

Entropy of the Mixture of Sources and Entropy Dimension

Marek Śmieja

Jacek Tabor

Preprint 2011/03

Abstract

Suppose that we are given two sources S_1, S_2 and an “error-control” family \mathcal{Q} . We assume that we lossy-code S_1 with \mathcal{Q} -acceptable alphabet \mathcal{P}_1 and S_2 with \mathcal{Q} -acceptable alphabet \mathcal{P}_2 . Consider a new source S which sends a signal produced by source S_1 with probability a_1 and by source S_2 with probability $a_2 = 1 - a_1$. We provide a simple greedy algorithm which constructs a \mathcal{Q} -acceptable coding alphabet \mathcal{P} of S such that the entropy $h(\mathcal{P})$ satisfies:

$$h(\mathcal{P}) \leq a_1 h(\mathcal{P}_1) + a_2 h(\mathcal{P}_2) + 1.$$

In the proof of the above formula the basic role is played by a new equivalent definition of entropy based on measures instead of partitions.

As a consequence we obtain an estimation of the entropy and Rényi entropy dimension of the convex combination of measures. In particular if probability measures μ_1, μ_2 have entropy dimension then

$$\dim_E(a_1\mu_1 + a_2\mu_2) = a_1\dim_E(\mu_1) + a_2\dim_E(\mu_2).$$

In the case of probability measures in \mathbb{R}^N this allows to link the upper local dimension at point with the upper entropy dimension of a measure by an improved version of Young estimation:

$$\overline{\dim}_E(\mu) \leq \int_{\mathbb{R}^N} \overline{D}_\mu(x) d\mu(x),$$

where $\overline{D}_\mu(x)$ stands for upper local dimension of μ at point x .

1 Introduction

The classical entropy introduced by C. E. Shannon [1] and the entropy dimension¹ defined by A. Rényi [2] play a crucial role in information theory, coding, study of statistical and physical systems [3–6]. In information theory, the entropy is understood as an absolute limit of the best possible lossless compression of any communication. The entropy dimension in turn can be interpreted as a rate of convergence of the minimal amount of information needed to encode randomly chosen element with respect to maximal error decreasing to zero.

1.1 Motivation

To explain our results, let us first recall that given a probability measure μ on a space X and a countable partition \mathcal{P} of X into measurable sets, we define the *entropy of μ with respect to \mathcal{P}* by the formula

$$h(\mu; \mathcal{P}) := \sum_{P \in \mathcal{P}} sh(\mu(P)), \quad (1)$$

where $sh(x) := -x \log_2 x$. As we know the entropy corresponds to the statistical amount of information given by optimal lossy-coding of X by elements of partition \mathcal{P} , where \mathcal{P} plays the role of the coding alphabet. Motivated by the idea of Rényi realized by the entropy dimension, we generalise the above formula for arbitrary measurable cover \mathcal{Q} of X by

$$H(\mu; \mathcal{Q}) := \inf\{h(\mu; \mathcal{P}) : \mathcal{P} \text{ is a partition of } X \text{ and } \mathcal{P} \prec \mathcal{Q}\}. \quad (2)$$

The family \mathcal{Q} is interpreted as a maximal error we are allowed to make in the lossy-coding. We accept only such coding alphabets \mathcal{P} , in which every element of \mathcal{P} is a subset of a certain element of \mathcal{Q} (if this is the case we say that \mathcal{P} is \mathcal{Q} -acceptable).

Remark 1.1. The simplest natural case of such error-control family \mathcal{Q} for classical random variables is given by the set \mathcal{B}_δ of all intervals in \mathbb{R} with length δ . Then to find $H(\mu; \mathcal{B}_\delta)$ we need to consider the infimum of entropies of all lossy-codings $h(\mu; \mathcal{P})$, where the elements of \mathcal{P} have length not greater than δ .

¹It is sometimes called Rényi information dimension.

A. Rényi considered the above error-control family \mathcal{B}_δ in his definition of entropy dimension [2] (he also studied the more general case of metric spaces when \mathcal{B}_δ denoted the family of all balls with radius δ). One can also encounter in the general metric spaces the family of sets with diameter δ or in the case of \mathbb{R}^N of cubes with edge-length δ .

Our basic motivation in the paper was the following problem:

Problem 1.1. Suppose that we are given an error-control family \mathcal{Q} and two sources S_1, S_2 in X (represented by probability measures μ_1, μ_2 on X). Let us consider a new source S which sends a signal produced by source S_1 with probability a_1 and by source S_2 with probability $a_2 = 1 - a_1$. Source S is a mixture of S_1 and S_2 . The question is what is the entropy of source S with respect to the error \mathcal{Q} ?

In other words we are interested in estimation of $H(a_1\mu_1 + a_2\mu_2; \mathcal{Q})$ in terms of $H(\mu_1; \mathcal{Q})$ and $H(\mu_2; \mathcal{Q})$.

Observation 1.1. *Observe that if elements of \mathcal{Q} are pairwise disjoint then the answer to the above problem is trivial as by the subadditivity of the function sh we have*

$$H(\mu; \mathcal{Q}) = h(\mu; \mathcal{Q}) = \sum_{Q \in \mathcal{Q}} sh(\mu(Q)) \quad (3)$$

$$= \sum_{Q \in \mathcal{Q}} sh(a_1\mu_1(Q) + a_2\mu_2(Q)) \leq \sum_{Q \in \mathcal{Q}} sh(a_1\mu_1(Q)) + sh(a_2\mu_2(Q)) \quad (4)$$

$$= a_1 H(\mu_1; \mathcal{Q}) + a_2 H(\mu_2; \mathcal{Q}) + sh(a_1) + sh(a_2). \quad (5)$$

To see that the above estimation is sharp it is sufficient to consider a source S_1 which sends only signal 0 and source S_2 which sends signal 1. Clearly, $H(S_1) = H(S_2) = 0$. Then the entropy of the source S which sends signal generated by S_1 with probability a_1 and S_2 with probability a_2 is exactly $a_1 H(S_1) + a_2 H(S_2) + sh(a_1) + sh(a_2)$.

1.2 Main Results

In our main result, Theorem 3.1, we show that the formula calculated in the above observation:

$$H(a_1\mu_1 + a_2\mu_2; \mathcal{Q}) \leq a_1 H(\mu_1; \mathcal{Q}) + a_2 H(\mu_2; \mathcal{Q}) + sh(a_1) + sh(a_2) \quad (6)$$

is valid in the general case, that is when \mathcal{Q} is an arbitrary measurable cover of X . The proof of our main result relies on a new definition of entropy based on measures instead of partitions, which we call *weighted entropy*. We provide an algorithm, which for given alphabets $\mathcal{P}_1, \mathcal{P}_2$ and measures μ_1, μ_2 allows to construct “joint” alphabet \mathcal{P} satisfying above inequality.

Remark 1.2. We would like to add here that our idea of weighted entropy is indebted to the notion of weighted Hausdorff measures considered by J. Howroyd [7, 8]. The advantage of weighted Hausdorff measures over the classical ones is well-summarised by words of K. Falconer [9, Introduction]: “Recently, a completely different approach was introduced by Howroyd using weighted Hausdorff measures to enable the use of powerful techniques from functional analysis, such as the Hahn-Banach and Krein-Milman theorems.” Making use of weighted Hausdorff measures Howroyd proves that

$$\dim_H(X) + \dim_H(Y) \leq \dim_H(X \times Y), \quad (7)$$

where $\dim_H(X)$ is the Hausdorff-Besicovitch dimension of X .

For the precise definition of weighted entropy we refer the reader to the next section. We would only like to mention that, roughly speaking, weighted entropy 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). This operation is crucial in the proof of formula (6), whereas the second important part is played by Theorem 2.1, which proves that the weighted entropy is equal to the classical one.

As an easy consequence of (6) in Theorem 4.1 we obtain an estimation of the entropy dimension of the convex combination of measures. This result can be summarised as follows (see Corollary 4.1):

Let μ_1 and μ_2 be probability measures which have entropy dimension and let $a_1, a_2 \in (0, 1)$ be such that $a_1 + a_2 = 1$. Then $a_1\mu_1 + a_2\mu_2$ has entropy dimension and

$$\dim_E(a_1\mu_1 + a_2\mu_2) = a_1\dim_E(\mu_1) + a_2\dim_E(\mu_2), \quad (8)$$

where $\dim_E(\cdot)$ stands for the entropy dimension of a given measure.

In the case of measures in \mathbb{R}^N this allows to combine the local upper dimension $\overline{D}_\mu(\cdot)$ with the upper entropy dimension $\overline{\dim}_E(\cdot)$ and improve Young estimation of the upper entropy dimension [10]:

$$\overline{\dim}_E(\mu) \leq \int_{\mathbb{R}^N} \overline{D}_\mu(x) d\mu(x). \quad (9)$$

2 Weighted Entropy

From now on, if not stated otherwise, we assume that (X, Σ, μ) is a probability space. The set of probability measures on (X, Σ) will be denoted by $M_1(X, \Sigma)$. When we consider a set of all measures then we will write $M(X, \Sigma)$.

2.1 Shannon Entropy and Deterministic Coding

We begin with the definition of μ -partitions, which will play a role of a coding alphabet.

Definition 2.1. Let $\mathcal{P} \subset \Sigma$. We say that \mathcal{P} is a μ -partition (of X) if \mathcal{P} is countable family of disjoint sets and

$$\mu(X \setminus \bigcup_{P \in \mathcal{P}} P) = 0. \quad (10)$$

Consequently every element $x \in X$, which can be randomly drawn (except for possibly elements of measure zero), is coded deterministically by the unique $P \in \mathcal{P}$ such that $x \in P$.

Then the entropy [1] of μ -partition is defined as follows:

Definition 2.2. Let $\mathcal{P} \subset \Sigma$ be a μ -partition of X . We define μ -entropy of \mathcal{P} by

$$h(\mu; \mathcal{P}) := \sum_{P \in \mathcal{P}} sh(\mu(P)), \quad (11)$$

where $sh : [0, 1] \rightarrow \mathbb{R}_+$ is the *Shannon function*, i.e.

$$sh(x) := \begin{cases} -x \cdot \log_2(x) & \text{for } x \in (0, 1], \\ 0 & \text{for } x = 0. \end{cases} \quad (12)$$

Let us mention that sh is a continuous, concave and subadditive function.

Classical μ -entropy is defined with use of disjoint sets, which is a very restrictive condition. It implies that we have fixed one alphabet \mathcal{P} in our lossy-coding. However, this alphabet does not have to be optimal. In other words, there may exist another \mathcal{Q} -acceptable alphabet \mathcal{P}' , which provides less entropy than \mathcal{P} (we assume that \mathcal{P} is also \mathcal{Q} -acceptable). Thus it would be better to make a coding with use of \mathcal{P}' rather than with \mathcal{P} . Therefore we will generalise the entropy for any error-control family. The error-control family can be an arbitrary family of measurable subsets of X .

We say that family \mathcal{P} is finer than \mathcal{Q} (which we write $\mathcal{P} \prec \mathcal{Q}$) if for every $P \in \mathcal{P}$ there exists $Q \in \mathcal{Q}$ such that $P \subset Q$. When \mathcal{P} is interpreted as a coding alphabet we may simply say that \mathcal{P} is \mathcal{Q} -acceptable.

Definition 2.3. Let $\mathcal{Q} \subset \Sigma$. We define the μ -entropy of \mathcal{Q} by

$$H(\mu; \mathcal{Q}) := \inf\{h(\mu; \mathcal{P}) \in [0, \infty] : \mathcal{P} \text{ is a } \mu\text{-partition and } \mathcal{P} \prec \mathcal{Q}\}. \quad (13)$$

Observe that if there is no μ -partition finer than \mathcal{Q} then directly from the definition² $H(\mu; \mathcal{Q}) = \infty$. Moreover, if \mathcal{Q} itself is a μ -partition of X then trivially³ $H(\mu; \mathcal{Q}) = h(\mu; \mathcal{Q})$. This observation implies that μ -entropy H of \mathcal{Q} is defined properly for μ -partitions as well as for families of measurable subsets of X .

2.2 Weighted Entropy and Random Coding

Motivation of the weighted entropy is the following observation. Given error-control family \mathcal{Q} in the classical approach we consider only \mathcal{Q} -acceptable deterministic codings \mathcal{P} . More precisely we always code a point $x \in X$ by the unique $P_x \in \mathcal{P}$ such that $x \in P_x$.

However, if we do not insist on being deterministic in our coding, we could alternatively encode point x by another set $P' \in \Sigma$ such that $x \in P'$ and for which there exists $Q' \in \mathcal{Q} : P' \subset Q'$. In this subsection we formalise this idea, namely we do not fix a \mathcal{Q} -acceptable alphabet \mathcal{P} but we allow any random coding demanding only that x can be encoded by $Q \in \mathcal{Q}$ iff $x \in Q$. Such a random coding might theoretically give lower entropy than the original one.

²We put $\inf(\emptyset) = \infty$.

³We can consider another μ -partition $\mathcal{P} \prec \mathcal{Q}$ of X but due to subadditivity of sh we get $h(\mu; \mathcal{Q}) \leq h(\mu; \mathcal{P})$.

We make it precise in the following way. We define the space of functions from a family of measurable subsets of X into a set of measures on X :

$$W(\mu; \mathcal{Q}) := \{\mathbf{m} : \mathcal{Q} \ni Q \rightarrow \mathbf{m}_Q \in M(X, \Sigma) : \mathbf{m}_Q(X \setminus Q) = 0 \text{ for every } Q \in \mathcal{Q} \text{ and } \sum_{Q \in \mathcal{Q}} \mathbf{m}_Q = \mu\}. \quad (14)$$

Thus given $\mathbf{m} \in W(\mu; \mathcal{Q})$ and $Q \in \mathcal{Q}$, the value of $\mathbf{m}_Q(X)$ denotes the probability that an arbitrary point $x \in X$ is coded by Q (and in that case $x \in Q$ with probability one). Observe also that every function $\mathbf{m} \in W(\mu; \mathcal{Q})$ is non-zero on at most countable sets of \mathcal{Q} .

Finally we define weighted μ -entropy of a given $\mathbf{m} \in W(\mu; \mathcal{Q})$:

Definition 2.4. Let $\mathcal{Q} \subset \Sigma$. We define the *weighted μ -entropy* of $\mathbf{m} \in W(\mu; \mathcal{Q})$ by

$$h_W(\mu; \mathbf{m}) := \sum_{Q \in \mathcal{Q}} sh(\mathbf{m}_Q(X)). \quad (15)$$

The weighted μ -entropy of \mathcal{Q} is

$$H_W(\mu; \mathcal{Q}) := \inf\{h_W(\mu; \mathbf{m}) \in [0, \infty] : \mathbf{m} \in W(\mu, \mathcal{Q})\}. \quad (16)$$

The following remark explains the importance of the formulation of weighted entropy.

Remark 2.1. Given functions $\mathbf{m}_1, \mathbf{m}_2 \in W(\mu; \mathcal{Q})$ and numbers $a_1, a_2 \in [0, 1]$ such that $a_1 + a_2 = 1$ we are allowed to perform convex combinations $a_1\mathbf{m}_1 + a_2\mathbf{m}_2$ in the space $W(\mu; \mathcal{Q})$. Therefore we can compute the weighted μ -entropy of a combination $h_W(\mu; a_1\mathbf{m}_1 + a_2\mathbf{m}_2)$ while the symbol $h(\mu; a_1\mathcal{P}_1 + a_2\mathcal{P}_2)$ does not make sense for μ -partitions $\mathcal{P}_1, \mathcal{P}_2$. This property will help us to find an estimation of entropy of convex combination of measures $H(a_1\mu_1 + a_2\mu_2; \mathcal{Q})$ for $\mathcal{Q} \subset \Sigma$.

2.3 Classical Entropy Equals Weighted

In this section we show that the classical μ -entropy of a family of measurable sets \mathcal{Q} equals to the weighted μ -entropy of \mathcal{Q} , i.e.

$$H_W(\mu; \mathcal{Q}) = H(\mu; \mathcal{Q}). \quad (17)$$

It seems natural that every deterministic coding is a particular case of a random one. We will show it in the following proposition.

Let us denote the restriction of measure μ to $A \in \Sigma$ by

$$\mu|_A(B) := \mu(A \cap B) \quad (18)$$

for every $B \in \Sigma$.

Proposition 2.1. *Random way of coding allows possibly more freedom than the deterministic one, i.e.*

$$H_W(\mu; \mathcal{Q}) \leq H(\mu; \mathcal{Q}) \quad (19)$$

for every family $\mathcal{Q} \subset \Sigma$.

Proof. Let us first observe that if there is no μ -partition finer than \mathcal{Q} then $H(\mu; \mathcal{Q}) = \infty$ and the inequality holds trivially.

Thus let \mathcal{P} be a μ -partition finer than \mathcal{Q} . As $\mathcal{P} \prec \mathcal{Q}$, for every $P \in \mathcal{P}$ there exists $Q \in \mathcal{Q}$ such that $P \subset Q$. Hence we obtain a mapping $\pi : \mathcal{P} \rightarrow \mathcal{Q}$ satisfying $P \subset \pi(P)$. We define the family

$$\mathcal{P}_{\mathcal{Q}} := \{P_Q\}_{Q \in \mathcal{Q}}, \quad (20)$$

where $P_Q := \bigcup_{P: \pi(P)=Q} P$. Let us notice that $\mathcal{P}_{\mathcal{Q}}$ is a μ -partition and $\mathcal{P} \prec \mathcal{P}_{\mathcal{Q}} \prec \mathcal{Q}$. Finally, we put $\mathbf{m} : \mathcal{Q} \ni Q \rightarrow \mu|_{P_Q} \in M(X, \Sigma)$.

Since $\mathcal{P}_{\mathcal{Q}}$ is a μ -partition and $P_Q \subset Q$ for every $Q \in \mathcal{Q}$ then

$$\sum_{Q \in \mathcal{Q}} \mathbf{m}_Q(X) = \sum_{Q \in \mathcal{Q}} \mu|_{P_Q}(Q) = \sum_{Q \in \mathcal{Q}} \mu(P_Q) = \mu(X). \quad (21)$$

Moreover, for every $Q \in \mathcal{Q}$

$$\mathbf{m}_Q(X \setminus Q) = \mu|_{P_Q}(X \setminus Q) \leq \mu|_Q(X \setminus Q) = 0. \quad (22)$$

Thus $\mathbf{m} \in W(\mu; \mathcal{Q})$. Making use of subadditivity of sh we obtain

$$h_W(\mu; \mathbf{m}) = \sum_{Q \in \mathcal{Q}} sh(\mathbf{m}_Q(X)) = \sum_{Q \in \mathcal{Q}} sh(\mu|_{P_Q}(X)) \quad (23)$$

$$= \sum_{Q \in \mathcal{Q}} sh(\mu(P_Q)) = \sum_{Q \in \mathcal{Q}} sh(\mu(\bigcup_{P: \pi(P)=Q} P)) \quad (24)$$

$$\leq \sum_{Q \in \mathcal{Q}} \sum_{P: \pi(P)=Q} sh(\mu(P)) = \sum_{P \in \mathcal{P}} sh(\mu(P)) = h(\mu; \mathcal{P}). \quad (25)$$

We conclude that $H_W(\mu; \mathcal{Q}) \leq H(\mu; \mathcal{Q})$. \square

The opposite inequality is more difficult to prove. To do this we will need an additional proposition. Given $\mathbf{m} \in W(\mu; \mathcal{Q})$ we will construct a μ -partition \mathcal{P} finer than \mathcal{Q} with not greater entropy.

Proposition 2.2. *Let $\mathcal{Q} = \{Q_i\}_{i \in I}$ be a family of measurable subsets of X , where either $I = \mathbb{N}$ or $I = \{1, \dots, N\}$ for a certain $N \in \mathbb{N}$. Let $\mathbf{m} \in W(\mu; \mathcal{Q})$. We assume that*

- $\mu(X \setminus \bigcup_{i \in I} Q_i) = 0$,
- *the sequence $I \ni i \rightarrow \mathbf{m}_{Q_i}(X)$ is nonincreasing.*

We define the family $\mathcal{P} = \{P_i\}_{i \in I} \subset \Sigma$ by the formula

$$P_1 := Q_1, P_i := Q_i \setminus \bigcup_{k=1}^{i-1} Q_k \text{ for } i \in I, i \geq 2. \quad (26)$$

Then \mathcal{P} is a μ -partition, $\mathcal{P} \prec \mathcal{Q}$ and

$$h_W(\mu; \mathbf{m}) \geq h(\mu; \mathcal{P}). \quad (27)$$

Proof. Let us observe that by the definition of \mathcal{P} , we have $\mathcal{P} \prec \mathcal{Q}$. Moreover, since $\mu(X \setminus \bigcup_{i \in I} Q_i) = 0$ and $\bigcup_{i \in I} P_i = \bigcup_{i \in I} Q_i$, we get that \mathcal{P} is a μ -partition.

To prove (27) we define sequences $(x_i)_{i \in I} \subset [0, 1]$ and $(y_i)_{i \in I} \subset [0, 1]$ by the formulas

$$x_i := \mathbf{m}_{Q_i}(X) = \mathbf{m}_{Q_i}(Q_i), y_i := \mu(P_i) \quad (28)$$

for $i \in I$. Then

$$\sum_{i \in I} x_i = \mu(X) = \sum_{i \in I} y_i. \quad (29)$$

Directly from the assumption we conclude that $(x_i)_{i \in I}$ is a nonincreasing sequence. Moreover, for every $n \in I$:

$$\sum_{i=1}^n x_i = \sum_{i=1}^n \mathbf{m}_{Q_i}(Q_i) = \left(\sum_{i=1}^n \mathbf{m}_{Q_i} \right) (Q_1 \cup \dots \cup Q_n) \quad (30)$$

$$\leq \mu(Q_1 \cup \dots \cup Q_n) = \sum_{i=1}^n \mu(P_i) = \sum_{i=1}^n y_i. \quad (31)$$

We have obtained that

$$\sum_{i=1}^n x_i \leq \sum_{i=1}^n y_i \text{ for } n \in I. \quad (32)$$

By applying the version of Hardy-Polya-Littlewood Theorem (see Appendix A for details) for sequences $(x_i)_{i \in I}$, $(y_i)_{i \in I}$ and the concave function sh we conclude that

$$h_W(\mu; \mathbf{m}) = \sum_{i \in I} sh(\mathbf{m}_{Q_i}(X)) = \sum_{i \in I} sh(x_i) \quad (33)$$

$$\geq \sum_{i \in I} sh(y_i) = \sum_{i \in I} sh(\mu(P_i)) = h(\mu; \mathcal{P}). \quad (34)$$

□

As a direct corollary we obtain that both random and deterministic coding provide the same entropy.

Theorem 2.1. *Let $\mathcal{Q} \subset \Sigma$. Then weighted entropy coincides with the classical entropy, i.e.*

$$H_W(\mu; \mathcal{Q}) = H(\mu; \mathcal{Q}). \quad (35)$$

Proof. Clearly by Proposition 2.1, we get $H_W(\mu; \mathcal{Q}) \leq H(\mu; \mathcal{Q})$.

To obtain the opposite inequality, let us first observe that if $W(\mu; \mathcal{Q}) = \emptyset$ then $H_W(\mu; \mathcal{Q}) = \infty$ and trivially $H_W(\mu; \mathcal{Q}) \geq H(\mu; \mathcal{Q})$.

We discuss the case when $W(\mu; \mathcal{Q}) \neq \emptyset$. Let $\mathbf{m} \in W(\mu; \mathcal{Q})$ be arbitrary. We define the family of measurable subsets of X by

$$\tilde{\mathcal{Q}} := \{Q \in \mathcal{Q} : \mathbf{m}_Q(X) > 0\}. \quad (36)$$

Let us notice that $\tilde{\mathcal{Q}}$ is a countable family since $\sum_{Q \in \tilde{\mathcal{Q}}} \mathbf{m}_Q(X) = 1$. Clearly,

$\tilde{\mathbf{m}} := \mathbf{m}|_{\tilde{\mathcal{Q}}} \in W(\mu; \tilde{\mathcal{Q}})$. Moreover, $\tilde{\mathcal{Q}} \prec \mathcal{Q}$ and $h_W(\mu; \tilde{\mathbf{m}}) = h_W(\mu; \mathbf{m})$.

As $\tilde{\mathcal{Q}}$ is countable, we may find a set of indices $I \subset \mathbb{N}$ such that $\tilde{\mathcal{Q}} = \{Q_i\}_{i \in I}$ and the sequence $I \ni i \rightarrow \mathbf{m}_{Q_i}(X)$ is nonincreasing. Making use of Proposition 2.2 we construct a μ -partition $\mathcal{P} \prec \tilde{\mathcal{Q}}$, which satisfies

$$h_W(\mu; \tilde{\mathbf{m}}) \geq h(\mu; \mathcal{P}) \quad (37)$$

This completes the proof since $\mathcal{P} \prec \tilde{\mathcal{Q}} \prec \mathcal{Q}$ and $h_W(\mu; \mathbf{m}) = h_W(\mu; \tilde{\mathbf{m}}) \geq h(\mu; \mathcal{P})$. □

As we proved the equality between classical and weighted entropy, we will use one notation $H(\mu; \mathcal{Q})$ to denote both classical and weighted μ -entropy of $\mathcal{Q} \subset \Sigma$.

3 Entropy of the Mixture of Sources

3.1 Estimation of the Entropy

We return to Problem 1.1. We are given two sources S_1, S_2 , which are represented by probability measures μ_1, μ_2 respectively. Suppose that we have fixed error-control family $\mathcal{Q} \subset \Sigma$, which defines the precision in the lossy-coding elements of X . Let us consider a new source S which sends a signal produced by S_1 with probability a_1 and produced by S_2 with probability a_2 . We are interested in estimation of the entropy of S (mixture of S_1 and S_2) with respect to \mathcal{Q} in terms of $H(\mu_1; \mathcal{Q})$ and $H(\mu_2; \mathcal{Q})$. In other words we would like to measure how much memory we need to reserve for information from source S providing that we know the mean amount of information needed to encode elements sent by S_1 and S_2 separately.

We will consider a general case: we assume $n \in \mathbb{N}$ sources S_1, \dots, S_n . Let us begin with a proposition.

Proposition 3.1. *Let $n \in \mathbb{N}$ and let $a_k \in (0, 1)$ for $k \in \{1, \dots, n\}$ be such that $\sum_{k=1}^n a_k = 1$. Let $\{\mu_k\}_{k=1}^n \subset M_1(X, \Sigma)$. We put $\mu := \sum_{k=1}^n a_k \mu_k \in M_1(X, \Sigma)$.*

- *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*

$$h(\mu; \mathcal{P}) \geq \sum_{k=1}^n a_k h(\mu_k; \mathcal{P}). \quad (38)$$

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

$$h_W(\mu; \mathbf{m}) \leq \sum_{k=1}^n a_k h_W(\mu_k; \mathbf{m}^k) + \sum_{k=1}^n sh(a_k). \quad (39)$$

Proof. Clearly, \mathcal{P} is a μ_k -partition for every $k \in \{1, \dots, n\}$. As a direct consequence of the concavity of the Shannon function we obtain that

$$h(\mu; \mathcal{P}) = \sum_{P \in \mathcal{P}} sh(\mu(P)) = \sum_{P \in \mathcal{P}} sh\left(\sum_{k=1}^n a_k \mu_k(P)\right) \quad (40)$$

$$\geq \sum_{P \in \mathcal{P}} \sum_{k=1}^n a_k sh(\mu_k(P)) = \sum_{k=1}^n a_k h(\mu_k; \mathcal{P}) \quad (41)$$

which proves (38).

It is easy to verify that $\mathbf{m} \in W(\mu; \mathcal{Q})$. To prove (39) we use subadditivity of the Shannon function and property: $sh(ax) = a sh(x) + x sh(a)$.

$$\begin{aligned} h_W(\mu; \mathbf{m}) &= \sum_{Q \in \mathcal{Q}} sh\left(\sum_{k=1}^n a_k \mathbf{m}_Q^k(X)\right) \leq \sum_{Q \in \mathcal{Q}} \sum_{k=1}^n sh(a_k \mathbf{m}_Q^k(X)) \quad (42) \\ &= \sum_{k=1}^n \sum_{Q \in \mathcal{Q}} [a_k sh(\mathbf{m}_Q^k(X)) + sh(a_k) \mathbf{m}_Q^k(X)] = \sum_{k=1}^n a_k h_W(\mu_k; \mathbf{m}^k) + \sum_{k=1}^n sh(a_k). \end{aligned} \quad (43)$$

□

Making use of Proposition 3.1 we can estimate the entropy of convex combination of measures, which is the main result of the paper:

Theorem 3.1. *Let $n \in \mathbb{N}$ and let $a_k \in [0, 1]$ for $k \in \{1, \dots, n\}$ be such that $\sum_{k=1}^n a_k = 1$. Let $\{\mu_k\}_{k=1}^n \subset M_1(X, \Sigma)$. If $\mathcal{Q} \subset \Sigma$ then*

$$H\left(\sum_{k=1}^n a_k \mu_k; \mathcal{Q}\right) \geq \sum_{k=1}^n a_k H(\mu_k; \mathcal{Q}) \quad (44)$$

and

$$H\left(\sum_{k=1}^n a_k \mu_k; \mathcal{Q}\right) \leq \sum_{k=1}^n a_k H(\mu_k; \mathcal{Q}) + \sum_{k=1}^n sh(a_k). \quad (45)$$

Proof. We consider the case when all considered entropies are finite because if $H(\mu_k; \mathcal{Q}) = \infty$ for a certain $k \in \{1, \dots, n\}$ then also $H(\mu; \mathcal{Q}) = \infty$ and the proof is completed. Moreover, without loss of generality, we may assume that $a_k \neq 0$ for every $k \in \{1, \dots, n\}$.

We denote $\mu := \sum_{k=1}^n a_k \mu_k$. Let $\varepsilon > 0$ be arbitrary. By the definition of entropy, we find a μ -partition \mathcal{P} finer than \mathcal{Q} such that

$$H(\mu; \mathcal{Q}) \geq h(\mu; \mathcal{P}) - \varepsilon. \quad (46)$$

Then by Proposition 3.1, we have

$$h(\mu; \mathcal{P}) = h\left(\sum_{k=1}^n a_k \mu_k; \mathcal{P}\right) \geq \sum_{k=1}^n a_k h(\mu_k; \mathcal{P}) \geq \sum_{k=1}^n a_k H(\mu_k; \mathcal{Q}). \quad (47)$$

Consequently by (46),

$$H(\mu; \mathcal{Q}) \geq h(\mu; \mathcal{P}) - \varepsilon \geq \sum_{k=1}^n a_k H(\mu_k; \mathcal{Q}) - \varepsilon. \quad (48)$$

We prove the second inequality. Again by the definition, for each $k \in \{1, \dots, n\}$ we find $\mathbf{m}^k \in W(\mu_k; \mathcal{Q})$ such that

$$h_W(\mu_k; \mathbf{m}^k) \leq H(\mu_k; \mathcal{Q}) + \frac{\varepsilon}{n}. \quad (49)$$

Then by Proposition 3.1 and (49), we obtain

$$H(\mu; \mathcal{Q}) \leq h_W\left(\mu; \sum_{k=1}^n a_k \mathbf{m}^k\right) \leq \sum_{k=1}^n [a_k h_W(\mu_k; \mathbf{m}^k) + sh(a_k)] \quad (50)$$

$$\leq \sum_{k=1}^n [a_k H(\mu_k; \mathcal{Q}) + sh(a_k)] + \varepsilon, \quad (51)$$

which completes the proof as $\varepsilon > 0$ was an arbitrary number. \square

Clearly, $\sum_{k=1}^n sh(a_k) \leq \log_2(n)$. Thus the assertion (45) of Theorem 3.1 can be also rewritten as

$$H\left(\sum_{k=1}^n a_k \mu_k; \mathcal{Q}\right) \leq \sum_{k=1}^n a_k H(\mu_k; \mathcal{Q}) + \log_2(n). \quad (52)$$

When we consider a combination of two probability measures then we get directly:

Corollary 3.1. *Let $a_1, a_2 \in (0, 1)$ be such that $a_1 + a_2 = 1$. Given probability measures μ_1, μ_2 and a family of measurable subsets \mathcal{Q} of X , we have*

$$H(a_1 \mu_1 + a_2 \mu_2; \mathcal{Q}) \geq a_1 H(\mu_1; \mathcal{Q}) + a_2 H(\mu_2; \mathcal{Q}), \quad (53)$$

$$H(a_1 \mu_1 + a_2 \mu_2; \mathcal{Q}) \leq a_1 H(\mu_1; \mathcal{Q}) + a_2 H(\mu_2; \mathcal{Q}) + 1. \quad (54)$$

3.2 Practical Algorithm for Finding “Joint” Coding Alphabet of the Mixture of Sources

A practical question is how to construct \mathcal{Q} -acceptable coding alphabet \mathcal{P} from given alphabets \mathcal{P}_1 and \mathcal{P}_2 such that

$$h(a_1\mu_1 + a_2\mu_2; \mathcal{P}) \leq a_1 h(\mu_1; \mathcal{P}_1) + a_2 h(\mu_2; \mathcal{P}_2) + sh(a_1) + sh(a_2). \quad (55)$$

For the case of simplicity we consider only the case when \mathcal{P}_1 and \mathcal{P}_2 are finite families.

Based on Propositions 2.2 and 3.1 it is not difficult to deduce the following simple, but general, greedy algorithm for constructing such an alphabet \mathcal{P} .

ALGORITHM:

1. $i = 0$;
 $\mathcal{P}^0 = \mathcal{P}_1 \cup \mathcal{P}_2$;
2. IF \mathcal{P}^i is empty GOTO STEP 4;
 ELSE find a set $\bar{P}_i \in \mathcal{P}^i$ which maximises the value of

$$\mathcal{P}^i \ni P \rightarrow a_1\mu_1(P) + a_2\mu_2(P);$$

IF maximum equals zero GOTO STEP 4;

3. $\mathcal{P}^{i+1} = \{P \setminus \bar{P}_i : P \in \mathcal{P}^i\}$;
 $i = i + 1$;
 GOTO STEP 2;
4. $\mathcal{P} = \{\bar{P}_0, \bar{P}_1, \dots, \bar{P}_{i-1}\}$;
 END.

Clearly, this algorithm can be directly adopted for more than two sources in X .

Let us look how the above algorithm works in practice.

Example 3.1. Let $X = [0, 2]$. We consider two measures $\mu_1 : [0, 1] \rightarrow \mathbb{R}$ and $\mu_2 : [\frac{1}{10}, \frac{11}{10}] \rightarrow \mathbb{R}$ given by

$$\mu_1(A) = \int_A 1 dx, \quad \mu_2(A) = 2 \int_A (x - \frac{1}{10}) dx. \quad (56)$$

As an error-control family \mathcal{Q} we take the family of all intervals contained in $[0, 2]$ with length not greater than $\frac{2}{5}$. We consider coding alphabets:

$$\begin{aligned}\mathcal{P}_1 &= \{[0, \frac{2}{5}), [\frac{2}{5}, \frac{3}{5}), [\frac{3}{5}, \frac{4}{5}), [\frac{4}{5}, 1]\}, \\ \mathcal{P}_2 &= \{[\frac{1}{10}, \frac{1}{2}), [\frac{1}{2}, \frac{7}{10}), [\frac{7}{10}, \frac{9}{10}), [\frac{9}{10}, \frac{11}{10}]\}.\end{aligned}\tag{57}$$

Mixture of sources is given by probabilities $a_1 = 2/5$ and $a_2 = 3/5$.

The algorithm presented above produces following \mathcal{Q} -acceptable alphabet of the mixture:

$$\mathcal{P} = \{[0, \frac{1}{10}), [\frac{1}{10}, \frac{1}{2}), [\frac{1}{2}, \frac{3}{5}), [\frac{3}{5}, \frac{4}{5}), [\frac{4}{5}, 1], (1, \frac{11}{10}]\}.\tag{58}$$

We get the entropies:

$$a_1 h(\mu_1; \mathcal{P}_1) + a_2 h(\mu_2; \mathcal{P}_2) \approx 1.93,\tag{59}$$

$$a_1 h(\mu_1; \mathcal{P}_1) + a_2 h(\mu_2; \mathcal{P}_2) + sh(a_1) + sh(a_2) \approx 2.9,\tag{60}$$

$$h(a_1\mu_1 + a_2\mu_2; \mathcal{P}) \approx 2.36.\tag{61}$$

As we see, we have obtained a reasonable coding method for finding joint alphabet of the mixture of sources.

4 Rényi Entropy Dimension

From now on we always assume that X is a metric space and Σ contains all Borel subsets of X .

4.1 Entropy Dimension of Convex Combination of Measures

Entropy of a probability measure μ with respect to the error-control family $\mathcal{Q} \in \Sigma$ identifies minimal amount of information needed to encode an arbitrary element of X with error \mathcal{Q} . Rényi entropy dimension in turn gives the rate of convergence of this quantity when error is decreasing. Thus it is also important to estimate the entropy dimension of convex combination of measures. Making use of Theorem 3.1 it is quite simple.

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

$$\mathcal{B}_\delta := \{B(x, \delta) : x \in X\},\tag{62}$$

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

We consider \mathcal{B}_δ as an error-control family. If we want to code a point $x \in X$ by a certain ball $B(q, \delta)$ then we may code it in fact by its centre q . Thus the error we make, simply equals to the distance between x and q . Consequently, the family \mathcal{B}_δ allows to code points from X with error not greater than δ .

For the convenience of the reader let us recall the definition of the entropy dimension [2].

Definition 4.1. The *upper and lower entropy dimension* of measure $\mu \in M_1(X, \Sigma)$ are defined by

$$\overline{\dim}_E(\mu) := \limsup_{\delta \rightarrow 0} \frac{H(\mu; \mathcal{B}_\delta)}{-\log_2(\delta)}, \quad (63)$$

$$\underline{\dim}_E(\mu) := \liminf_{\delta \rightarrow 0} \frac{H(\mu; \mathcal{B}_\delta)}{-\log_2(\delta)}. \quad (64)$$

If the above are equal we say that μ has the *entropy dimension* and denote it by $\dim_E(\mu)$.

We apply Theorem 3.1 for estimation of Rényi entropy dimension of convex combination of measures.

Theorem 4.1. Let $n \in \mathbb{N}$ and let $a_k \in [0, 1]$ for $k \in \{1, \dots, n\}$ be such that $\sum_{k=1}^n a_k = 1$. If $\{\mu_k\}_{k=1}^n \subset M_1(X, \Sigma)$ then

$$\underline{\dim}_E\left(\sum_{k=1}^n a_k \mu_k\right) \geq \sum_{k=1}^n a_k \underline{\dim}_E(\mu_k), \quad (65)$$

$$\overline{\dim}_E\left(\sum_{k=1}^n a_k \mu_k\right) \leq \sum_{k=1}^n a_k \overline{\dim}_E(\mu_k). \quad (66)$$

Proof. Let $\delta \in (0, 1)$ be given. By Theorem 3.1, we have

$$H\left(\sum_{k=1}^n a_k \mu_k; \delta\right) \geq \sum_{k=1}^n a_k H(\mu_k; \delta) \quad (67)$$

and

$$H\left(\sum_{k=1}^n a_k \mu_k; \delta\right) \leq \sum_{k=1}^n a_k H(\mu_k; \delta) + \sum_{k=1}^n sh(a_k). \quad (68)$$

Dividing by $-\log_2(\delta)$ and taking respective limits as $\delta \rightarrow 0$, we obtain assertion of the theorem. \square

Corollary 4.1. *Let $n \in \mathbb{N}$ and let $a_k \in [0, 1]$ for $k \in \{1, \dots, n\}$ be such that $\sum_{k=1}^n a_k = 1$. Let $\{\mu_k\}_{k=1}^n \subset M_1(X, \Sigma)$. If every μ_k has entropy dimension for $k \in \{1, \dots, n\}$ then $\sum_{k=1}^n a_k \mu_k$ also has entropy dimension and*

$$\dim_E\left(\sum_{k=1}^n a_k \mu_k\right) = \sum_{k=1}^n a_k \dim_E(\mu_k). \quad (69)$$

We generalise Theorem 4.1 for the case of countable families of measures under an additional assumption that the upper box dimension of X is finite. It will allow us to prove stronger version (see Corollary 4.2) of theorem proved by A. Rényi [2, page 196] concerning the entropy dimension of discrete measure. It is worth mentioning first the definition of upper box dimension [11].

The upper box dimension of any non-empty bounded subset S of X is defined by

$$\overline{\dim}_B(S) := \limsup_{\delta \rightarrow 0} \frac{\log N_\delta(S)}{-\log \delta}, \quad (70)$$

where $N_\delta(S)$ denotes the smallest number of closed balls of radius δ that cover S .

Theorem 4.2. *We assume that $\overline{\dim}_B(X) < \infty$. Let $\{\mu_k\}_{k=1}^\infty \subset M_1(X, \Sigma)$ and let a sequence $(a_k)_{k=1}^\infty \subset [0, 1]$ be such that $\sum_{k=1}^\infty a_k = 1$. Then*

$$\underline{\dim}_E\left(\sum_{k=1}^\infty a_k \mu_k\right) \geq \sum_{k=1}^\infty a_k \underline{\dim}_E(\mu_k) \quad (71)$$

and

$$\overline{\dim}_E\left(\sum_{k=1}^\infty a_k \mu_k\right) \leq \sum_{k=1}^\infty a_k \overline{\dim}_E(\mu_k). \quad (72)$$

Proof. To prove first inequality we use Theorem 4.1. For every $N \in \mathbb{N}$ we have:

$$\underline{\dim}_E\left(\sum_{k=1}^\infty a_k \mu_k\right) = \quad (73)$$

$$= \underline{\dim}_E \left(\left(\sum_{i=1}^N a_i \right) \sum_{k=1}^N \frac{a_k}{\sum_{j=1}^N a_j} \mu_k + \left(\sum_{i=N+1}^{\infty} a_i \right) \sum_{k=N+1}^{\infty} \frac{a_k}{\sum_{j=N+1}^{\infty} a_j} \mu_k \right) \quad (74)$$

$$\geq \left(\sum_{i=1}^N a_i \right) \underline{\dim}_E \left(\sum_{k=1}^N \frac{a_k}{\sum_{j=1}^N a_j} \mu_k \right) + \left(\sum_{i=N+1}^{\infty} a_i \right) \underline{\dim}_E \left(\sum_{k=N+1}^{\infty} \frac{a_k}{\sum_{j=N+1}^{\infty} a_j} \mu_k \right) \quad (75)$$

$$\geq \left(\sum_{i=1}^N a_i \right) \sum_{k=1}^N \frac{a_k}{\sum_{j=1}^N a_j} \underline{\dim}_E(\mu_k) = \sum_{k=1}^N a_k \underline{\dim}_E(\mu_k). \quad (76)$$

Since $N \in \mathbb{N}$ was arbitrary then

$$\underline{\dim}_E \left(\sum_{k=1}^{\infty} a_k \mu_k \right) \geq \sum_{k=1}^{\infty} a_k \underline{\dim}_E(\mu_k). \quad (77)$$

We prove second inequality. It is well known that if $\nu \in M_1(X, \Sigma)$ then

$$\overline{\dim}_E(\nu) \leq \overline{\dim}_B(X). \quad (78)$$

As $\overline{\dim}_B(X) < \infty$, for every $\varepsilon > 0$ we find $N \in \mathbb{N}$ such that

$$\sum_{k=N+1}^{\infty} a_k \leq \frac{\varepsilon}{\overline{\dim}_B(X)}. \quad (79)$$

Thus by Theorem 4.1, we get:

$$\overline{\dim}_E \left(\sum_{k=1}^{\infty} a_k \mu_k \right) \leq \quad (80)$$

$$\leq \left(\sum_{i=1}^N a_i \right) \overline{\dim}_E \left(\sum_{k=1}^N \frac{a_k}{\sum_{j=1}^N a_j} \mu_k \right) + \left(\sum_{i=N+1}^{\infty} a_i \right) \overline{\dim}_E \left(\sum_{k=N+1}^{\infty} \frac{a_k}{\sum_{j=N+1}^{\infty} a_j} \mu_k \right) \quad (81)$$

$$\leq \sum_{k=1}^N a_k \overline{\dim}_E(\mu_k) + \varepsilon \leq \sum_{k=1}^{\infty} a_k \overline{\dim}_E(\mu_k) + \varepsilon. \quad (82)$$

□

Given a point $x \in X$, let δ_x be an atomic measure at x , i.e.

$$\delta_x(A) := \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A \end{cases} \quad \text{for every } A \in \Sigma. \quad (83)$$

Clearly, $\dim_E(\delta_x) = 0$ for every $x \in X$. Making use of Theorem 4.2 we obtain the following corollary:

Corollary 4.2. *We assume that $\overline{\dim}_B(X) < \infty$. Let $(x_k)_{k=1}^\infty \subset X$ and let $(a_k)_{k=1}^\infty \subset [0, 1]$ be sequence such that $\sum_{k=1}^\infty a_k = 1$. Then $\dim_E(\sum_{k=1}^\infty a_k \delta_{x_k}) = 0$.*

4.2 Improved Version of Young Theorem

Finding the Rényi entropy dimension of a given measure is quite hard task in practice. It is much easier to calculate its local dimension.

The local upper dimension of $\mu \in M_1(X, \Sigma)$ at point $x \in X$, is defined by

$$\overline{D}_\mu(x) := \limsup_{\delta \rightarrow 0} \frac{\log \mu(B(x, \delta))}{\log \delta}. \quad (84)$$

Fan [10] obtained an estimation of upper entropy dimension of Borel probability measure on \mathbb{R}^N by the supremum of local upper dimension, which can be seen as a version of Young Theorem [12]:

Consequence of Young Theorem (see [10, Theorem 1.3.]) *For a Borel probability measure μ on \mathbb{R}^N , we have*

$$\overline{\dim}_E(\mu) \leq \text{ess sup } \overline{D}_\mu(x). \quad (85)$$

We show that this estimation can be improved:

Theorem 4.3. *For a Borel probability measure μ on \mathbb{R}^N , we have*

$$\overline{\dim}_E(\mu) \leq \int_{\mathbb{R}^N} \overline{D}_\mu(x) d\mu(x). \quad (86)$$

Proof. Let us first observe that $\overline{D}_\mu(x)$ is a measurable function, as the mapping $x \rightarrow \mu(B(x, \delta))$ is measurable.

Since for μ -almost all $x \in \mathbb{R}^N$: $\overline{D}_\mu(x) \leq N$ then we divide the segment $[0, N]$ into $n \in \mathbb{N}$ parts and denote sets

$$A_k^n := \{x : \overline{D}_\mu(x) \in (\frac{k-1}{n-1}N, \frac{k}{n-1}N]\} \quad (87)$$

for $n \in \mathbb{N}$ and $k \in \{0, \dots, n-1\}$. Let us define probability measures

$$\mu_i^n := \begin{cases} \frac{1}{\mu(A_i^n)} \mu|_{A_i^n} & , \text{ if } \mu(A_i^n) > 0, \\ 0 & , \text{ if } \mu(A_i^n) = 0 \end{cases} \quad (88)$$

for $n \in \mathbb{N}$ and $i \in \{0, \dots, n-1\}$. Since $A_i^n \subset X$ then

$$\overline{D}_{\mu_i^n}(x) \leq \overline{D}_\mu(x) \leq \frac{i}{n-1}N \quad (89)$$

for μ -almost all the points $x \in A_i^n$. Making use of Consequence of Young Theorem and (89), we have

$$\overline{\dim}_E(\mu_i^n) \leq \text{ess sup} \overline{D}_{\mu_i^n}(x) \leq \frac{i}{n-1}N. \quad (90)$$

By the definition of μ_i^n , we represent measure μ as a convex combination of μ_i^n , i.e.

$$\mu = \sum_{i=0}^{n-1} \mu(A_i^n) \mu_i^n \quad (91)$$

for each $n \in \mathbb{N}$. Applying Theorem 4.1 and (90), we get

$$\overline{\dim}_E(\mu) = \overline{\dim}_E\left(\sum_{i=0}^{n-1} \mu(A_i^n) \mu_i^n\right) \quad (92)$$

$$\leq \sum_{i=0}^{n-1} \mu(A_i^n) \overline{\dim}_E(\mu_i^n) \leq \sum_{i=0}^{n-1} \mu(A_i^n) \frac{i}{n-1}N. \quad (93)$$

Finally taking limits as $n \rightarrow \infty$, we obtain

$$\overline{\dim}_E(\mu) \leq \int_{\mathbb{R}^N} \overline{D}_\mu(x) d\mu(x). \quad (94)$$

□

We were unable to verify whether a similar estimation holds for the lower entropy dimension, i.e. if $\int_{\mathbb{R}^N} \underline{D}_\mu(x) d\mu(x) \leq \underline{\dim}_E(\mu)$.

5 Conclusion

Our paper investigates the problem of joint lossy-coding of information from combined sources. The main result gives the estimation of the entropy of mixture of sources by the combination of their entropies. The proof is based on the new equivalent definition of the entropy, which allows to obtain a convex combination of partitions contrary to the classical definition. We also present a practical and easy to implement algorithm of constructing joint coding alphabet for above problem. As a corollary we generalise some results concerning the Rényi entropy dimension.

6 Appendix A: Hardy-Polya-Littlewood Theorem

We generalise the classical Hardy-Littlewood-Polya Theorem [13, Theorem 1.5.4.] for infinite sequences.

Hardy-Littlewood-Polya Theorem. *Let $a > 0$ and let $\varphi : [0, a] \rightarrow \mathbb{R}_+$, $\varphi(0) = 0$ be a continuous concave function. Let $(x_i)_{i \in I}, (y_i)_{i \in I} \subset [0, a]$ be given sequences where either $I = \mathbb{N}$ or $I = \{1, \dots, N\}$ for a certain $N \in \mathbb{N}$. We assume that*

$$\sum_{i=1}^n x_i \leq \sum_{i=1}^n y_n \text{ for } n \in I \quad (95)$$

and

$$\sum_{i \in I} x_i = \sum_{i \in I} y_i. \quad (96)$$

If $(x_i)_{i \in I}$ is nonincreasing sequence then

$$\sum_{i \in I} \varphi(x_i) \geq \sum_{i \in I} \varphi(y_j). \quad (97)$$

Proof. The classical Hardy-Littlewood-Polya Theorem [13, Theorem 1.5.4] covers exactly the finite sequence case, that is when $I = \{1, \dots, N\}$ for a certain $N \in \mathbb{N}$. We will show that the case when $I = \mathbb{N}$ follows from the case when I is finite.

To prove (97) it is sufficient to show that for every $n \in \mathbb{N}$ there exist $k_n \in \mathbb{N}$ such that

$$\sum_{i=1}^{k_n} \varphi(x_i) \geq \sum_{i=1}^n \varphi(y_i), \quad (98)$$

since all sequences under considerations are nonnegative. Let $n \in \mathbb{N}$ be arbitrary and let $k_n > n$ be chosen so that

$$r_{n+1} := \sum_{i=1}^{k_n} x_i - \sum_{i=1}^n y_i \geq 0. \quad (99)$$

Such a choice is possible since $(x_i)_{i \in I}$ and $(y_i)_{i \in I}$ are nonnegative sequences which have equal sum.

Consider two finite sequences of equal length k_n :

$$\tilde{x} = (x_1, \dots, x_{k_n}) \text{ and } \tilde{y} = (y_1, \dots, y_n, r_{n+1}, 0, \dots, 0).$$

Observe that the above sequences have equal sum and that \tilde{x} is nonincreasing. We show that for every $k \leq k_n$

$$\sum_{i=1}^k \tilde{x}_i \leq \sum_{i=1}^k \tilde{y}_i. \quad (100)$$

If $k \leq n$, this follows from the assumptions made on sequences $(x_i)_{i \in I}$ and $(y_i)_{i \in I}$. If $k > n$ then

$$\sum_{i=1}^k \tilde{x}_i \leq \sum_{i=1}^{k_n} \tilde{x}_i = \sum_{i=1}^{k_n} x_i = \sum_{i=1}^n y_i + r_{n+1} = \sum_{i=1}^k \tilde{y}_i. \quad (101)$$

Since $(x_i)_{i \in I}$ is a nonincreasing we can apply to sequences \tilde{x} , \tilde{y} and function φ the finite sequence version of the classical Hardy-Littlewood-Polya and obtain that

$$\sum_{i=1}^{k_n} \varphi(x_i) = \sum_{i=1}^{k_n} \varphi(\tilde{x}_i) \geq \sum_{i=1}^{k_n} \varphi(\tilde{y}_i) \quad (102)$$

$$= \sum_{i=1}^n \varphi(y_i) + \varphi(r_{n+1}) + (k_n - (n + 1))\varphi(0) \geq \sum_{i=1}^n \varphi(y_i). \quad (103)$$

□

References

- [1] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.
- [2] A. Rényi, “On the dimension and entropy of probability distributions,” *Acta Mathematica Hungarica*, vol. 10, no. 1–2, pp. 193–215, 1959.
- [3] R. S. Ellis, *Entropy, large deviations, and statistical mechanics*, 1st ed. Springer, 1985.
- [4] R. G. Gray, *Entropy and Information Theory*, 2nd ed. Springer, 2011.

- [5] P. Seibt, *Algorithmic Information Theory*. Berlin Heidelberg: Springer-Verlag, 2006.
- [6] Y. Wu and S. Verdú, “Rényi information dimension: Fundamental limits of almost lossless analog compression,” *IEEE Transactions On Information Theory*, vol. 56, no. 8, pp. 3721–3748, 2010.
- [7] J. D. Howroyd, “On dimension and on existence of sets of finite positive hausdorff measure,” *Proc. London Math. Soc.*, vol. 70, no. 3, pp. 581–604, 1995.
- [8] ———, *On the theory of Hausdorff measure in metric space*. London: Ph.D. Thesis, University Collage, 1994.
- [9] C. A. Rogers, *Hausdorff measures*, 2nd ed. Cambridge University Press, 1998.
- [10] A. Fan, K. Lau, and H. Rao, “Relationships between different dimensions of a measure,” *Monatsh. Math.*, vol. 135, pp. 191–201, 2002.
- [11] K. Falconer, *Fractal Geometry - Mathematical Foundations and Applications*, 2nd ed. John Wiley, 2003.
- [12] L. Young, “Dimension, entropy and lyapunov exponents,” *Ergod. Th. & Dynam. Sys.*, vol. 2, pp. 109–124, 1982.
- [13] C. P. Niculescu and L. Persson, *Convex Functions and Their Applications - A Contemporary Approach*. Springer, 2006.