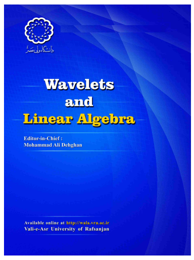
Wavelets and Linear Algebra 7(2) (2020) 1-9

Vali-e-Asr University of Rafsanjan

Wavelets and Linear Algebra

http://wala.vru.ac.ir



New Bounds for Entropy of Information Sources

Yamin Sayyari[a]

[a]*Department of Mathematics, Sirjan University Of Technology, Sirjan, Islamic Republic of Iran.*

Article Info

Article history:
Received 24 July 2019
Accepted 8 July 2020
Available online 6 November 2020
Communicated by Abbas Salemi

Keywords:
Entropy, Shannon's entropy, information source, convex function, random variable.

2000 MSC:
37B40
94A17
26A51.

Abstract

Shannon's entropy plays an important role in information theory, dynamical systems and thermodynamics. In this paper we applying Jensen's inequality in information theory and we obtain some results for the Shannon's entropy of random variables and Shannon's entropy of stochastic process. Also we obtain upper bound and lower bound for Shannon's entropy of information sources.

1. Introduction and Preliminaries

The concept of entropy is one of the most important mathematical concepts used in statistics, thermodynamics, code theory, information theory, physics and dynamical systems. The entropy

*Corresponding author
Email address: ysayyari@gmail.com (Yamin Sayyari)

http://doi.org/10.22072/wala.2020.111881.1240

actually measures the degree of irregularities of a dynamic system, and researchers have done so much to calculate this concept, which is often successful [1, 3], but numerical calculations of entropy are still difficult. In [6, 5], the authors presented a strong upper bound for the classical Shannon entropy. In [4], the authors presented the algebraic and Shannon entropies for hypergroupoids and commutative hypergroups, respectively, and studies their fundamental properties. In this paper, we tried to find a high bound and a lower bound for a stochastic process using convex functions for Shannon entropy. Entropy is one of the interesting and applicable concepts in many fields including mechanics, physics, statistics, dynamical systems and measure theory. The concept of entropy has evolved along different ways: Tsallis entropy, metric entropy, topological entropy, sequence entropy, directional entropy, permutation entropy, Rnyi entropy, epsilontau entropy, etc. We call three $(\Omega, \mathcal{F}, \mu)$ a measure probability space if Ω is a non-empty set, $\mathcal{F}$ is an σ-algebra of subsets of Ω, μ is a measure on Ω, and $\mu(\Omega) = 1$. In this paper $(\Omega, \mathcal{F}, \mu)$ is a probability space.

Definition 1.1. [7] Let $(\Omega, \mathcal{F}, \mu)$ be a measure probability space. A finite set of measurable sets $\alpha = \{A_1, \ldots, A_n\}$ is called a finite partition if the following properties are fulfilled:

$$\bigcup_{i=1}^{n} A_i = \Omega, \text{ and } A_i \cap A_j = \emptyset \text{ for every } i, j (1 \leq i \neq j \leq n).$$

Definition 1.2. [7] Let $\alpha = \{A_1, \ldots, A_n\}$ be a finite partition of the measure probability space $(\Omega, \mathcal{F}, \mu)$. The entropy of α is defined by

$$H_\mu(\alpha) := -\sum_{i=1}^{n} \mu(A_i) \log(\mu(A_i)).$$

Definition 1.3. [2] We assume that S is a random variable on Ω with discrete finite state space $A = \{a_1, ..., a_N\}$. In this case we can define $p : A \to [0, 1]$ by $p(s) = \mu\{\omega \in \Omega : S(\omega) = s\}$. The Shannon's entropy of S is defined by

$$H_\mu(S) := -\sum_{s \in A,\ p(s) \neq 0} p(s) \log p(s).$$

Definition 1.4. [1, 2] A finite space stochastic process or an information source **S** is a sequence $(S_n)_{n=1}^{\infty}$ of the random variables $S_n : \Omega \longrightarrow A$, where $n \in \mathbb{N}$. For given $L \geq 1$ we define a mapping $p : A^L \to [0, 1]$ by

$$p(s_1, ..., s_L) = p(s_1^L) = \mu\{\omega \in \Omega : S_1(\omega) = s_1, ..., S_L(\omega) = s_L\}.$$

The Shannon entropy of order L and the Shannon entropy of source **S** are respectively defined by

$$H_\mu(S_1^L) = -\frac{1}{L} \sum_{s_1^L \in A^L} p(s_1, ..., s_L) \log p(s_1, ..., s_L), \text{ and } h_\mu(\mathbf{S}) = \lim_{L \to \infty} H_\mu(S_1^L),$$

where the the summation is taken over the collection $\{s_1^L \in A^L : p(s_1^L) \neq 0\}$.

In this paper we use the symbol s_1^L instead of notation $(s_1, ..., s_L)$ and Let $p(s_1^L) \neq 0$ for every $L \in \mathbb{N}$.

2. main results

In this section we applying Jensen's inequality in information theory and obtain upper and lower bounds for Shannon's entropy of information sources.

Theorem 2.1. *Let $I = [a,b]$ be an interval, $H : A^L \longrightarrow I$ be a function, and $f : I \longrightarrow \mathbb{R}$ be a convex function, then*

$$\begin{aligned}&\sum_{s_1^L\in A^L} p(s_1^L)f(H(s_1^L)) - f(\sum_{s_1^L\in A^L} p(s_1^L)H(s_1^L))\\ &\geq \max\{p(r_1^L)f(H(r_1^L)) + p(t_1^L)f(H(t_1^L))\\ &-(p(r_1^L)+p(t_1^L))f(\frac{p(r_1^L)H(r_1^L)+p(t_1^L)H(t_1^L)}{p(r_1^L)+p(t_1^L)})\},\end{aligned} \tag{2.1}$$

where the maximum is taken over all $r_1^L \neq t_1^L \in A^L$.

Proof. Choose arbitrary $t_1^L, r_1^L \in A^L$. So,

$$\begin{aligned}&f(\sum_{s_1^L\in A^L} p(s_1^L)H(s_1^L)) = f(\sum_{s_1^L\neq r_1^L, t_1^L\in A^L} p(s_1^L)H(s_1^L))\\ &+(p(r_1^L)+p(t_1^L))(\frac{p(r_1^L)H(r_1^L)+p(t_1^L)H(t_1^L)}{p(r_1^L)+p(t_1^L)})\\ &\leq \sum_{s_1^L\neq r_1^L, t_1^L\in A^L} p(s_1^L)f(H(s_1^L)) + (p(r_1^L)+p(t_1^L))f(\frac{p(r_1^L)H(r_1^L)+p(t_1^L)H(t_1^L)}{p(r_1^L)+p(t_1^L)}).\end{aligned}$$

Hence,

$$\begin{aligned}&\sum_{s_1^L\in A^L} p(s_1^L)f(H(s_1^L)) - f(\sum_{s_1^L\in A^L} p(s_1^L)H(s_1^L))\\ &\geq p(r_1^L)f(H(r_1^L)) + p(t_1^L)f(H(t_1^L))\\ &-(p(r_1^L)+p(t_1^L))f(\frac{p(r_1^L)H(r_1^L)+p(t_1^L)H(t_1^L)}{p(r_1^L)+p(t_1^L)}),\end{aligned}$$

Since $s_1^L, t_1^L \in A^L$ are arbitrary, the proof is completed. □

Lemma 2.2. *[5] Let $f : [a,b] \longrightarrow \mathbb{R}$ be a convex function, $0 \leq p, q \leq 1$ and $p+q = 1$, then*

$$pf(a) + qf(b) - f(pa+qb) \leq f(a) + f(b) - 2f(\frac{a+b}{2}).$$

Theorem 2.3. *Let $I = [a,b]$ be an interval, $H : A^L \longrightarrow I$ be a function, and $f : I \longrightarrow \mathbb{R}$ be a convex function, then*

$$\sum_{s_1^L\in A^L} p(s_1^L)f(H(s_1^L)) - f(\sum_{s_1^L\in A^L} p(s_1^L)H(s_1^L)) \leq f(a) + f(b) - 2f(\frac{a+b}{2}).$$

Proof. Since $H(s_1^L) \in [a, b]$, so there is some $\lambda(s_1^L) \in [0, 1]$ such that

$$H(s_1^L) = \lambda(s_1^L)a + (1 - \lambda(s_1^L))b.$$

Thus,

$$\begin{aligned}
&\sum_{s_1^L \in A^L} p(s_1^L) f(H(s_1^L)) - f(\sum_{s_1^L \in A^L} p(s_1^L) H(s_1^L)) \\
&= \sum_{s_1^L \in A^L} p(s_1^L) f(\lambda(s_1^L)a + (1 - \lambda(s_1^L))b) \\
&- f(\sum_{s_1^L \in A^L} p(s_1^L)(\lambda(s_1^L)a + (1 - \lambda(s_1^L))b)) \\
&\leq \sum_{s_1^L \in A^L} p(s_1^L)(\lambda(s_1^L) f(a) + (1 - \lambda(s_1^L)) f(b)) \\
&- f(a \sum_{s_1^L \in A^L} p(s_1^L)\lambda(s_1^L) + b \sum_{s_1^L \in A^L} p(s_1^L)(1 - \lambda(s_1^L)) \\
&= f(a) \sum_{s_1^L \in A^L} p(s_1^L)\lambda(s_1^L) + f(b)(1 - \sum_{s_1^L \in A^L} p(s_1^L)\lambda(s_1^L)) \\
&- f(a \sum_{s_1^L \in A^L} p(s_1^L)\lambda(s_1^L) + b \sum_{s_1^L \in A^L} p(s_1^L)(1 - \lambda(s_1^L)).
\end{aligned}$$

Suppose that $p_0 = \sum_{s_1^L \in A^L} p(s_1^L)\lambda(s_1^L)$ and $q_0 = 1 - p_0$, we have

$$\begin{aligned}
&\sum_{s_1^L \in A^L} p(s_1^L) f(H(s_1^L)) - f(\sum_{s_1^L \in A^L} p(s_1^L) H(s_1^L)) \\
&\leq p_0 f(a) + q_0 f(b) - f(p_0 a + q_0 b).
\end{aligned}$$

On the other hand, by the use of Lemma 2.2 have

$$p_0 f(a) + q_0 f(b) - f(p_0 a + q_0 b) \leq f(a) + f(b) - 2f(\frac{a+b}{2}).$$

So,

$$\sum_{s_1^L \in A^L} p(s_1^L) f(H(s_1^L)) - f(\sum_{s_1^L \in A^L} p(s_1^L) H(s_1^L)) \leq f(a) + f(b) - 2f(\frac{a+b}{2}).$$

□

Theorem 2.4. *Let* $A = \{a_1, ..., a_N\}$ *be a finite alphabet,* $H : A^L \longrightarrow I$ *be a function, and* $f : I \longrightarrow \mathbb{R}$ *be a convex function. If* $\mu_L = \min_{s_1^L \in A^L}\{H(s_1^L)\}$, $\nu_L = \max_{s_1^L \in A^L}\{H(s_1^L)\}$ *and* $[\mu_L, \nu_L] \subset I$*, then*

$$\begin{aligned}&\frac{f(\mu_L)}{N^L} + \frac{f(\nu_L)}{N^L} - \frac{2}{N^L} f(\frac{\mu_L + \nu_L}{2})\\ &\leq \frac{\sum_{s_1^L \in A^L} f(H(s_1^L))}{N^L} - f(\frac{\sum_{s_1^L \in A^L} H(s_1^L)}{N^L})\\ &\leq f(\mu_L) + f(\nu_L) - 2f(\frac{\mu_L + \nu_L}{2}).\end{aligned}$$

Proof. By Theorem 2.1 with $p(s_1^L) = \frac{1}{N^L}$, $\mu_L = \min_{s_1^L \in A^L}\{H(s_1^L)\}$ and $\nu_L = \max_{s_1^L \in A^L}\{H(s_1^L)\}$ have

$$\begin{aligned}&\frac{\sum_{s_1^L \in A^L} f(H(s_1^L))}{N^L} - f(\frac{\sum_{s_1^L \in A^L} H(s_1^L)}{N^L})\\ &\geq \frac{f(\mu_L)}{N^L} + \frac{f(\nu_L)}{N^L} - \frac{2}{N^L} f(\frac{\mu_L + \nu_L}{2}).\end{aligned} \tag{2.2}$$

Also Theorem 2.3 with $p(s_1^L) = \frac{1}{N^L}$, $b = \nu_L = \max_{s_1^L \in A^L}\{H(s_1^L)\}$ and $a = \mu_L = \min_{s_1^L \in A^L}\{H(s_1^L)\}$ yields

$$\begin{aligned}&\frac{\sum_{s_1^L \in A^L} f(H(s_1^L))}{N^L} - f(\frac{\sum_{s_1^L \in A^L} H(s_1^L)}{N^L})\\ &\leq f(\mu_L) + f(\nu_L) - 2f(\frac{\mu_L + \nu_L}{2}).\end{aligned} \tag{2.3}$$

□

Lemma 2.5. *Define* $\mu_L = \min_{s_1^L \in A^L}\{p(s_1^L)\}$ *and* $\nu_L = \max_{s_1^L \in A^L}\{p(s_1^L)\}$*. Then*

$$\mu_L \log(\sqrt[L]{\frac{2\mu_L}{\mu_L + \nu_L}}) + \nu_L \log(\sqrt[L]{\frac{2\nu_L}{\mu_L + \nu_L}}) \leq \log(N) - H_\mu(S_1^L).$$

Proof. If $f(x) = \log(\frac{1}{x})$ and $H(s_1^L) = \frac{1}{p(s_1^L)}$, then

$$\begin{aligned}&\sum_{s_1^L \in A^L} p(s_1^L) f(H(s_1^L)) - f(\sum_{s_1^L \in A^L} p(s_1^L) H(s_1^L))\\ &= -\sum_{s_1^L \in A^L} p(s_1^L) \log(\frac{1}{p(s_1^L)}) + \log(\sum_{s_1^L \in A^L} 1)\\ &= -LH_\mu(S_1^L) + \log(N^L) = L\log(N) - LH_\mu(S_1^L).\end{aligned}$$

With the use of Theorem 2.1 we obtain

$$\begin{aligned}&p(r_1^L)\log(p(r_1^L)) + p(t_1^L)\log(p(t_1^L)) + (p(r_1^L) + p(t_1^L))\log(\frac{1+1}{p(r_1^L) + p(t_1^L)})\\ &= p(r_1^L)\log(\frac{2p(r_1^L)}{ptr_1^L) + p(r_1^L)}) + p(t_1^L)\log(\frac{2p(t_1^L)}{p(t_1^L) + p(r_1^L)})\\ &\leq L\log(N) - LH_\mu(S_1^L).\end{aligned}$$

Select $\mu_L = \min_{s_1^L \in A^L}\{p(s_1^L)\}$ and $\nu_L = \max_{s_1^L \in A^L}\{p(s_1^L)\}$, have

$$L\log(N) - LH_\mu(S_1^L) \geq \mu_L \log(\frac{2\mu_L}{\mu_L + \nu_L}) + \nu_L \log(\frac{2\nu_L}{\nu_L + \mu_L}).$$

So,

$$\log(N) - H_\mu(S_1^L) \geq \mu_L \log(\sqrt[L]{\frac{2\mu_L}{\mu_L + \nu_L}}) + \nu_L \log\sqrt[L]{(\frac{2\nu_L}{\nu_L + \mu_L}}).$$

□

Lemma 2.6. *Suppose that* $\mu_L = \min_{s_1^L \in A^L}\{p(s_1^L)\}$ *and* $\nu_L = \max_{s_1^L \in A^L}\{p(s_1^L)\}$*, then*

$$\log(N) - H_\mu(S_1^L) \leq \log(\sqrt[L]{\frac{(\mu_L + \nu_L)^2}{4\mu_L\nu_L}}).$$

Proof. By $f(x) = \log(\frac{1}{x})$ and $H(s_1^L) = \frac{1}{p(s_!^L)}$ and use of Theorem 2.3 have

$$L\log(N) - LH_\mu(S_1^L) \leq \log(\frac{1}{\mu_L}) + \log(\frac{1}{\nu_L}) - 2\log(\frac{2}{\mu_L + \nu_L}).$$

Thus,

$$\begin{aligned}\log(N) - H_\mu(S_1^L) &\leq \frac{1}{L}(\log(\frac{1}{\mu_L}) + \log(\frac{1}{\nu_L}) - 2\log(\frac{2}{\mu_L + \nu_L})) \\ &= \log(\sqrt[L]{\frac{(\mu_L + \nu_L)^2}{4\mu_L\nu_L}}).\end{aligned}$$

□

Remark 2.7. Let $A = \{a_1, \ldots, a_N\}$ be a finite alphabet and $\mathbf{S} = (S_n)_{n=1}^\infty$ be an information source, then

$$\begin{aligned}m(\mu_L, \nu_L) &:= \mu_L \log(\sqrt[L]{\frac{2\mu_L}{\mu_L + \nu_L}}) + \nu_L \log(\sqrt[L]{\frac{2\nu_L}{\mu_L + \nu_L}}) \\ &\leq \log(N) - H_\mu(S_1^L) \leq \log(\sqrt[L]{\frac{(\mu_L + \nu_L)^2}{4\mu_L\nu_L}}) := M(\mu_L, \nu_L).\end{aligned}$$

Example 2.8. If $S = \{1, ..., N\}$, and the probability distribution S is given by

$$p(S = i) = p_i, \quad p_i > 0, \ \forall i = 1, ..., N, \quad \sum_{i=1}^N p_i = 1,$$

then $\mu_1 \log(\frac{2\mu_1}{\mu_1+\nu_1}) + \nu_1 \log(\frac{2\nu_1}{\mu_1+\nu_1}) \leq \log(N) - H(S) \leq \log(\frac{(\mu_1+\nu_1)^2}{4\mu_1\nu_1})$, where $H(S) = -\sum_{i=1}^N p_i \log p_i$, $\mu_1 = \min_{1\leq i\leq N}\{p_i\}$ and $\nu_1 = \max_{1\leq i\leq N}\{p_i\}$.

Remark 2.9. Let $A = \{a_1, \dots, a_N\}$ be a finite alphabet and $\mathbf{S} = (S_n)_{n=1}^{\infty}$ be an information source, then

$$0 \leq \log(N) - H_\mu(S_1^L) \leq \frac{(\mu_L - \nu_L)^2}{4L\mu_L\nu_L} := D(\mu_L, \nu_L).$$

Proof. For each $x > 0$, $\log(\sqrt[L]{1+x}) \leq \frac{x}{L}$. By the use of Remark 2.7 and $x = \frac{(\mu_L-\nu_L)^2}{4\mu_L\nu_L}$, have $M(\mu_L, \nu_L) \leq D(\mu_L, \nu_L)$. □

Lemma 2.10. *Under the notation of Theorem 2.4, have*

$$m(\mu_L, \nu_L) \leq \log(N) - H_\mu(S_1^L) \leq N^L m(\mu_L, \nu_L).$$

Proof. Put in Theorem 2.4 $f(x) = x\log(x)$ and $H(s_1^L) = p(s_1^L)$. Thus,

$$\begin{aligned}
&\frac{\mu_L \log(\mu_L)}{N^L} + \frac{\nu_L \log(\nu_L)}{N^L} - (\frac{\mu_L + \nu_L}{N^L}) \log(\frac{\mu_L + \nu_L}{2}) \\
&\leq \frac{\sum_{s_1^L \in A^L} p(s_1^L) \log(p(s_1^L))}{N^L} - \frac{1}{N^L} \log(\frac{1}{N^L}) \\
&\leq \mu_L \log(\mu_L) + \nu_L \log(\nu_L) - (\mu_L + \nu_L) \log(\frac{\mu_L + \nu_L}{2}).
\end{aligned}$$

Hence,

$$\begin{aligned}
\mu_L \log(\frac{2\mu_L}{\mu_L + \nu_L}) + \nu_L \log(\frac{2\nu_L}{\nu_L + \mu_L}) &\leq L\log(N) - LH_\mu(S_1^L) \\
&\leq N^L(\mu_L \log(\frac{2\mu_L}{\mu_L + \nu_L}) + \nu_L \log(\frac{2\nu_L}{\nu_L + \mu_L})),
\end{aligned}$$

which completes the proof. □

Example 2.11. [2] Let $\alpha_0 = [0, \frac{1}{2}]$ and $\alpha_1 = (\frac{1}{2}, 1]$. We consider the tent map (see Figure 1) $\Lambda : [0, 1] \longrightarrow [0, 1]$ by

$$\Lambda(x) = \begin{cases} 2x & 0 \leq x \leq \frac{1}{2}, \\ 2 - 2x & \frac{1}{2} \leq x \leq 1, \end{cases}$$

and we define $S_L : [0, 1] \longrightarrow \{0, 1\}$ by

$$S_L(x) = \begin{cases} 0 & \text{if } \Lambda^L(x) \in \alpha_0, \\ 1 & \text{if } \Lambda^L(x) \in \alpha_1, \end{cases}$$

for every $L \geq 1$ and $\mathbf{S} = \{S_n\}_{n=1}^{\infty}$. Let

$$\begin{aligned}
\alpha_{00} = [0, \frac{1}{4}] = \{x \in \alpha_0 : \Lambda(x) \in \alpha_0\}, \alpha_{01} = (\frac{1}{4}, \frac{1}{2}] = \{x \in \alpha_0 : \Lambda(x) \in \alpha_1\}, \\
\alpha_{10} = [\frac{3}{4}, 1] = \{x \in \alpha_1 : \Lambda(x) \in \alpha_0\}, \alpha_{11} = (\frac{1}{2}, \frac{3}{4}) = \{x \in \alpha_1 : \Lambda(x) \in \alpha_1\}.
\end{aligned}$$

Also let $\alpha_{i_1\ldots i_L}$ is defined, we define $\alpha_{i_1\ldots i_L j}$ by

$$\alpha_{i_1\ldots i_L 0} = \alpha_{i_1\ldots i_L} \cap \{x \in [0,1] : \Lambda^L(x) \in \alpha_0\},$$
$$\alpha_{i_1\ldots i_L 1} = \alpha_{i_1\ldots i_L} \cap \{x \in [0,1] : \Lambda^L(x) \in \alpha_1\}.$$

Therefore

$$\alpha_{000} = [0, \frac{1}{8}],\ \alpha_{001} = (\frac{1}{8}, \frac{1}{4}],\ \alpha_{010} = [\frac{3}{8}, \frac{1}{2}],\ \alpha_{011} = (\frac{1}{4}, \frac{3}{8}),$$
$$\alpha_{100} = [\frac{7}{8}, 1],\ \alpha_{101} = [\frac{3}{4}, \frac{7}{8}), \alpha_{110} = (\frac{1}{2}, \frac{5}{8}],\ \alpha_{111} = (\frac{5}{8}, \frac{3}{4}),$$
$$\alpha_{0000} = [0, \frac{1}{16}],\ \alpha_{0001} = (\frac{1}{16}, \frac{1}{8}],\ \alpha_{0010} = [\frac{3}{16}, \frac{1}{4}],\ \alpha_{0011} = (\frac{1}{8}, \frac{3}{16}),$$
$$\alpha_{0100} = [\frac{7}{16}, \frac{1}{2}],\ \alpha_{0101} = [\frac{3}{8}, \frac{7}{16}),\ \alpha_{0110} = (\frac{1}{4}, \frac{5}{16}],\ \alpha_{0111} = (\frac{5}{16}, \frac{3}{8}),$$
$$\alpha_{1000} = [\frac{15}{16}, 1],\ \alpha_{1000} = [\frac{7}{8}, \frac{15}{16}),\ \alpha_{1010} = [\frac{3}{4}, \frac{13}{16}],\ \alpha_{1011} = (\frac{13}{16}, \frac{7}{8}),$$
$$\alpha_{1100} = (\frac{1}{2}, \frac{9}{16}],\ \alpha_{1101} = (\frac{9}{16}, \frac{5}{8}], \alpha_{1110} = [\frac{11}{16}, \frac{3}{4}),\ \alpha_{1111} = (\frac{5}{8}, \frac{3}{4}).$$

The sets $\{\alpha_{s_1\ldots s_L}\}$ are identical to the binary sequences of 0 , 1 in length L. If μ denotes the Lebesgue

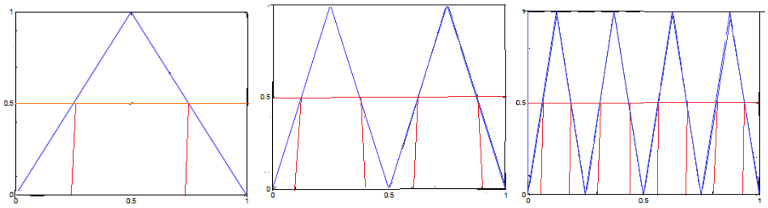


Figure 1: Λ, Λ^2 and Λ^3

measure, then $\mu(\alpha_{s_1,\ldots,s_L}) = p(s_1^L)$ and $\mu(\alpha_{s_1\ldots s_L}) = \frac{1}{2^L}$, thus $p(s_1^L) = \frac{1}{2^L}$ for each $\{s_1, \ldots, s_L\} \subseteq \{0, 1\}$ and $L \in \mathbb{N}$. So $\mu_L = \nu_L = \frac{1}{2^L}$, for every natural number L. Hence,

$$m(\mu_L, \nu_L) = \mu_L \log(\sqrt[L]{\frac{2\mu_L}{\mu_L + \nu_L}}) + \nu_L \log(\sqrt[L]{\frac{2\nu_L}{\mu_L + \nu_L}}) = 0,$$

for every $L \in \mathbb{N}$, and by the use of Lemma 2.10 we obtain $h_\mu(\mathbf{S}) = \log 2$.

Remark 2.12. Under the notation of Lemma 2.7,

$$m(\mu_L, \nu_L) \le \log(N) - H_\mu(S_1^L) \le \min\{N^L m(\mu_L, \nu_L), M(\mu_L, \nu_L)\}.$$

Proof. From the Remark 2.7 and Lemma 2.10, it is easy to see. □

Theorem 2.13. *Let* $\mathbf{S} = (S_k)_{k=1}^{\infty}$ *be a stochastic process on the measure probability space* $(\Omega, \mathcal{F}, \mu)$ *with the alphabet A, also*

$$\mu_L = \min_{s_1^L \in A^L}\{p(s_1^L)\}, \quad \nu_L = \max_{s_1^L \in A^L}\{p(s_1^L)\} \ \textit{and}$$

$$m(\mu_L, \nu_L) := \mu_L \log(\sqrt[L]{\frac{2\mu_L}{\mu_L + \nu_L}}) + \nu_L \log(\sqrt[L]{\frac{2\nu_L}{\mu_L + \nu_L}})$$

then,

$$\log(N) - \liminf_{L\to\infty} N^L m(\mu_L, \nu_L) \leq h_\mu(\mathbf{S}) \leq \log(N) - \limsup_{L\to\infty} m(\mu_L, \nu_L).$$

Proof. Since, $\log(N) - N^L m(\mu_L, \nu_L) \leq H_\mu(S_1^L) \leq \log(N) - m(\mu_L, \nu_L)$ for every $L \geq 1$, $\log(N) - \liminf_{L\to\infty} N^L m(\mu_L, \nu_L) \leq h_\mu(\mathbf{S}) \leq \log(N) - \limsup_{L\to\infty} m(\mu_L, \nu_L)$. □

Theorem 2.14. *Under the notation of Theorem 2.13 and*

$$M(\mu_L, \nu_L) := \log(\sqrt[L]{\frac{(\mu_L + \nu_L)^2}{4\mu_L \nu_L}}),$$

we have

$$\log(N) - \liminf_{L\to\infty}(M(\mu_L, \nu_L)) \leq h_\mu(\mathbf{S}) \leq \log(N) - \limsup_{L\to\infty}(m(\mu_L, \nu_L)).$$

Proof. By the use of Remark 2.7, we have

$$\log(N) - M(\mu_L, \nu_L) \leq H_\mu(S_1^L) \leq \log(N) - m(\mu_L, \nu_L)$$

for every $L \geq 1$. Hence,

$$\log(N) - \liminf_{L\to\infty}(M(\mu_L, \nu_L)) \leq h_\mu(\mathbf{S}) \leq \log(N) - \limsup_{L\to\infty}(m(\mu_L, \nu_L)).$$

□

References

[1] J.M. Amigo and M.B. Kennel, Topological permutation entropy, *Physica D*, **231** (2007), 137–142.

[2] J.M. Amigo, *Permutation Complexity in Dynamical Systems "Ordinal Patterns, Permutation Entropy, and All That"*, Springer-Verlag, Berlin, 2010.

[3] Ch. Corda, M. FatehiNia, M.R. Molaei and Y. Sayyari, Entropy of iterated function systems and their relations with black holes and Bohr-Like black holes entropies, *Entropy*, **20**(56) (2018), 2–17.

[4] A. Mehrpooya, Y. Sayyari and M.R. Molaei, Algebraic and Shannon entropies of commutative hypergroups and their connection with information and permutation entropies and with calculation of entropy for chemical algebras, *Soft Computing*, **23**(24) (2019), 13035–13053.

[5] S. SIMIC, Jensens inequality and new entropy bounds, *Appl. Math. Lett.*, **22**(8) (2009), 1262–1265.

[6] N. Tapus and P.G. Popescu, A new entropy upper bound, *Appl. Math. Lett.*, **25**(11) (2012), 1887–1890.

[7] P. Walters, *An Introduction to Ergodic Theory*, Springer Verlag, New York, 2000.