Journal of Mathematical Extension
Vol. 14, No. 3, (2020), 97-115
ISSN: 1735-8299
URL: http://www.ijmex.com
Original Research Paper

Multi-Dimensional Observers and Relative Entropy of Dynamical Systems

M. H. Asadian*
Bandar Abbas Branch, Islamic Azad University

M. R. Molaei
Bandar Abbas Branch, Islamic Azad University

Abstract. In this paper the notion of relative probability measure of a set E is considered with respect to a multi-dimensional observer of a set X as a superset of E. Relative entropy of a multi-dimensional observer for the partitions is defined and the properties of relative entropy is extended to multi-dimensional observers. It is shown that the observer of a set plays a role in uncertainty of a partition of it. Relative conditional entropy is also considered and its main properties are proved. Moreover, the relative entropy off a relative measure preserving map is studied as well.

AMS Subject Classification: 37B51; 83A05
Keywords and Phrases: Multi-dimensional observer, fuzzy information system, relative entropy, relative conditional entropy, dynamical system

1. Introduction

In this section, multi-dimensional observer and also the notions of relative probability measure of E with respect to multi-dimensional observer μ are defined. The notion of multi-dimensional observer as an extension of one dimensional observer [11] has been introduced first in 2009

Received: June 2018; Accepted: April 2019
*Corresponding author

[12]. The notion of one dimensional observer is very important in physics and has been applied in information theory [5, 6]. A new concept of topological entropy has been presented in [12] via multi-dimensional observer. If X is a set, then a multi-dimensional observer of X is a mapping $\mu : X \to \prod_{i\in I}[0,1]$ where I is an index set and $\prod_{i\in I}[0,1] = \{g : I \to [0,1] \mid g$ is a mapping $\}$.

For example each information system can be considered as a multi-dimensional observer. We recall that [9, 13] an information system is a triple (X, I, F) where X, and I are non-empty finite sets, and $F = \{f_i \mid f_i \textit{ is a map on } X, \textit{ and } i \in I\}$. Since X is finite, then the image of each f_i is a finite set, so it can be correspond to a finite subset of the interval $[0,1]$ via a one to one mapping g_i. Thus up to these correspondences an information system (X, I, F) can be denoted by a finite dimensional observer $\mu : X \to \prod_{i\in I}[0,1]$ defined by $\mu(x) = (g_1(f_1(x)), g_2(f_2(x)), ..., g_{|I|}(f_{|I|}(x))$, where X and I are finite sets, and $x \in X$.

Fuzzy information system is considered an information system by allowing each f_i to take it's values in the interval $[0,1]$ [10]. Hence it is also an example of a multi-dimensional observer.

It is known that many nature processes are modeled by stochastic (or random) process on finite spaces. In fact a stochastic process [2, 3] on a finite space X is a sequence $\mathbf{S} = (S_n)$, where (S_n) is a random variable on X with values in $A = \{a_1, ..., a_{|A|}\}$ where $n \in I$, and I is N or $N_0 = N \bigcup \{0\}$ or Z. If we correspond the image of each S_n with a finite subset of the interval $[0,1]$ by a one to one map g_n, then a stochastic process $\mathbf{S} = (S_n)$ can be considered as a multi-dimensional observer $\mu : X \to \prod_{i\in I}[0,1]$ defined by $\mu = (g_n o S_n)$, where I is N or $N_0 = N \bigcup \{0\}$ or Z.

Therefore an information system, a fuzzy information system and a stochastic process are examples of multi-dimensional observers. The reader must pay attention to this point that the index set I in the definition of a multi-dimensional observer can be an uncountable set.

Let X be a set and $f : X \to X$ is a mapping. We denote $f^1 = f$, $f^2 = fof, \ldots, f^{n-1} = fo...of$, so we assume that $\{f^n \ : \ n \, is \, a \, natural \, number\}$ is a semi-dynamical system on X. This dynamics on the set X creates a new kind of measurement from an observer viewpoint which is mentioned in the next definition.

Definition 1.1. *If E is a subset of X, then the relative probability measure of E with respect to a multi-dimensional observer μ is the mapping $\vec{m}^f_\mu(E) : X \to \prod_{i \in I}[0,1]$ defined by:*
$\vec{m}^f_\mu(E)(x) = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_E(f^i(x))\mu(f^i(x))$, where χ_E is the characteristic function of E.

Thus, the image of $x \in X$ given by $\vec{m}^f_\mu$ is the function $g : X \to [0,1]$. Relative probability measure is an extension of probability measure. To show this: we assume that (X, B, m) is a probability space, and $f : (X, B, m) \to (X, B, m)$ is an ergodic map. If E is a member of the σ-algebra B, and if the one dimensional observer $\mu : X \to [0,1]$ is the characteristic function of X, then Birkhoff ergodic theorem [14] implies that

$$\vec{m}^f_\mu(E)(x) = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \chi_E(f^i(x))\chi_X(f^i(x)) = m(E) \; a.e., \; for \; all \; x \in X.$$

Thus in the crisp case the relative probability measure of measurable sets is equal to the probability measure of them, almost everywhere.

The researchers show that relative probability measure has subadditivity property, and if a sequence of intricate subsets of X tend to it, then their relative probability measures tend to the relative probability measure of X. An equivalence relation on partitions which preserve the relative entropy of a multi-dimensional observer is found. Conditional entropy from an observer viewpoint is considered in section four. If we denote the relative entropy of a multi-dimensional observer of partition $\mathcal{A}$ given $\mathcal{C}$ at $x \in X$ by $\vec{H}^f_\mu(\mathcal{A}/\mathcal{C})(x)$, then in section 4 we prove that:

(i) $\vec{H}^f_\mu(\mathcal{A} \vee \mathcal{C}/\mathcal{D})(x) = \vec{H}^f_\mu(\mathcal{A}/\mathcal{D})(x) + \vec{H}^f_\mu(\mathcal{C}/\mathcal{A} \vee \mathcal{D})(x)$.

(ii) If $\mathcal{A} \subseteq \mathcal{C}$ then $\vec{H}^f_\mu(\mathcal{A})(x) \leqslant \vec{H}^f_\mu(\mathcal{C})(x)$.

(iii) If $\mathcal{C} \subseteq \mathcal{D}$ then $\vec{H}^f_\mu(\mathcal{A}/\mathcal{C})(x) \geqslant \vec{H}^f_\mu(\mathcal{A}/\mathcal{D})(x)$.

(iv) $\vec{H}^f_\mu(\mathcal{A} \vee \mathcal{C}/\mathcal{D})(x) \leqslant \vec{H}^f_\mu(\mathcal{A}/\mathcal{D})(x) + \vec{H}^f_\mu(\mathcal{C}/\mathcal{D})(x)$.

The researchers also consider independent finite partitions.

In section five relative entropy of a relative measure preserving map is defined and the main properties of it are deduced. Some invetigations concerning entropy of dynamical systems and the fuzzy case can be seen in [7, 8].

2. Properties of Relative Probability Measure

In this section, the researchers provide a study of some properties of relative probability measure and provide examples. Let μ be a multi-dimensional observer of X and let $\vec{m}^f_\mu$ be the relative probability measure created by μ and a dynamics $f : X \to X$. Moreover let A, B be two subsets of X with $A \cap B \neq \emptyset$. Then for $x \in X$ we have:

$$\vec{m}^f_\mu(A \cup B)(x) = \limsup \frac{1}{n} \Sigma_{i=0}^{n-1} \chi_{A\cup B}(f^i(x)).\mu(f^i(x)) =$$

$$\limsup \frac{1}{n} \Sigma_{i=0}^{n-1} (\chi_A(f^i(x)) + \chi_B(f^i(x))).\mu(f^i(x)) =$$

$$\limsup(\frac{1}{n} \Sigma_{i=0}^{n-1} \chi_A(f^i(x)).\mu(f^i(x)) + \frac{1}{n} \Sigma_{i=0}^{n-1} \chi_B(f^i(x)).\mu(f^i(x))) =$$

$$\vec{m}^f_\mu(A)(x) + \vec{m}^f_\mu(B)(x).$$

Assume that E, F are two subsets of X and $E \subseteq F$ then $F = E \cup (F - E)$. Hence the above discussion implies that $\vec{m}^f_\mu(F - E)(x) = \vec{m}^f_\mu(F)(x) - \vec{m}^f_\mu(E)(x)$, for all $x \in X$. Now we define an order on the values of relative probability measures. For this purpose we use the norm of bounded functions. In fact if $h : X \to R$ is a bounded function then it's norm is defined by $\|h\| = \sup_{x\in X} |h(x)|$.

If E and F are two subsets of X and μ and η are two multi-dimensional observers of X then we say that $\vec{m}^f_\mu(E)(x) \leqslant \vec{m}^f_\eta(F)(x)$ if and only if $\|\chi_E(x)\mu(x)\| \leqslant \|\chi_F(x)\eta(x)\|$.

The straightforward calculations imply that relative probability measure $\vec{m}^f_\mu(.)$ has the subadditivity property. Moreover if $E \subseteq F$ then $\vec{m}^f_\mu(E)(x) \leqslant \vec{m}^f_\mu(F)(x)$, for all $x \in X$.

Theorem 2.1. *If $A_1 \subseteq A_2 \subseteq ... \subseteq X$ and $A = \cup_{i=1}^{\infty} A_n$, then for all $x \in X$,$\lim_{n\to\infty} \vec{m}^f_\mu(A_n)(x) = \vec{m}^f_\mu(A)(x)$.*

Proof. If $B_1 = A_1$ and $B_n = A_n - A_{n-1} (n \geqslant 2)$ then$\cup_{n=1}^{\infty} B_n = \cup_{n=1}^{\infty} A_n = A$ and $B_n \cap B_m = \emptyset, (n \neq m)$. Hence

$$\lim_{n\to\infty} \vec{m}^f_\mu(A_n)(x) = \lim_{n\to\infty} \vec{m}^f_\mu(\cup_{i=1}^{n} B_i)(x) = \lim_{n\to\infty} \Sigma_{i=1}^{n} \vec{m}^f_\mu(B_i)(x)$$

$$= \Sigma_{i=1}^{\infty} \vec{m}^f_\mu(B_i)(x) \geqslant \vec{m}^f_\mu(A)(x).$$

Since $A_n \subseteq A$ then $\vec{m}^f_\mu(A_n)(x) \leqslant \vec{m}^f_\mu(A)(x)$. Thus $\lim_{n\to\infty} \vec{m}^f_\mu(A_n)(x) = \vec{m}^f_\mu(A)(x)$. $\square$

Example 2.2. Suppose $X = R\backslash\{0,1\}$ and $f : X \to X$ is defined by $x \mapsto \frac{1}{x}$. If $E = [0,1] \cap Q$ and $\mu : X \to \prod_{i \in I}[0,1]$ is defined by $x \mapsto \mu(x)$ where $\mu(x) : I \to [0,1]$ is the map $\mu(x)(i) = \frac{1}{i+x+1}$ then

$$\vec{m}^f_\mu(E)(x) = \begin{cases} \frac{1}{2}\mu(\frac{1}{x}) & if\ x \in (1,\infty) \cap Q \\ \frac{1}{2}\mu(x) & if\ x \in (0,1) \cap Q \\ 0 & otherwise. \end{cases}$$

3. Relative Entropy of a Multi-Dimensional Observer for Partitions

In this section, the notion of relative entropy of a multi-dimensional observer for partition $\mathcal{A}$ is defined and an example is provided. A partition of X is a disjoin collection of elements of $P(X)$ that its union is X, whenever $P(X)$ is a power set of X [11]. If ξ is a finite partition of X, then the collection of elements of $P(X)$ which are unions of elements of ξ is a finite subset of $P(X)$. We denote it by $\mathcal{A}(\xi)$ [11]. Conversely, if $\mathcal{C}$ is a finite subset of $P(X)$, then the non-empty sets of the form $B_1 \cap \cdots \cap B_n$, whenever $B_i = C_j$ or $X - C_j$ for all C_j belong to $\mathcal{C}$, form a finite partition of X. We denote it by $\xi(\mathcal{C})$. Suppose ξ and η

are two finite partitions of X. We write $\xi \leqslant \eta$ to mean that each element of ξ is a union of elements of η [11]. Let $\mathcal{A} = \{A_1, ..., A_n\}$, and $\mathcal{C} = \{C_1, ..., C_n\}$ be two finite partitions of X. Then their join is the partition: $\mathcal{A} \vee \mathcal{C} = \{A_i \cap C_j : 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant k\}$ [11]. A map $T : X \to X$ is measure-preserving if $\vec{m}^f_\mu(T^{-1}E)(x) = \vec{m}^f_\mu(E)(x)$for all $x \in X$ [11]. The map T can be called as a discrete dynamical system.

Let $\mathcal{C}, \mathcal{D}$ be two subsets of $P(X)$ we write $\mathcal{C} \subset^f_\mu \mathcal{D}$ if for every $C \in \mathcal{C}$ there exists $D \in \mathcal{D}$ with $\vec{m}^f_\mu(D \Delta C)(x) = 0$.

Definition 3.1. *Let $\mathcal{A}$ be a finite subset of $P(X)$ with $\xi(\mathcal{A}) = \{A_1, ..., A_n\}$. Then the relative entropy of a multi-dimensional observer of $\mathcal{A}$ is the number* $\vec{H}^f_\mu(\mathcal{A})(x) = -\sum_{i=1}^n \vec{m}^f_\mu(A_i)(x) \log \|\vec{m}^f_\mu(A_i)(x)\|$.

$\vec{H}^f_\mu(\mathcal{A})(x)$ is a multi-dimensional measure of the uncertainty removed by performing the experiment with outcome $\{A_1, ..., A_n\}$ when a multi-dimensional observer pointed to x. We assume $0 \log 0 = 0$.

If $\mathcal{A} = \{X\}$ then $\xi(\mathcal{A}) = \{X, \emptyset\}$. Thus

$$\vec{H}^f_\mu(X)(x) = -\vec{m}^f_\mu(\emptyset)(x) \log \|(\vec{m}^f_\mu(\emptyset)(x))\| - \vec{m}^f_\mu(X)(x) \log \|(\vec{m}^f_\mu(X)(x))\|$$
$$= -\vec{m}^f_\mu(X)(x) \log \|(\vec{m}^f_\mu(X)(x))\|$$

and if $\|\vec{m}^f_\mu(X)(x)\| = 1$ then $\vec{H}^f_\mu(X) = 0$. Here $\mathcal{A}$ represents the outcome of a certain experiment so there is no uncertainty about the outcome. Therefore the uncertainty of the observer is zero.

Example 3.2. Let (X, I, F) be a fuzzy information system, where $X = \{x_1, ...x_5\}$ and $I = I_{Age} \cup I_{Morality}$ [10], such that
$I_{Age} = I_{Ag} = \{Young(d_1), Middle(d_2), Between20and25(d_3), About50(d_4)\}$
$I_{Morality} = I_{Mo} = \{Good(d_5), Averge(d_6), Verygood(d_7), Outstanding(d_8)\}$.
Let $\Pi(I_{(Ag)j}) = \{\Pi_{ij} | \Pi_{ij} : I_{Ag} \longrightarrow [0, 1], i = 1, ..., 5\}$ and
$\Pi(I_{(Mo)j}) = \{\Pi_{ij} | \Pi_{ij} : I_{Mo} \longrightarrow [0, 1], i = 1, ..., 5\}$, where $j \in \{1, ..., m\}$. Moreover let $F = \{h_1, ..., h_m\}$, where $h_j : X \longrightarrow \Pi(I) = \Pi(I_{Ag}) \cup (\Pi(I_{Mo}))$ is defined by $h_j(x_i) = (\Pi_{ij}(d_1), \Pi_{ij}(d_2), ..., \Pi_{ij}(d_8))$.
$(\Pi_{ij}(d_k))$ means that: the possibility of d_k in $h_j(x_i)$ is equal to $\Pi_{ij}(d_k)$. The non-zero values of $\Pi_{ij}(d_k)$ are determined in table 1. We define an observer $\mu : X \to \Pi_{i \in I}[0, 1]$ by

$$\mu(x_i) = (\max\{\Pi_{ij}(d_k) |\ k = 1, ..., 4\}, \max\{\Pi_{ij}(d_k) |\ k = 5, ..., 8\})$$

and we define $f : X \longrightarrow X$ by $f(x_t) = x_{<t^2>_5}$, where $< a >_5$ is the remaining a to 5. If we assume $E = \{x_1, x_2\}$ then

$$\vec{m}^f_\mu(E)(x_2) = \limsup \frac{1}{n} \Sigma \chi_E(f^i(x_2))\mu(f^i(x_2))$$
$$= \limsup \frac{1}{n}(\mu(x_2) + (n-2)\mu(x_1)) = \mu(x_1)$$
$$= (\max\{0.8_{(d_1)}, 1_{(d_3)}\}, \max\{0.8_{(d_6)}, 0.4_{(d_5)}\}) = (1_{(d_1)}, 0.8_{(d_6)}).$$

Let $\mathcal{A} = \{\{x_1, x_2\}, \{x_3\}, \{x_4\}, \{x_5\}\}$. Then

$$\vec{H}^f_\mu(\mathcal{A})(x_2) = -\sum_{i=1}^{5} \vec{m}^f_\mu(A_i)(x_2) \log \|\vec{m}^f_\mu(A_i)(x_2)\| =$$
$$-\vec{m}^f_\mu(A_1)(x_2) \log \|\vec{m}^f_\mu(A_1)(x_2)\| - \vec{m}^f_\mu(A_2)(x_2) \log \|\vec{m}^f_\mu(A_2)(x_2)\| - ...$$
$$-\vec{m}^f_\mu(A_5)(x_2) \log \|\vec{m}^f_\mu(A_5)(x_2)\| =$$
$$-\mu(x_2) \log \|\mu(x_2)\| = -(0.7_{(d_2)}, 0.6_{(d_7)}) \log(0.7) = (0.25_{(d_2)}, 0.22_{(d_7)}).$$

Table 1: Fuzzy information system

X	AGE(AG)	Morality(MO)
x_1	$0.8_{(d_1)}, 1_{([20,25])}$	$0.8_{(d_6)}, 0.4_{(d_5)}$
x_2	$0.7_{(d_1)}, 0.2_{(d_2)}$	$0.6_{(d_6)}, 0.5_{(d_5)}$
x_3	$0.7_{(d_2)}, 0.2_{(d_4)}$	$1_{(d_7)}$
x_4	$0.7_{(d_3)}$	$0.9_{(d_6)}, 0.4_{(d_8)}$
x_5	$1_{(d_4)}$	$0.9_{(d_8)}$

Example 3.3. Suppose (X, I, F) is an information system about the grade marks of some students where $X = \{x_1, ..., x_5\}$ is a set of students in a college,

MA(mathematics), Ch(chemistry) and PH(physics),

$$I = \{d^k_i \mid d^k_i \text{ is the mark of } x_i \text{ in the college} k\}.$$

The values of d^k_i are shown in table (2) [10]. The researchers define a finite dimensional observer $\mu : X \longrightarrow \Pi^3_1[0,1]$ by

$$\mu(x_i) = (\frac{d^{MA}_i}{100}, \frac{d^{CH}_i}{100}, \frac{d^{PH}_i}{100})$$

and assume that $f : X \longrightarrow X$ is the identity function. If $E = \{x_1, x_2\}$ then

$$\vec{m}^f_\mu(E)(x_1) = \limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \chi_E(f^i(x_1))\mu(f^i(x_1)) = \mu(x_1) = (0.85, 0.9, 0.75).$$

For $j = 1, ..., 5$ we have

$$\vec{m}^f_\mu(X)(x_j) = \limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \chi_X(f^i(x_j))\mu(f^i(x_j)) =$$

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mu(f^i(x_j)) = \mu(x_j)$$

Let $\mathcal{A} = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_5\}\}$. Then

$$\vec{H}^f_\mu(\mathcal{A})(x_1) = -\sum_{i=1}^{5} \vec{m}^f_\mu(A_i)(x_1) \log \|\vec{m}^f_\mu(A_i)(x_1)\| =$$

$$-\vec{m}^f_\mu(A_1)(x_1) \log \|\vec{m}^f_\mu(A_1)(x_1)\| =$$

$$-\mu(x_1) \log \|\mu(x_1)\| = -\mu(x_1) \log(0.9) = 0.11\mu(x_1) = (0.09, 0.1, 0.08).$$

Also we have

$$\vec{H}^f_\mu(\mathcal{A})(x_4) = -\sum_{i=1}^{5} \vec{m}^f_\mu(A_i)(x_4) \log \|\vec{m}^f_\mu(A_i)(x_4)\| =$$

$$-\vec{m}^f_\mu(A_4)(x_4) \log \|\vec{m}^f_\mu(A_4)(x_4)\| =$$

$$-\mu(x_4) \log \|\mu(x_4)\| = -\mu(x_4) \log(0.86) = 0.15\mu(x_4) = (0.13, 0.13, 0.08).$$

If we compare $\vec{H}^f_\mu(\mathcal{A})(x_1)$ and $\vec{H}^f_\mu(\mathcal{A})(x_4)$, then we see that the relative entropy of $\mathcal{A}$ when observer is pointed to x_4 is more than the case in which observer is pointed to x_1. Thus the student x_1 is more rulable than x_4.

Let $\mathcal{A}$ be a finite partition of $P(X)$. Since $\vec{m}^f_\mu(A_i)(x) \leqslant \vec{m}^f_\mu(X)(x)$ then $\|\vec{m}^f_\mu(A_i)(x)\| \leqslant \|\vec{m}^f_\mu(X)(x)\|$.

Table 2: Fuzzy information system of success in lessons.

X	MA	CH	PH
x_1	85	90	75
x_2	86	90	75
x_3	90	87	100
x_4	86	86	50
x_5	87	85	25

Hence $\log \|\vec{m}_\mu^f(A_i)(x)\| \leqslant \log \|\vec{m}_\mu^f(X)(x)\|$.
Thus

$$\vec{m}_\mu^f(A_i)(x) \log \|\vec{m}_\mu^f(A_i)(x)\| \leqslant \vec{m}_\mu^f(A_i)(x) \log \|\vec{m}_\mu^f(X)(x)\|$$

$$\leqslant \vec{m}_\mu^f(X)(x) \log \|\vec{m}_\mu^f(X)(x)\|.$$

Hence $\vec{H}_\mu^f(\mathcal{A})(x) \geqslant \vec{H}_\mu^f(X)(x)$.

One can easily prove that if $T : X \to X$ is a relative probability measure-preserving map then $\vec{H}_\mu^f(T^{-1}\mathcal{A})(x) = \vec{H}_\mu^f(\mathcal{A})(x)$, for all $x \in X$.

Theorem 3.4. *If $\mathcal{A} =_\mu^f \mathcal{C}$ then $\vec{H}_\mu^f(\mathcal{A})(x) = \vec{H}_\mu^f(\mathcal{C})(x)$ for all $x \in X$.*

Proof. Since $\mathcal{A} =_\mu^f \mathcal{C}$, then for all $A \in \mathcal{A}$, there exists $C \in \mathcal{C}$ such that $\vec{m}_\mu^f(A \Delta C)(x) = 0$ for all $x \in X$.

We have $A = (A \bigcap C) \bigcup (A - C)$. Thus

$A \bigcup (C - A) = (A \bigcap C) \bigcup (C - A) \bigcup (A - C) = (A \bigcap C) \bigcup (A \Delta C)$. Therefore $\vec{m}_\mu^f(A)(x) + \vec{m}_\mu^f(C - A)(x) = \vec{m}_\mu^f(A \bigcap C)(x)$. Hence $\vec{m}_\mu^f(A)(x) = \vec{m}_\mu^f(A \bigcap C)(x)$, because $0 \leqslant \vec{m}_\mu^f(C - A)(x) \leqslant \vec{m}_\mu^f(A \Delta C)(x) = 0$.

If we change A with C, then $\vec{m}_\mu^f(C)(x) = \vec{m}_\mu^f(A \bigcap C)(x)$.

Thus for all $A \in \mathcal{A}$ there exists $C \in \mathcal{C}$ such that $\vec{m}_\mu^f(A)(x) = \vec{m}_\mu^f(C)(x)$. Hence $\vec{H}_\mu^f(\mathcal{A})(x) = \vec{H}_\mu^f(\mathcal{C})(x)$. □

4. Relative Conditional Entropy of a Multi-Dimensional Observer

In this section, the researchers define the concept relative conditional entropy of a multi-dimensional observer for partitions $\mathcal{A}, \mathbf{B}$ and provide the proof of some ergodic properties of the suggested measures.

We know the conditional probability play an important role in stochastic processes, so in this section it is considered from the viewpoint of an observer. Suppose $\mathcal{A}, \mathcal{C}$ are subsets of $P(X)$ with k and p members respectively.

Definition 4.1. *We define the relative entropy of a multi-dimensional observer of $\mathcal{A}$ given $\mathcal{C}$(called relative conditional entropy) by:*

$$\vec{H}^f_\mu(\mathcal{A}/\mathcal{C})(x) =$$

$$-\sum_{j=1}^{p}\sum_{i=1}^{k}[\vec{m}^f_\mu(A_i \cap C_j)(x)\log\|\vec{m}^f_\mu(A_i \cap C_j)(x)\| - \vec{m}^f_\mu(A_i \cap C_j)(x)\log\|\vec{m}^f_\mu(C_j)(x)\|].$$

In the case $\mathcal{C} = \{\emptyset, X\}$ we have

$$\vec{H}^f_\mu(\mathcal{A}/\mathcal{C})(x) =$$

$$-\sum_{i=1}^{k}[[\vec{m}^f_\mu(A_i \cap \emptyset)(x)\log\|\vec{m}^f_\mu(A_i \cap \emptyset)(x)\| - \vec{m}^f_\mu(A_i \cap \emptyset)(x)\log\|\vec{m}^f_\mu(\emptyset)(x)\|] +$$

$$[\vec{m}^f_\mu(A_i \cap X)(x)\log\|\vec{m}^f_\mu(A_i \cap X)(x)\| - \vec{m}^f_\mu(A_i \cap X)(x)\log\|\vec{m}^f_\mu(X)(x)\|]]$$

$$= -\sum_{i=1}^{k}[\vec{m}^f_\mu(A_i)(x)\log\|\vec{m}^f_\mu(A_i)(x)\| - \vec{m}^f_\mu(A_i)(x)\log\|\vec{m}^f_\mu(X)(x)\|]$$

$$= \vec{H}^f_\mu(\mathcal{A})(x) + \sum_{i=1}^{k}\vec{m}^f_\mu(X)(x)\log\|\vec{m}^f_\mu(X)(x)\| = \vec{H}^f_\mu(\mathcal{A})(x) - \vec{H}^f_\mu(X)(x).$$

In the crisp case we have $\|\vec{m}^f_\mu(X)(x)\| = 1$. Hence in this case $\vec{H}^f_\mu(\mathcal{A}/\mathcal{C})(x) = \vec{H}^f_\mu(\mathcal{A})(x)$. If $\xi(\mathcal{A}) = \{A_i\},\ \xi(\mathcal{C}) = \{C_j\},\ \xi(\mathcal{D}) = \{D_k\}$ are finite partitions of X, then $\{A_i \cap C_j\}$ for all and $\{D_k \cap C_j\}$ are partitions of X. If $\mathcal{A} =^f_\mu \mathcal{D}$ then$\mathcal{A} \cap \mathcal{C} =^f_\mu \mathcal{D} \cap \mathcal{C}$. Thus for given $A_i \cap C_j \in \mathcal{A} \cap \mathcal{C}$ the exists $D_k \cap C_m \in \mathcal{D} \cap \mathcal{C}$ such that $\vec{m}^f_\mu(A_i \cap C_j)(x) = \vec{m}^f_\mu(D_k \cap C_m)(x)$. Hence

$$\vec{H}^f_\mu(\mathcal{A}/\mathcal{C})(x) =$$

$$-\sum_{i,j}[\vec{m}^f_\mu(A_i \cap C_j)(x)\log\|\vec{m}^f_\mu(A_i \cap C_j)(x)\| - \vec{m}^f_\mu(A_i \cap C_j)(x)\log\|\vec{m}^f_\mu(C_j)(x)\|]$$

$$= -\sum_{k,j}[\vec{m}_\mu^f(D_k \cap C_j)(x)\log\|\vec{m}_\mu^f(D_k \cap C_j)(x)\| - \vec{m}_\mu^f(D_k \cap C_j)(x)\log\|\vec{m}_\mu^f(C_j)(x)\|]$$

$$= \vec{H}_\mu^f(\mathcal{D}/\mathcal{C})(x).$$

Also in relative conditional entropy we have $\vec{H}_\mu^f(\mathcal{A}/\mathcal{C})(x) = \vec{H}_\mu^f(\mathcal{A}/\mathcal{D})(x)$.

Theorem 4.2. *Suppose that $\mathcal{A}, \mathcal{C}, \mathcal{D}$ are finite partitions of X. Then $\vec{H}_\mu^f(\mathcal{A} \vee \mathcal{C}/\mathcal{D})(x) = \vec{H}_\mu^f(\mathcal{A}/\mathcal{D})(x) + \vec{H}_\mu^f(\mathcal{C}/\mathcal{A} \vee \mathcal{D})(x)$.*

Proof. Let $\xi(\mathcal{A}) = \{A_i\},\ \xi(\mathcal{C}) = \{C_j\},\ \xi(\mathcal{D}) = \{D_k\}$. Then

$$\vec{H}_\mu^f(\mathcal{A} \vee \mathcal{C}/\mathcal{D})(x) =$$

$$-\sum_{i,j,k}[\vec{m}_\mu^f(A_i \cap C_j \cap D_k)(x)\log\|\vec{m}_\mu^f(A_i \cap C_j \cap D_k)(x)\|$$

$$-\vec{m}_\mu^f(A_i \cap C_j \cap D_k)(x)\log\|\vec{m}_\mu^f(D_k)(x)\|]$$

$$= -\sum_{i,j,k}[\vec{m}_\mu^f(A_i \cap C_j \cap D_k)(x)\log\|\vec{m}_\mu^f(A_i \cap D_k)(x)\|$$

$$-\vec{m}_\mu^f(A_i \cap C_j \cap D_k)(x)\log\|\vec{m}_\mu^f(D_k)(x)\|]$$

$$-\sum_{i,j,k}[\vec{m}_\mu^f(A_i \cap C_j \cap D_k)(x)\log\|\vec{m}_\mu^f(A_i \cap C_j \cap D_k)(x)\|$$

$$-\vec{m}_\mu^f(A_i \cap C_j \cap D_k)(x)\log\|\vec{m}_\mu^f(A_i \cap D_k)(x)\|]$$

$$= -\sum_{i,k}[\vec{m}_\mu^f(A_i \cap D_k)(x)\log\|\vec{m}_\mu^f(A_i \cap D_k)(x)\|$$

$$-\vec{m}_\mu^f(A_i \cap D_k)(x)\log\|\vec{m}_\mu^f(D_k)(x)\|] + \vec{H}_\mu^f(\mathcal{C}/\mathcal{A} \vee \mathcal{D})(x)$$

$$= \vec{H}_\mu^f(\mathcal{A}/\mathcal{D})(x) + \vec{H}_\mu^f(\mathcal{C}/\mathcal{A} \vee \mathcal{D})(x). \quad \Box$$

The previous theorem implies if $\mathcal{A}, \mathcal{C}$ are finite partitions of X then $\vec{H}_\mu^f(\mathcal{A} \vee \mathcal{C})(x) = \vec{H}_\mu^f(\mathcal{A})(x) + \vec{H}_\mu^f(\mathcal{C}/\mathcal{A})(x)$.

Theorem 4.3. *If $\mathcal{A}, \mathcal{C}$ are finite partitions of X, and $\mathcal{A} \subseteq \mathcal{C}$ then $\vec{H}_\mu^f(\mathcal{A})(x) \leqslant \vec{H}_\mu^f(\mathcal{C})(x)$.*

Proof. Let $\xi(\mathcal{A}) = \{A_i\}$, $\xi(\mathcal{C}) = \{C_j\}$. Then $\xi(\mathcal{A}) \leqslant \xi(\mathcal{C})$. For given $C_j \in \mathcal{C}$ there exists $A_i \in \mathcal{A}$ such that $C_j \subseteq A_i$. Hence $\vec{m}^f_\mu(C_j)(x) \leqslant \vec{m}^f_\mu(A_i)(x)$. Thus $\log \|\vec{m}^f_\mu(C_j)(x)\| \leqslant \log \|\vec{m}^f_\mu(A_i)(x)\|$. Hence

$$-\vec{m}^f_\mu(A_i)(x) \log \|\vec{m}^f_\mu(C_j)(x)\| \geqslant -\vec{m}^f_\mu(A_i)(x) \log \|\vec{m}^f_\mu(A_i)(x)\|.$$

Therefore

$$-\sum_i \vec{m}^f_\mu(A_i)(x) \log \|\vec{m}^f_\mu(A_i)(x)\| \leqslant -\sum_i \vec{m}^f_\mu(A_i)(x) \log \|\vec{m}^f_\mu(C_j)(x)\|.$$

Thus

$$\vec{H}^f_\mu(\mathcal{A})(x) \leqslant -\sum_i \vec{m}^f_\mu(A_i)(x) \log \|\vec{m}^f_\mu(C_j)(x)\| =$$

$$-\vec{m}^f_\mu(X)(x) \log \|\vec{m}^f_\mu(C_j)(x)\| = -\vec{m}^f_\mu(\cup_j C_j)(x) \log \|\vec{m}^f_\mu(C_j)(x)\| =$$

$$-\sum_j \vec{m}^f_\mu(C_j)(x) \log \|\vec{m}^f_\mu(C_j)(x)\| = \vec{H}^f_\mu(\mathcal{C})(x). \quad \square$$

If $\mathcal{A}$, $\mathcal{C}$, $\mathcal{D}$ are finite partitions of X and $\mathcal{A} \subseteq \mathcal{C}$ then $\xi(\mathcal{A}) \leqslant \xi(\mathcal{C})$. Since $\mathcal{A} \vee \mathcal{C} = \mathcal{C}$ then

$$\vec{H}^f_\mu(\mathcal{C}/\mathcal{D})(x) = \vec{H}^f_\mu(\mathcal{A} \vee \mathcal{C}/\mathcal{D})(x) =$$

$$\vec{H}^f_\mu(\mathcal{A}/\mathcal{D})(x) + \vec{H}^f_\mu(\mathcal{C}/\mathcal{A} \vee \mathcal{D})(x) \geqslant \vec{H}^f_\mu(\mathcal{A}/\mathcal{D})(x).$$

Theorem 4.4. *If $\mathcal{A}$, $\mathcal{C}$, $\mathcal{D}$ are finite partitions of X and $\mathcal{C} \subseteq \mathcal{D}$ then $\vec{H}^f_\mu(\mathcal{A}/\mathcal{C})(x) \geqslant \vec{H}^f_\mu(\mathcal{A}/\mathcal{D})(x)$.*

Proof. We first assume $\vec{m}^f_\mu(C_j)(x) \neq 0$ and $\vec{m}^f_\mu(D_k)(x) \neq 0$ for all $x \in X$. Let i, j, be given and let $\phi : [0, \infty] \to \mathbb{R}$ be defined by

$$\phi(x) = \begin{cases} 0 & textif \quad x = 0, \\ x \log x & \text{if} \quad x \neq 0 \end{cases}$$

then theorem 4.2 of [14] implies

$$\phi\Big(\sum_k \frac{\|\vec{m}^f_\mu(D_k \cap C_j)(x)\|}{\|\vec{m}^f_\mu(C_j)(x)\|} \times \frac{\|\vec{m}^f_\mu(A_i \cap D_k)(x)\|}{\|\vec{m}^f_\mu(D_k)(x)\|}\Big) \leqslant$$

$$\sum_k \frac{\|\vec{m}_\mu^f(D_k \cap C_j)(x)\|}{\|\vec{m}_\mu^f(C_j)(x)\|}\phi(\frac{\|\vec{m}_\mu^f(A_i \cap D_k)(x)\|}{\|\vec{m}_\mu^f(D_k)(x)\|}).$$

Thus

$$\phi(\frac{\|\vec{m}_\mu^f(A_i \cap C_j)(x)\|}{\|\vec{m}_\mu^f(C_j)(x)\|}) \leqslant$$

$$\sum_k \frac{\|\vec{m}_\mu^f(D_k)(x)\|}{\|\vec{m}_\mu^f(C_j)(x)\|}\phi(\frac{\|\vec{m}_\mu^f(A_i \cap D_k)(x)\|}{\|\vec{m}_\mu^f(D_k)(x)\|}).$$

Thus

$$\frac{\|\vec{m}_\mu^f(A_i \cap C_j)(x)\|}{\|\vec{m}_\mu^f(C_j)(x)\|}\log\frac{\|\vec{m}_\mu^f(A_i \cap C_j)(x)\|}{\|\vec{m}_\mu^f(C_j)(x)\|} \leqslant$$

$$\sum_k \frac{\|\vec{m}_\mu^f(D_k)(x)\|}{\|\vec{m}_\mu^f(C_j)(x)\|}\frac{\|\vec{m}_\mu^f(A_i \cap D_k)(x)\|}{\|\vec{m}_\mu^f(D_k)(x)\|}\log\frac{\|\vec{m}_\mu^f(A_i \cap D_k)(x)\|}{\|\vec{m}_\mu^f(D_k)(x)\|}.$$

Hence

$$\sum_{i,j} \|\vec{m}_\mu^f(A_i \cap C_j)(x)\|\log\frac{\|\vec{m}_\mu^f(A_i \cap C_j)(x)\|}{\|\vec{m}_\mu^f(C_j)(x)\|} \leqslant$$

$$\sum_{i} \|\vec{m}_\mu^f(A_i \cap C_j)(x)\|\log\frac{\|\vec{m}_\mu^f(A_i \cap C_j)(x)\|}{\|\vec{m}_\mu^f(C_j)(x)\|} \leqslant$$

$$\sum_{i,k} \|\vec{m}_\mu^f(A_i \cap D_k)(x)\|\log\frac{\|\vec{m}_\mu^f(A_i \cap D_k)(x)\|}{\|\vec{m}_\mu^f(D_k)(x)\|}.$$

Therefore

$$\sum_{i,j} \vec{m}_\mu^f(A_i \cap C_j)(x)\log\frac{\|\vec{m}_\mu^f(A_i \cap C_j)(x)\|}{\|\vec{m}_\mu^f(C_j)(x)\|} \leqslant$$

$$\sum_{i,j} \|\vec{m}_\mu^f(A_i \cap C_j)(x)\|\log\frac{\|\vec{m}_\mu^f(A_i \cap C_j)(x)\|}{\|\vec{m}_\mu^f(C_j)(x)\|} \leqslant$$

$$\sum_{i,k} \|\vec{m}_\mu^f(A_i \cap D_k)(x)\|\log\frac{\|\vec{m}_\mu^f(A_i \cap D_k)(x)\|}{\|\vec{m}_\mu^f(D_k)(x)\|} \leqslant$$

$$\sum_{i,k} \|\vec{m}_\mu^f(A_i \cap D_k)(x)\| \log \frac{\|\vec{m}_\mu^f(A_i \cap D_k)(x)\|}{\|\vec{m}_\mu^f(D_k)(x)\|}.$$

Thus

$$\sum_{i,j}[\vec{m}_\mu^f(A_i \cap C_j)(x) \log \|\vec{m}_\mu^f(A_i \cap C_j)(x)\| - \vec{m}_\mu^f(A_i \cap C_j)(x) \log \|\vec{m}_\mu^f(C_j)(x)\|] \leqslant$$

$$\sum_{i,k}[\vec{m}_\mu^f(A_i \cap D_k)(x) \log \|\vec{m}_\mu^f(A_i \cap D_k)(x)\| - \vec{m}_\mu^f(A_i \cap D_k)(x) \log \|\vec{m}_\mu^f(D_k)(x)\|].$$

Hence

$$-\vec{H}_\mu^f(\mathcal{A}/\mathcal{C})(x) \leqslant -\vec{H}_\mu^f(\mathcal{A}/\mathcal{D})(x).$$

Thus

$$\vec{H}_\mu^f(\mathcal{A}/\mathcal{C})(x) \geqslant \vec{H}_\mu^f(\mathcal{A}/\mathcal{D})(x).$$

If $\vec{m}_\mu^f(D_k)(x) = 0$ for some k then

$$\vec{H}_\mu^f(\mathcal{A}/\mathcal{C})(x) \geqslant \vec{H}_\mu^f(\mathcal{A}/\mathcal{D})(x) = 0.$$

If $\vec{m}_\mu^f(C_j)(x) = 0$ for some j then there exists $D_k \in \mathcal{D}$ such that $\vec{m}_\mu^f(D_k)(x) = 0$. Thus

$$\vec{H}_\mu^f(\mathcal{A}/\mathcal{C})(x) = 0 \geqslant \vec{H}_\mu^f(\mathcal{A}/\mathcal{D})(x) = 0. \quad \square$$

In a special case if we take $\mathcal{C} = \{\emptyset, X\}$, then $\vec{H}_\mu^f(\mathcal{A})(x) - \vec{H}_\mu^f(X)(x) \geqslant \vec{H}_\mu^f(\mathcal{A}/\mathcal{D})(x)$. Thus $\vec{H}_\mu^f(\mathcal{A})(x) \geqslant \vec{H}_\mu^f(\mathcal{A}/\mathcal{D})(x)$.

Theorem 4.5. *If $\mathcal{A}$, $\mathcal{C}$, $\mathcal{D}$ are finite partitions of X then* $\vec{H}_\mu^f(\mathcal{A} \vee \mathcal{C}/\mathcal{D})(x) \leqslant \vec{H}_\mu^f(\mathcal{A}/\mathcal{D})(x) + \vec{H}_\mu^f(\mathcal{C}/\mathcal{D})(x)$.

Proof. Since $\mathcal{A} \vee \mathcal{D} \supseteq \mathcal{D}$ then $\vec{H}_\mu^f(\mathcal{C}/\mathcal{A} \vee \mathcal{D})(x) \leqslant \vec{H}_\mu^f(\mathcal{C}/\mathcal{D})(x)$. Hence

$$\vec{H}_\mu^f(\mathcal{A} \vee \mathcal{C}/\mathcal{D})(x) = \vec{H}_\mu^f(\mathcal{A}/\mathcal{D})(x) + \vec{H}_\mu^f(\mathcal{C}/\mathcal{A} \vee \mathcal{D})(x) \leqslant \vec{H}_\mu^f(\mathcal{A}/\mathcal{D})(x) + \vec{H}_\mu^f(\mathcal{C}/\mathcal{D})(x). \ \square$$

If we take $\mathcal{D} = \{\emptyset, X\}$ then the above theorem implies $\vec{H}_\mu^f(\mathcal{A} \vee \mathcal{C})(x) - \vec{H}_\mu^f(X)(x) \leqslant \vec{H}_\mu^f(\mathcal{A})(x) - \vec{H}_\mu^f(X)(x) + \vec{H}_\mu^f(\mathcal{C})(x) - \vec{H}_\mu^f(X)(x)$.
Thus $\vec{H}_\mu^f(\mathcal{A} \vee \mathcal{C})(x) \leqslant \vec{H}_\mu^f(\mathcal{A})(x) + \vec{H}_\mu^f(\mathcal{C})(x) - \vec{H}_\mu^f(X)(x)$.
Therefore $\vec{H}_\mu^f(\mathcal{A} \vee \mathcal{C})(x) \leqslant \vec{H}_\mu^f(\mathcal{A})(x) + \vec{H}_\mu^f(\mathcal{C})(x)$.

Let $\mathcal{A}$, and $\mathcal{C}$ be finite subsets of $P(X)$ and $\mathcal{A} \subset_{\mu}^{f} \mathcal{C}$. Then for given $A_i \in \mathcal{A}$, there exists $C_j \in \mathcal{C}$ such that $\vec{m}_{\mu}^{f}(A_i \Delta C_j)(x) = 0$. Hence $\vec{m}_{\mu}^{f}(A_i \cap C_j)(x) = \vec{m}_{\mu}^{f}(C_j)(x)$ or $\vec{m}_{\mu}^{f}(A_i \cap C_j)(x) = 0$. Thus $\vec{H}_{\mu}^{f}(\mathcal{A}/\mathcal{C})(x) = 0$.

If $\mathcal{A}$, and $\mathcal{C}$ are finite partitions of $P(X)$ then we say $\mathcal{A}$ and $\mathcal{C}$ are independent if $\|\vec{m}_{\mu}^{f}(A \cap C)(x)\| = \|\vec{m}_{\mu}^{f}(A)(x)\|.\|\vec{m}_{\mu}^{f}(C)(x)\|$, where $A \in \mathcal{A},\ C \in \mathcal{C}$, and $x \in X$.

Theorem 4.6. *Assume $\mathcal{A}, \mathcal{C}$ are finite partitions of $(X, \vec{m}_{\mu}^{f})$ and $\vec{m}_{\mu}^{f}(X)(x) = \prod 1$ for all $x \in X$, then $\vec{H}_{\mu}^{f}(\mathcal{A}/\mathcal{C})(x) = \vec{H}_{\mu}^{f}(\mathcal{A})(x)$ if and only if $\mathcal{A}$ and $\mathcal{C}$ are independent.*

Proof. If $\mathcal{A}$ and $\mathcal{C}$ are independent then it is clear that $\vec{H}_{\mu}^{f}(\mathcal{A}/\mathcal{C})(x) = \vec{H}_{\mu}^{f}(\mathcal{A})$. Conversely: Let $\vec{H}_{\mu}^{f}(\mathcal{A}/\mathcal{C})(x) = \vec{H}_{\mu}^{f}(\mathcal{A})(x)$ for all $x \in X$, and let ϕ be the function which is defined in the proof of theorem 4.3.Then
$\phi(\sum_{j=1}^{k} \|\vec{m}_{\mu}^{f}(C_j)(x)\|.\frac{\|\vec{m}_{\mu}^{f}(A_i \cap C_j)(x)\|}{\|\vec{m}_{\mu}^{f}(C_j)(x)\|}) \leqslant \sum_{j=1}^{k} \|\vec{m}_{\mu}^{f}(C_j)(x)\| \phi(\frac{\|\vec{m}_{\mu}^{f}(A_i \cap C_j)(x)\|}{\|\vec{m}_{\mu}^{f}(C_j)(x)\|})$.
Hence $(\sum_{j=1}^{k} \|\vec{m}_{\mu}^{f}(A_i \cap C_j)(x)\|) \log(\sum_{j=1}^{k} \|\vec{m}_{\mu}^{f}(A_i \cap C_j)(x)\|)$
$\leqslant \sum_{j=1}^{k} [\|\vec{m}_{\mu}^{f}(C_j)(x)\|.\frac{\|\vec{m}_{\mu}^{f}(A_i \cap C_j)(x)\|}{\|\vec{m}_{\mu}^{f}(C_j)(x)\|} \log \frac{\|\vec{m}_{\mu}^{f}(A_i \cap C_j)(x)\|}{\|\vec{m}_{\mu}^{f}(C_j)(x)\|})]$.

Thus
$\|\vec{m}_{\mu}^{f}(A_i)(x)\| \log \|\vec{m}_{\mu}^{f}(A_i)(x)\| \leqslant \sum_{j=1}^{k} \|\vec{m}_{\mu}^{f}(A_i \cap C_j)(x)\| \log \frac{\|\vec{m}_{\mu}^{f}(A_i \cap C_j)(x)\|}{\|\vec{m}_{\mu}^{f}(C_j)(x)\|}$.
Hence $-\|\vec{m}_{\mu}^{f}(A_i)(x)\| \log \|\vec{m}_{\mu}^{f}(A_i)(x)\| \geqslant -\sum_{j=1}^{k} [\|\vec{m}_{\mu}^{f}(A_i \cap C_j)(x)\| \log \|\vec{m}_{\mu}^{f}(A_i \cap C_j)(x)\| - \|\vec{m}_{\mu}^{f}(A_i \cap C_j)(x)\| \log \|\vec{m}_{\mu}^{f}(C_j)(x)\|]$.
The equality occures only when $\|\vec{m}_{\mu}^{f}(A_i \cap C_j)(x)\| / \|\vec{m}_{\mu}^{f}(C_j)(x)\|$ does not depend to j, i.e. $\|\vec{m}_{\mu}^{f}(A_i \cap C_j)(x)\| / \|\vec{m}_{\mu}^{f}(C_j)(x)\| = a_i$ or $\|\vec{m}_{\mu}^{f}(A_i \cap C_j)(x)\| = a_i.\|\vec{m}_{\mu}^{f}(C_j)(x)\|$. Thus
$\| \sum_j \vec{m}_{\mu}^{f}(A_i \cap C_j)(x)\| = a_i. \sum_j \|\vec{m}_{\mu}^{f}(C_j)(x)\|$. Therefore $a_i = \|\vec{m}_{\mu}^{f}(A_i)(x)\|$.
Hence $\|\vec{m}_{\mu}^{f}(A_i \cap C_j)(x)\| = \|\vec{m}_{\mu}^{f}(A_i)(x)\|.\|\vec{m}_{\mu}^{f}(C_j)(x)\|$ for all i, j. □

5. Relative Entropy of a Multi-Dimensional Observer of a Measure Preserving Transformation

In this section the notion of entropy is extended. [1, 4] The researchers assume that $T : X \to X$ is a relative measure-preserving transformation

(relative dynamical system), and $\mathcal{U}_0$ is the set of all finite partitions of X. For $\mathcal{A} \in \mathcal{U}_0$ the function $\vec{h}^f_\mu(T,\mathcal{A})(x) = \lim_{n\to\infty} -\frac{1}{n}\vec{H}^f_\mu(\vee_{i=0}^{n-1}T^{-i}\mathcal{A})(x)$ is called the relative entropy of T with respect to $\mathcal{A}$ and μ.
$\vec{h}^f_\mu(T)(x) = \sup_{\mathcal{A}\in\mathcal{U}_0} \vec{h}^f_\mu(T,\mathcal{A})(x)$ is called the relative entropy of T at X.

Theorem 5.1. *If $T : X \to X$ is a relative dynamical system and $\mathcal{A} \in \mathcal{U}_0$ then $\lim_{n\to\infty}(\frac{1}{n})\vec{H}^f_\mu(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A})(x)$ exists.*

Proof. Let $a_n = \sup[\vec{H}^f_\mu(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A})(x)] \geqslant 0$, then
$a_{n+p} = \sup[\vec{H}^f_\mu(\bigvee_{i=0}^{n+p-1} T^{-i}\mathcal{A})(x)] \leqslant \sup[\vec{H}^f_\mu(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A})(x)]$
$+\sup[\vec{H}^f_\mu(\bigvee_{i=n}^{n+p-1} T^{-i}\mathcal{A})(x)] = a_n + \sup[\vec{H}^f_\mu(\bigvee_{i=0}^{p-1} T^{-i}\mathcal{A})(x)] = a_n + a_p$.
Thus $\lim \frac{a_n}{n}$ exists. □

Theorem 5.2. *If $\mathcal{A} \in \mathcal{U}_0$ then $\frac{1}{n+1}\vec{H}^f_\mu(\vee_{i=0}^{n}T^{-i}\mathcal{A})(x) \leqslant \frac{1}{n}\vec{H}^f_\mu(\vee_{i=0}^{n-1}T^{-i}\mathcal{A})(x)$.*

Proof. By induction we have $\vec{H}^f_\mu(\vee_{i=0}^{n-1}T^{-i}\mathcal{A})(x) = \vec{H}^f_\mu(\mathcal{A})(x) +$
$\sum_{j=1}^{n-1} \vec{H}^f_\mu(\mathcal{A}/ \vee_{i=1}^{j} T^{-i}\mathcal{A})(x)$.
Hence $\vec{H}^f_\mu(\vee_{i=0}^{n-1}T^{-i}\mathcal{A})(x) \geqslant n\vec{H}^f_\mu(\mathcal{A}/ \vee_{i=0}^{n} T^{-i}\mathcal{A})(x)$.
Thus $n\vec{H}^f_\mu(\vee_{i=0}^{n}T^{-i}\mathcal{A})(x) = n[\vec{H}^f_\mu(\vee_{i=0}^{n-1}T^{-i}\mathcal{A})(x)+\vec{H}^f_\mu(\mathcal{A}/\vee_{i=0}^{n}T^{-i}\mathcal{A})(x)] \leqslant$
$(n+1)\vec{H}^f_\mu(\vee_{i=0}^{n-1}T^{-i}\mathcal{A})(x)$.
Therefore $\frac{1}{n+1}\vec{H}^f_\mu(\vee_{i=0}^{n}T^{-i}\mathcal{A})(x) \leqslant \frac{1}{n}\vec{H}^f_\mu(\vee_{i=0}^{n-1}T^{-i})(x)$. □

For $\mathcal{A},\mathcal{C} \in \mathcal{U}_0$ we have $\frac{1}{n}\vec{H}^f_\mu(\vee_{i=0}^{n-1}T^{-i}\mathcal{A})(x) \leqslant \frac{1}{n}\sum_{i=0}^{n-1} \vec{H}^f_\mu(T^{-i}\mathcal{A})(x) = \frac{1}{n}\sum_{i=0}^{n-1} \vec{H}^f_\mu(\mathcal{A})(x) = \vec{H}^f_\mu(\mathcal{A})(x)$ and so $\vec{h}^f_\mu(T,\mathcal{A})(x) \leqslant \vec{H}^f_\mu(\mathcal{A})(x)$.

Also since $\vec{H}^f_\mu(\bigvee_{i=0}^{n-1} T^{-i}(\mathcal{A}\vee\mathcal{C}))(x) = \vec{H}^f_\mu(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A} \vee \bigvee_{i=0}^{n-1} T^{-i}\mathcal{C})(x)$
$\leqslant \vec{H}^f_\mu(\vee_{i=0}^{n-1}T^{-i}\mathcal{A})(x) + \vec{H}^f_\mu(\vee_{i=0}^{n-1}T^{-i}\mathcal{C})(x)$, then
$\frac{1}{n}\vec{H}^f_\mu(\vee_{i=0}^{n-1}T^{-i}(\mathcal{A}\vee\mathcal{C}))(x) \leqslant \frac{1}{n}\vec{H}^f_\mu(\vee_{i=0}^{n-1}T^{-i}\mathcal{A})(x)+\frac{1}{n}\vec{H}^f_\mu(\vee_{i=0}^{n-1}T^{-i}\mathcal{C})(x)$.
Hence $\vec{h}^f_\mu(T,\mathcal{A}\vee\mathcal{C})(x) \leqslant \vec{h}^f_\mu(T,\mathcal{A})(x) + \vec{h}^f_\mu(T,\mathcal{C})(x)$.

Suppose $\mathcal{A} \subseteq^f_\mu \mathcal{C}$ then for $n \geqslant 1$ we have $\vee_{i=0}^{n-1}T^{-i}\mathcal{A} \subseteq^f_\mu \vee_{i=0}^{n-1}T^{-i}\mathcal{C}$. Thus $\vec{H}^f_\mu(\vee_{i=0}^{n-1}T^{-i}\mathcal{A})(x) \leqslant \vec{H}^f_\mu(\vee_{i=0}^{n-1}T^{-i}\mathcal{C})(x)$. Hence $\frac{1}{n}\vec{H}^f_\mu(\vee_{i=0}^{n-1}T^{-i}\mathcal{A})(x) \leqslant \frac{1}{n}\vec{H}^f_\mu(\vee_{i=0}^{n-1}T^{-i}\mathcal{C})(x)$ so $\vec{h}^f_\mu(T,\mathcal{A})(x) \leqslant \vec{h}^f_\mu(T,\mathcal{C})(x)$.

Theorem 5.3. *If $\mathcal{A},\mathcal{C} \in \mathcal{U}_0$ then $\vec{h}^f_\mu(T,\mathcal{A})(x) \leqslant \vec{h}^f_\mu(T,\mathcal{C})(x)+\vec{H}^f_\mu(\mathcal{A}/\mathcal{C})(x)$.*

Proof. $\vec{H}^f_\mu(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A})(x) \leqslant \vec{H}^f_\mu((\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A}) \vee (\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C}))(x)$
$= \vec{H}^f_\mu(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C})(x) + \vec{H}^f_\mu((\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A})/(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C}))(x)$.
Since $\vec{H}^f_\mu((\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A})/(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C}))(x) \leqslant \sum_{i=0}^{n-1} \vec{H}^f_\mu((T^{-i}\mathcal{A})/(\bigvee_{j=0}^{n-1} T^{-j}\mathcal{C}))(x)$

$\leqslant \sum_{i=0}^{n-1} \vec{H}_\mu^f(T^{-i}\mathcal{A}/T^{-i}\mathcal{C})(x) = n\vec{H}_\mu^f(\mathcal{A}/\mathcal{C})(x)$ then
$\vec{H}_\mu^f(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A})(x) \leqslant \vec{H}_\mu^f(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C})(x) + n\vec{H}_\mu^f(\mathcal{A}/\mathcal{C})(x)$.
Therefore $\frac{1}{n}\vec{H}_\mu^f(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A})(x) \leqslant \frac{1}{n}\vec{H}_\mu^f(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{C})(x) + \vec{H}_\mu^f(\mathcal{A}/\mathcal{C})(x)$.
Hence

$\vec{h}_\mu^f(T,\mathcal{A})(x) \leqslant \vec{h}_\mu^f(T,\mathcal{C})(x) + \vec{H}_\mu^f(\mathcal{A}/\mathcal{C})(x)$. □

Theorem 5.4. *If $\mathcal{A} \in \mathcal{U}_0$, then $\vec{h}_\mu^f(T,\mathcal{A})(x) = \vec{h}_\mu^f(T, \vee_{i=0}^k T^{-i}\mathcal{A})(x)$, where $x \in X$ and $k \geqslant 1$.*

Proof. We have $\vec{h}_\mu^f(T, \vee_{i=0}^k T^{-i}\mathcal{A})(x) = \lim_{n\to\infty} \frac{1}{n}\vec{H}_\mu^f(\bigvee_{j=0}^{n-1} T^{-j}(\bigvee_{i=0}^{k} T^{-i}\mathcal{A}))(x)$
$= \lim_{n\to\infty} \frac{1}{n}\vec{H}_\mu^f(\bigvee_{i=0}^{k+n-1} T^{-i}\mathcal{A})(x)$
$= \lim_{n\to\infty}(\frac{k+n-1}{n})(\frac{1}{k+n-1})\vec{H}_\mu^f(\bigvee_{i=0}^{k+n-1} T^{-i}\mathcal{A})(x) = \vec{h}_\mu^f(T,\mathcal{A})(x)$. □

In the above theorem if we assume T is invertible, then for $k \geqslant 1$ we have $\vec{h}_\mu^f(T,\mathcal{A})(x) = \vec{h}_\mu^f(T, \vee_{i=-k}^k T^{i}\mathcal{A})(x)$.

Theorem 5.5. *If $k > 0$, and $x \in X$ then*
(i) $\vec{h}_\mu^f(T^k)(x) = k\vec{h}_\mu^f(T)(x)$.
(ii) $\vec{h}_\mu^f(T^k)(x) = |k|\vec{h}_\mu^f(T)(x)$ when T is invertible.

Proof. (i) We have
$\vec{h}_\mu^f(T^k, \bigvee_{i=0}^{k-1} T^{-i}\mathcal{A})(x) = \lim_{n\to\infty} \frac{1}{n}\vec{H}_\mu^f(\bigvee_{j=0}^{n-1} T^{-kj}(\bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}))(x)$
$= \lim_{n\to\infty} \frac{1}{n}\vec{H}_\mu^f(\bigvee_{j=0}^{n-1}\bigvee_{i=0}^{k-1} T^{-kj-i}\mathcal{A})(x) = \lim_{n\to\infty} \frac{k}{nk}\vec{H}_\mu^f(\bigvee_{i=0}^{nk-1} T^{-i}\mathcal{A})(x)$
$= k\vec{h}_\mu^f(T,\mathcal{A})(x)$. Thus

$k\vec{h}_\mu^f(T)(x) = k\sup_{\mathcal{A}\in\mathcal{U}_0}(\vec{h}_\mu^f(T,\mathcal{A})(x)) = \sup_{\mathcal{A}\in\mathcal{U}_0}(k\vec{h}_\mu^f(T,\mathcal{A})(x))$
$= \sup_{\mathcal{A}\in\mathcal{U}_0}\vec{h}_\mu^f(T^k, \bigvee_{i=0}^{k-1} T^{-i}\mathcal{A})(x) \leqslant \sup_{\mathcal{A}\in\mathcal{U}_0}\vec{h}_\mu^f(T^k,\mathcal{C})(x) = \vec{h}_\mu^f(T^k)(x)$.
We also have $\vec{h}_\mu^f(T^k,\mathcal{A})(x) \leqslant \vec{h}_\mu^f(T^k, \bigvee_{i=0}^{k-1} T^{-i}\mathcal{A})(x) = k\vec{h}_\mu^f(T,\mathcal{A})(x)$.
Thus $\vec{h}_\mu^f(T^k)(x) \leqslant k\vec{h}_\mu^f(T)(x)$.

(ii) This part follows part (i) and this fact which $\vec{h}_\mu^f(T^{-1})(x) = \vec{h}_\mu^f(T)(x)$. □

As a result of the previous theorem one can construct relative measure preserving maps with entropies larger than any positive real number.

6. Conclusions

In this paper the researchers considered the multi-dimensional observer as a model for realistic phenomena. Relative probability measures, relative entropy of partitions, relative conditional entropy, and relative entropy of a relative measure preserving map via multi-dimensional observers were studied and examples were provided.

Acknowledgments
The authors thank the editor and the referees for their valuable comments and suggestions.

References

[1] J. Abellan and S. Moral, Difference of entropies as a non-specificity function on credal sets, *International Journal of General Systems,* 34 (3) (2005), 201-214.

[2] J. M. Amigo, *Permutation Complexity in Dynamical Systems-Ordinal Patterns, Permutation Entropy, and all that*, Springer Verlag, 2010.

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

[4] J. Dai, Q. Xu, W. Wang, and H. Tian, Conditional entropy for incomplete decision systems and its application in data mining, *International Journal of General Systems,* 41 (7) (2012), 713-728.

[5] A. Ebrahimzadeh, The Observational Modelling of information Function with Finite Partitions , *Journal of mathematical extension,* 10 (2) (2016), 47-57.

[6] A. Ebrahimzadeh, S. Sargazi, and Z. Eslami Giski, Observational modeling of logical entropy of dynamical systems with local approach, *J. Math. Ext.*, 12 (3) (2018), 1-19.

[7] A. Ebrahimzadeh and J. Jamalzadeh, Conditional logical entropy of fuzzy σ-algebras, *Journal of Intelligent and Fuzzy Systems,* 33 (2017), 1019-1026.

[8] A. Ebrahimzadeh and Z. Eslami Giski, Tsallis entropy of dynamical systems-A general scheme, *accepted to Journal of Intelligent and Fuzzy Systems.*

[9] R. M. Gray, *Entropy and Information Theory*, Springer Verlag, 1990.

[10] R. Intan, *Rarity-Based Similarity Relations in a Generalized Fuzzy Information System,* Conference on Cybernetics and Intelligent System, Singapore, 2004.

[11] M. R. Molaei, Relative semi–dynamical systems, International Journal of Uncertainty, *Fuzziness and Knowledge-Based Systems,* 12 (2) (2004), 237-243.

[12] M. R. Molaei, Observational modeling of topological spaces, *Chaos Solitons and Fractals,* 42 (2009), 615–619. doi: 10.1016/j.chaos.2009.01.035, 2009.

[13] W. Xu, X. Zhang, and W. Zhang, Knowledge granulation, knowledge entropy and knowledge uncertainty measure in ordered information systems, *Applied Soft Computing,* 9 (2009), 1244-1251.

[14] P. Walters, *An Introduction to Ergodic Theory*, Springer Verlag, 1982.

Mohammad Hossein Asadian
Assistant professor of Mathematics
Department of Mathematics
Bandar Abbas Branch, Islamic Azad University
Bandar Abbas, Iran
E-mail: mh.asadian@iauba.ac.ir

Mohammad Reza Molaei
Assistant Professor of Mathematics
Department of Mathematics
Bandar Abbas Branch, Islamic Azad University
Bandar Abbas, Iran
E-mail: mrmolaei@uk.ac.ir