EPJ manuscript No.
(will be inserted by the editor)

Permutation approach to finite-alphabet stationary stochastic processes based on the duality between values and orderings

Taichi Haruna[1,a] and Kohei Nakajima[2]

[1] Department of Earth & Planetary Sciences, Graduate School of Science, Kobe University, 1-1 Rokkodaicho, Nada, Kobe 657-8501, Japan
[2] Department of Mechanical and Process Engineering, ETH Zurich, Leonhardstrasse 27, 8092 Zurich, Switzerland

Abstract. The duality between values and orderings is a powerful tool to discuss relationships between various information-theoretic measures and their permutation analogues for discrete-time finite-alphabet stationary stochastic processes (SSPs). Applying it to output processes of hidden Markov models with ergodic internal processes, we have shown in our previous work that the excess entropy and the transfer entropy rate coincide with their permutation analogues. In this paper, we discuss two permutation characterizations of the two measures for general ergodic SSPs not necessarily having the Markov property assumed in our previous work. In the first approach, we show that the excess entropy and the transfer entropy rate of an ergodic SSP can be obtained as the limits of permutation analogues of them for the N-th order approximation by hidden Markov models, respectively. In the second approach, we employ the modified permutation partition of the set of words which considers equalities of symbols in addition to permutations of words. We show that the excess entropy and the transfer entropy rate of an ergodic SSP are equal to their modified permutation analogues, respectively.

1 Introduction

This paper concerns permutation characterizations of information-theoretic measures for discrete-time finite-alphabet stationary stochastic processes (SSPs). The permutation entropy was first introduced by Bandt and Pompe [1] as a simple complexity measure for time series. It has been shown that significant amount of the information contained in time series can be captured by orderings of values through both applications to real-world time series data ([2] and references therein) and theoretical developments [3–8].

In our previous work, we showed that a combinatorial discussion on the relationship between the set of all words and the set of all permutations reveals a certain aspect of the relationship between information-theoretic measures and their permu-

[a] e-mail: tharuna@penguin.kobe-u.ac.jp

tation analogues for SSPs. It is called the duality between values and orderings [9]. By making use of the duality, we obtained a new proof of the equality between the entropy rate and the permutation entropy rate [9] for all SSPs. For ergodic Markov SSPs, we proved the equality between the excess entropy (a measure of complexity [10]) and the permutation excess entropy [9], the mutual information expression of the permutation excess entropy [11], and the equality between the transfer entropy rate (a measure of magnitude and direction of information flow [12,13]) and the symbolic transfer entropy rate [14]. Later these results were extended to SSPs that are output processes of discrete-time finite-state finite-alphabet stationary hidden Markov models (HMMs) with ergodic internal processes [15].

The aim of this paper is to give permutation characterizations of the excess entropy and the symbolic transfer entropy rate for general ergodic SSPs not necessarily having the Markov property assumed in our previous work. We take two approaches for the purpose. The first approach is based on the N-th order approximation of an SSP by HMM. We show that the excess entropy for an ergodic SSP can be calculated as the limit of the permutation excess entropy of the N-th order approximation of the SSP by HMM. The similar result for the symbolic transfer entropy rate is also proved. In the second approach, we employ the modified permutation partition of the set of words which considers equalities of symbols in addition to permutations of words. In recent years, several types of modification on the permutation entropies have been proposed [16–18]. The one considered here is based on [16]. We show that the excess entropy and the transfer entropy rate of an ergodic SSP coincide with their modified permutation analogues, respectively.

This paper is organized as follows. In Section 2, we summarize necessary background on information-theoretic measures for SSPs. In Section 3, we review the duality between values and orderings. In Section 4, we discuss the relationship between the three information-theoretic measures mentioned above and their permutation analogues for ergodic SSPs in terms of the approximation by HMM. In Section 5, we show that the three information-theoretic measures are equal to their modified permutation analogues, respectively, for ergodic SSPs. Finally, in Section 6, we indicate a future direction of the study.

2 Information-Theoretic Measures for Discrete-Time Finite-Alphabet Stationary Stochastic Processes

In this section, we review information-theoretic measures for discrete-time finite-alphabet stationary stochastic processes considered in this paper.

2.1 Discrete-Time Finite-Alphabet Stationary Stochastic Processes

Let $\mathbf{X} = \{X_1, X_2, \cdots\}$ be a discrete-time finite-alphabet stationary stochastic processes over a finite alphabet A (in short, SSP $\mathbf{X}$ over A). $X_1, X_2, \cdots$ are stochastic variables on a probability space taking their values in A. Since $\mathbf{X}$ is assumed to be stationary, namely, we have

$$\Pr\{X_1 = x_1, X_2 = x_2, \cdots, X_L = x_L\} = \Pr\{X_k = x_1, X_{k+1} = x_2, \cdots, X_{L+k-1} = x_L\}$$

for any word $x_{1:L} := x_1 x_2 \cdots x_L \in A^L = \underbrace{A \times \cdots \times A}_{L}$ and for $L, k \in \mathbb{N}$, we can define the probability of the occurrence of the each word of length L by

$$p(x_{1:L}) = \Pr\{X_1 = x_1, X_2 = x_2, \cdots, X_L = x_L\}.$$

An SSP $\mathbf{X}$ over A is said to be *ergodic* [19] if the relative frequency of each word $x_{1:L}$ converges to the probability $p(x_{1:L})$ in probability, namely, for any word $x_{1:k}$ of length $k \geq 1$, any $\epsilon > 0$ and any $\delta > 0$, there exists a natural number K such that if $L > K$, then

$$\Pr\{|F_{x_{1:k},L} - p(x_{1:k})| < \delta\} > 1 - \epsilon,$$

where $F_{x_{1:k},L}$ is the stochastic variable defined by the number of the occurrence of the word $x_{1:k}$ in the sequence $X_1, X_2, \cdots, X_L$ divided by $L - k + 1$.

Let $\{\mathbf{X}_N\}_{N\in\mathbb{N}}$ be a sequence of SSPs over a common alphabet A and $\mathbf{X}$ an SSP over A. The probability of the occurrence of the words for $\mathbf{X}_N$ is denoted by p_N and that for $\mathbf{X}$ by p. $\{\mathbf{X}_N\}_{N\in\mathbb{N}}$ is said to *converge weakly to* $\mathbf{X}$ [20,21] if

$$\lim_{N\to\infty} p_N(x_{1:L}) = p(x_{1:L})$$

for any $L \geq 1$ and word $x_{1:L} \in A^L$. We write $\mathbf{X}_N \Rightarrow \mathbf{X}$ when $\mathbf{X}_N$ converges weakly to $\mathbf{X}$.

2.2 Discrete-Time Finite-State Finite-Alphabet Stationary Hidden Markov Models

A *discrete-time finite-state finite-alphabet stationary hidden Markov model* (in short, HMM) [22] is a quadruple $(\Sigma, A, \{T^{(a)}\}_{a\in A}, \mu)$, where Σ and A are finite sets called *state set* and *alphabet*, respectively, $\{T^{(a)}\}_{a\in A}$ is a family of $|\Sigma|\times|\Sigma|$ matrices indexed by elements of A where $|\Sigma|$ is the size of state set Σ, and μ is a probability distribution on the state set Σ satisfying the following three conditions:

(i) $T^{(a)}_{ss'} \geq 0$ for any $s, s' \in \Sigma$ and $a \in A$,
(ii) $\sum_{s',a} T^{(a)}_{ss'} = 1$ for any $s \in \Sigma$,
(iii) and $\mu(s') = \sum_{s,a} \mu(s) T^{(a)}_{ss'}$ for any $s' \in \Sigma$.

Here, $T^{(a)}_{ss'}$ is the probability of the transition emitting symbol a from state s to state s'. $\mu(s)$ is the probability that the system is in state s. Any probability distribution satisfying the condition (iii) is called a *stationary distribution*. The $|\Sigma| \times |\Sigma|$ matrix $T := \sum_{a\in A} T^{(a)}$ is called *state transition matrix*. The ternary (Σ, T, μ) defines the *underlying Markov chain*. The condition (iii) is equivalent to the following condition:

(iii') $\mu(s') = \sum_s \mu(s) T_{ss'}$.

Two SSPs are induced by an HMM $(\Sigma, A, \{T^{(a)}\}_{a\in A}, \mu)$. The first one comes from the underlying Markov chain and is called an *internal process*. We denote it by $\mathbf{S} = \{S_1, S_2, \cdots\}$. The alphabet for $\mathbf{S}$ is Σ. The joint probability distributions are given by

$$\Pr\{S_1 = s_1, S_2 = s_2, \cdots, S_L = s_L\} := \mu(s_1) T_{s_1 s_2} \cdots T_{s_{L-1} s_L}$$

for any $s_1, \cdots, s_L \in \Sigma$ and $L \geq 1$. The second process $\mathbf{X} = \{X_1, X_2, \cdots\}$ with the alphabet A is defined by the joint probability distributions

$$\Pr\{X_1 = x_1, X_2 = x_2, \cdots, X_L = x_L\} := \sum_{s,s'} \mu(s) \left(T^{(x_1)} \cdots T^{(x_L)}\right)_{ss'}$$

for any $x_1, \cdots, x_L \in A$ and $L \geq 1$. We call it *output process*. Both the internal and output processes are stationary because μ satisfies (iii).

The internal process $\mathbf{S}$ of an HMM $(\Sigma, A, \{T^{(a)}\}_{a\in A}, \mu)$ is ergodic if and only if the state transition matrix T is *irreducible* [23]: for any $s, s' \in \Sigma$ there exists $k > 0$ such that $(T^k)_{ss'} > 0$. If the internal process $\mathbf{S}$ is ergodic, then the stationary distribution μ is uniquely determined by the state transition matrix T by the condition (iii'). It is known that if the internal process $\mathbf{S}$ is ergodic, then the output process $\mathbf{X}$ is so, but not vice versa [24].

There are two types of hidden Markov models depending on the way how outputs are emitted: from edges or states. The HMM defined here is edge emitting type. However, we can convert a hidden Markov model in one of the two classes to the one in the other class without changing the output process [22]. In particular, any discrete-time finite-alphabet finite-order stationary Markov process can be described as an output process of an HMM defined here.

2.3 Approximation of SSP by HMM

Any SSP can be approximated by a sequence of HMMs in the sense of weak convergence. Let $\mathbf{X}$ be an SSP over A and p the corresponding probability of the occurrence of words. For each natural number N, we construct an HMM $(\Sigma^{(N)}, A, \{T^{(N),(a)}\}_{a\in A}, \mu^{(N)})$ by setting $\Sigma^{(N)} = A^N$,

$$\left(T^{(N),(a)}\right)_{ss'} = p(x_{N+1}|x_{1:N})$$

for $s = x_{1:N}$, $s' = x_{2:N+1}$ and $a = x_{N+1}$, where

$$p(x_{N+1}|x_{1:N}) = \begin{cases} p(x_{1:N+1})/p(x_{1:N}) & \text{if } p(x_{1:N}) \neq 0 \\ 0 & \text{otherwise,} \end{cases}$$

and $\mu^{(N)}(x_{1:N}) = p(x_{1:N})$. It is straightforward to confirm that the conditions (i), (ii) and (iii) for HMM hold. We denote the output process of $(\Sigma^{(N)}, A, \{T^{(N),(a)}\}_{a\in A}, \mu^{(N)})$ and the corresponding probability of the occurrence of the words by $\mathbf{X}^{(N)}$ and $p^{(N)}$, respectively. We call $\mathbf{X}^{(N)}$ the *N-th order approximation of* $\mathbf{X}$.

By construction, we have

$$p^{(N)}(x_{1:L}) = p(x_{1:L}) \tag{1}$$

for any word $x_{1:L}$ when $L \leq N$, and

$$p^{(N)}(x_{L+1}|x_{1:L}) = p^{(N)}(x_{L+1}|x_{L-N+1:L}) = p(x_{L+1}|x_{L-N+1:L}) \tag{2}$$

for any word $x_{1:L+1}$ when $L \geq N$. From (1), we have $\mathbf{X}^{(N)} \Rightarrow \mathbf{X}$.

It can be shown that if $\mathbf{X}$ is ergodic, then $\mathbf{X}^{(N)}$ is also ergodic for any N [19].

2.4 Entropy Rate, Excess Entropy and Transfer Entropy Rate

Let $\mathbf{X}$ be an SSP over A. Its *entropy rate* is defined by

$$h(\mathbf{X}) = \lim_{L\to\infty} H(X_{1:L})/L, \tag{3}$$

where

$$H(X_{1:L}) = -\sum_{x_{1:L}\in A^L} p(x_{1:L}) \log_2 p(x_{1:L})$$

is the Shannon entropy with respect to the occurrence of words of length L. It is well-known that $h(\mathbf{X})$ exists for any SSP [25]. It is also well-known that

$$h(\mathbf{X}) = \lim_{L\to\infty} H(X_{L+1}|X_{1:L})$$

and the convergence is monotone-decreasing. The entropy rate is a measure of average uncertainty of values per unit time.

If $\mathbf{X}^{(N)}$ is the N-th order approximation of $\mathbf{X}$, then we have [19]

$$h(\mathbf{X}^{(N)}) = H(X_{N+1}|X_{1:N}). \tag{4}$$

It follows that

$$h(\mathbf{X}) = \lim_{N\to\infty} h(\mathbf{X}^{(N)}). \tag{5}$$

The *excess entropy* is a measure of complexity or global correlation existing in a process [26–31]. It is defined by [10]

$$\begin{aligned} \mathbf{E}(\mathbf{X}) &= \lim_{L\to\infty} \left(H(X_{1:L}) - h(\mathbf{X})L\right) \\ &= \sum_{L=1}^{\infty} \left(H(X_L|X_{1:L-1}) - h(\mathbf{X})\right). \end{aligned} \tag{6}$$

Since each term in the infinite sum in (6) is non-negative, either $\mathbf{E}(\mathbf{X})$ exists or it diverges. $\mathbf{E}(\mathbf{X})$ always exists if $\mathbf{X}$ is an output process of an HMM [24].

The excess entropy can be written as the mutual information between the past and future [10]:

$$\mathbf{E}(\mathbf{X}) = \lim_{L\to\infty} I(X_{1:L}; X_{L+1:2L}), \tag{7}$$

where $I(\mathbf{Y};\mathbf{Z})$ is the mutual information between two stochastic variables $\mathbf{Y}$ and $\mathbf{Z}$.

Since the excess entropy can be written as

$$\mathbf{E}(\mathbf{X}) = \sup_{L\geq 1}\{I(X_{1:L}; X_{L+1:2L})\} \tag{8}$$

and $I(X_{1:L}; X_{L+1:2L})$ is continuous with respect to the weak convergence of SSP, the excess entropy is lower-semi continuous with respect to the weak convergence of SSP [24]. Namely, if $\mathbf{X}_N \Rightarrow \mathbf{X}$, then we have

$$\mathbf{E}(\mathbf{X}) \leq \liminf_{N\to\infty} \mathbf{E}(\mathbf{X}_N). \tag{9}$$

Now, let $\mathbf{X}^{(N)}$ be the N-th order approximation of an SSP $\mathbf{X}$. By (1), (2) and (4), we have

$$\begin{aligned} \mathbf{E}(\mathbf{X}^{(N)}) &= \sum_{L=1}^{N} \left(H(X_{L+1}|X_{1:L}) - H(X_{N+1}|X_{1:N})\right) \\ &\leq \sum_{L=1}^{N} \left(H(X_{L+1}|X_{1:L}) - h(\mathbf{X})\right) \\ &\leq \mathbf{E}(\mathbf{X}). \end{aligned} \tag{10}$$

Since $\mathbf{E}(\mathbf{X}^{(N)})$ is a monotone increasing function of N by (10), if $\mathbf{E}(\mathbf{X})$ exists, then so does $\lim_{N\to\infty}\mathbf{E}(\mathbf{X}^{(N)})$ and

$$\lim_{N\to\infty}\mathbf{E}(\mathbf{X}^{(N)}) \le \mathbf{E}(\mathbf{X}). \tag{11}$$

Since $\mathbf{X}^{(N)} \Rightarrow \mathbf{X}$, we have

$$\lim_{N\to\infty}\mathbf{E}(\mathbf{X}^{(N)}) = \mathbf{E}(\mathbf{X}) \tag{12}$$

by (9) and (11). It also holds that if one side of the equality (12) diverges, then the other side also diverges.

The transfer entropy rate is a measure of the magnitude and direction of information flow between two jointly distributed stochastic processes [12,13]. Let

$$(\mathbf{X},\mathbf{Y}) = \{(X_1,Y_1),(X_2,Y_2),\cdots\}$$

be a bivariate SSP over a finite alphabet $A_X \times A_Y$, where X_i takes its values in A_X and Y_i takes its values in A_Y. We denote the joint probability of the occurrence of words $x_{1:L_1} \in A_X^{L_1}$ and $y_{1:L_2} \in A_Y^{L_2}$ for $L_1, L_2 \ge 1$ by $p(x_{1:L_1}, y_{1:L_2})$.

The *transfer entropy rate* from $\mathbf{Y}$ to $\mathbf{X}$ is defined by

$$t(\mathbf{X}|\mathbf{Y}) = h(\mathbf{X}) - h(\mathbf{X}|\mathbf{Y}), \tag{13}$$

where $h(\mathbf{X}|\mathbf{Y})$ is the *conditional entropy rate* of $\mathbf{X}$ given $\mathbf{Y}$ defined by

$$h(\mathbf{X}|\mathbf{Y}) = \lim_{L\to\infty} H(X_{L+1}|X_{1:L}, Y_{1:L}), \tag{14}$$

which always exists. It is straightforward to see that $t(\mathbf{X}|\mathbf{Y})$ satisfies

$$0 \le t(\mathbf{X}|\mathbf{Y}) \le h(\mathbf{X})$$

and $t(\mathbf{X}|\mathbf{Y}) = 0$ if $X_{1:L}$ is independent of $Y_{1:L}$ for all $L \ge 1$.

Let $(\mathbf{X}^{(N)}, \mathbf{Y}^{(N)})$ be the N-th order approximation of $(\mathbf{X},\mathbf{Y})$. By the similar way as in the univariate case, we can show that

$$h(\mathbf{X}|\mathbf{Y}) = \lim_{N\to\infty} h(\mathbf{X}^{(N)}|\mathbf{Y}^{(N)}). \tag{15}$$

Consequently, we have

$$t(\mathbf{X}|\mathbf{Y}) = \lim_{N\to\infty} t(\mathbf{X}^{(N)}|\mathbf{Y}^{(N)}). \tag{16}$$

3 Duality between Values and Orderings

In this section, we summarize the results in our previous work on the duality between values and orderings [9] which will be used in later sections.

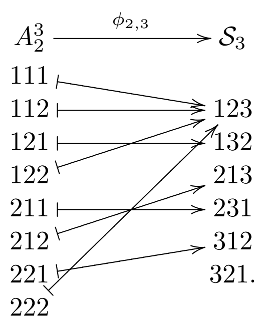
3.1 Partition Induced by Permutations

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

A permutation of length L is a bijective map from the set $\{1, 2, \cdots, L\}$ to itself. The set of all permutations of length $L \geq 1$ is denoted by $\mathcal{S}_L$. We denote a permutation π of length L by the sequence $\pi(1)\pi(2)\cdots\pi(L)$. The number of *ascents*, places with $\pi(i) < \pi(i+1)$, of a permutation $\pi \in \mathcal{S}_L$ is denoted by $\mathrm{Asc}(\pi)$. For example, if $\pi \in \mathcal{S}_5$ is given by $\pi(1)\pi(2)\pi(3)\pi(4)\pi(5) = 24153$, then $\mathrm{Asc}(\pi) = 2$.

Consider a word $x_{1:L}$ of length L over the alphabet A_n, namely, $x_{1:L} \in A_n^L$. We define the *permutation type* of the word $x_{1:L}$ as a unique permutation $\pi \in \mathcal{S}_L$ such that $x_{\pi(i)} \leq x_{\pi(i+1)}$ and $\pi(i) < \pi(i+1)$ when $x_{\pi(i)} = x_{\pi(i+1)}$ for $i = 1, 2, \cdots, L-1$. Thus, the permutation type of $x_{1:L}$ is the sequence of indexes when $x_1, x_2, \cdots, x_L$ are re-ordered in the increasing order. For example, the permutation type of $x_{1:5} = 23121 \in A_3^5$ is $\pi(1)\pi(2)\pi(3)\pi(4)\pi(5) = 35142$ because $x_3x_5x_1x_4x_2 = 11223$.

By sending each word of length L over the alphabet A_n to its permutation type, we define the map $\phi_{n,L} : A_n^L \to \mathcal{S}_L$. For example, the map $\phi_{2,3} : A_2^3 \to \mathcal{S}_3$ is given by



The map $\phi_{n,L}$ gives rise to a partition of the set A_n^L by its fibers, namely, two words $x_{1:L}$ and $y_{1:L}$ are contained in the same block of the partition if and only if $\phi_{n,L}(x_{1:L}) = \phi_{n,L}(y_{1:L})$. The size of each block, namely, the size of the inverse image of each permutation $\pi \in \mathcal{S}_L$ is given by the following binomial coefficient [15, 32]:

$$|\phi_{n,L}^{-1}(\pi)| = \binom{n + \mathrm{Asc}(\pi)}{L}. \tag{17}$$

Thus, the size of the inverse image of a permutation of length L is determined solely by the number of ascents. Since the number of permutations of length L with k ascents is given by the *Eulerian number* $\left\langle {L \atop k} \right\rangle$, we have the following equality with respect to the partition of A_n^L by permutation:

$$n^L = \sum_{k=L-n}^{L-1} \left\langle {L \atop k} \right\rangle \binom{n+k}{L} \tag{18}$$

which is known as *Worpitzky's identity* [33].

The equality (18) should be compared with the similar equality (37) which is obtained from the modified permutation partition defined in Section 5. In the former, the size of each block in the partition of A_n^L depends on L, the length of words. In contrast, it is independent of L in the latter. As we will see in subsequent sections, this difference leads to that in the asymptotic behaviors of the differences between the block entropy and its permutation/modified permutation analogues.

3.2 Duality

We define the map $\mu_{n,L} : \phi_{n,L}\left(A_n^L\right) \subseteq \mathcal{S}_L \to A_n^L$ by the following procedure: first, given a permutation $\pi \in \phi_{n,L}\left(A_n^L\right) \subseteq \mathcal{S}_L$, we decompose the sequence $\pi(1)\cdots\pi(L)$ of length L into *maximal ascending subsequences* defined as follows: 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, $i_j \le i_{j+1} \le \cdots \le i_{j+k}$, and neither $i_{j-1}i_j\cdots i_{j+k}$ nor $i_j i_{j+1}\cdots i_{j+k+1}$ is ascending. Second, if $\pi(1)\cdots\pi(i_1),\ \pi(i_1+1)\cdots\pi(i_2),\cdots,\pi(i_{h-1}+1)\cdots\pi(L)$ is the decomposition of $\pi(1)\cdots\pi(L)$ into maximal ascending subsequences, then we define word $x_{1:L} \in A_n^L$ by

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

Note that $h = L - \mathrm{Asc}(\pi) - 1$. Finally, we define $\mu_{n,L}(\pi) = x_{1:L}$.

By construction, we have $\phi_{n,L} \circ \mu_{n,L}(\pi) = \pi$ for all permutations $\pi \in \phi_{n,L}\left(A_n^L\right)$. For example, the permutation type of $x_{1:5} = 21313 \in A_3^5$ is $\pi(1)\pi(2)\pi(3)\pi(4)\pi(5) = 24135$. The decomposition of 24135 into maximal ascending subsequences is $24, 135$. We obtain $\mu_{n,L}(\pi) = x_1x_2x_3x_4x_5 = 21212$ by putting $x_2x_4x_1x_3x_5 = 11222$.

Let us put

$$B_{n,L} := \{x_{1:L} \in A_n^L : \text{there exists } \pi \in \mathcal{S}_L \text{ such that } \phi_{n,L}^{-1}(\pi) = \{x_{1:L}\}\}$$

and

$$C_{n,L} := \{\pi \in \mathcal{S}_L : |\phi_{n,L}^{-1}(\pi)| = 1\}.$$

Then, the duality between values and orderings can be stated as follows [9](Lemma 5 and Theorem 9 (iv); see also Theorem 1 in [15]):

Theorem 1 *$\phi_{n,L}$ restricted on $B_{n,L}$ is a map into $C_{n,L}$, $\mu_{n,L}$ restricted on $C_{n,L}$ is a map into $B_{n,L}$, and they form a pair of mutually inverse maps. Furthermore, we have*

$$B_{n,L} = \{x_{1:L} \in A_n^L : \textit{for any } 1 \le i \le n-1 \textit{ there exist } 1 \le j < k \le L \\ \textit{such that } x_j = i+1, x_k = i\}, \quad (19)$$

and

$$C_{n,L} = \{\pi \in \mathcal{S}_L : \mathrm{Asc}(\pi) = L - n\}.$$

When $n = 2$ and $L = 3$, the duality

$$B_{2,3} \underset{\mu_{2,3}}{\overset{\phi_{2,3}}{\rightleftarrows}} C_{2,3}$$

is given by

$$\begin{array}{lcl} 121 & \leftrightsquigarrow & 132 \\ 211 & \leftrightsquigarrow & 231 \\ 212 & \leftrightsquigarrow & 213 \\ 221 & \leftrightsquigarrow & 312. \end{array}$$

Let $\mathbf{X}$ be an SSP over the alphabet A_n and p the corresponding probability of the occurrence of the words. The probability of the occurrence of a permutation $\pi \in \mathcal{S}_L$ is given by

$$p_*(\pi) = \sum_{x_{1:L} \in \phi_{n,L}^{-1}(\pi)} p(x_{1:L}).$$

Let $\alpha_{\mathbf{X},L}$ be the probability that we observe in $\mathbf{X}$ a permutation of length L which is not a member of the set $C_{n,L}$, namely,

$$\alpha_{\mathbf{X},L} = \sum_{\pi \notin C_{n,L}} p_*(\pi),$$

and $\beta_{x,\mathbf{X},L}$ be the probability that symbol x does not appear in a word of length $\lfloor L/2 \rfloor$ realized in $\mathbf{X}$, namely,

$$\beta_{x,\mathbf{X},L} = \sum_{\substack{x_j \neq x, \\ 1 \leq j \leq \lfloor L/2 \rfloor}} p(x_{1:\lfloor L/2 \rfloor}),$$

where $L \geq 1$, $x \in A_n$ and $\lfloor a \rfloor$ is the largest integer not greater than a.

We have the following lemma on the relationship between $\alpha_{\mathbf{X},L}$ and $\beta_{x,\mathbf{X},L}$ [9] (Lemma 12):

Lemma 1 *Let $\boldsymbol{X}$ be an SSP over A_n and ϵ be a positive real number. If $\beta_{x,\boldsymbol{X},L} < \epsilon$ for any $x \in A_n$, then we have $\alpha_{\boldsymbol{X},L} < 2n\epsilon$.*

3.3 A Lemma for HMM

If $\mathbf{X}$ is the output process of an HMM, then we have the following expression for $\beta_{x,\mathbf{X},L}$:

$$\beta_{x,\mathbf{X},L} = \langle \mu \left(T - T^{(x)} \right)^N, \mathbf{1} \rangle,$$

where $\mathbf{1} = (1, 1, \cdots, 1)$ and $\langle \cdots, \cdots \rangle$ is the usual inner product in the Euclidean space. If the internal process of the HMM is ergodic, then it can be shown that the largest eigenvalue of the matrix $T_{(x)} := T - T^{(x)}$ is less than 1 [15]. Thus, we have the following lemma [15].

Lemma 2 *Let $\boldsymbol{X}$ be the output process of an HMM over the alphabet A_n. Then, if the internal process $\boldsymbol{S}$ of the HMM is ergodic, then for any $x \in A_n$ there exists $0 < \gamma_x < 1$ and $C_x > 0$ such that $\beta_{x,\boldsymbol{X},L} < C_x \gamma_x^L$ for any $L \geq 1$.*

4 Permutation Entropies

In this section, we discuss the relationship between information-theoretic measures and their permutation analogues for SSPs. In particular, we focus on three measures, entropy rate, excess entropy and transfer entropy rate.

4.1 Fundamental Inequalities

We can give a bound for the difference between the block Shannon entropy with respect to words of length L and that with respect to permutations of length L for an SSP $\mathbf{X}$ over A_n. The crucial point is that the size of inverse image of each permutation is given by the binomial coefficient (17). This fact leads to the following bound [9]:

Lemma 3 *Let $\mathbf{X}$ be an SSP over A_n and p the corresponding probability of the occurrence of the words. Then, we have*

$$0 \le H(X_{1:L}) - H^*(X_{1:L}) \le \alpha_{\boldsymbol{X},L} n \log_2(L+n), \tag{20}$$

where

$$H(X_{1:L}) = -\sum_{x_{1:L} \in A_n^L} p(x_{1:L}) \log_2 p(x_{1:L})$$

and

$$H^*(X_{1:L}) = -\sum_{\pi \in \mathcal{S}_L} p_*(\pi) \log_2 p_*(\pi).$$

The bivariate version is as follows:

Lemma 4 *Let $(\mathbf{X}, \mathbf{Y})$ be a bivariate SSP over $A_{n_1} \times A_{n_2}$ and p the corresponding joint probability of the occurrence of the words. Then, we have*

$$\begin{aligned} &0 \le H(X_{a:b}, Y_{c:d}) - H^*(X_{a:b}, Y_{c:d}) \\ &\le (\alpha_{\boldsymbol{X},b-a+1} + \alpha_{\boldsymbol{Y},d-c+1}) \left(n_1 \log_2(b-a+1+n_1) + n_2 \log_2(d-c+1+n_2)\right), \end{aligned} \tag{21}$$

where

$$H(X_{a:b}, Y_{c:d}) = -\sum_{(x_{a:b}, y_{c:d}) \in A_{n_1}^{b-a+1} \times A_{n_2}^{d-c+1}} p(x_{a:b}, y_{c:d}) \log_2 p(x_{a:b}, y_{c:d}),$$

$$H^*(X_{a:b}, Y_{c:d}) = -\sum_{(\pi_1, \pi_2) \in \mathcal{S}_{b-a+1} \times \mathcal{S}_{d-c+1}} p_*(\pi_1, \pi_2) \log_2 p_*(\pi_1, \pi_2).$$

and

$$p_*(\pi_1, \pi_2) = \sum_{\substack{x_{a:b} \in \phi^{-1}_{n_1, b-a+1}(\pi_1), \\ y_{c:d} \in \phi^{-1}_{n_2, d-c+1}(\pi_2)}} p(x_{a:b}, y_{c:d}).$$

A proof for the general multivariate version of the inequality is given in [15].

4.2 Permutation Entropy Rate

For an SSP $\mathbf{X}$ over A_n, the *permutation entropy rate* is defined by

$$h^*(\mathbf{X}) = \lim_{L \to \infty} H^*(X_{1:L})/L. \tag{22}$$

$h^*(\mathbf{X})$ is a measure of average uncertainty of orderings per unit time. Amigó et al. [3, 5] proved the equality

$$h(\mathbf{X}) = h^*(\mathbf{X}) \tag{23}$$

for any SSP $\mathbf{X}$ over A_n by applying the ergodic decomposition of the entropy rate. As an alternative proof, we pointed out that the equality follows immediately from the inequality (20) in our previous work [9].

4.3 Permutation Excess Entropy

The *permutation excess entropy* of an SSP $\mathbf{X}$ over A_n is defined by [1]

$$\mathbf{E}^*(\mathbf{X}) = \limsup_{L\to\infty} \left(H^*(X_{1:L}) - h^*(\mathbf{X})L\right) \\ = \limsup_{K\to\infty} \left(\sum_{L=1}^{K} \left(H^*(X_L|X_{1:L-1}) - h^*(\mathbf{X})\right) \right), \tag{24}$$

where $H^*(X_L|X_{1:L-1}) = H^*(X_{1:L}) - H^*(X_{1:L-1})$.

We have

$$\mathbf{E}^*(\mathbf{X}) \le \mathbf{E}(\mathbf{X}) \tag{25}$$

for any SSP $\mathbf{X}$ over A_n. However, the equality cannot hold in general if the process is not ergodic [9]. At present, we do not know whether the equality holds for all ergodic SSPs or not. However, if $\mathbf{X}$ is the output process of an HMM over A_n with an ergodic internal process, then we have the following equalities by Lemma 1, Lemma 2 and Lemma 3 [15]:

$$\mathbf{E}(\mathbf{X}) = \mathbf{E}^*(\mathbf{X}) = \lim_{L\to\infty} I^*(X_{1:L}; X_{L+1:2L}), \tag{26}$$

where $I^*(X_{1:L}; X_{L+1:2L}) = H^*(X_{1:L}) + H^*(X_{L+1:2L}) - H^*(X_{1:L}, X_{L+1:2L})$.

Now, let $\mathbf{X}$ be an ergodic SSP over A_n and $\mathbf{X}^{(N)}$ its N-th order approximation. Since $\mathbf{X}^{(N)}$ is the output process of a HMM over A_n with an ergodic internal process, we have

$$\mathbf{E}(\mathbf{X}^{(N)}) = \mathbf{E}^*(\mathbf{X}^{(N)}) \tag{27}$$

by (26). By (12) and (27), we obtain

$$\mathbf{E}(\mathbf{X}) = \lim_{N\to\infty} \mathbf{E}(\mathbf{X}^{(N)}) = \lim_{N\to\infty} \mathbf{E}^*(\mathbf{X}^{(N)}). \tag{28}$$

Thus, $\mathbf{E}(\mathbf{X})$ can be calculated as the limit of $\mathbf{E}^*(\mathbf{X}^{(N)})$ when $\mathbf{X}$ is an ergodic SSP over A_n.

4.4 Symbolic Transfer Entropy Rate

Let $(\mathbf{X}, \mathbf{Y})$ be a bivariate SSP over the alphabet $A_{n_1} \times A_{n_2}$. The *symbolic transfer entropy rate* [14,15] from $\mathbf{Y}$ to $\mathbf{X}$ is defined by

$$t^*(\mathbf{X}|\mathbf{Y}) = h^*(\mathbf{X}) - h^*(\mathbf{X}|\mathbf{Y}), \tag{29}$$

where

$$h^*(\mathbf{X}|\mathbf{Y}) = \limsup_{L\to\infty} H^*(X_{L+1}|X_{1:L}, Y_{1:L})$$

and

$$H^*(X_{L+1}|X_{1:L}, Y_{1:L}) = H^*(X_{1:L+1}, Y_{1:L}) - H^*(X_{1:L}, Y_{1:L}).$$

[1] In [9] we defined $\mathbf{E}^*(\mathbf{X}) = \lim_{L\to\infty} \left(H^*(X_{1:L}) - h^*(\mathbf{X})L\right)$ when it exists. Here, we define it as the limsup which always exists including infinity. The same remark holds for the definition of the symbolic transfer entropy rate.

Note that the symbolic transfer entropy rate is not the rate of the original symbolic transfer entropy [34] but that of the transfer entropy on rank vectors introduced by [35] to improve its performance.

If $(\mathbf{X}, \mathbf{Y})$ is a bivariate ergodic SSP over $A_{n_1} \times A_{n_2}$, then we have [14]

$$h(\mathbf{X}|\mathbf{Y}) \leq h^*(\mathbf{X}|\mathbf{Y}), \tag{30}$$

which implies that

$$t^*(\mathbf{X}|\mathbf{Y}) \leq t(\mathbf{X}|\mathbf{Y}). \tag{31}$$

If $(\mathbf{X}, \mathbf{Y})$ is the output process of an HMM over $A_{n_1} \times A_{n_2}$ with an ergodic internal process, we have the following equality by Lemma 1 and Lemma 2 and Lemma 4 [15]

$$t(\mathbf{X}|\mathbf{Y}) = t^*(\mathbf{X}|\mathbf{Y}). \tag{32}$$

Combining (16) and (32), if $(\mathbf{X}, \mathbf{Y})$ is a bivariate ergodic SSP over $A_{n_1} \times A_{n_2}$ and $(\mathbf{X}^{(N)}, \mathbf{Y}^{(N)})$ is its N-th order approximation, then we have

$$t(\mathbf{X}|\mathbf{Y}) = \lim_{N\to\infty} t(\mathbf{X}^{(N)}|\mathbf{Y}^{(N)}) = \lim_{N\to\infty} t^*(\mathbf{X}^{(N)}|\mathbf{Y}^{(N)}). \tag{33}$$

Namely, as in the case of permutation excess entropy, $t(\mathbf{X}|\mathbf{Y})$ can be calculated as the limit of $t^*(\mathbf{X}^{(N)}|\mathbf{Y}^{(N)})$ when $(\mathbf{X}, \mathbf{Y})$ is a bivariate ergodic SSP over $A_{n_1} \times A_{n_2}$.

5 Modified Permutation Entropies

In this section, we discuss modified permutation entropies. In recent years, several types of modified permutation entropies have been proposed for suitable entropy estimation for specific real-world time series data [16–18]. Here, we focus on that introduced by [16] and study its theoretical aspect.

5.1 Partition Induced by Permutations and Equalities

Let us consider the partition of the set of all words of length L over the alphabet A_n induced by permutations and equalities. We define the map

$$\eta_{n,L} : A_n^L \to \mathcal{T}_L := \mathcal{S}_L \times \{0,1\}^{L-1} \tag{34}$$

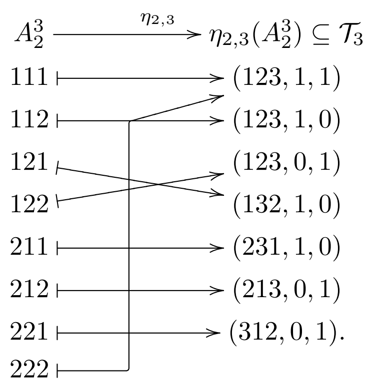
by $\eta_{n,L}(x_1^L) = (\pi, e_1, \cdots, e_{L-1})$ where π is the permutation type of $x_{1:L}$ and

$$e_i = \begin{cases} 1 & \text{if } x_{\pi(i)} = x_{\pi(i+1)}, \\ 0 & \text{otherwise} \end{cases}$$

for $i = 1, 2, \cdots, L-1$. If $\mathrm{Proj} : \mathcal{T}_L \to \mathcal{S}_L$ is the projection onto $\mathcal{S}_L$, we have

$$\phi_{n,L} = \mathrm{Proj} \circ \eta_{n,L}. \tag{35}$$

Thus, $\eta_{n,L}$ defines a finer partition of the set A_n^L than that defined by $\phi_{n,L}$. For example, when $n = 2$ and $L = 3$, the map $\eta_{2,3} : A_2^3 \to \mathcal{T}_3$ is given by



For any $\sigma = (\pi, e_1, \cdots, e_{L-1}) \in \eta_{n,L}(A_n^L)$, if $\sum_{i=1}^{L-1} e_i = L - k$, then we have

$$|\eta_{n,L}^{-1}(\sigma)| = \binom{n}{k}. \tag{36}$$

Indeed, if $\sum_{i=1}^{L-1} e_i = L - k$, then there are k blocks for which $x_{\pi(i)}$ have the same values in each block and have different values between blocks. Since the set of possible values to be assigned to each block is A_n, the number of words of length L over the alphabet A_n that are mapped to the same σ with $\sum_{i=1}^{L-1} e_i = L - k$ is the number of combinations choosing k values from A_n.

The number of ways to divide the set $\{1, 2, \cdots, L\}$ into k nonempty blocks is given by the Stirling number of the second kind $\left\{ {L \atop k} \right\}$ and the number of permutations of k blocks is $k!$. Consequently, we obtain the following equality with respect to the partition of A_n^L by the fibers of $\eta_{n,L}$ [32]:

$$n^L = \sum_{k=1}^{n} \left\{ {L \atop k} \right\} k! \binom{n}{k}, \tag{37}$$

where $\left\{ {L \atop k} \right\} k!$ is the number of blocks with size $\binom{n}{k}$ in the partition.

5.2 Modified Permutation Entropy Rate

Let $\mathbf{X}$ be an SSP over A_n and p the corresponding probability of the occurrence of the words. The *modified permutation entropy rate* is defined by

$$h^m(\mathbf{X}) = \lim_{L \to \infty} H^m(X_{1:L})/L, \tag{38}$$

where

$$H^m(X_{1:L}) = -\sum_{\sigma \in \mathcal{T}_L} p_m(\sigma) \log_2 p_m(\sigma)$$

and

$$p_m(\sigma) = \sum_{x_{1:L} \in \eta_{n,L}^{-1}(\sigma)} p(x_{1:L})$$

for $\sigma \in \mathcal{T}_L$. By definition, $H^*(X_{1:L}) = H(\phi_{n,L}(X_{1:L}))$ and $H^m(X_{1:L}) = H(\eta_{n,L}(X_{1:L}))$. Hence, by (35), we have

$$H^*(X_{1:L}) \le H^m(X_{1:L}) \le H(X_{1:L}).$$

Consequently, by (23), we obtain

$$h^*(\mathbf{X}) = h^m(\mathbf{X}) = h(\mathbf{X}). \tag{39}$$

5.3 Fundamental Inequalities

Lemma 5 *Let $\mathbf{X}$ be an SSP over A_n and p the corresponding probability of the occurrence of the words. Then, we have*

$$0 \le H(X_{1:L}) - H^m(X_{1:L}) \le \alpha_{\mathbf{X},L} n \log_2 n. \tag{40}$$

Proof. We have

$$\begin{aligned} 0 &\le H(X_{1:L}) - H^m(X_{1:L}) \\ &= \sum_{\substack{\sigma \in \mathcal{T}_L, p_m(\sigma)>0, \\ |\eta_{n,L}^{-1}(\sigma)|>1}} p_m(\sigma) \left(- \sum_{x_{1:L} \in \eta_{n,L}^{-1}(\sigma)} \frac{p(x_{1:L})}{p_m(\sigma)} \log_2 \frac{p(x_{1:L})}{p_m(\sigma)} \right). \end{aligned}$$

By (36) and $\binom{n}{k} \le n^n$, we have

$$- \sum_{x_{1:L} \in \eta_{n,L}^{-1}(\sigma)} \frac{p(x_{1:L})}{p_m(\sigma)} \log_2 \frac{p(x_{1:L})}{p_m(\sigma)} \le \log_2 \binom{n}{k} \le n \log_2 n$$

for each $\sigma \in \mathcal{T}_L$ such that $\sum_{i=1}^{L-1} e_i = L - k$. Since

$$\sum_{\substack{\sigma \in \mathcal{T}_L, \\ |\eta_{n,L}^{-1}(\sigma)|>1}} p_m(\sigma) \le \sum_{\substack{\pi \in \mathcal{S}_L, \\ |\phi_{n,L}^{-1}(\pi)|>1}} p_*(\pi) = \alpha_{\mathbf{X},L},$$

the claim follows.

□

Similarly, we have the following bivariate version.

Lemma 6 *Let $(\mathbf{X}, \mathbf{Y})$ be a bivariate SSP over $A_{n_1} \times A_{n_2}$ and p the corresponding joint probability of the occurrence of the words. Then, we have*

$$\begin{aligned} 0 &\le H(X_{a:b}, Y_{c,d}) - H^m(X_{a:b}, Y_{c:d}) \\ &\le (\alpha_{\boldsymbol{X},b-a+1} + \alpha_{\boldsymbol{Y},d-c+1}) (n_1 \log_2 n_1 + n_2 \log_2 n_2), \end{aligned} \tag{41}$$

where

$$H^m(X_{a:b}, Y_{c:d}) = - \sum_{(\sigma_1,\sigma_2) \in \mathcal{T}_{b-a+1} \times \mathcal{T}_{d-c+1}} p_m(\sigma_1, \sigma_2) \log_2 p_m(\sigma_1, \sigma_2)$$

and

$$p_m(\sigma_1, \sigma_2) = \sum_{\substack{x_{a:b} \in \eta^{-1}_{n_1, b-a+1}(\sigma_1), \\ y_{c:d} \in \eta^{-1}_{n_2, d-c+1}(\sigma_2)}} p(x_{a:b}, y_{c:d}).$$

Lemma 5 and Lemma 6 seem to be similar to Lemma 3 and Lemma 4 in their apparent forms, respectively, however, there is a significant difference with respect to the asymptotic behavior. Namely, say, the bound given in Lemma 3 has $O(\log L)$ diverging term, while that in Lemma 5 is at most a constant. This distinct property of the modified permutation entropy allows to establish equalities between the usual information-theoretic measures and their modified permutation analogues for ergodic SSPs as we illustrate in the next subsection.

5.4 Modified Permutation Excess Entropy and Modified Symbolic Transfer Entropy Rate

The modified permutation excess entropy and the modified symbolic transfer entropy rate are defined similarly with the permutation excess entropy and the symbolic transfer entropy rate, respectively: let $\mathbf{X}$ be an SSP over A_n. The *modified permutation excess entropy* is defined by

$$\begin{aligned} \mathbf{E}^m(\mathbf{X}) &= \limsup_{L\to\infty} \left(H^m(X_{1:L}) - h^m(\mathbf{X})L\right) \\ &= \limsup_{K\to\infty} \left(\sum_{L=1}^{K} \left(H^m(X_L|X_{1:L-1}) - h^m(\mathbf{X})\right) \right), \end{aligned} \tag{42}$$

where $H^m(X_L|X_{1:L-1}) = H^m(X_{1:L}) - H^m(X_{1:L-1})$.

Let $(\mathbf{X}, \mathbf{Y})$ be a bivariate SSP over $A_{n_1} \times A_{n_2}$. The *modified symbolic transfer entropy rate* from $\mathbf{Y}$ to $\mathbf{X}$ is defined by

$$t^m(\mathbf{X}|\mathbf{Y}) = h^m(\mathbf{X}) - h^m(\mathbf{X}|\mathbf{Y}), \tag{43}$$

where

$$h^m(\mathbf{X}|\mathbf{Y}) = \limsup_{L\to\infty} H^m(X_{L+1}|X_{1:L}, Y_{1:L})$$

and

$$H^m(X_{L+1}|X_{1:L}, Y_{1:L}) = H^m(X_{1:L+1}, Y_{1:L}) - H^m(X_{1:L}, Y_{1:L}).$$

Lemma 7 *Let $\mathbf{X}$ be an ergodic SSP over A_n and p the corresponding probability of the occurrence of the words. Then, for any $x \in A_n$ such that $p(x) > 0$, we have $\beta_{x,\mathbf{X},L} \to 0$ as $L \to \infty$.*

Proof. Let $x \in A_n$ satisfy $p(x) > 0$. By the ergodicity of $\mathbf{X}$, we have for any $\epsilon > 0$

$$\Pr\{|F_{x,N} - p(x)| < \epsilon\} \to 1$$

as $L \to \infty$, where $N = \lfloor L/2 \rfloor$ and $F_{x,N}$ is the stochastic variable defined by the number of the occurrence of the symbol x in the sequence $X_1, X_2, \cdots, X_N$ divided by N. Recalling

$$\beta_{x,\mathbf{X},L} = \sum_{\substack{x_j \neq x, \\ 1 \leq j \leq N}} p(x_{1:N}),$$

we can write

$$\beta_{x,\mathbf{X},L} = \Pr\left\{\frac{F_{x,N}}{N} = 0\right\}.$$

Hence, by taking $\epsilon = p(x)/2$, we obtain

$$\beta_{x,\mathbf{X},L} \leq 1 - \Pr\left\{\frac{p(x)}{2} < \frac{F_{x,N}}{N} < \frac{3p(x)}{2}\right\} \to 0$$

as $L \to \infty$.

□

By combining Lemma 1, Lemma 5 and Lemma 7, we obtain the equality between the excess entropy and the modified permutation excess entropy for ergodic SSPs.

Theorem 2 *Let* $\mathbf{X}$ *be an ergodic SSP over* A_n*. Then, we have*

$$\mathbf{E}(\mathbf{X}) = \mathbf{E}^m(\mathbf{X}) = I^m(X_{1:L}; X_{L+1:2L}), \tag{44}$$

where $I^m(X_{1:L}; X_{L+1:2L}) = H^m(X_{1:L}) + H^m(X_{L+1:2L}) - H^m(X_{1:L}, X_{L+1:2L})$.

Similarly, the combination of Lemma 1, Lemma 6 and Lemma 7 establishes the equality between the transfer entropy rate and the modified symbolic transfer entropy rate for bivariate ergodic SSPs.

Theorem 3 *Let* $(\mathbf{X}, \mathbf{Y})$ *be a bivariate ergodic SSP over* $A_{n_1} \times A_{n_2}$*. Then, we have*

$$t(\mathbf{X}|\mathbf{Y}) = t^m(\mathbf{X}|\mathbf{Y}). \tag{45}$$

6 Discussions

In this final section, we shall indicate a future direction of the study along the present work.

As we have shown in Section 5.1, the equality (37) holds for the partition of the set of all words of length L over a finite alphabet induced by permutations and equalities. This can be seen as an instance of the umbral composition in the theory of the sequences of binomial type [36,37]. The partition itself is a set-theoretic interpretation of (37). It is known that the set-theoretic interpretation of the equality like (37) can be systematically given [38] based on the theory of combinatorial species [39]. Thus, the result obtained here on the modified permutation entropies suggests a possibility of *general pattern analysis* extending the ordinal pattern analysis, namely, there might be a modified permutation entropy for a certain class of sequences of binomial type.

We would like to thank the anonymous referee for giving invaluable comments which significantly improve the manuscript.

References

1. C. Bandt, B. Pompe, Phys. Rev. Lett. **88**, 174102 (2002)

2. J. M. Amigó, *Permutation Complexity in Dynamical Systems.* (Springer-Verlag Berlin Heidelberg, 2010)
3. J. M. Amigó, M. B. Kennel, L. Kocarev, Physica D **210**, 77 (2005)
4. J. M. Amigó, M. B. Kennel, Physica D **231**, 137 (2007)
5. J. M. Amigó, Physica D **241**, 789 (2012)
6. C. Bandt, G. Keller, B. Pompe, Nonlinearity **15**, 1595 (2002)
7. K. Keller, M. Sinn, Physica D **239**, 997 (2010)
8. K. Keller, A. M. Unakafov, V. A. Unakafova, Physica D **241**, 1477 (2012)
9. T. Haruna, K. Nakajima, Physica D **240**, 1370 (2011)
10. J. P. Crutchfield, D. P. Feldman, Chaos **15**, 25 (2003)
11. T. Haruna, K. Nakajima, Int. J. Comput. Ant. Sys. (in press)
12. T. Schreiber, Phys. Rev. Lett. **85**, 461 (2000)
13. P. O. Amblard, O. J. J. Michel, J. Comput. Neurosci. **30**, 7 (2011)
14. T. Haruna, K. Nakajima, arXiv:1112.2493
15. T. Haruna, K. Nakajima, arXiv:1204.1821
16. C. Bian, C. Qin, Q. D. Y. Ma, Q. Shen, Phys. Rev. E **85**, 021906 (2012)
17. M. C. Ruiz, A. Guillamón, A. Gabaldón, Entropy **14**, 74 (2012)
18. X. Sun, Y. Zou, V. Nikiforova, V. Kurths, D. Walther, BMC Biofinformatics **11**, 607 (2010)
19. R. B. Ash, *Information Theory.* (Interscience Publishers, 1965)
20. P. Billingsley, *Convergence of Probability Measures, Second Edition.* (John Wiley & Sons, Inc., 1999)
21. A. Kehagias, PhD thesis, Brown University, 1992
22. D. R. Upper, PhD thesis, University of California, Berkeley, 1997
23. P. Walters, *An Introduction to Ergodic Theory.* (Springer-Verlag, New York, 1982)
24. W. Löhr, PhD thesis, Max Planck Institute for Mathematics in the Sciences, Leipzig, 2010
25. T. M. Cover, J. A. Thomas, *Elements of Information Theory.* (John Wiley & Sons, Inc., 1991)
26. D. V. Arnold, Complex Systems **10**, 143 (1996)
27. W. Bialek, I. Nemenman, N. Tishby, Neural Computation **13**, 2409 (2001)
28. D. P. Feldman, C. S. McTague, J. P. Crutchfield, Chaos **18**, 043106 (2008)
29. P. Grassberger, Int. J. Theor. Phys. **25**, 907 (1986)
30. W. Li, Complex Systems **5**, 381 (1991)
31. R. Shaw, *The Dripping Faucet as a Model Chaotic System.* (Aerial Press, Santa Cruz, California, 1984)
32. R. E. Knop, J. Combi. Theor. A **15**, 338 (1973)
33. R. L. Graham, D. E. Knuth, O. Patashnik, *Concrete Mathematics, Second Edition.* (Addison-Wesley Publishing Company, Inc., 1994)
34. M. Staniek, K. Lehnertz, Phys. Rev. Lett. **100**, 158101 (2008)
35. D. Kugiumtzis, J. Nonlin. Sys. Appl. **3**, 73 (2012)
36. R. Mullin, G. C. Rota, in *Graph Theory and its Applications*, (Academic Press, 1970), p. 167
37. S. Roman, G. C. Rota, Adv. in Math. **27**, 95 (1978)
38. D. Senato, A. Venezia, Comput. Math. Appl. **41**, 1109 (2001)
39. A. Joyal, Adv. in Math. **42**, 1 (1981)