Permutation Excess Entropy and Mutual Information between the Past and Future

Taichi Haruna [1], Kohei Nakajima [2]
[1] Department of Earth & Planetary Sciences, Graduate School of Science, Kobe University, 1-1 Rokkodaicho, Nada, Kobe 657-8501, Japan
[2] Artificial Intelligence Laboratory, Department of Informatics, University of Zurich, Andreasstrasse 15, 8050 Zurich, Switzerland
tharuna@penguin.kobe-u.ac.jp (T. Haruna)

Abstract We address the excess entropy, which is a measure of complexity for stationary time series, from the ordinal point of view. We show that the permutation excess entropy is equal to the mutual information between two adjacent semi-infinite blocks in the space of orderings for finite-state stationary ergodic Markov processes. This result may spread a new light on the relationship between complexity and anticipation.
Keywords : Permutation Entropy, Excess Entropy, Mutual Information, Complexity, Duality

1 Introduction

Recently, it was found that much of the information contained in stationary time series can be captured by orderings of values, not the values themselves [1]. The permutation entropy rate which was first introduced in [5, 6] quantifies the average uncertainty of orderings of values per time unit. This is in contrast to the usual entropy rate which quantifies the average uncertainty of values per time unit. However, surprisingly, it is known that the permutation entropy rate is equal to the entropy rate for finite-state stationary stochastic processes [1, 2]. Similar results for dynamical systems are also known [2, 3, 6, 15, 18].

In our previous work [14], we found a new proof of the equality between the permutation entropy rate and the entropy rate based on the duality between values and orderings, which can be seen as a Galois connection [11] (categorical adjunction [17] for partially ordered sets, however, we do not refer to the Galois connection explicitly in this paper). By making use of the duality, we also proved that the permutation excess entropy is equal to the excess entropy for finite-state stationary ergodic Markov processes.

The excess entropy has attracted interest from the complex systems community for decades [4, 7, 9, 10, 12, 13, 16, 19]. By definition, the excess entropy is the sum of entropy over-estimates over finite length of words [10]. However, it can be expressed as the mutual information between the past and future, namely, the mutual information between two adjacent semi-infinite blocks of stochastic variables. Thus, the excess entropy can be interpreted as a measure of global correlation present in a system.

In this paper, based on the duality between values and orderings, we show that the permutation excess entropy also admit the mutual information expression in the space of orderings when the process is finite-state stationary ergodic Markov one. This result partially justifies the claim that the permutation excess entropy measures global correlation at the level of orderings of values present in stationary time series.

This paper is organized as follows. In Section 2, we review the duality between values and orderings. In Section 3, we explain the permutation excess entropy. In Section 4, we present a proof of the claim that the permutation excess entropy has the mutual information expression for finite-state stationary ergodic Markov processes. In Section 5, we give conclusions.

2 Duality between Values and Orderings Explained

Let $A_n = \{1, 2, \cdots, n\}$ be a finite alphabet consisting of natural numbers from 1 to n. We consider A_n as a totally ordered set ordered by the usual 'less-than-or-equal-to' relationship.

We denote the set of all permutations of length $L \geq 1$ by $\mathcal{S}_L$. Namely, each element $\pi \in \mathcal{S}_L$ is a bijection on the set of indexes $\{1, 2, \cdots, L\}$. For convenience, we denote each permutation $\pi \in \mathcal{S}_L$ by the string $\pi(1) \cdots \pi(L)$.

For each *word* $s_1^L := s_1 \cdots s_L := (s_1, \cdots, s_L) \in A_n^L = \underbrace{A_n \times \cdots \times A_n}_{L}$ of length $L \geq 1$, we define its *permutation type* $\pi \in \mathcal{S}_L$ by re-ordering symbols $s_1, \cdots, s_L$ in increasing order: s_1^L is of type π if we have $s_{\pi(i)} \leq s_{\pi(i+1)}$ and $\pi(i) < \pi(i+1)$ when $s_{\pi(i)} = s_{\pi(i+1)}$ for $i = 1, 2, \cdots, L-1$. For example, $\pi(1)\pi(2)\pi(3)\pi(4) = 3142$ for $s_1^4 = 2312$ because $s_3 s_1 s_4 s_2 = 1223$.

We introduce the map $\phi : A_n^L \to \mathcal{S}_L$ that sends each word s_1^L to its unique permutation type $\pi = \phi(s_1^L)$. This map ϕ classifies or coarse-grains words of length L by the criterion whether they have the same permutation type. In general, ϕ is many-to-one map. For example, all of $111, 112, 122, 222 \in A_2^3$ have the same permutation type $\pi \in \mathcal{S}_3$ defined by $\pi(1)\pi(2)\pi(3) = 123$ (identity on $\{1, 2, 3\}$).

Now, we list the properties of the map ϕ which will be used later.

Lemma 1 *For $s_1^L, t_1^L \in A_n^L$, $\phi(s_1^L) = \phi(t_1^L)$ if and only if $s_k \leq s_j \Leftrightarrow t_k \leq t_j$ for all $1 \leq j \leq k \leq L$.*

Proof. See Corollary 4 in [14]. □

Lemma 2 *Let $n \geq i \geq 1$. Fix $\pi \in \mathcal{S}_L$. Assume that there is no $s_1^L \in A_{i-1}^L$ such that $\phi(s_1^L) = \pi$, but there exists $s_1^L \in A_i^L$ such that $\phi(s_1^L) = \pi$ (When $i = 1$ we define $A_{i-1} = A_0 = \emptyset$). Then,*

(i) there exists a unique $s_1^L \in A_i^L$ such that $\phi(s_1^L) = \pi$. Moreover, if $\phi(t_1^L) = \pi$ for $t_1^L \in A_n^L$, then there exist $c_1, \cdots, c_L$ such that $s_k + c_k = t_k$ for $k = 1, \cdots, L$ and $0 \leq c_{\pi(1)} \leq \cdots \leq c_{\pi(L)} \leq n - i$.

(ii) $|\phi^{-1}(\pi)| = \binom{L+n-i}{n-i}$, where $|X|$ denotes the cardinality of a set X.

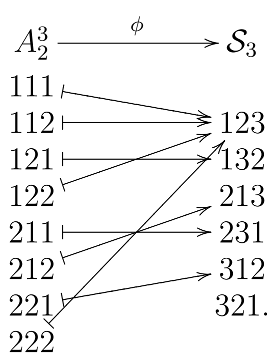
Proof. See Lemma 5 in [14]. (ii) follows from the fact that the number of sequences $a_1 \cdots a_L$ satisfying $0 \leq a_1 \leq a_2 \leq \cdots \leq a_L \leq n-i$ is given by the binomial coefficient $\binom{L+n-i}{n-i}$. $\square$

For example, let $\phi \in \mathcal{S}_5$ be given by $\pi(1)\pi(2)\pi(3)\pi(4)\pi(5) = 24315$. We have $\phi(s_1^5) = \pi$ for $s_1^5 = s_1 s_2 s_3 s_4 s_5 = 31213 \in A_3^5$. Consider $t_1^5 = t_1 t_2 t_3 t_4 t_5 = 41325 \in A_5^5$ and $c_1 c_2 c_3 c_4 c_5 = 10112$. We have $\phi(t_1^5) = \pi$ and $t_2 t_4 t_3 t_1 t_5 = 12345 = 11233 + 01112 = s_2 s_4 s_3 s_1 s_5 + c_2 c_4 c_3 c_1 c_5$.

As a more thorough illustration of Lemma 2, let us write down how ϕ sends each word to its permutation type for $L = 3$ and $n = 1, 2$.

When $n = 1$, the unique element $111 \in A_1^3$ is mapped to $123 \in \mathcal{S}_3$.

When $n = 2$, we have



For example, there is no $s_1^3 \in A_1^3$ such that $\phi(s_1^3) = 132 \in \mathcal{S}_3$. On the other hand, $\phi^{-1}(132) = \{121\}$ for $\phi : A_2^3 \to \mathcal{S}_3$. We have $\phi^{-1}(123) = \{111, 112, 122, 222\}$ for $\phi : A_2^3 \to \mathcal{S}_3$. Note that $|\phi^{-1}(123)| = 4 = \binom{3+2-1}{2-1}$.

Let us introduce the map $\mu : \mathcal{S}_L \to \mathbb{N}^L$, where $\mathbb{N} = \{1, 2, \cdots\}$ is the set of all natural numbers, by the following procedure:

(i) given a permutation $\pi \in \mathcal{S}_L$, we decompose the sequence $\pi(1) \cdots \pi(L)$ into *maximal ascending subsequences*. A subsequence $i_j \cdots i_{j+k}$ of a sequence $i_1 \cdots i_L$ is called a *maximal ascending subsequence* if it is ascending, namely, $i_j \leq i_{j+1} \leq \cdots \leq i_{j+k}$, and neither $i_{j-1} i_j \cdots i_{j+k}$ nor $i_j i_{j+1} \cdots i_{j+k+1}$ is ascending.

(ii) If $\pi(1)\cdots\pi(i_1),\ \pi(i_1{+}1)\cdots\pi(i_2),\cdots,\pi(i_{k-1}{+}1)\cdots\pi(L)$ is the decomposition of $\pi(1)\cdots\pi(L)$ into maximal ascending subsequences, then we define the word $s_1^L \in \mathbb{N}^L$ by

$$s_{\pi(1)} = \cdots = s_{\pi(i_1)} = 1, s_{\pi(i_1+1)} = \cdots = s_{\pi(i_2)} = 2, \cdots,$$
$$s_{\pi(i_{k-1})+1} = \cdots = s_{\pi(L)} = k.$$

(iii) We define $\mu(\pi) = s_1^L$.

By construction, we have $\phi \circ \mu(\pi) = \pi$ when $\mu(\pi) \in A_n^L$ for all $\pi \in \mathcal{S}_L$.

For example, the decomposition of $15423 \in \mathcal{S}_5$ into maximal ascending subsequences is $15, 4, 23$. We obtain $\mu(\pi) = s_1s_2s_3s_4s_5 = 13321$ by putting $s_1s_5s_4s_2s_3 = 11233$.

The map μ can be seen as the dual to the map ϕ in the following sense:

Theorem 3 *Let us put*

$$\begin{aligned} X &= \{s_1^L \in A_n^L | \phi^{-1}(\pi) = \{s_1^L\} \text{ for some } \pi \in \mathcal{S}_L\}, && (1)\\ Y &= \{\pi \in \mathcal{S}_L || \phi^{-1}(\pi)| = 1\}. && (2) \end{aligned}$$

Then, ϕ restricted on X is a map into Y, μ restricted on Y is a map into X, and they form a pair of mutually inverse maps. Furthermore, we have

$$X = \{s_1^L \in A_n^L | \text{for any } 1 \le i \le n-1 \text{ there exist } 1 \le j < k \le L \text{ such that } s_j = i+1, s_k = i\} \qquad (3)$$

Proof. See Theorem 9 in [14]. □

For the map $\phi : A_2^3 \to \mathcal{S}_3$, the duality

$$\begin{tikzcd} X \arrow[r, shift left, "\phi"] & Y \arrow[l, shift left, "\mu"] \end{tikzcd} \qquad (4)$$

is given by

$$\begin{tikzcd}[row sep=tiny] 121 \arrow[r, leftrightsquigarrow] & 132 \\ 211 \arrow[dr, leftrightsquigarrow] & 213 \\ 212 \arrow[ur, leftrightsquigarrow] & 231 \\ 221 \arrow[r, leftrightsquigarrow] & 312. \end{tikzcd}$$

3 Permutation Excess Entropy

Let $\mathbf{S} = \{S_1, S_2, \cdots\}$ be a finite-state stationary stochastic process, where each stochastic variable S_i takes its value in A_n. By stationarity, we mean

$$\Pr\{S_1 = s_1, \cdots, S_L = s_L\} = \Pr\{S_{k+1} = s_1, \cdots, S_{k+L} = s_L\}$$

for any $k, L \geq 1$ and $s_1, \cdots, s_L \in A_n$. Hence, we can define the probability of the occurrence of each word $s_1^L \in A_n^L$ by $p(s_1^L) := p(s_1 \cdots s_L) := \Pr\{S_1 = s_1, \cdots, S_L = s_L\}$.

The *entropy rate* $h(\mathbf{S})$ of a finite-state stationary stochastic process $\mathbf{S}$, which quantifies the average uncertainty of values per time unit, is defined by

$$h(\mathbf{S}) = \lim_{L\to\infty} \frac{1}{L} H(S_1^L), \tag{5}$$

where $H(S_1^L) = H(S_1, \cdots, S_L) = -\sum_{s_1^L \in A_n^L} p(s_1^L) \log_2 p(s_1^L)$. The limit exists for any finite-state stationary stochastic process [8].

The *permutation entropy rate* quantifies the average uncertainty of orderings of values per time unit. It is defined by

$$h^*(\mathbf{S}) = \lim_{L\to\infty} \frac{1}{L} H^*(S_1^L) \tag{6}$$

if the limit exists, where $H^*(S_1^L) = H^*(S_1, \cdots, S_L) = -\sum_{\pi \in \mathcal{S}_L} p(\pi) \log_2 p(\pi)$ and $p(\pi)$ is the probability that π is realized in $\mathbf{S}$, namely, $p(\pi) = \sum_{s_1^L \in \phi^{-1}(\pi)} p(s_1^L)$ for $\pi \in \mathcal{S}_L$.

Theorem 4 *For any finite-state stationary stochastic process* $\mathbf{S}$*, the permutation entropy rate* $h^*(\mathbf{S})$ *exists and*

$$h^*(\mathbf{S}) = h(\mathbf{S}). \tag{7}$$

Proof. The proof appealing to ergodic theory is found in [1, 2]. For an alternative proof based on the duality between values and orderings, see [14]. □

The entropy rate can be seen as a measure of randomness of a finite-state stationary stochastic process. Meanwhile the excess entropy can be interpreted as a measure of complexity [12]. More precisely, it measures global correlation present in a system. The *excess entropy* $\mathbf{E}(\mathbf{S})$ of a finite-state stationary stochastic process $\mathbf{S}$ is defined by [10]

$$\mathbf{E}(\mathbf{S}) = \lim_{L\to\infty} \left(H(S_1^L) - h(\mathbf{S})L \right) \tag{8}$$

if the limit exists. If $\mathbf{E}(\mathbf{S})$ exists, then we have [10]

$$\mathbf{E}(\mathbf{S}) = \sum_{L=1}^{\infty} \left(H(S_L|S_1^{L-1}) - h(\mathbf{S}) \right) = \lim_{L \to \infty} I(S_1^L; S_{L+1}^{2L}), \tag{9}$$

where $H(Y|X)$ is the conditional entropy of Y given X and $I(X;Y)$ is the mutual information between X and Y for stochastic variables X and Y.

The permutation excess entropy was introduced in [14] by imitating the definition of the excess entropy. The *permutation excess entropy* $\mathbf{E}^*(\mathbf{S})$ of a finite-state stationary stochastic process $\mathbf{S}$ is defined by

$$\mathbf{E}^*(\mathbf{S}) = \lim_{L \to \infty} \left(H^*(S_1^L) - h^*(\mathbf{S})L \right), \tag{10}$$

if the limit exists. However, it is unclear what form of correlation the permutation excess entropy quantifies from this expression. In the following discussion, we partially resolve this problem. We will show that the equality

$$\mathbf{E}^*(\mathbf{S}) = \lim_{L \to \infty} I(\phi(S_1^L); \phi(S_{L+1}^{2L})) \tag{11}$$

holds for any *finite-state stationary ergodic Markov process* $\mathbf{S}$. Recall that the entropy rate and the excess entropy of a finite-state stationary Markov process $\mathbf{S}$ are given by

$$h(\mathbf{S}) = -\sum_{i,j=1}^{n} p_i p_{ij} \log_2 p_{ij}$$

and

$$\mathbf{E}(\mathbf{S}) = -\sum_{i=1}^{n} p_i \log_2 p_i + \sum_{i,j=1}^{n} p_i p_{ij} \log_2 p_{ij},$$

respectively, where $P = (p_{ij})$ is the *transition matrix* and $\mathbf{p} = (p_1, \cdots, p_n)$ is the *stationary distribution*. P and $\mathbf{p}$ satisfy the following properties:

(i) $p_{ij} \geq 0$ for all $1 \leq i, j \leq n$.

(ii) $\sum_{j=1}^{n} p_{ij} = 1$ for all $1 \leq i \leq n$.

(iii) $p_i \geq 0$ for all $1 \leq i \leq n$.

(iv) $\sum_{i=1}^{n} p_i = 1$.

(v) $\sum_{i=1}^{n} p_i p_{ij} = p_j$ for all $1 \leq j \leq n$.

p_{ij} is the probability of transition from state i to state j. p_i is the probability of observing state i. Thus, the probability of the occurrence of each word $s_1^L \in A_n^L$ is given by $p(s_1^L) = p_{s_1} p_{s_1 s_2} \cdots p_{s_{L-1} s_L}$. A finite-state stationary Markov process $\mathbf{S}$ is *ergodic* if and only if its transition matrix P is *irreducible* [20]: a matrix P is *irreducible* if for all $1 \leq i, j \leq n$ there exists $l > 0$ such that $p_{ij}^{(l)} > 0$, where $p_{ij}^{(l)}$ is the (i,j)-th element of P^l. For an irreducible non-negative matrix P, stationary distribution $\mathbf{p} = (p_1, \cdots, p_n)$ exists uniquely and satisfies $p_i > 0$ for all $1 \leq i \leq n$.

In our previous work [14], we showed that the equality

$$\mathbf{E}^*(\mathbf{S}) = \mathbf{E}(\mathbf{S}) \tag{12}$$

holds for any finite-state stationary ergodic Markov process. The key point of the proof is that the probability

$$q_L = \sum_{\substack{\pi \in \mathcal{S}_L, \\ |\phi^{-1}(\pi)| > 1}} p(\pi) = \sum_{\pi \notin Y} p(\pi) \tag{13}$$

diminishes exponentially fast as $L \to \infty$ for any finite-state stationary ergodic Markov process, where the set Y is given by (2) in Theorem 3. For the proof of the equality (11), we also appeal to this fact. Hence, we shortly review the reason why this fact follows.

Let L be a positive integer. We introduce the following probability β_s for each symbol $s \in A_n$:

$$\beta_s = \Pr\{s_1^N | s_j \neq s \text{ for any } 1 \leq j \leq N\}, \tag{14}$$

where $N = \lfloor L/2 \rfloor$ and $\lfloor x \rfloor$ is the largest integer not greater than x.

Lemma 5 (Lemma 12 in [14]) *Let* $\mathbf{S}$ *be a finite-state stationary stochastic process and* ϵ *be a positive real number. If* $\beta_s \leq \epsilon$ *for any* $s \in A_n$*, then* $q_L \leq 2n\epsilon$*.*

Proof. We shall prove $\sum_{\pi \in Y} p(\pi) \geq 1 - 2n\epsilon$, where the set Y is given by (2) in Theorem 3. Let us consider words $s_1^L \in A_n^L$ satisfying the following two conditions:

(i) Each symbol $s \in A_n$ appears in s_1^N at least once.

(ii) Each symbol $s \in A_n$ appears in s_{N+1}^L at least once.

By the assumption of the lemma, we have

$$\Pr\{s_1^N | \text{(i) holds}\} \geq 1 - n\epsilon,$$

because

$$\Pr\{s_1^N | \text{(i) holds}\} + \sum_{s=1}^{n} \Pr\{s_1^N | s_j \neq s \text{ for any } 1 \leq j \leq N\} \geq 1.$$

Similarly,

$$\Pr\{s_{N+1}^L | \text{(ii) holds}\} \geq 1 - n\epsilon$$

holds because of the stationarity. Hence, we obtain

$$\Pr\{s_1^L | \text{both (i) and (ii) hold}\} \geq 1 - 2n\epsilon.$$

Since any word $s_1^L \in A_n^L$ satisfying both (i) and (ii) is a member of the set X given by (1) in Theorem 3, we obtain

$$\sum_{\pi \in Y} p(\pi) = \sum_{s_1^L \in X} p(s_1^L) \geq \Pr\{s_1^L | \text{both (i) and (ii) hold}\} \geq 1 - 2n\epsilon.$$

□

Let $\mathbf{S}$ be a finite-state stationary ergodic Markov process whose transition matrix is P and stationary distribution is $\mathbf{p}$. We can write β_s in the following form by using Markov property:

$$\beta_s = \sum_{\substack{s_j \neq s, \\ 1 \leq j \leq N}} p(s_1 \cdots p_N) = \sum_{\substack{s_j \neq s, \\ 1 \leq j \leq N}} p_{s_1} p_{s_1 s_2} \cdots p_{s_{N-1} s_N} = \langle (P_s)^{N-1} \mathbf{u}_s, \mathbf{p} \rangle, \qquad (15)$$

where the matrix P_s is defined by

$$(P_s)_{ij} = \begin{cases} 0 & \text{if } i = s \\ p_{ij} & \text{otherwise,} \end{cases}$$

the vector $\mathbf{u}_s = (u_1, \cdots, u_n)$ is defined by $u_i = 0$ if $i = s$ otherwise $u_i = 1$ and $\langle \cdots, \cdots \rangle$ is the usual inner product in the n-dimensional Euclidean space.

We can prove that the non-negative largest eigenvalue λ of P_s is strictly less than 1 and absolute value of any other eigenvalue of P_s is not greater than λ by using Perron-Frobenius Theorem for non-negative matrices and the irreducibility of P (Lemma 13 in [14]). Hence, by decomposing P_s into a sum of a diagonalizable matrix and a nilpotent matrix, we obtain the following lemma:

Lemma 6 *Let* $\mathbf{S}$ *be a finite-state stationary ergodic Markov process. There exists* $0 \leq \alpha < 1$, $C > 0$ *and a positive integer* k *such that* $\beta_s \leq C \alpha^L L^k$ *for any* $s \in A_n$ *and sufficiently large* L.

4 Mutual Information Expression of Permutation Excess Entropy

In this section, we give a proof of the equality (11) for finite-state stationary ergodic Markov processes. We make use of the notions of *rank sequences* and *rank variables* which are introduced in [2].

Rank sequences of length L are words $r_1^L \in \mathbb{N}^L$ satisfying $1 \le r_i \le i$ for $i = 1, \cdots, L$. We denote the set of all rank sequences of length L by $\mathcal{R}_L$. Clearly, $|\mathcal{R}_L| = L! = |\mathcal{S}_L|$.

We can transform each word $s_1^L \in A_n^L$ into a rank sequence $r_1^L \in \mathcal{R}_L$ by defining

$$r_i = \sum_{j=1}^{i} \delta(s_j \le s_i),\ i = 1, \cdots, L, \tag{16}$$

where $\delta(X) = 1$ if the proposition X is true, otherwise $\delta(X) = 0$. Namely, r_i is the number of indexes j $(1 \le j \le i)$ such that $s_j \le s_i$. Thus, we obtain the map $\varphi : A_n^L \to \mathcal{R}_L$ such that $\varphi(s_1^L) = r_1^L$.

We can show that the map $\varphi : A_n^L \to \mathcal{R}_L$ is compatible with the map $\phi : A_n^L \to \mathcal{S}_L$. Namely, there exists a bijection $\iota : \mathcal{R}_L \to \mathcal{S}_L$ satisfying $\iota \circ \varphi = \phi$ [14].

Given a stationary stochastic process $\mathbf{S} = \{S_1, S_2, \cdots\}$, its associated *rank variables* are defined by $R_i = \sum_{j=1}^{n} \delta\left(S_j \le S_i\right)$ for $i = 1, 2, \cdots$. Note that rank variables R_i $(i = 1, 2, \cdots)$ are not stationary stochastic variables in general. By the compatibility between ϕ and φ, we have

$$H(R_1^L) = H^*(S_1^L) = H(\phi(S_1^L)) \tag{17}$$

for $L \ge 1$.

Now, let $\mathbf{S}$ be a finite-state stationary ergodic Markov process. By (12), we know that the permutation excess entropy $\mathbf{E}^*(\mathbf{S})$ exists. By (17) and the chain rule, we have

$$\begin{aligned}\mathbf{E}^*(\mathbf{S}) = \lim_{L\to\infty}\left(H^*(S_1^L) - h^*(\mathbf{S})L\right) &= \lim_{L\to\infty}\left(H(R_1^L) - h^*(\mathbf{S})L\right) \\ &= \sum_{L=1}^{\infty}\left(H(R_L|R_1^{L-1}) - h^*(\mathbf{S})\right).\end{aligned} \tag{18}$$

Since the infinite sum in (18) converges, we obtain

$$\left|H(R_{L+1}^{2L}|R_1^L) - h^*(\mathbf{S})L\right| = \left|\sum_{i=1}^{L}\left(H(R_{L+i}|R_1^{L+i-1}) - h^*(\mathbf{S})\right)\right| \underset{L\to\infty}{\to} 0. \tag{19}$$

By the definition of mutual information, we have

$$I(\phi(S_1^L); \phi(S_{L+1}^{2L})) = H(\phi(S_{L+1}^{2L})) - H(\phi(S_{L+1}^{2L})|\phi(S_1^L)).$$

In addition, we have $H(\phi(S_{L+1}^{2L})) = H(\phi(S_1^L)) = H^*(S_1^L)$ by the stationarity of $\mathbf{S}$. Hence, it is sufficient to show that

$$\left|H(\phi(S_{L+1}^{2L})|\phi(S_1^L)) - h^*(\mathbf{S})L\right| \underset{L\to\infty}{\to} 0 \tag{20}$$

to prove the equality (11). However, by (19), this reduces to showing that

$$\left|H(\phi(S_{L+1}^{2L})|\phi(S_1^L)) - H(R_{L+1}^{2L}|R_1^L)\right| \underset{L\to\infty}{\to} 0, \tag{21}$$

which is equivalent to showing that

$$\left|H(\phi(S_1^L),\phi(S_{L+1}^{2L}))-H(\phi(S_1^{2L}))\right| \underset{L\to\infty}{\longrightarrow} 0 \tag{22}$$

by (17).

Lemma 7 *For $s_1^{2L}, t_1^{2L} \in A_n^{2L}$, if $\phi(s_1^{2L}) = \phi(t_1^{2L})$, then $\phi(s_1^L) = \phi(t_1^L)$ and $\phi(s_{L+1}^{2L}) = \phi(t_{L+1}^{2L})$. Namely, the partition of A_n^{2L} by the map $\phi : A_n^{2L} \to \mathcal{S}_{2L}$ is a refinement of the partition of $A_n^L \times A_n^L = A_n^{2L}$ by the map $\phi \times \phi : A_n^L \times A_n^L \to \mathcal{S}_L \times \mathcal{S}_L$.*

Proof. The claim follows immediately from Lemma 1.

□

Lemma 8

$$\begin{aligned} 0 &\leq H(\phi(S_1^{2L}))-H(\phi(S_1^L),\phi(S_{L+1}^{2L})) \\ &\leq \left(\sum_{\substack{\pi',\pi''\in\mathcal{S}_L,\\ |\phi^{-1}(\pi')|>1 \text{ or } |\phi^{-1}(\pi'')|>1}} p(\pi',\pi'')\right) 2n\log_2(L+n) \end{aligned} \tag{23}$$

holds for any finite-state stationary stochastic process $\mathbf{S}$*, where*

$$p(\pi',\pi'') = \sum_{\substack{s_1^L\in\phi^{-1}(\pi'),\\ s_{L+1}^{2L}\in\phi^{-1}(\pi'')}} p(s_1^{2L})$$

for $\pi',\pi''\in\mathcal{S}_L$.

Proof. By Lemma 7, we can write

$$\begin{aligned} &H(\phi(S_1^{2L}))-H(\phi(S_1^L),\phi(S_{L+1}^{2L})) \\ =& -\sum_{\pi\in\mathcal{S}_{2L}} p(\pi)\log_2 p(\pi) + \sum_{\pi',\pi''\in\mathcal{S}_L} p(\pi',\pi'')\log_2 p(\pi',\pi'') \\ =& \sum_{\pi',\pi''\in\mathcal{S}_L}\left(-\sum_{\substack{\phi^{-1}(\pi)\subseteq\\ (\phi\times\phi)^{-1}(\pi',\pi'')}} p(\pi)\log_2 p(\pi) + p(\pi',\pi'')\log_2 p(\pi',\pi'')\right) \\ =& \sum_{\pi',\pi''\in\mathcal{S}_L}\left(-\sum_{\substack{\phi^{-1}(\pi)\subseteq\\ (\phi\times\phi)^{-1}(\pi',\pi'')}} p(\pi)\log_2 p(\pi) + \sum_{\substack{\phi^{-1}(\pi)\subseteq\\ (\phi\times\phi)^{-1}(\pi',\pi'')}} p(\pi)\log_2 p(\pi',\pi'')\right) \\ =& \sum_{\substack{\pi',\pi''\in\mathcal{S}_L,\\ p(\pi',\pi'')>0}} p(\pi',\pi'')\left(-\sum_{\substack{\phi^{-1}(\pi)\subseteq\\ (\phi\times\phi)^{-1}(\pi',\pi'')}} \frac{p(\pi)}{p(\pi',\pi'')}\log_2\frac{p(\pi)}{p(\pi',\pi'')}\right). \end{aligned}$$

By Lemma 2 (ii), we have

$$0 \leq - \sum_{\substack{\phi^{-1}(\pi) \subseteq \\ (\phi\times\phi)^{-1}(\pi',\pi'')}} \frac{p(\pi)}{p(\pi',\pi'')} \log_2 \frac{p(\pi)}{p(\pi',\pi'')} \leq 2n \log_2(L+n).$$

If $|\phi^{-1}(\pi')| = 1$ and $|\phi^{-1}(\pi'')| = 1$ hold for $(\pi', \pi'') \in \mathcal{S}_L \times \mathcal{S}_L$, then $|(\phi \times \phi)^{-1}(\pi', \pi'')| = 1$. In this case, if $p(\pi', \pi'') > 0$, then we have

$$- \sum_{\substack{\phi^{-1}(\pi) \subseteq \\ (\phi\times\phi)^{-1}(\pi',\pi'')}} \frac{p(\pi)}{p(\pi',\pi'')} \log_2 \frac{p(\pi)}{p(\pi',\pi'')} = 0.$$

□

Lemma 9 *(22) holds for any finite-state stationary ergodic Markov process* $\mathbf{S}$*.*

Proof. We have

$$\begin{aligned}
\sum_{\substack{\pi',\pi'' \in \mathcal{S}_L, \\ |\phi^{-1}(\pi')|>1 \text{ or } |\phi^{-1}(\pi'')|>1}} p(\pi', \pi'') &\leq \sum_{\substack{|\phi^{-1}(\pi')|>1, \\ \pi'' \in \mathcal{S}_L}} p(\pi', \pi'') + \sum_{\substack{|\phi^{-1}(\pi'')|>1, \\ \pi' \in \mathcal{S}_L}} p(\pi', \pi'') \\
&= 2 \sum_{|\phi^{-1}(\pi')|>1} p(\pi') = 2q_L.
\end{aligned}$$

By Lemma 5 and Lemma 6, there exist $0 \leq \alpha < 1$, $C > 0$ and $k > 0$ such that $q_L \leq C\alpha^L L^k$ for sufficiently large L if $\mathbf{S}$ is a finite-state stationary ergodic Markov process. The claim follows from Lemma 8.

□

Thus, we get our main theorem in this paper:

Theorem 10 *The equality (11)*

$$\mathbf{E}^*(\mathbf{S}) = \lim_{L\to\infty} I(\phi(S_1^L); \phi(S_{L+1}^{2L}))$$

holds for any finite-state stationary ergodic Markov process $\mathbf{S}$*.*

5 Conclusions

In this paper, we showed that the permutation excess entropy is equal to the mutual information between the past and future in the space of orderings for finite-state stationary ergodic Markov processes.

Combining this result with the equality between the excess entropy and the permutation excess entropy for finite-state stationary ergodic Markov processes proved in our previous work [14], we can see that the global correlation measured by the mutual information between the past and future in a given finite-state stationary ergodic Markov process is fully captured by a different coding of the process, namely, coding by the orderings of values.

We hope that our result gives rise to a new insight into the relationship between complexity and anticipation.

Acknowledgements TH was supported by JST PRESTO program. The authors acknowledge anonymous referees for their helpful comments to improve the manuscript.

References

[1] J. M. Amigó (2010). Permutation Complexity in Dynamical Systems. Springer-Verlag Berlin Heidelberg.

[2] J. M. Amigó, M. B. Kennel, L. Kocarev (2005). The permutation entropy rate equals the metric entropy rate for ergodic information sources and ergodic dynamical systems. Physica D 210, 77-95.

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

[4] D. V. Arnold (1996). Information-theoretic analysis of phase transitions. Complex Systems 10, 143-155.

[5] C. Bandt, B. Pompe (2002). Permutation entropy: a natural complexity measure for time series. Physical Review Letters 88, 174102.

[6] C. Bandt, G. Keller, B. Pompe (2002). Entropy of interval maps via permutations. Nonlinearity 15, 1595-1602.

[7] W. Bialek, I. Nemenman, N. Tishby (2001). Predictability, complexity, and learning. Neural Computation 13, 2409-2463.

[8] T. M. Cover, J. A. Thomas (1991). Elements of Information Theory. John Wiley & Sons, Inc.

[9] J. P. Crutchfield, N. H. Packard (1983). Symbolic dynamics of noisy chaos. Physica D 7, 201-223.

[10] J. P. Crutchfield, D. P. Feldman (2003). Regularities unseen, randomness observed: Levels of entropy convergence. Chaos 15, 25-54.

[11] B. A. Davey, H. A. Priestley (2002). Introduction to Lattices and Order, second edition. Cambridge Univ. Press, Cambridge.

[12] D. P. Feldman, C. S. McTague, J. P. Crutchfield (2008). The organization of intrinsic computation: complexity-entropy diagrams and the diversity of natural information processing. Chaos 18, 043106.

[13] P. Grassberger (1986). Toward a quantitative theory of self-generated complexity. International Journal of Theoretical Physics 25, 907-938.

[14] T. Haruna, K. Nakajima (2011). Permutation Complexity via Duality between Values and Orderings. Physica D 240, 1370-1377.

[15] K. Keller, M. Sinn (2010). Kolmogorov-Sinai entropy from the ordinal viewpoint. Physica D 239, 997-1000.

[16] W. Li (1991). On the relationship between complexity and entropy for Markov chains and regular languages. Complex Systems 5, 381-399.

[17] S. MacLane (1998). Categories for the Working Mathematician, second edition. Springer-Verlag, New York.

[18] M. Misiurewicz (2003). Permutations and topological entropy for interval maps. Nonlinearity 16, 971-976.

[19] R. Shaw (1983). The Dripping Faucet as a Model Chaotic System. Aerial Press, Santa Cruz, California.

[20] P. Walters (1982). An Introduction to Ergodic Theory. Springer-Verlag New York, Inc.