entropy

Article

Logical Entropy of Information Sources

Peng Xu [1], Yamin Sayyari [2] and Saad Ihsan Butt [3,*]

[1] School of Computer Science of Information Technology, Qiannan Normal University for Nationalities, Duyun 558000, China
[2] Department of Mathematics, Sirjan University of Technology, Sirjan 7813733385, Iran
[3] Department of Mathematics, COMSATS University Islamabad, Lahore Campus, Islamabad 54000, Pakistan
* Correspondence: saadihsanbutt@gmail.com

Abstract: In this paper, we present the concept of the logical entropy of order m, logical mutual information, and the logical entropy for information sources. We found upper and lower bounds for the logical entropy of a random variable by using convex functions. We show that the logical entropy of the joint distributions X_1 and X_2 is always less than the sum of the logical entropy of the variables X_1 and X_2. We define the logical Shannon entropy and logical metric permutation entropy to an information system and examine the properties of this kind of entropy. Finally, we examine the amount of the logical metric entropy and permutation logical entropy for maps.

Keywords: entropy; logical entropy; random variable; information source; convex function

MSC: 94A17; 37B40; 26A51; 81P10

check for updates

Citation: Xu, P.; Sayyari, Y.; Butt, S.I. Logical Entropy of Information Sources. *Entropy* **2022**, *24*, 1174. https://doi.org/10.3390/e24091174

Academic Editor: Ercan Kuruoglu

Received: 27 July 2022
Accepted: 18 August 2022
Published: 23 August 2022

1. Introduction and Basic Notions

Entropy is an influential quantity that has been explored in a wide range of studies, from applied to physical sciences. In the 19th century, Carnot and Clausius diversified the concept of entropy into three main directions—entropy associated with heat engines (where it behaves similar to a thermal charge), statistical entropy, and (according to Boltzmann and Shannon) entropy in communications channels and information security. Thus, the theory of entropy plays a key role in mathematics, statistics, dynamical systems (where complexity is mostly measured by entropy), information theory [1], chemistry [2], and physics [3] (see also [4–6]).

In recent years, other information source entropies have been studied [7–9]. Butt et al. in [10,11] introduced new bounds for Shannon, relative, and Mandelbrot entropies via interpolating polynomials. Amig and colleagues defined entropy as a random process and the permutation entropy of a source [1,12].

Ellerman [13] was the first to take credit for introducing a detailed introduction to the concept of logical entropy and establishing its relationship with the renowned Shannon entropy. In recent years, many researchers have focused on extending the notion of logical entropy in new directions/perspectives. Markechová et al. [14] proposed the study of logical entropy and logical mutual information of experiments in the intuitionistic fuzzy case. Ebrahimzadeh [15] proposed the logical entropy of a quantum dynamical system and investigated its ergodic properties. However, the logical entropy of a fuzzy dynamical system was investigated in [7] (see also [16]). Tamir et al. [17] extended the idea of logical entropy over the quantum domain and expressed it in terms of the density matrix. In [18], Ellerman defined logical conditional entropy and logical relative entropy. In fact, logical entropy is a particular case of Tsallis entropy when $q = 2$. Logical entropy resembles the information measure introduced by Brukner and Zeilinger [19]. In [13], Ellerman introduced the concept of logical entropy for a random variable. He studied the logical entropy of the joint distribution $p(x,y)$ over $X \times Y$ as:

$$h(x,y) = 1 - \sum_{x,y}[p(x,y)]^2.$$

The motive of this study was to extend the concept of logical entropy presented in [13] to information sources. Since estimating entropy from the information source can be difficult [20], we defined the logical metric permutation entropy of a map and used it to apply for an information source.

In the article, $(\Gamma, \mathcal{G}, \mu)$ is a measurable probability space (i.e., $\Gamma \neq \emptyset$ and $\mathcal{G}$ enjoys the structure of σ-algebra of subsets of Γ with $\mu(\Gamma) = 1$). Further, if X is a random variable of Γ possessing discrete finite state space $A = \{a_1, \ldots, a_n\}$, then the function $p : A \to [0,1]$ defined by

$$p(x) = \mu\{\gamma \in \Gamma : X(\gamma) = x\}$$

is a probability function. $H_\mu(X) = -\sum_{x \in A} p(x) \log p(x)$ denotes the Shannon entropy of X [1]. If $(X_n)_{n=1}^{\infty}$ is a sequence of the random variables on Γ, the sequence X_n is called an information source (also called the stochastic process [**S**.**P**]). Similarly, if $m \geq 1$, then we define $p : A^m \to [0,1]$ by

$$p(x_1, \ldots, x_m) = \mu\{\gamma \in \Gamma : X_1(\gamma) = x_1, \ldots, X_m(\gamma) = x_m\}.$$

We know that

$$\sum_{x_1,\ldots,x_m \in A} p(x_1, \ldots, x_m) = \mu(\Gamma) = 1$$

for every natural number m. A finite space **S**.**P**, $\mathbf{X} = (X_n)_{n=1}^{\infty}$ can be recalled as a stationary finite space **S**.**P** if

$$p(x_1, \ldots, x_m) = \mu\{\gamma \in \Gamma : X_{k+1}(\gamma) = x_1, \ldots, X_{k+m}(\gamma) = x_m\},$$

for every $m, k \in \mathbb{N}$. In an information–theoretical setting, one may assume a stationary **S**.**P**, **X** as a data source. A finite space **S**.**P**, **X** is strictly a stationary finite space **S**.**P** if

$$p(x_1, \ldots, x_m) = \mu\{\gamma \in \Gamma : X_{k_1}(\gamma) = x_1, \ldots, X_{k_m}(\gamma) = x_m\},$$

for every $\{k_1, \ldots, k_m\} \subseteq \mathbb{N}$. The Shannon entropy of order m of source **X** is defined by [1,12]

$$H_\mu(X_1^m) = -\sum_{x_1,\ldots,x_m \in A} p(x_1, \ldots, x_m) \log p(x_1, \ldots, x_m).$$

The Shannon entropy of source **X** is defined by $h_\mu(\mathbf{X}) = \lim_{m\to\infty}(\frac{1}{m} H_\mu(X_1^m))$. If we assume that the alphabet A from source **X** accepts an order $\leq$, so that $(A, \leq)$ is a totally ordered set, then define another order $\prec$ on A by [1]

$$t_i \prec t_j \Leftrightarrow t_i < t_j \text{ or } (t_i = t_j \text{ and } i < j).$$

We say that a length-m sequence $t_k^{k+m-1} = (t_k, \ldots, t_{k+m-1})$ has an order pattern π if, $t_{k+\pi(0)} \prec t_{k+\pi(1)} \prec \ldots \prec t_{k+\pi(m-1)}$, where $t_i, t_j \in A$, $k \in N$ and $i \neq j$. To a **S**.**P**, $\mathbf{X} = (X_n)_{n \in N_0}$ we associate a probability process $\mathbf{R} = (R_n)_{n \in N_0}$ defined by $R_m(\gamma) = \sum_{i=0}^{m} \delta(X_i(\gamma) \leq X_m(\gamma))$. The sequence **R** defines a discrete-time process that is non-stationary. The metric permutation entropy of order m and the metric permutation entropy of source **X** are, respectively, defined by [1,12]

$$H^{\star}_{\mu}(X_0^{m-1}) = H_{\mu}(R_0^{m-1}) = \frac{-1}{m-1}\sum_{r_0,\dots,r_{m-1}} p(r_0^{m-1})\log p(r_0^{m-1}),$$

and $h^{\star}_{\mu}(\mathbf{X}) = \limsup_{m\to\infty} H^{\star}_{\mu}(X_0^{m-1})$.

2. Main Results

In this section, we use the symbol x_1^m for $(x_1,\dots,x_m)$ to simplify the notation.

Definition 1. *Reference [13]. Let X be a random variable on Γ with discrete finite state space $A = \{a_1,\dots,a_n\}$. Then,*

$$H_{\mu l}(X) = \sum_{x\in A} p(x)[1-p(x)] = 1 - \sum_{x\in A}[p(x)]^2$$

is called the logical Shannon entropy of X.

Theorem 1. *Reference [21] If f is convex on I and $\zeta = \min_{1\le i\le n}\{y_i\}$, $\eta = \max_{1\le i\le n}\{y_i\}$, then*

$$\frac{f(\zeta)+f(\eta)-2f(\frac{\zeta+\eta}{2})}{n} \le \frac{\sum_{i=1}^n f(y_i)}{n} - f(\frac{\sum_{i=1}^n y_i}{n}) \le f(\zeta)+f(\eta)-2f(\frac{\zeta+\eta}{2}).$$

Theorem 2. *Suppose that X is a random variable on Γ with a discrete finite state space $A = \{a_1,\dots,a_n\}$, $\zeta = \min_{1\le i\le n}\{p(a_i)\}$ and $\eta = \max_{1\le i\le n}\{p(a_i)\}$, then*

$$0 \le \Delta(\zeta,\eta) := \frac{(\zeta-\eta)^2}{4} \le \frac{n-1}{n} - H_{\mu l}(X) \le n\frac{(\zeta-\eta)^2}{4} = n\Delta(\zeta,\eta).$$

Proof. Applying Theorem 1 with $f(x) = x^2 - x$, we obtain

$$\begin{aligned}
&\frac{1}{n}((\zeta^2-\zeta)+(\eta^2-\eta)-2((\frac{\zeta+\eta}{2})^2-\frac{\zeta+\eta}{2}))\\
&\le \frac{1}{n}\sum_{i=1}^n (x_i^2 - x_i) - ((\frac{\sum_1^n x_i}{n})^2 - (\frac{\sum_1^n x_i}{n}))\\
&\le (\zeta^2-\zeta)+(\eta^2-\eta)-2((\frac{\zeta+\eta}{2})^2-\frac{\zeta+\eta}{2}).
\end{aligned}$$

Putting $y_i = p(a_i)$, it follows that

$$\begin{aligned}
&\frac{1}{n}((\zeta^2-\zeta)+(\eta^2-\eta)-2((\frac{\zeta+\eta}{2})^2-\frac{\zeta+\eta}{2}))\\
&\le \frac{1}{n}\sum_{i=1}^n ((p(a_i))^2 - p(a_i)) - ((\frac{\sum_1^n p(a_i)}{n})^2 - (\frac{\sum_1^n p(a_i)}{n}))\\
&\le (\zeta^2-\zeta)+(\eta^2-\eta)-2((\frac{\zeta+\eta}{2})^2-\frac{\zeta+\eta}{2}).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\frac{1}{n}((\zeta^2-\zeta)+(\eta^2-\eta)-2((\frac{\zeta+\eta}{2})^2-\frac{\zeta+\eta}{2}))\\
&\le \frac{1}{n}\sum_{i=1}^n (p(a_i))^2 - \frac{1}{n}\sum_{i=1}^n p(a_i) - (\frac{1}{n^2} - \frac{1}{n})\\
&\le (\zeta^2-\zeta)+(\eta^2-\eta)-2((\frac{\zeta+\eta}{2})^2-\frac{\zeta+\eta}{2}).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\frac{1}{n}((\zeta^2-\zeta)+(\eta^2-\eta)-2((\frac{\zeta+\eta}{2})^2-\frac{\zeta+\eta}{2})) \\
&\leq \frac{1}{n}(1-H_{\zeta l}(X))-\frac{1}{n}-(\frac{1}{n^2}-\frac{1}{n}) \\
&\leq (\zeta^2-\zeta)+(\eta^2-\eta)-2((\frac{\zeta+\eta}{2})^2-\frac{\zeta+\eta}{2}).
\end{aligned}$$

After some calculations, it turns out that

$$\Delta(\zeta,\eta) := \frac{(\zeta-\eta)^2}{4} \leq \frac{n-1}{n} - H_{\mu l}(X) \leq n\frac{(\zeta-\eta)^2}{4}.$$

□

Lemma 1. *Let X be a random variable with alphabet $A=\{a_1,\ldots,a_n\}$. Then, $0 \leq H_{\mu l}(X) \leq \frac{n-1}{n}$, and equality holds if and only if $p(a_i)=p(a_j)$ for every $1\leq i,j\leq n$.*

Proof. Using Theorem 2, we obtain $0 \leq H_{\mu l}(X) \leq \frac{n-1}{n}$. Now, let $H_{\mu l}(X) = \frac{n-1}{n}$, by the use of Theorem 2, we have $M(\zeta,\eta) = \frac{(\zeta-\eta)^2}{4} = 0$ and, thus, $\zeta=\eta$. Therefore, $\max_{1\leq i\leq n}\{p(a_i)\} = \min_{1\leq i\leq n}\{p(a_i)\}$. Thus, $p(a_i)=p(a_j)$ for every $1\leq i,j\leq n$. On the other hand, if $p(a_i)=p(a_j)$ for every $1\leq i,j\leq n$, then $\zeta=\eta$, so $M(\zeta,\eta)=0$ and by the use of Theorem 2, we obtain $H_{\mu l}(X)-\frac{n-1}{n}=0$. Hence, $H_{\mu l}(X)=\frac{n-1}{n}$. □

Definition 2. *The logical Shannon entropy of order m of source* **X** *is defined by*

$$\begin{aligned}
H_{\mu l}(X_1^m) = H_{\mu l}(X_1,\ldots,X_m) &:= \sum_{x_1,\ldots,x_m\in A} p(x_1,\ldots,x_m)(1-p(x_1,\ldots,x_m)), \\
&= 1-\sum_{x_1,\ldots,x_m\in A}(p(x_1,\ldots,x_m))^2
\end{aligned}$$

It is easy to see that may be $p(x_1,x_2)\neq p(x_2,x_1)$ but for every two random variables x_1,x_2 we have $H_{\mu l}(x_1,x_2)=H_{\mu l}(x_2,x_1)$.

Definition 3. *Let m be a natural number and $1\leq i_1,\ldots,i_m\leq n$. We define the sets $\mathcal{A}_{i_1i_2\ldots i_m}$ by*

$$\mathcal{A}_{i_1i_2\ldots i_m} = \{\gamma\in\Gamma : X_1(\gamma)=a_{i_1}, X_2(\gamma)=a_{i_2},\ldots,X_m(\gamma)=a_{i_m}\}.$$

and $\mu(\mathcal{A}_{i_1i_2\ldots i_m}) := a_{i_1i_2\ldots i_m}$.

Moreover, $\mathcal{A}_{i_1i_2\ldots i_m}\cap\mathcal{A}_{j_1j_2\ldots j_m}=\emptyset$ for every $(i_1,i_2,\ldots,i_m)\neq(j_1,j_2,\ldots,j_m)$ and for every $m\in\mathbb{N}$. Furthermore, if $\gamma\in\bigcup_{j=1}^n\mathcal{A}_{i_1i_2\ldots i_mj}$, then $\gamma\in\mathcal{A}_{i_1i_2\ldots i_mj_0}$ for some $j_0\in\{1,\ldots,n\}$. Hence,

$$X_1(\gamma)=a_{i_1},\ldots,X_m(\gamma)=a_{i_m},X_{m+1}(\gamma)=a_{j_0}$$

for some $j_0\in\{1,\ldots,n\}$ and, thus, $\gamma\in\mathcal{A}_{i_1i_2\ldots i_m}$. Moreover, if $\gamma\in\mathcal{A}_{i_1i_2\ldots i_m}$, then

$$X_1(\gamma)=a_{i_1},\ldots,X_m(\gamma)=a_{i_m}.$$

Define $X_{m+1}(\gamma)=a_{j_0}$. Therefore,

$$X_1(\gamma)=a_{i_1},\ldots,X_m(\gamma)=a_{i_m},X_{m+1}(\gamma)=a_{j_0}.$$

Hence, $\gamma \in \mathcal{A}_{i_1 i_2 \ldots i_m j_0}$ for some $j_0 \in \{1, \ldots, n\}$ and, thus, $\gamma \in \bigcup_{j=1}^{n} \mathcal{A}_{i_1 i_2 \ldots i_m j}$. So,

$$\mathcal{A}_{i_1 i_2 \ldots i_m} = \bigcup_{j=1}^{n} \mathcal{A}_{i_1 i_2 \ldots i_m j}$$

and, therefore, $\Gamma = \bigcup_{i_1, i_2, \ldots, i_m} \mathcal{A}_{i_1 i_2 \ldots i_m}$. Hence, we obtain

$$\sum_{i_1 i_2 \ldots i_m} a_{i_1 i_2 \ldots i_m} = 1$$

and

$$\begin{aligned} a_{i_1 i_2 \ldots i_m} &= \mu(\mathcal{A}_{i_1 i_2 \ldots i_m}) = \mu(\cup_{j=1}^{n} \mathcal{A}_{i_1 i_2 \ldots i_m j}) \\ &= \sum_{j=1}^{n} \mu(\mathcal{A}_{i_1 i_2 \ldots i_m j}) = \sum_{j=1}^{n} a_{i_1 i_2 \ldots i_m j} \end{aligned} \tag{1}$$

for every $1 \leq i_1, i_2, \ldots, i_m \leq n$.

We now prove the following Theorem by employing Lemma A1 (see Appendix A):

Theorem 3. *If X_1 and X_2 are two random variables on Γ, then*

$$\max\{H_{\mu l}(X_1), H_{\mu l}(X_2)\} \leq H_{\mu l}(X_1, X_2) \leq H_{\mu l}(X_1) + H_{\mu l}(X_2). \tag{2}$$

Proof. Suppose $A = \{a_1, \ldots, a_n\}$. For every $1 \leq i, j \leq n$, we consider

$$\begin{aligned} &\mathcal{B}_i = \{\gamma \in \Gamma : X_1(\gamma) = a_i\},\ \mathcal{C}_j = \{\gamma \in \Gamma : X_2(\gamma) = a_j\},\ \mathcal{A}_{ij} = \mathcal{B}_i \cap \mathcal{C}_j, \\ &b_i = \mu(\mathcal{B}_i),\ c_j := \mu(\mathcal{C}_j),\ a_{ij} = \mu(\mathcal{A}_{ij}). \end{aligned}$$

Moreover, $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset$ and $\mathcal{B}_i \cap \mathcal{B}_j = \emptyset$ for every $1 \leq i \neq j \leq n$; thus, $\mathcal{A}_{ij} \cap \mathcal{A}_{kl} = \emptyset$ for every ordered pair $(i, j) \neq (k, l)$. For obvious reasons, $\mathcal{B}_i = \cup_{j=1}^{n} \mathcal{A}_{ij}$ for each $1 \leq i \leq n$ and $\mathcal{C}_j = \cup_{i=1}^{n} \mathcal{A}_{ij}$ for each $1 \leq j \leq n$, and $\Gamma = \cup_{i,j} \mathcal{A}_{ij}$. So, we have $\sum_{i,j} a_{ij} = 1$ and for every $1 \leq i, j \leq n$,

$$b_i = \mu(\mathcal{B}_i) = \mu(\bigcup_{j=1}^{n} \mathcal{A}_{ij}) = \sum_{j=1} \mu(\mathcal{A}_{ij}) = \sum_{j=1}^{n} a_{ij},$$

and

$$c_j = \mu(\mathcal{C}_j) = \mu(\cup_{i=1}^{n} \mathcal{A}_{ij}) = \sum_{i=1} \mu(\mathcal{A}_{ij}) = \sum_{i=1}^{n} a_{ij}.$$

With the use of Lemma A1, we have

$$\sum_{i=1}^{n} (\sum_{j=1}^{n} a_{ij})^2 + \sum_{j=1}^{n} (\sum_{i=1}^{n} a_{ij})^2 \leq 1 + \sum_{i,j} a_{ij}^2.$$

Therefore,

$$\sum_{i=1}^{n} b_i^2 + \sum_{j=1}^{n} c_j^2 \leq 1 + \sum_{i,j} a_{ij}^2.$$

Consequently,

$$\sum_{i=1}^{n}(\mu(\mathcal{B}_i))^2 + \sum_{j=1}^{n}(\mu(\mathcal{C}_j))^2 \leq 1 + \sum_{i,j}(\mu(\mathcal{A}_{ij}))^2,$$

and

$$-\sum_{i,j}(\mu(\mathcal{A}_{ij}))^2 \leq 1 - \sum_{i=1}^{n}(\mu(\mathcal{B}_i))^2 - \sum_{j=1}^{n}(\mu(\mathcal{C}_j))^2.$$

Hence,

$$1 - \sum_{i,j}(\mu(\mathcal{A}_{ij}))^2 \leq (1 - \sum_{i=1}^{n}(\mu(\mathcal{B}_i))^2) + (1 - \sum_{j=1}^{n}(\mu(\mathcal{C}_j))^2),$$

it follows that $H_{\mu l}(X_1, X_2) \leq H_{\mu l}(X_1) + H_{\mu l}(X_2)$.

Now, we prove the left-hand inequality. Since

$$b_i = \mu(\mathcal{B}_i) = \mu(\bigcup_{j=1}^{n} \mathcal{A}_{ij}) = \sum_{j=1} \mu(\mathcal{A}_{ij}) = \sum_{j=1}^{n} a_{ij}$$

for every $1 \leq i \leq n$, $b_i^2 = (\sum_{j=1}^{n} a_{ij})^2 \geq \sum_{j=1}^{n} a_{ij}^2$. Therefore,

$$(\mu(\mathcal{B}_i))^2 \geq \sum_{j=1}^{n}(\mu(\mathcal{A}_{ij}))^2,$$

and, thus,

$$\sum_{i=1}^{n}(\mu(\mathcal{B}_i))^2 \geq \sum_{i=1}^{n}\sum_{j=1}^{n}(\mu(\mathcal{A}_{ij}))^2.$$

So, $H_{\mu l}(X_1) \leq H_{\mu l}(X_1, X_2)$.

Similarly, $H_{\mu l}(X_2) \leq H_{\mu l}(X_1, X_2)$. Consequently,

$$\max\{H_{\mu l}(X_1), H_{\mu l}(X_2)\} \leq H_{\mu l}(X_1, X_2).$$

□

Corollary 1. *If* **X** *is an information source, then*

$$\max\{H_{\mu l}(X_i) : 1 \leq i \leq k\} \leq H_{\mu l}(X_1, \ldots, X_k) \leq \sum_{i=1}^{k} H_{\mu l}(X_i),\ (\forall k \in \mathbb{N}).$$

Proof. This follows from Theorem 3. □

Definition 4. *The logical metric permutation entropy of order m of source* $\mathbf{X} = \{X_0, X_1, \ldots\}$ *defined by*

$$H^{\star}_{\mu l}(X_0^{m-1}) = H_{\mu l}(R_0^{m-1}) = 1 - \sum_{r_0,\ldots,r_{m-1}} (p(r_0^{m-1}))^2.$$

Lemma 2. *For a* **S**.**P**, **X**, *the sequence of* $\{H_{\mu l}(X_1^m)\}_m$ *increases. Thus,* $\lim_{m\to\infty} H_{\mu l}(X_1^m)$ *exists.*

Proof. According to (1),

$$p(x_1, \ldots, x_m) = \sum_{x_{m+1}} p(x_1, \ldots, x_m, x_{m+1})$$

for every $m \in \mathbb{N}$. Therefore,

$$\begin{aligned}(p(x_1, \ldots, x_m))^2 &= (\sum_{x_{m+1}} p(x_1, \ldots, x_m, x_{m+1}))^2 \\ &\leq \sum_{x_{m+1}} (p(x_1, \ldots, x_m, x_{m+1}))^2,\end{aligned}$$

and

$$\sum_{x_1^m} (p(x_1, \ldots, x_m))^2 \leq \sum_{x_1^{m+1}} (p(x_1, \ldots, x_m, x_{m+1}))^2.$$

This means that

$$\begin{aligned}H_{\mu l}(X_1^m) &= 1 - \sum_{x_1^m} (p(x_1, \ldots, x_m))^2 \\ &\geq 1 - \sum_{x_1^{m+1}} (p(x_1, \ldots, x_m, x_{m+1}))^2 = H_{\mu l}(X_1^{m+1}).\end{aligned}$$

□

Definition 5. *The logical Shannon entropy of source* $\mathbf{X} = \{X_1, X_2, \ldots\}$ *is defined by*

$$h_{\mu l}(\mathbf{X}) = \lim_{m \to \infty} (H_{\mu l}(X_1^m)).$$

Definition 6. *The logical metric permutation entropy of source* $\mathbf{X} = \{X_0, X_1, \ldots\}$ *is defined by*

$$h^{\star}_{\mu l}(\mathbf{X}) = \lim_{m \to \infty} H^{\star}_{\mu l}(X_0^{m-1}).$$

Remark 1. *Let m be a positive integer number. Then* $0 \leq H_{\mu l}(X_1^m) \leq 1$ *and* $0 \leq h_{\mu l}(\mathbf{X}) \leq 1$.

Lemma 3. *Let* $\mathbf{X} = (X_1, X_1, X_1, \ldots)$ *be an information source. Then the following holds:*

1. $H_{\mu l}(\underbrace{X_1, X_1, \ldots, X_1}_{m\ times}) = H_{\mu l}(X_1)$, *for every* $m \in \mathbb{N}$.
2. $h_{\mu l}(\mathbf{X}) = H_{\mu l}(X_1)$.

Proof.

1. If $\mathbf{X} = (X_1, X_1, X_1, \ldots)$, then

$$p(x_1, x_2, \ldots, x_m) = \begin{cases} p(x_1) & x_1 = x_2 = \ldots = x_m \\ 0 & x_i \neq x_j, \text{ for some } 1 \leq i \neq j \leq m. \end{cases}$$

Hence,

$$\begin{aligned}H_{\mu l}(X_1, \ldots, X_m) &= \sum_{x_1, \ldots, x_m \in A} p(x_1, \ldots, x_m)(1 - p(x_1, \ldots, x_m)) \\ &= \sum_{x_1 \in A} p(x_1)(1 - p(x_1)) \\ &= H_{\mu l}(X_1).\end{aligned} \tag{3}$$

2. We derive from (3) that

$$\begin{aligned} h_{\mu l}(\mathbf{X}) &= \lim_{m\to\infty} H_{\mu l}(X_1,\ldots,X_m) \\ &= \lim_{m\to\infty} H_{\mu l}(X_1,\ldots,X_1) \\ &= \lim_{m\to\infty} H_{\mu l}(X_1) = H_{\mu l}(X_1). \end{aligned}$$

□

Theorem 4. *Suppose that* $\mathbf{X}$ *represents an information source on* Γ *with the discrete finite state space* $A = \{a_1, \ldots, a_n\}$.

1. *If* $\zeta_m = \min_{x_1^m \in A}\{p(x_1^m)\}$ *and* $\eta_m = \max_{x_1^m \in A}\{p(x_1^m)\}$, *then*

$$0 \le \Delta(\zeta_m, \eta_m) \le \frac{n^m - 1}{n^m} - H_{\mu l}(x_1^m) \le n^m \Delta(\zeta_m, \eta_m), \tag{4}$$

2. $\lim_{m\to\infty} \Delta(\zeta_m, \eta_m) \le 1 - h_{\mu l}(\mathbf{X}) \le \lim_{m\to\infty} n^m \Delta(\zeta_m, \eta_m)$.

Proof.

1. The result follows from Theorem 2.
2. Taking the limit as $m \to \infty$ in (4), consequently (2) holds.

□

Lemma 4. *Let* $\mathbf{X}$ *represent an information source on* Γ *with the discrete finite state space* $A = \{a_1, \ldots, a_n\}$, *then* $0 \le H_{\mu l}(X_1^m) \le \frac{n^m-1}{n^m}$, *and equality holds if and only if* $p(x_1^m) = p(t_1^m)$ *for every* $x_1^m, t_1^m \in A^m$.

Proof. By Theorem 4, $0 \le H_{\mu l}(X_1^m) \le \frac{n^m-1}{n^m}$. If $H_{\mu l}(X_1^m) = \frac{n^m-1}{n^m}$, then by the use of Theorem 4 we obtain $\Delta(\zeta_m, \eta_m) = \frac{(\zeta_m - \eta_m)^2}{4} = 0$. Hence $\zeta_m = \eta_m$. Therefore $\max_{x_1^m \in A}\{p(x_1^m)\} = \min_{x_1^m \in A}\{p(x_1^m)\}$. Thus $p(x_1^m) = p(t_1^m)$ for every $x_1^m, t_1^m \in A^m$. On the other hand if $p(x_1^m) = p(t_1^m)$ for every $x_1^m, t_1^m \in A^m$, then $\zeta_m = \eta_m$. Therefore $\Delta(\zeta_m, \eta_m) = 0$ and by Theorem 4 has $H_{\mu l}(X_1^m) - \frac{n^m-1}{n^m} = 0$ and thus $H_{\mu l}(X_1^m) = \frac{n^m-1}{n^m}$. □

Definition 7. *Let* $p(x) \neq 0$, *the conditional probability function defined by* $p(y|x) := \frac{p(x,y)}{p(x)}$. *In general, for* $p(x_1, \ldots, x_n) \neq 0$, *the conditional probability function is defined by* $p(x_1|x_2, \ldots, x_{n+1}) := \frac{p(x_1, x_2, \ldots, x_{n+1})}{p(x_2, x_3, \ldots, x_n)}$.

Lemma 5. *Let* $x_1, x_2, \ldots, x_{n+1}$ *be a word. Then*

$$p(x_{m+1}, x_m, \ldots, x_1) = \prod_{i=1}^{m+1} p(x_i|x_{i-1}, \ldots, x_1),$$

where $m \in \mathbb{N}$ *and* $p(x_1|x_0) := p(x_1)$.

Proof. We prove the lemma by induction. If $m = 1$, have $p(x_1, x_2) = p(x_1) \times p(x_1|x_2)$. Thus, the statement is true for $m = 1$. Now suppose the statement is true for $m = k - 1$, we give reasons for $m = k$.

$$\begin{aligned}\prod_{i=1}^{k+1} p(x_i|x_{i-1}, \ldots, x_1) &= \prod_{i=1}^{k} p(x_i|x_{i-1}, \ldots, x_1) \times p(x_{k+1}|x_k, \ldots, x_1) \\ &= p(x_k, x_{k-1}, \ldots, x_1) \times p(x_{k+1}|x_k, \ldots, x_1) \\ &= p(x_k, x_{k-1}, \ldots, x_1) \times \frac{p(x_{k+1}, x_k, \ldots, x_1)}{p(x_k, x_{k-1}, \ldots, x_1)} \\ &= p(x_{k+1}, x_k, \ldots, x_1),\end{aligned}$$

which completes the proof. □

Definition 8. *Let X_1 and X_2 be two random variables on Γ. We define the conditional logical entropy of X_2 given X_1 by*

$$H_{\mu l}(X_2|X_1) := \sum_{x_1, x_2} (p(x_1))^2 (p(x_2) - (p(x_2|x_1))^2).$$

Note: if $p(x_1) = 0$, define $(p(x_1))^2(p(x_2) - (p(x_2|x_1))^2 = 0$.

Definition 9. *Suppose $X_1, X_2, \ldots, X_m$ are m random variables on Γ. Define the conditional logical entropy of X_m given $X_1, \ldots, X_{m-1}$ by*

$$\begin{aligned}H_{\mu l}(X_m|X_{m-1}, \ldots, X_2, X_1) : = \sum_{x_1^m} &(p(x_{m-1}, \ldots, x_2, x_1))^2 \\ &[p(x_m) - (p(x_m|x_{m-1}, \ldots, x_1))^2].\end{aligned}$$

Lemma 6. *Suppose $X_1, X_2, \ldots, X_m$ are m random variables on Γ, then*

$$\begin{aligned}&H_{\mu l}(X_m|X_{m-1}, \ldots, X_2, X_1) \\ &= \sum_{x_1^{m-1}} (p(x_{m-1}, \ldots, x_2, x_1))^2 - \sum_{x_1^m} (p(x_m, \ldots, x_2, x_1))^2 \\ &= H_{\mu l}(X_m, X_{m-1}, \ldots, X_2, X_1) - H_{\mu l}(X_{m-1}, \ldots, X_2, X_1).\end{aligned}$$

Proof. According to Definition 9, we obtain

$$\begin{aligned}&H_{\mu l}(X_n|X_{m-1}, \ldots, X_2, X_1) \\ &= \sum_{x_1^m} (p(x_{m-1}, \ldots, x_2, x_1))^2 (p(x_m) - (p(x_m|x_{m-1}, \ldots, x_1))^2) \\ &= \sum_{x_1^m} p(x_{m-1}, \ldots, x_2, x_1))^2 p(x_m) \\ &- \sum_{x_1^m} p(x_{m-1}, \ldots, x_2, x_1))^2 (p(x_m|x_{m-1}, \ldots, x_1))^2 \\ &= (\sum_{x_1^{m-1}} p(x_{m-1}, \ldots, x_2, x_1))^2)(\sum_{x_m} p(x_m)) \\ &- \sum_{x_1^m} p(x_{m-1}, \ldots, x_2, x_1))^2 (\frac{p(x_m, x_{m-1}, \ldots, x_1)}{p(x_{m-1}, \ldots, x_1)})^2 \\ &= \sum_{x_1^{m-1}} (p(x_{m-1}, \ldots, x_2, x_1))^2 - \sum_{x_1^m} (p(x_m, \ldots, x_2, x_1))^2.\end{aligned}$$

□

Lemma 7. *Let* **X** *be a stationary finite space* **S.P**, *then*

$$\sum_{x_2^n}(p(x_n,\ldots,x_2))^2 = \sum_{x_1^{n-1}}(p(x_{n-1},\ldots,x_2,x_1))^2. \tag{5}$$

Proof. Since **X** is stationary,

$$\begin{aligned}\sum_{x_2^n}(p(x_n,\ldots,x_2))^2 &= \sum_{x_2^n}(\mu(\{\gamma\in\Gamma : X_n(\gamma)=x_n,\ldots,X_2(\gamma)=x_2\}))^2\\ &= \sum_{x_2^n}(\mu(\{\gamma\in\Gamma : X_{n-1}(\gamma)=x_n,\ldots,X_1(\gamma)=x_2\}))^2\\ &= \sum_{x_1^{n-1}}(p(x_{n-1},\ldots,x_2,x_1))^2,\end{aligned}$$

which yields (5). □

Theorem 5. *Let* **X** *be a stationary finite space* **S.P**, *with discrete finite state space* $A=\{a_1,\ldots,a_n\}$. *Then the sequence of conditional logical entropies* $H_{\mu l}(X_m|X_{m-1},\ldots,X_1)$ *decreases.*

Proof. Under the notation of Definition 3, define

$$\{\mathcal{A}_{i_1 i_2\ldots i_m} : 1\le i_1,i_2,\ldots,i_m\le n\} = \{\mathcal{D}_1,\mathcal{D}_2,\ldots,\mathcal{D}_M\},$$

and $\mu(\mathcal{D}_r)=d_r$ where $M=n^m$. Furthermore, assume that

$$\mathcal{D}_{ij} = \mathcal{D}_i \bigcap \{\gamma\in\Gamma : x_j(\gamma)=a_j\},\ \mu(\mathcal{D}_{ij})=d_{ij},$$

$$\mathcal{D}_{ijk} = \mathcal{D}_{ij} \bigcap \{\gamma\in\Gamma : x_k(\gamma)=a_k\},\ \mu(\mathcal{D}_{ijk})=d_{ijk},$$

where $1\le i\le M$ and $1\le j,k\le n$. It is easy to see that $\mathcal{D}_i\cap\mathcal{D}_j=\emptyset$ for every $1\le i\ne j\le n$, and $\mathcal{D}_{ij}\cap\mathcal{D}_{rs}=\emptyset$ for every ordered pair $(i,j)\le(r,s)$. Therefore, $\mathcal{D}_{ijk}\cap\mathcal{D}_{rst}=\emptyset$ for every $(i,j,k)\ne(r,s,t)$. For obvious reasons, $\mathcal{D}_i=\cup_{j=1}^n\mathcal{D}_{ij}$ for each $1\le i\le n$, $\mathcal{D}_{ij}=\cup_{k=1}^n\mathcal{D}_{ijk}$ for every $1\le i,j\le n$ and $\Gamma=\cup_{i,j,k}\mathcal{D}_{ijk}$. Consequently, $\sum_{i,j,k}d_{ijk}=1$ and

$$d_i=\mu(\mathcal{D}_i)=\mu(\bigcup_{j=1}^n\mathcal{D}_{ij})=\sum_{j=1}^n\mu(\mathcal{D}_{ij})=\sum_{j=1}^n d_{ij}$$

and

$$d_{ij}=\mu(\mathcal{D}_{ij})=\mu(\bigcup_{k=1}^n\mathcal{D}_{ijk})=\sum_{k=1}^n\mu(\mathcal{D}_{ijk})=\sum_{k=1}^n d_{ijk}$$

for every $1\le j\le M, 1\le i\le n$.

Using Theorem A1 and Lemma 7, we deduce that

$$\sum_{x_1^{m+2}} (p(x_{m+2},\ldots,x_2,x_1))^2 = \sum_{i=1}^{M}\sum_{j=1}^{n}\sum_{k=1}^{n} d_{ijk}^2 = \sum_{i,j,k} d_{ijk}^2$$
$$\sum_{x_1^{m+1}} (p(x_{m+1},\ldots,x_2,x_1))^2 = \sum_{i=1}^{M}\sum_{j=1}^{n} d_{ij}^2 = \sum_{i,j} d_{ij}^2 = \sum_{i,j}(\sum_{k=1}^{n} d_{ijk})^2$$
$$\sum_{x_1^{m}} (p(x_{m},\ldots,x_2,x_1))^2 = \sum_{i=1}^{M} d_{i}^2 = \sum_{i=1}^{M}(\sum_{j,k=1}^{n} d_{ijk})^2$$
$$\sum_{x_2^{m+2}} (p(x_{m+2},\ldots,x_2))^2 = \sum_{x_1^{m+1}} (p(x_{m+1},\ldots,x_2,x_1))^2 = \sum_{i,j} d_{ij}^2.$$

With the use of Theorem A1, we obtain

$$\begin{aligned}
&H_{\mu l}(X_{m+2}|X_{m+1},\ldots,X_2,X_1) \\
&= \sum_{x_1^{m+1}} (p(x_{m+1},\ldots,x_2,x_1))^2 - \sum_{x_1^{m+2}} (p(x_{m+2},\ldots,x_2,x_1))^2. \\
&= \sum_{i,j,k} d_{ijk}^2 - \sum_{i,j}(\sum_{k=1}^{n} d_{ijk})^2 \\
&\geq \sum_{i,j}(\sum_{k=1}^{n} d_{ijk})^2 - \sum_{i=1}^{M}(\sum_{j,k=1}^{n} d_{ijk})^2 \\
&= \sum_{x_1^{m}} (p(x_{m},\ldots,x_2,x_1))^2 - \sum_{x_1^{m+1}} (p(x_{m+1},\ldots,x_2,x_1))^2 \\
&= H_{\mu l}(X_{m+1}|X_m,\ldots,X_2,X_1),
\end{aligned}$$

this means that the sequence of conditional logical entropies

$$H_{\mu l}(X_m|X_{m-1},\ldots,X_1)$$

is decreasing, so

$$0 \leq \ldots \leq H_{\mu l}(X_{m+1}|X_m,\ldots,X_1) \leq H_{\mu l}(X_m|X_{m-1},\ldots,X_1) \leq \ldots \leq H_{\mu l}(X_1).$$

□

Corollary 2. *Let* $\mathbf{X} = (X_1, X_2, X_3, \ldots)$ *be a source. Then the limit* $\lim_{n\to\infty} H_{\mu l}(X_n|X_{n-1},\ldots,X_1)$ *exists.*

Lemma 8. *Let* $\mathbf{X} = (X_m)_{m=1}^{\infty}$ *be a stationary finite space* **S.P**. *Then*

$$\sum_{x_2^{m+1}} (p(x_{m+1}|x_m,\ldots,x_2))^2 = \sum_{x_1^{m}} (p(x_m|x_{m-1},\ldots,x_1))^2.$$

Proof. Since **X** is stationary,

$$\begin{aligned}
&\sum_{x_2^{m+1}} (p(x_{m+1}|x_m,\ldots,x_2))^2 \\
&= \sum_{x_2^{m+1}} (\frac{p(x_{m+1},x_m,\ldots,x_2)}{p(x_m,\ldots,x_2)})^2 \\
&= \sum_{x_2^{m+1}} (\frac{\mu(\{\gamma\in\Gamma : x_{m+1}(\gamma)=x_{m+1},\ldots,x_2(\Gamma)=x_2\})}{\mu(\{\Gamma\in\Gamma : x_m(\Gamma)=x_m,\ldots,x_2(\gamma)=x_2\})})^2 \\
&= \sum_{x_2^{m+1}} (\frac{\mu(\{\gamma\in\Gamma : x_m(\gamma)=x_{m+1},\ldots,x_1(\gamma)=x_2\})}{\mu(\{\gamma\in\Gamma : x_{m-1}(\gamma)=x_m,\ldots,x_1(\gamma)=x_2\})})^2 \\
&= \sum_{x_1^m} (p(x_m|x_{m-1},\ldots,x_1))^2,
\end{aligned}$$

which completes the proof. □

Theorem 6. *Let* $\mathbf{X}=(X_n)_{n=1}^{\infty}$ *be a stationary finite space* **S.P**. *Then*

$$H_{\mu l}(X_{m+1}|X_m,\ldots,X_2) = H_{\mu l}(X_m|X_{m-1},\ldots,X_1)$$

Proof. According to Lemma 7,

$$\begin{aligned}
H_{\mu l}(X_{m+1}|X_m,\ldots,X_2) &= \sum_{x_2^m}(p(x_m,\ldots,x_2))^2 - \sum_{x_2^{m+1}}(p(x_m,\ldots,x_2))^2 \\
&= \sum_{x_1^{m-1}}(p(x_{m-1},\ldots,x_1))^2 - \sum_{x_1^m}(p(x_m,\ldots,x_1))^2 \\
&= H_{\mu l}(X_m|X_{m-1},\ldots,X_2,X_1).
\end{aligned}$$

Theorem 6 is thus proved. □

Theorem 7. *Let* X_1 *and* X_2 *be two random variables on* Γ. *Then the following hold:*

1. $H_{\mu l}(X_2|X_1) = H_{\mu l}(X_1,X_2) - H_{\mu l}(X_1)$.
2. $H_{\mu l}(X_2|X_1) + H_{\mu l}(X_1) = H_{\mu l}(X_1|X_2) + H_{\mu l}(X_2)$.

Proof.

1. Using the definition of condition logical entropy, we deduce

$$\begin{aligned}
H_{\mu l}(X_2|X_1) &= (\sum_{x_1}(p(x_1))^2) - \sum_{x_1,x_2}(p(x_1,x_2))^2 \\
&= (1-\sum_{x_1,x_2}(p(x_1,x_2))^2) - (1-\sum_{x_1}(p(x_1))^2) \\
&= H_{\mu l}(X_1,X_2) - H_{\mu l}(X_1),
\end{aligned}$$

   which completes the proof.
2. From the previous part, and since $H_{\mu l}(X_1,X_2) = H_{\mu l}(X_2,X_1)$, we have

$$\begin{aligned}
H_{\mu l}(X_2|X_1) + H_{\mu l}(X_1) &= H_{\mu l}(X_1,X_2) \\
&= H_{\mu l}(X_2,X_1) \\
&= H_{\mu l}(X_1|X_2) + H_{\mu l}(X_2).
\end{aligned}$$

□

Theorem 8. *Let* $\mathbf{X} = (X_1, X_1, X_1, \ldots)$ *be an information source. Then*

$$H_{\mu l}(X_1, \ldots, X_m) = \sum_{i=1}^{m} H_{\mu l}(X_i | X_{i-1}, \ldots, X_1),$$

where $H_{\mu l}(X_1 | X_0) := H_{\mu l}(X_1)$.

Proof. According to Lemma 6, we obtain

$$\begin{aligned}
&\sum_{i=1}^{m} H_{\mu l}(X_i | X_{i-1}, \ldots, X_1) \\
&= H_{\mu l}(X_1) + \sum_{i=2}^{m} (H_{\mu l}(X_i, \ldots, X_1) - H_{\mu l}(X_{i-1}, \ldots, X_1)) \\
&= H_{\mu l}(X_1, \ldots, X_m),
\end{aligned}$$

hence the theorem is proven. □

Theorem 9. *Let* $\mathbf{X} = (X_1, X_2, X_3, \ldots)$ *be an information source. Then*

$$h_{\mu l}(\mathbf{X}) = \sum_{i=1}^{\infty} H_{\mu l}(X_i | X_{i-1}, \ldots, X_1).$$

Proof. By the use of Theorem 8, we obtain

$$h_{\mu l}(\mathbf{X}) = \lim_{n \to \infty} \sum_{i=1}^{n} H_{\mu l}(X_i | X_{i-1}, \ldots, X_1) = \sum_{i=1}^{\infty} H_{\mu l}(X_i | X_{i-1}, \ldots, X_1),$$

which completes the proof. □

Definition 10. *An independent information source,* $\mathbf{X} = (X_1, X_2, X_3, \ldots)$*, is a source with the following property*

$$p(x_1, x_2, \ldots, x_m) = \prod_{i=1}^{m} p(x_i)$$

for all x_1^m.

Theorem 10. *Let* $\mathbf{X} = (X_1, X_2, X_3, \ldots)$ *be an independent information source. Then*

$$H_{\mu l}(X_{m+1} | X_m, \ldots, X_1) = (1 - H_{\mu l}(X_m, \ldots, X_1)) H_{\mu l}(X_{m+1})$$

for every $m \in \mathbb{N}$.

Proof. Since $\mathbf{X} = (X_1, X_2, X_3, \dots)$ is an independent random variables, we have

$$\begin{aligned}
&H_{\mu l}(X_{m+1}|X_m, \dots, X_1) \\
&= \sum_{x_1^{m+1}} (p(x_m, \dots, x_1))^2 p(x_{m+1}) - \sum_{x_1^{m+1}} (p(x_{m+1}, \dots, x_1))^2 \\
&= \sum_{x_1^{m+1}} (p(x_m, \dots, x_1))^2 p(x_{m+1}) - \sum_{x_1^{m+1}} (p(x_{m+1}, \dots, x_1))^2 \\
&= \sum_{x_1^{m+1}} (p(x_m, \dots, x_1))^2 p(x_{m+1}) - \sum_{x_1^{m+1}} (p(x_m, \dots, x_1))^2 (p(x_{m+1}))^2 \\
&= \sum_{x_1^{m+1}} (p(x_m, \dots, x_1))^2 (p(x_{m+1}) - (p(x_{m+1}))^2) \\
&= (\sum_{x_1^m} (p(x_m, \dots, x_1))^2)(\sum_{x_{m+1}} (p(x_{m+1}) - (p(x_{m+1}))^2)) \\
&= (1 - H_{\mu l}(X_m, \dots, X_1)) H_{\mu l}(X_{m+1}).
\end{aligned} \tag{6}$$

The result follows from (6). □

Theorem 11. *Suppose that* $\mathbf{X} = (X_1, X_2, X_3, \dots)$ *is an independent information source and* $\lim_n H_{\mu l}(X_n) \neq 0$. *Then* $h_{\mu l}(\mathbf{X}) = 1$.

Proof. In view of Theorem 10 and Lemma A2, we conclude that

$$\begin{aligned}
\lim_{n \to \infty} H_{\mu l}(X_{n+1}|X_n, \dots, X_1) &= \lim_{n \to \infty} (1 - H_{\mu l}(X_n, \dots, X_1)) H_{\mu l}(X_{n+1}) \\
&= \lim_n (1 - H_{\mu l}(X_n, \dots, X_1)) \times \lim_n H_{\mu l}(X_{n+1}) = 0.
\end{aligned}$$

Since $\lim_n H_{\mu l}(X_n) \neq 0$, $\lim_n (1 - H_{\mu l}(X_n, \dots, X_1)) = 0$. Hence,

$$h_{\mu l}(\mathbf{X}) = \lim_{n \to \infty} H_{\mu l}(X_n, \dots, X_1) = 1.$$

□

Theorem 12. *Let* $\mathbf{X} = (X_1, X_2, X_3, \dots)$ *be an independent information source. Then*

$$H_{\mu l}(X_m, \dots, X_1) = 1 - \prod_{i=1}^m (1 - H_{\mu l}(X_i))$$

for every $m \in \mathbb{N}$.

Proof. Since $\mathbf{X}$ is an independent source,

$$\begin{aligned}
H_{\mu l}(X_m, \dots, X_1) &= 1 - \sum_{x_1, \dots, x_m} (p(x_1, \dots, x_m))^2 \\
&= 1 - \sum_{x_1, \dots, x_m} (\prod_{i=1}^m p(x_i))^2 = 1 - \sum_{x_1, \dots, x_m} (\prod_{i=1}^m (p(x_i))^2) \\
&= 1 - \prod_{i=1}^m (\sum_{x_i} (p(x_i))^2) = 1 - \prod_{i=1}^m (1 - H_{\mu l}(X_i)),
\end{aligned}$$

which is the desired result. □

Theorem 13. *If* $\mathbf{X} = (X_1, X_2, X_3, \dots)$ *is an independent information source, then*

1. $\lim_{n \to \infty} \prod_{i=1}^n (1 - H_{\mu l}(X_i)) = 1 - h_{\mu l}(\mathbf{X})$.
2. *If there exists* $k \in \mathbb{N}$, *such that* $H_{\mu l}(X_k) = 1$, *then* $h_{\mu l}(\mathbf{X}) = 1$.

Proof.

1. This follows from Theorem 12.
2. Let $H_{\mu l}(X_k) = 1$ for some $k \in \mathbb{N}$. Since $H_{\mu l}(X_k) = 1$,

$$1 - h_{\mu l}(\mathbf{X}) = \lim_{n\to\infty} \prod_{i=1}^{n} (1 - H_{\mu l}(X_i)) = 0.$$

Hence, $h_{\mu l}(\mathbf{X}) = 1$. □

Definition 11. *Let X_1 and X_2 be two random variables on Γ. Define the logical mutual information of X_2 and X_1 by*

$$I_{\mu l}(X_1, X_2) := H_{\mu l}(X_1) - H_{\mu l}(X_1|X_2).$$

Lemma 9. *Let X_1 and X_2 be two random variables on Γ. Then the following hold:*

1. $I_{\mu l}(X_1, X_2) = H_{\mu l}(X_2) - H_{\mu l}(X_2|X_1)$.
2. $I_{\mu l}(X_1, X_2) = H_{\mu l}(X_1) + H_{\mu l}(X_2) - H_{\mu l}(X_1, X_2)$.
3. $I_{\mu l}(X_1, X_2) = I_{\mu l}(X_2, X_1)$.
4. $I_{\mu l}(X_1, X_1) = H_{\mu l}(X_1)$.
5. *If X_1 and X_2 are independent random variables, then*

$$I_{\mu l}(X_1, X_2) = H_{\mu l}(X_1)H_{\mu l}(X_2).$$

Proof.

1–3. follows from Definition 11 and Theorem 7.

4. According to Lemma 3, $H_{\mu l}(X_1, X_1) = H_{\mu l}(X_1)$. Therefore,

$$\begin{aligned} I_{\mu l}(X_1, X_1) &= H_{\mu l}(X_1) + H_{\mu l}(X_1) - H_{\mu l}(X_1, X_1) \\ &= 2H_{\mu l}(X_1) - H_{\mu l}(X_1)) = H_{\mu l}(X_1). \end{aligned}$$

5. It follows from Lemma 12 that

$$\begin{aligned} H_{\mu l}(X_1, X_2) &= 1 - (1 - H_{\mu l}(X_1))(1 - H_{\mu l}(X_2)) \\ &= H_{\mu l}(X_1) + H_{\mu l}(X_2) - H_{\mu l}(X_1)H_{\mu l}(X_2). \end{aligned}$$

Hence, the result follows from 2. □

Definition 12. *Let $\mathbf{X} = (X_1, X_2, X_3, \ldots)$ be an information source. Define the logical mutual information of $X_1,\ldots,X_m$ by*

$$I_{\mu l}(X_1, \ldots, X_m) := \sum_{i=1}^{m} H_{\mu l}(X_i) - H_{\mu l}(X_1, \ldots, X_m).$$

Lemma 10. *Let X_1 and X_2 be two random variables on Γ. Then*

$$0 \le H_{\mu l}(X_2|X_1) \le H_{\mu l}(X_2).$$

Proof. It follows from Theorem 8 that

$$H_{\mu l}(X_1, X_2) = H_{\mu l}(X_1) + H_{\mu l}(X_2|X_1)$$

and from Theorem 3 that $H_{\mu l}(X_1, X_2) \leq H_{\mu l}(X_1) + H_{\mu l}(X_2)$. Hence,

$$H_{\mu l}(X_1) + H_{\mu l}(X_2|X_1) \leq H_{\mu l}(X_1) + H_{\mu l}(X_2).$$

This means that $H_{\mu l}(X_2|X_1) \leq H_{\mu l}(X_2)$. □

Theorem 14. *Let X_1 and X_2 be two random variables on Γ. Then the following holds:*

$$0 \leq I_{\mu l}(X_1, X_2) \leq \min\{H_{\mu l}(X_1), H_{\mu l}(X_1)\}.$$

Proof. From Lemma 9, it follows that

$$I_{\mu l}(X_1, X_2) = H_{\mu l}(X_1) + H_{\mu l}(X_2) - H_{\mu l}(X_1, X_2).$$

Furthermore, Theorem 3 yields $H_{\mu l}(X_2) \leq H_{\mu l}(X_1, X_2)$. Hence,

$$\begin{aligned} I_{\mu l}(X_1, X_2) &= H_{\mu l}(X_1) + H_{\mu l}(X_2) - H_{\mu l}(X_1, X_2) \\ &\leq H_{\mu l}(X_1) + H_{\mu l}(X_1, X_2) - H_{\mu l}(X_1, X_2) \\ &= H_{\mu l}(X_1). \end{aligned}$$

Similarly, $I_{\mu l}(X_1, X_2) \leq H_{\mu l}(X_2)$; therefore,

$$I_{\mu l}(X_1, X_2) \leq \min\{H_{\mu l}(X_1), H_{\mu l}(X_1)\}.$$

On the other hand, (2) yields

$$H_{\mu l}(X_2) + H_{\mu l}(X_2) - H_{\mu l}(X_1, X_2) \geq 0.$$

Therefore, $I_{\mu l}(X_1, X_2) \geq 0$ and, thus,

$$0 \leq I_{\mu l}(X_1, X_2) \leq \min\{H_{\mu l}(X_1), H_{\mu l}(X_1)\}.$$

□

3. Logical Entropy of Maps

Definition 13. *Let $f : \Gamma \longrightarrow \Gamma$ be a measurable function and $\alpha = \{\alpha_1, \ldots, \alpha_n\}$ be a partition of Γ. The logical metric entropy of order m of f with respect to the partition α is defined by*

$$h_{\mu l,m}(f, \alpha) = 1 - \sum_{1 \leq x_0, \ldots, x_m \leq n} (\mu(\alpha_{x_0} \cap f^{-1}(\alpha_{x_1}) \cap \ldots \cap f^{-m}((\alpha_{x_m})))^2, \tag{7}$$

and the logical metric entropy of f with respect to the partition α is defined by

$$h_{\mu l}(f, \alpha) = \lim_{m \to \infty} h_{\mu l,m}(f, \alpha). \tag{8}$$

The limits in (7) and (8) exist (see Theorem 15). The logical metric entropy of f is defined by $h_{\mu l}(f) = \sup_\alpha h_{\mu l}(f, \alpha)$.

Remark 2. $0 \leq h_{\mu l}(f) \leq 1$.

Let I be an interval, $h : I \longrightarrow I$ be a function and $x \in I$. For the finite orbit $\{h^n(x) : 0 \leq n \leq L-1\}$, we say that x is of type ordinal L-pattern $\pi = \pi(x) = (\pi_0, \ldots, \pi_{L-1})$ if

$$h^{\pi_0}(x) < h^{\pi_1}(x) < \ldots < h^{\pi_{L-1}}(x).$$

We denote P_π the set of $x \in I$ that are of type π.

Definition 14. *The logical metric permutation entropy of order m of f is defined by*

$$H^{\star}_{\mu l,m}(f) := 1 - \sum_{\pi \in \mathcal{S}_m} (\mu(p_\pi))^2,$$

and the logical metric permutation entropy of f is defined by

$$h^{*}_{\mu l}(f) := \lim_{m\to\infty} H^{\star}_{\mu l,m}(f) = 1 - \lim_{m\to\infty} \sum_{\pi \in \mathcal{S}_m} (\mu(p_\pi))^2.$$

Theorem 15. *Given $A = \{0, 1, \ldots, n-1\}$ with $X_m : [0,1] \longrightarrow A$, is defined as follows:*

$$X_m(x) = i \Longleftrightarrow f^m(x) \in \alpha_i.$$

Then $h_{\mu l}(f, \alpha) = h_{\mu l}(\mathbf{X})$ where $\mathbf{X}$ is a stationary process $(X_0, X_1, \ldots)$.

Proof. Since

$$\begin{aligned} H_{\mu l}(X_m) &= 1 - \sum_{i=0}^{n-1} (\mu(\{x : f^m(x) \in \alpha_i\}))^2 \\ &= 1 - \sum_{i=0}^{n-1} (\mu(f^{-m}\alpha_i))^2, \end{aligned}$$

we have

$$\begin{aligned} p(x_0, \ldots, x_m) &= \mu(\{x : x_0(x) = x_0, \ldots, x_m(x) = x_m\}) \\ &= \mu(\{x : x \in \alpha_{x_0}, \ldots, f^m(x) \in \alpha_{x_m}\}) \\ &= \mu(\alpha_{x_0} \bigcap f^{-1}(\alpha_{x_1}) \cap \ldots \bigcap f^{-m}((\alpha_{x_m})). \end{aligned}$$

Hence,

$$H_{\mu l}(X_0^m) = 1 - \sum_{x_0^m} (\mu(\alpha_{x_0} \bigcap f^{-1}(\alpha_{x_1}) \bigcap \ldots \bigcap f^{-m}((\alpha_{x_m})))^2, \tag{9}$$

and so (9) implies that $h_{\mu l}(f, \alpha) = h_{\mu l}(\mathbf{X})$. □

4. Examples and Applications in Logistic and Tent Maps

Example 1. *Let $g(x) = 4x(1-x) : [0,1] \longrightarrow [0,1]$ be the logistic map (see Figures 1 and 2 and Table 1). Then*

$$\begin{aligned} &p_{(0,1)} = (0, \frac{3}{4}),\ p_{(1,0)} = (\frac{3}{4}, 1), \\ &p_{(0,1,2)} = (0, \frac{1}{4}),\ p_{(0,2,1)} = (\frac{1}{4}, \frac{5-\sqrt{5}}{8}),\ p_{(2,0,1)} = (\frac{5-\sqrt{5}}{8}, \frac{3}{4}), \\ &p_{(1,0,2)} = (\frac{3}{4}, \frac{5+\sqrt{5}}{8}),\ p_{(1,2,0)} = (\frac{5+\sqrt{5}}{8}, 1),\ p_{(2,1,0)} = \emptyset. \end{aligned}$$

Therefore,

$$\begin{aligned} &\sum_{\pi \in \mathcal{S}_2} (\mu(p_\pi))^2 = (\frac{3}{4})^2 + (\frac{1}{4})^2 = \frac{5}{8}, \\ &H^{\star}_{\mu l,2}(g) = 1 - \frac{5}{8} = 0/375, \end{aligned}$$

$$\sum_{\pi \in \mathcal{S}_3} (\mu(p_\pi))^2 = (\frac{1}{4})^2 + (\frac{3-\sqrt{5}}{8})^2 + (\frac{1+\sqrt{5}}{8})^2 + (\frac{\sqrt{5}-1}{8})^2 + (\frac{3-\sqrt{5}}{8})^2,$$
$$= \frac{17-6\sqrt{5}}{32}$$

and

$$H^{\star}_{\mu l,3}(g) = 1 - \frac{17-6\sqrt{5}}{32} = \frac{15+6\sqrt{5}}{32} \simeq 0/888.$$

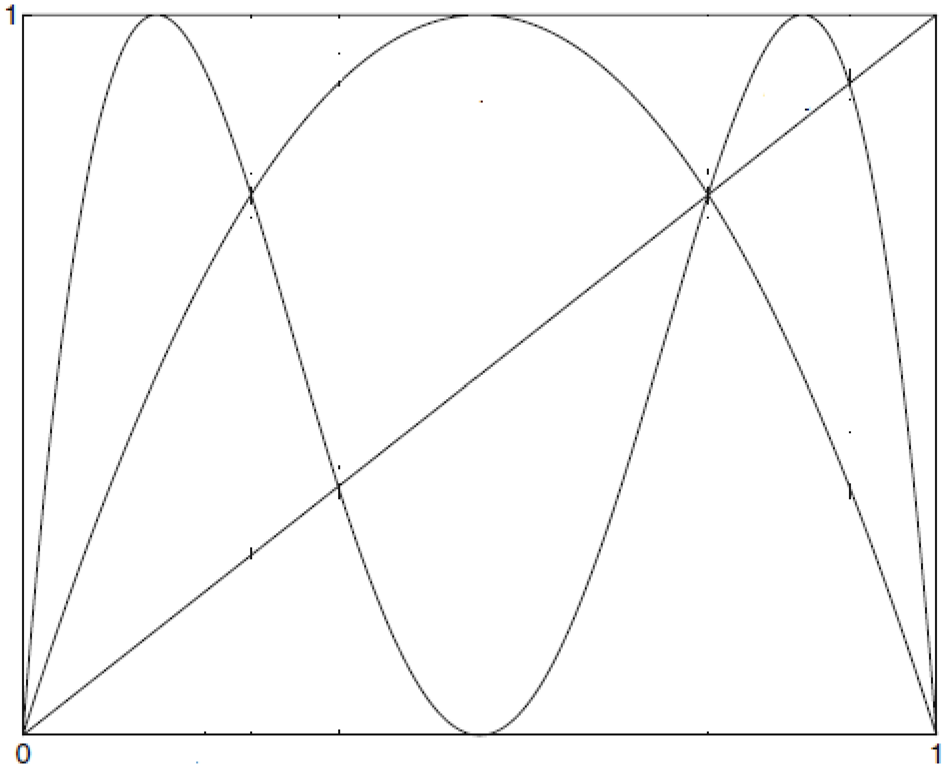

Figure 1. $x, g(x)$, and $g^2(x)$.

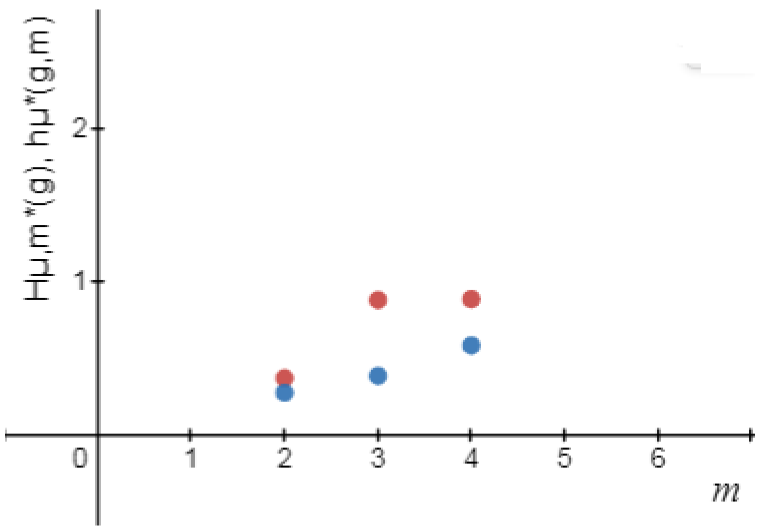

Figure 2. $H^{\star}_{\mu l,m}(g)$, and $h^{\star}_{\mu}(g,m)$.

Table 1. Logical metric permutation entropy and metric permutation entropy [1] for the logistic map up to order $m = 3$.

m	1	2	3
$H^{\star}_{\mu l,m}(g)$	0	0/375	0/888
$h^{\star}_{\mu}(g,m)$	0	0/28	0/39

Example 2. *Ref. [1] Let* $\alpha = \{\alpha_0 = [0,\frac{1}{2}], \alpha_1 = (\frac{1}{2},1]\}$*. We consider the tent map (see Figures 3 and 4 and Table 2)* $\Lambda : [0,1] \longrightarrow [0,1]$ *by*

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

Define $X_m : [0,1] \longrightarrow \{0,1\}$ *via*

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

for every $m \ge 0$*. Let*

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

Given $\alpha_{i_1 \ldots i_m}$*, where* $m \in \mathbb{N}$*, set*

$$\alpha_{i_1 \ldots i_m 0} = \alpha_{i_1 \ldots i_m} \bigcap \{x \in [0,1] : \Lambda^m(x) \in \alpha_0\},$$

$$\alpha_{i_1 \ldots i_m 1} = \alpha_{i_1 \ldots i_m} \bigcap \{x \in [0,1] : \Lambda^m(x) \in \alpha_1\},$$

and

$$\alpha_{i_0 i_1 \ldots i_m} = \bigcap_{k=0}^{m} \Lambda^{-k} \alpha_{i_k}.$$

Therefore,

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

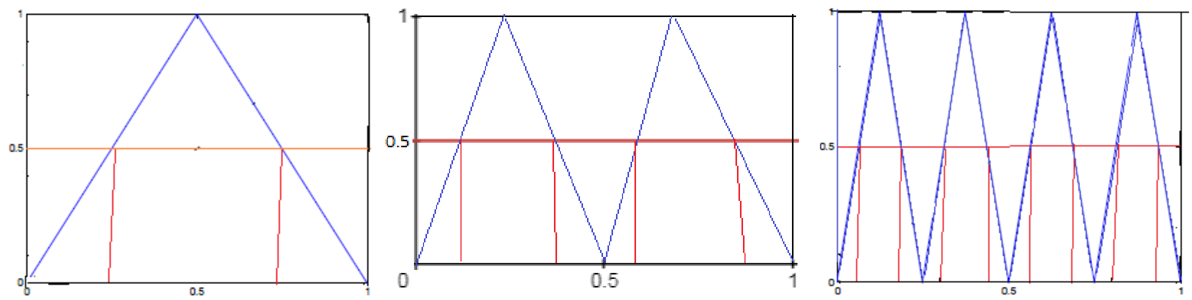

Figure 3. Λ, Λ^2 and Λ^3.

The sets $\{\alpha_{i_0 \dots i_{m-1}}\}$ *are identical to the binary sequences of* 0, 1 *in length* m. *Hence,* $\mu(\alpha_{i_0 \dots i_{m-1}}) = \frac{1}{2^m}$ *and, thus,*

$$H_{\mu l}(X_0^{m-1}) = 1 - \sum_{i_0 \dots i_{m-1}} (\mu(\alpha_{i_1 \dots i_m}))^2 = 1 - \sum_{i_0^{m-1}} (\frac{1}{2^m})^2$$
$$= 1 - 2^m \times \frac{1}{2^{2m}} = \frac{2^m - 1}{2^m}.$$

So, $h_{\mu l}(\mathbf{X}) = 1$, $h_{\mu l}(\Lambda, \alpha) = 1$ *and* $h_{\mu l}(\Lambda) = 1$. *Furthermore, if*

$$\mathbf{X} = (X_1, X_1, X_1, \dots),$$

then $H_{\mu l}(X_1, \dots, X_1) = H_{\mu l}(X_1)$ *and* $h_{\mu l}(\mathbf{X}) = \frac{1}{2}$.

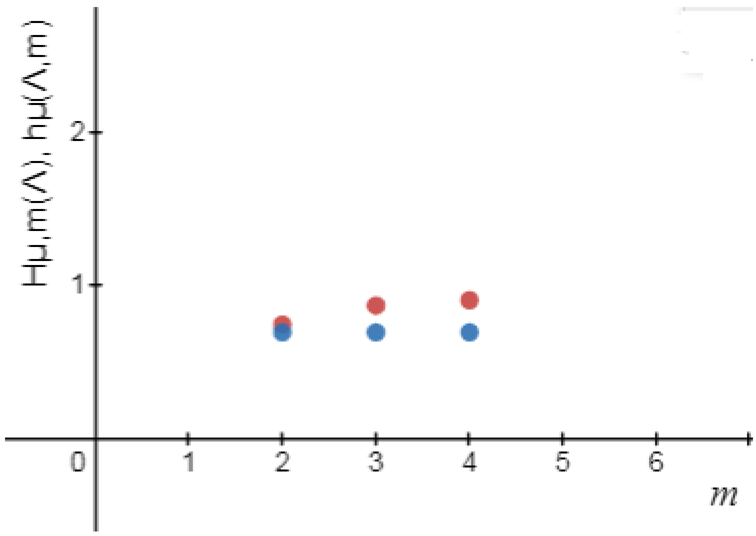

Figure 4. $H_{\mu l,m}(\Lambda)$, and $h_\mu(\Lambda, m)$.

Table 2. Logical metric entropy and metric entropy [1] for the tent map up to order $m = 3$.

m	1	2	3	...	m
$H_{\mu l,m}(\Lambda)$	0	0/75	0.875	...	$1 - \frac{1}{2^m}$
$h^\star_\mu(\Lambda, m)$	0	ln 2	ln 2	...	ln 2

Example 3. *Reference [1]. Consider the symmetric tent map in Example 2, we obtain (Figures 5 and 6 and Table 3)*

$$
\begin{aligned}
&p_{(0,1)} = (0, \tfrac{2}{3}),\ p_{(1,0)} = (\tfrac{2}{3}, 1),\\
&p_{(0,1,2)} = (0, \tfrac{1}{3}),\ p_{(0,2,1)} = (\tfrac{1}{3}, \tfrac{2}{5}),\ p_{(2,0,1)} = (\tfrac{2}{5}, \tfrac{2}{3}),\\
&p_{(1,0,2)} = (\tfrac{2}{3}, \tfrac{4}{5}),\ p_{(1,2,0)} = (\tfrac{4}{5}, 1),\ p_{(2,1,0)} = \emptyset,\\
&p_{(0,1,2,3)} = (0, \tfrac{1}{6}),\ p_{(0,1,3,2)} = (\tfrac{1}{6}, \tfrac{1}{5}),\ p_{(0,3,1,2)} = (\tfrac{1}{5}, \tfrac{2}{9}) \cup (\tfrac{2}{7}, \tfrac{1}{3}),\\
&p_{(3,0,1,2)} = (\tfrac{2}{9}, \tfrac{2}{7}),\ p_{(0,2,1,3)} = (\tfrac{1}{3}, \tfrac{2}{5}),\ p_{(2,0,3,1)} = (\tfrac{2}{5}, \tfrac{4}{9}) \cup (\tfrac{4}{7}, \tfrac{3}{5}),\\
&p_{(2,3,0,1)} = (\tfrac{4}{9}, \tfrac{4}{7}),\ p_{(2,0,1,3)} = (\tfrac{3}{5}, \tfrac{2}{3}),\ p_{(3,1,0,2)} = (\tfrac{2}{3}, \tfrac{4}{5}),\\
&p_{(1,3,2,0)} = (\tfrac{4}{5}, \tfrac{5}{6}),\ p_{(1,2,0,3)} = (\tfrac{6}{7}, \tfrac{8}{9}),\ p_{(1,2,3,0)} = (\tfrac{5}{6}, \tfrac{6}{7}) \cup (\tfrac{8}{9}, 1).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\sum_{\pi \in \mathcal{S}_2} (\mu(p_\pi))^2 = (\tfrac{2}{3})^2 + (\tfrac{1}{3})^2 = \tfrac{5}{9} \simeq 0/556,\\
&H^{\star}_{\mu l,2}(\Lambda) = 1 - \tfrac{5}{9} = \tfrac{4}{9} \simeq 0/444,\\
&\sum_{\pi \in \mathcal{S}_3} (\mu(p_\pi))^2 = (\tfrac{1}{3})^2 + (\tfrac{1}{15})^2 + (\tfrac{4}{15})^2 + (\tfrac{2}{15})^2 + (\tfrac{1}{5})^2 = \tfrac{11}{45} \simeq 0/244,\\
&H^{\star}_{\mu l,3}(\Lambda) = 1 - \tfrac{11}{45} = \tfrac{34}{45} \simeq 0/756,\\
&\sum_{\pi \in \mathcal{S}_4} (\mu(p_\pi))^2 \simeq 0.106,\\
&H^{\star}_{\mu l,4}(\Lambda) \simeq 0/894.
\end{aligned}
$$

Furthermore,

$$
h^{\star}_{\mu l}(\Lambda) = \lim_{m \to \infty} H^{\star}_{\mu l,m}(\Lambda) = 1 = h_{\mu l}(\Lambda).
$$

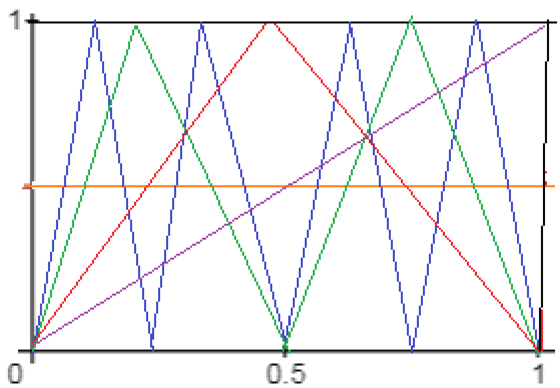

Figure 5. $x, \Lambda(x), \Lambda^2(x)$ and $\Lambda^3(x)$.

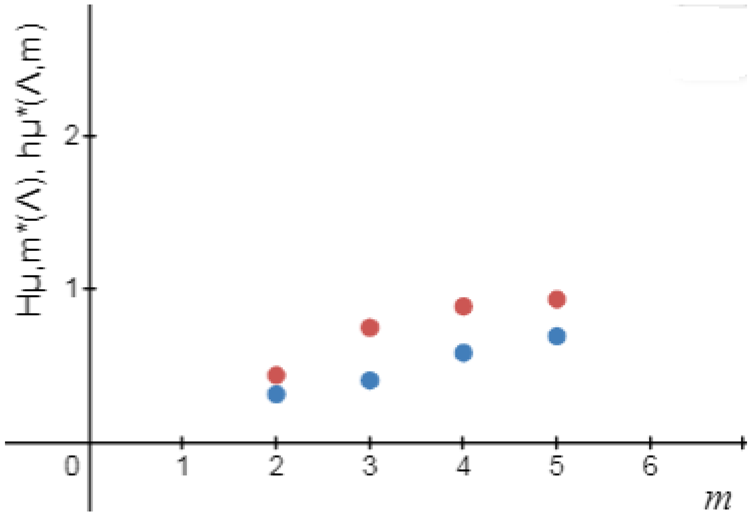

Figure 6. $H^*_{\mu l,m}(\Lambda)$, and $h^*_\mu(\Lambda, m)$.

Table 3. Logical metric permutation entropy and metric permutation entropy [1] for the tent map up to order $m = 4$.

m	1	2	3	4
$H^\star_{\mu l,m}(\Lambda)$	0	0/444	0/756	0/894
$h^\star_\mu(\Lambda, m)$	0	0/32	0/41	0/59

Example 4. *Let* $I = [0,1]$ *endowed with the measure* ν*,*

$$\nu(A) = \chi_A(\frac{1}{2}) = \begin{cases} 1 & if \ \frac{1}{2} \in A \\ 0 & if \ \frac{1}{2} \notin A, \end{cases}$$

and let $f : [0,1] \longrightarrow [0,1]$ *be a function. Then* $h_{\nu l}(f, \alpha) = 0$ *for every partition* α*. Hence,* $h_{\nu l}(f) = 0$*.*

5. Concluding Remarks

We introduced the concept of the logical entropy of random variables. In addition, we found a bound for the logical entropy of a random variable. We also extended the Shannon and permutation entropies to information sources. Finally, we used these results to estimate the logical entropy of the maps. In this article, we only introduced the concept of logical entropy for information systems. In future studies, researchers can find methods that calculate or estimate the numerical value of this type of entropy. It is pertinent to mention that, in the future, Rényi's metric entropy and Rényi's permutation entropy can be generalized for information sources. Another important problem is to extend this idea for quantum logical entropy, as it is a good direction to investigate the existence of such results.

Author Contributions: Conceptualization, Y.S.; Formal analysis, Y.S.; Funding acquisition, P.X.; Investigation, Y.S.; Methodology, Y.S. and S.I.B.; Supervision, P.X. and S.I.B.; Validation, P.X.; Writing—original draft, Y.S.; Writing—review & editing, P.X. and S.I.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded in part by the National Natural Science Foundation of China (grant no. 62002079).

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In this appendix, we prove the following results that we need in the paper.

Lemma A1. *Let $\mathcal{M} = [\theta_{ij}]_{n\times n}$ be a matrix that $0 \leq \theta_{ij} \leq 1$ for every $1 \leq i,j \leq n$ and $\sum_{i,j}\theta_{ij} = 1$, then*

$$\sum_{i=1}^{n}(\sum_{j=1}^{n}\theta_{ij})^2 + \sum_{j=1}^{n}(\sum_{i=1}^{n}\theta_{ij})^2 \leq 1 + \sum_{i,j}\theta_{ij}^2.$$

Proof. Since

$$\begin{aligned}\sum_{i=1}^{n}(\sum_{j=1}^{n}\theta_{ij})^2 + \sum_{j=1}^{n}(\sum_{i=1}^{n}\theta_{ij})^2 &= \sum_{i,j}\theta_{ij}^2 + (\sum_{i,j}\theta_{ij})^2 - 2\sum_{i\neq k, j<l}\theta_{ij}\theta_{kl} \\ &= \sum_{i,j}\theta_{ij}^2 + 1 - 2\sum_{i\neq k, j<l}\theta_{ij}\theta_{kl} \leq 1 + \sum_{i,j}\theta_{ij}^2,\end{aligned}$$

the assertion is proved. □

Theorem A1. *Let a_{ijk} be real numbers and $a_{ijk} \geq 0$ for every $1 \leq i \leq n_1$, $1 \leq i \leq n_2$ and $1 \leq i \leq n_3$. Then*

$$2\sum_{i,j}(\sum_{k}a_{ijk})^2 \leq \sum_{i,j,k}a_{ijk}^2 + \sum_{i}(\sum_{j,k}a_{ijk})^2. \quad \text{(A1)}$$

Proof. Since

$$\begin{aligned}2\sum_{i,j}(\sum_{k}a_{ijk})^2 &= 2\sum_{i,j}(\sum_{k,r}a_{ijk}a_{ijr}) = 2\sum_{i,j,k,r}a_{ijk}a_{ijr} \\ &\leq \sum_{i,j,t,k,r}a_{ijk}a_{itr} = \sum_{i}(\sum_{j,k}a_{ijk})^2 \\ &\leq \sum_{i,j,k}a_{ijk}^2 + \sum_{i}(\sum_{j,k}a_{ijk})^2,\end{aligned}$$

which completes the proof of the theorem. □

Lemma A2. *For an information source $\mathbf{X} = (X_1, X_2, X_3, \ldots)$,*

$$\lim_{n\to\infty} H_{\mu l}(X_n|X_{n-1}, \ldots, X_1) = 0.$$

Proof. According to Theorem 9, the series $\sum_{n=1}^{\infty} H_{\mu l}(X_n|X_{n-1}, \ldots, X_1)$ converges and, thus, $\lim_{n\to\infty} H_{\mu l}(X_n|X_{n-1}, \ldots, X_1) = 0$. □

References

1. Amigo, J.M. *Permutation Complexity in Dynamical Systems "Ordinal Patterns, Permutation Entropy, and All That"*; Springer: Berlin/Heidelberg, Germany, 2010.
2. Mehrpooya, A.; Sayyari, Y.; Molaei, M.R. Algebraic and Shannon entropies of commutative hypergroups and their connection with information and permutation entropies and with calculation of entropy for chemical algebras. *Soft Comput.* **2019**, *23*, 13035–13053. [CrossRef]
3. Corda, C.; FatehiNia, M.; Molaei, M.R.; Sayyari, Y. Entropy of iterated function systems and their relations with black holes and Bohr-like black holes entropies. *Entropy* **2018**, *20*, 56. [CrossRef] [PubMed]
4. Walters, P. *An Introduction to Ergodic Theory*; Springer: New York, NY, USA, 1982.
5. Ellerman, D. Logical Entropy: Introduction to Classical and Quantum Logical Information Theory. *Entropy* **2018**, *20*, 679. [CrossRef] [PubMed]
6. Natal, J.; Avila, I.; Tsukahara, V.B.; Pinheiro, M.; Maciel, C.D. Entropy, From Thermodynamics to Information Processing. *Entropy* **2021**, *23*, 1340. [CrossRef] [PubMed]
7. Markechová, D.; Riečan, B. Logical entropy of fuzzy dynamical systems. *Entropy* **2016**, *18*, 157. [CrossRef]

8. Ellerman, D. The logic of partitions: Introduction to the dual of the logic of subsets. *Rev. Symb. Log.* **2010**, *3*, 287–350. [CrossRef]
9. Sayyari, Y. An improvement of the upper bound on the entropy of information sources. *J. Math. Ext.* **2021**, *15*, 1–12.
10. Butt, S.I.; Mehmood, N.; Pečarić, D.; Pečarić, J. New bounds for Shannon, relative and Mandelbrot entropies via Abel-Gontscharoff interpolating polynomial. *Math. Inequal. Appl.* **2019**, *22*, 1283–1301. [CrossRef]
11. Mehmood, N.; Butt, S.I.;Pečarić, D.; Pečarić, J. New bounds for Shannon, relative and Mandelbrot entropies via Hermite interpolating polynomial. *Demonstr. Math.* **2018**, *51*, 112–130. [CrossRef]
12. Amigo, J.M.; Kennel, M.B. Topological Permutation Entropy. *Phys. Nonlinear Phenom.* **2007**, *231*, 137–142. [CrossRef]
13. Ellerman, D. An introduction to logical entropy and its relation to Shannon entropy. *Int. J. Semant. Comput.* **2013**, *7*, 121–145. [CrossRef]
14. Markechová, D.; Riečan, B. Logical Entropy and Logical Mutual Information of Experiments in the Intuitionistic Fuzzy Case. *Entropy* **2017**, *19*, 429. [CrossRef] [PubMed]
15. Ebrahimzadeh, A. Logical entropy of quantum dynamical systems. *Open Phys.* **2016**, *14*, 1–5. [CrossRef]
16. Markechová, D.; Ebrahimzadeh, A.; Giski, Z.E. Logical entropy of dynamical systems. *Adv. Differ. Equ.* **2018**, *2018*, 70. [CrossRef]
17. Tamir, B.; Cohen, E. Logical Entropy for Quantum States. *arXiv* **2014**, arXiv:1412.0616v2.
18. Ellerman, D. Information as Distinction: New Foundation for Information Theory. *arXiv* **2013**, arXiv:1301.5607.
19. Brukner, C.; Zeilinger, A. Conceptual Inadequacy of the Shannon Information in Quantum Measurements. *Phys. Rev. A* **2001**, *63*, 022113. [CrossRef]
20. Sayyari, Y. New entropy bounds via uniformly convex functions. *Chaos Solitons Fractals* **2020**, *141*, 110360. [CrossRef]
21. Simic, S. Jensen's inequality and new entropy bounds. *Appl. Math. Lett.* **2009**, *22*, 1262–1265. [CrossRef]