.1 let us observe that the following simple observation is valid:

Observation II.1.
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.*

Proof: (the sketch of the proof of Theorem II.1) The case when $H_\alpha(\mu;\mathcal{Q}) = \infty$ for a certain $k \in \{1,\dots,n\}$ is trivial so without loss of generality we assume that $H_\alpha(\mu_k;\mathcal{Q}) < \infty$ and $a_k \neq 0$ for every $k \in \{1,\dots,n\}$.

Let us first concentrate on the inequality (16) and observe that every μ-partition $\mathcal{P}$ is also a μ_k-partition for $k \in \{1,\dots,n\}$. Moreover, making use of the Observation II.1: 1a and 1c for $\alpha \in (0,1)$ or 2a and 2c for $\alpha \in (1,\infty)$, we get that for a μ-partition $\mathcal{P}$ the following inequality holds:

$$h_\alpha(\mu;\mathcal{P}) \geq g_\alpha^{-1}\Big[\sum_{k=1}^n a_k g_\alpha(h_\alpha(\mu_k;\mathcal{P}))\Big]. \tag{18}$$

Therefore, given an arbitrary $\varepsilon > 0$, we find a $\mathcal{Q}$-acceptable μ-partition $\mathcal{P}$ such that

$$H_\alpha(\mu;\mathcal{Q}) \geq h_\alpha(\mu;\mathcal{P}) - \varepsilon = h_\alpha\Big(\sum_{k=1}^n a_k\mu_k;\mathcal{P}\Big) - \varepsilon \tag{19}$$

$$\geq g_\alpha^{-1}\Big[\sum_{k=1}^n a_k g_\alpha(h_\alpha(\mu_k;\mathcal{P}))\Big] - \varepsilon \tag{20}$$

$$\geq g_\alpha^{-1}\Big[\sum_{k=1}^n a_k g_\alpha(H_\alpha(\mu_k;\mathcal{Q}))\Big] - \varepsilon, \tag{21}$$

which proves (16).

To obtain (17) let us fix $\varepsilon > 0$ and for every $k \in \{1,\dots n\}$ let us choose $\mathfrak{m}^k \in W(\mu_k;\mathcal{Q})$ satisfying

$$h_\alpha^W(\mu_k;\mathfrak{m}^k) \leq H_\alpha(\mu_k;\mathcal{Q}) + \frac{\varepsilon}{n}. \tag{22}$$

Then using Observation II.1: 1a and 1b for $\alpha \in (0,1)$ or 2a and 2b for $\alpha \in (1,\infty)$, we get that $\mathfrak{m} := \sum_{k=1}^n a_k\mathfrak{m}^k \in W(\mu;\mathcal{Q})$ and

$$h_\alpha^W(\mu;\mathfrak{m}) \leq g_\alpha^{-1}\Big[\sum_{k=1}^n a_k^\alpha g_\alpha(h_\alpha^W(\mu_k;\mathfrak{m}^k))\Big]. \tag{23}$$

By (22), we have

$$H_\alpha(\mu;\mathcal{Q}) \leq g_\alpha^{-1}\Big[\sum_{k=1}^n a_k^\alpha g_\alpha(h_\alpha^W(\mu_k;\mathfrak{m}^k))\Big] \tag{24}$$

$$\leq g_\alpha^{-1}\Big[\sum_{k=1}^n a_k^\alpha g_\alpha(H_\alpha(\mu_k;\mathcal{Q}))\Big] + \varepsilon, \tag{25}$$

which ends the proof. ∎

The bounds from Theorem II.1 cannot be improved. By an easy calculation we get that the right side of the formula (16) is attained on the combination of measures with pairwise disjoint supports (i.e. are non zero on pairwise disjoint subsets of X) while the right side of the expression (17) is obtained when we use identical measures [18, Example 4.1].

Moreover, these bounds converge to the corresponding ones of the Shannon μ-entropy as $\alpha \to 1$. For the details we refer the reader to [18] where this property was presented for Tsallis μ-entropy.

III. Bounds for the Rényi entropy dimension of the mixture of measures

Campbell calculated the bounds of the Rényi entropy of combined sources in $\mathbb{R}$ [14, Lemma 1] which leads to the approximation of the Rényi entropy dimension of the mixture of measures. We will consider the case of arbitrary probability metric space and show that the similar approximation holds. Our proof relies on applying Theorem II.1.

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

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

$$\mathcal{B}_\delta := \{B(x,\delta) : x \in X\}, \tag{26}$$

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

Definition III.1. 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

$$\overline{\dim}_\alpha(\mu) := \limsup_{\delta\to 0} \frac{H_\alpha(\mu;\mathcal{B}_\delta)}{-\log_2(\delta)}, \tag{27}$$

$$\underline{\dim}_\alpha(\mu) := \liminf_{\delta\to 0} \frac{H_\alpha(\mu;\mathcal{B}_\delta)}{-\log_2(\delta)}. \tag{28}$$

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 III.1. *Let $a_k \in (0,1)$ for $k = 1,\dots,n$ be such that $\sum_{k=1}^n a_k = 1$ where $n \in \mathbb{N}$ and let $\{\mu_k\}_{k=1}^n \subset M_1(X,\Sigma)$. If $\overline{\dim}_\alpha(\mu_k) < \infty$ for every $k \in \{1,\dots,n\}$ then:*

$$\overline{\dim}_\alpha\Big(\sum_{k=1}^n a_k\mu_k\Big) \le \begin{cases} \max\limits_{k=1,\dots,n} \overline{\dim}_\alpha(\mu_k), \textit{for } \alpha \in (0,1), \\ \min\limits_{k=1,\dots,n} \overline{\dim}_\alpha(\mu_k), \textit{for } \alpha \in (1,\infty) \end{cases} \tag{29}$$

and

$$\underline{\dim}_\alpha\Big(\sum_{k=1}^n a_k\mu_k\Big) \ge \begin{cases} \max\limits_{k=1,\dots,n} \underline{\dim}_\alpha(\mu_k), \textit{for } \alpha \in (0,1), \\ \min\limits_{k=1,\dots,n} \underline{\dim}_\alpha(\mu_k), \textit{for } \alpha \in (1,\infty). \end{cases} \tag{30}$$

Proof: For the sake of transparency let us show the first inequality from formula (30) for $n = 2$. Other inequalities can be proved in similar manner (for arbitrary $n \in \mathbb{N}$).

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

$$\liminf_{\delta\to 0} \frac{H_\alpha(\mu_k;\mathcal{B}_\delta)}{-\log_2(\delta)} = \underline{\dim}_\alpha(\mu_k), \text{ for } k = 1,2. \tag{31}$$

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

$$\frac{H_\alpha(\mu_k;\mathcal{B}_{\delta_k})}{-\log_2(\delta_k)} \ge \underline{\dim}_\alpha(\mu_k) - \varepsilon_k \tag{32}$$

and consequently

$$H_\alpha(\mu_k;\mathcal{B}_{\delta_k}) \ge -\log_2(\delta_k)(\underline{\dim}_\alpha(\mu_k) - \varepsilon_k), \tag{33}$$

for $k = 1,2$.

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

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

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

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

By Theorem II.1, we have:

$$H_\alpha(a_1\mu_1 + a_2\mu_2;\mathcal{B}_\delta) \tag{37}$$

$$\ge g_\alpha^{-1}\big\{a_1 g_\alpha\big[H_\alpha(\mu_1;\mathcal{B}_\delta)\big] + a_2 g_\alpha\big[H_\alpha(\mu_2;\mathcal{B}_\delta)\big]\big\} \tag{38}$$

$$\ge g_\alpha^{-1}\big\{a_1 g_\alpha\big[-\log_2(\delta)(\underline{\dim}_\alpha(\mu_1) - \varepsilon_1)\big] \tag{39}$$

$$+ a_2 g_\alpha\big[-\log_2(\delta)(\underline{\dim}_\alpha(\mu_2) - \varepsilon_2)\big]\big\} \tag{40}$$

$$= \frac{1}{1-\alpha}\log_2\big\{\delta^{-(1-\alpha)(\underline{\dim}_\alpha(\mu_1)-\varepsilon_1)} \tag{41}$$

$$\cdot\big[a_1 + a_2\delta^{(1-\alpha)(\underline{\dim}_\alpha(\mu_1)-\underline{\dim}_\alpha(\mu_2)-\varepsilon_1+\varepsilon_2)}\big]\big\}. \tag{42}$$

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

$$\liminf_{\delta\to 0} \frac{H_\alpha(a_1\mu_1 + a_2\mu_2;\mathcal{B}_\delta)}{-\log_2(\delta)} \ge \underline{\dim}_\alpha(\mu_1) - \varepsilon_1. \tag{43}$$

Since $\varepsilon_1, \varepsilon_2$ was the arbitrary numbers, then the first inequality of (30) holds. ∎

When all measures have Rényi entropy dimension of order α then the entropy dimension of the convex combination of measures is determined precisely.

Corollary III.1. *Let $a_k \in (0,1)$ for $k = 1,\dots,n$ be such that $\sum_{k=1}^n a_k = 1$ 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 $\sum_{k=1}^n a_k\mu_k$ also have Rényi entropy dimension. Moreover,*

$$\dim_\alpha\Big(\sum_{k=1}^n a_k\mu_k\Big) = \begin{cases} \max\limits_{k=1,\dots,n} \dim(\mu_k) & \textit{for } \alpha \in (0,1), \\ \min\limits_{k=1,\dots,n} \dim(\mu_k) & \textit{for } \alpha \in (1,\infty). \end{cases} \tag{44}$$

The following example illustrates Corollary III.1.

Example III.1. Let μ_1 and μ_2 be two measures such that $\dim_\alpha(\mu_1) = 1$ and $\dim_\alpha(\mu_2) = 2$, for $\alpha \in (0,\infty)$ (e.g. μ_1, μ_2 can be uniformly distributed measures on $[0,1]$ and $[0,1] \times [0,1]$, respectively). The value of Rényi entropy dimension of order α of

$$\mu = \frac{1}{3}\mu_1 + \frac{2}{3}\mu_2 \tag{45}$$

follows directly from Corollary III.1. As can be seen in Figure 2 the dimension as a function of α might not be continuous. In this example the dimension of the mixture is determined by the dimensions of μ_2 for $\alpha \in (0,1)$, of μ_1 for $\alpha \in (1,\infty)$ and of their combination for $\alpha = 1$. The value of Rényi entropy dimension of order $\alpha = 1$ is substituted by the dimension calculated using Shannon entropy [15, Corollary IV.1], i.e. $\dim(\mu) = \frac{1}{3}\dim(\mu_1) + \frac{2}{3}\dim(\mu_2)$.

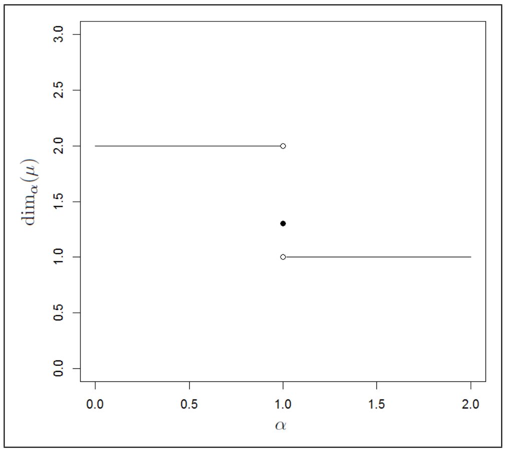


Fig. 2: Rényi entropy dimension of order α of the mixture of two measures $\mu = \frac{1}{3}\mu_1 + \frac{2}{3}\mu_2$, where $\dim_\alpha(\mu_1) = 1$ and $\dim_\alpha(\mu_2) = 2$, for $\alpha \in (0, \infty)$.

IV. Conclusion

The weighted form of the entropy is very useful for calculating the entropy and the entropy dimension of the combination of measures. In this paper it was applied for the case Rényi entropy of order α which determines the cost of coding in the case of the exponential dependence between the code length and the coding cost. Our result generalizes the estimation derived by Csiszár for general probability metrics spaces.

Acknowledgment

This research was supported by National Centre of Science (Poland), Grant No. 2011/01/B/ST6/01887.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

[2] L. L. Campbell, "A coding theorem and Rényi's entropy," *Inform. Control*, vol. 8, pp. 423–429, 1965.

[3] J. F. Bercher, "Source coding with escort distributions and Rényi entropy bounds," *Phys. Lett. A*, vol. 373, no. 36, pp. 3235–3238, 2009.

[4] F. Chapeau-Blondeau, A. Delahaies, and D. Rousseau, "Source coding with Tsallis entropy," *Electron. Lett*, vol. 47, no. 3, pp. 187–188, 2011.

[5] J. N. Kapur, *Measures of Information and Their Applications*. New York: John Wiley & Sonc, 1994.

[6] S. Kumar and R. Kumar, "Some noiseless coding theorem connected with Havrda and Charvat and Tsalliss entropy," *Kragujevac Journal of Mathematics*, vol. 35, no. 1, pp. 111–117, 2011.

[7] A. Rényi, "On the dimension and entropy of probability distributions," *Acta Math. Hungarica*, vol. 10, no. 1–2, pp. 193–215, 1959.

[8] Y. Heurteaux, "Dimension of measures: the probabilistic approach," *Publ. Mat.*, vol. 51, no. 2, pp. 243–290, 2007.

[9] A. Fan, K. Lau, and H. Rao, "Relationships between different dimensions of a measure," *Monatsh. Math.*, vol. 135, no. 3, pp. 191–201, 2002.

[10] T. Kawabata and A. Dembo, "The rate-distortion dimension of sets and measures," *IEEE Trans. Inf. Th.*, vol. 40, no. 5, pp. 1564–1572, 1994.

[11] Y. Wu and S. Verdú, "Rényi information dimension: fundamental limits of almost lossless analog compression," *IEEE Trans. Inf. Th.*, vol. 56, no. 8, pp. 3721–3748, 2010.

[12] A. Haghighatshoar and E. Abbe, "Polarization of the Rényi information dimension for single and multi terminal analog compression," *arXiv preprint arXiv:1301.6388*, 2013.

[13] B. C. Geiger and G. Kubin, "Relative information loss in the PCA," in *Information Theory Workshop (ITW), 2012 IEEE*. IEEE, 2012, pp. 562–566.

[14] I. Csiszár, "On the dimension and entropy of order α of the mixture of probability distributions," *Acta Math. Hungarica*, vol. 13, no. 3–4, pp. 245–255, 1962.

[15] M. Śmieja and J. Tabor, "Entropy of the mixture of sources and entropy dimension," *IEEE Trans. Inf. Th.*, vol. 58, no. 5, pp. 2719–2728, 2012.

[16] A. Rényi, "On measures of entropy an information," *Proc. Fourth Berkeley Symp. on Math. Statist.and Prob.*, vol. 1, pp. 647–561, 1961.

[17] E. C. Posner and E. R. Rodemich, "Epsilon entropy and data compression," *Ann. Math. Stat.*, vol. 42, pp. 2079–2125, 1971.

[18] M. Śmieja, "Weighted approach to general entropy function," *IMA Journal of Mathematical Control and Information, doi:10.1093/imamci/DNT044*, 2013.