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

Partially Ordered Permutation Entropies

Taichi Haruna

Department of Information and Sciences, Tokyo Woman's Christian University, 2-6-1 Zempukuji, Suginami-ku, Tokyo 167-8585, Japan, e-mail: tharuna@lab.twcu.ac.jp

Received: date / Revised version: date

Abstract. In the past decade, it has been shown through both theoretical and practical studies of permutation entropies that complexity of time series can be captured by order relations between numerical values. In this paper, we investigate a generalisation of permutation entropies in terms of the order structure for further understanding of their nature. To calculate conventional permutation entropies of time series, one needs to assume a total order on the alphabet. We generalise this to an arbitrary partial order; that is, the alphabet is assumed to be a partially ordered set, and we introduce partially ordered permutation entropies. The relationship between entropies and their partial-order analogues for discrete-time finite-alphabet stationary stochastic processes is theoretically studied. We will show that the entropy rate and its partial-order analogues are equal without restriction, whereas equalities between excess entropy and partial-order analogues depend on asymmetry of the order structure of the alphabet. As all finite totally ordered sets are asymmetric, our results explain one reason why conventional permutation entropies are so effective.

PACS. XX.XX.XX No PACS code given

1 Introduction

Permutation entropies quantify complexity of time series by using order relations between numerical values [1]. They have been shown to be easy-to-implement, robust measures of complexity of time series [2,3] and have been harnessed in various scientific fields [4–6].

One way to deepen our understanding of a mathematical concept is to study it in an ideal situation. Discrete-time finite-alphabet stationary stochastic processes (SSPs) are a simple mathematical model of stationary time series and are an appropriate starting point for theoretical investigation of the properties of permutation entropies. The first result on permutation entropies for SSP was given in Ref. 7 which showed that the entropy rate of any SSP is equal to its permutation analogue (but see also Ref. 8). This result was extended to other entropies such as the excess entropy [9] and the transfer entropy rate [10,11] under appropriate conditions in our previous work [12–16].

This paper is devoted to a theoretical investigation to generalise permutation entropies for further understanding of their nature. Permutations are related to two different mathematical structures. One is groups and the other is ordered sets. The group-theoretic aspect of permutations has been utilised to analyse coupled time series [17–19] under the name *transcript*. In Ref. 20, transcripts are applied to dimensional reduction of conditional multi-information. To complement the existing approach, which extends permutation entropies by focusing on the group structure, here we attempt to shed new light on them in terms of partially ordered sets. When calculating conventional permutation entropies, a total order is assumed on the alphabet. We generalise this to partial orders and study the relationship between entropies and their partial-order analogues called *partially ordered permutation entropies*. We consider two kinds of partially ordered permutation entropies. One is called *square partially ordered permutation entropies*, which distinguish ties (equality between occurrences of symbols) between numerical values as in the modified permutation entropies [21]. The other is called *triangular partially ordered permutation entropies*, which do not concern ties as in the original permutation entropies.

Partial orders naturally arise when one tries to extend the idea of permutation entropy to multivariate time series. Let us consider a multivariate time series consisting of $N \geq 2$ time series and suppose that each time series takes its numerical values in a totally ordered set A, for example, the set of real numbers as one typically encounters in real-world data. Then, we can introduce the pointwise order on the product set A^N, which is in general a partial order. The multivariate time series can be treated as if it is a univariate time series by considering that it takes values in the partially ordered set A^N. Partially ordered permutation entropies provide ways to calculate information-theoretic quantities of a given multivariate time series from this viewpoint. This idea can be utilised to define a complexity measure for coupling among multiple time series [22]. In this paper, we reveal a limitation of this approach under certain assumptions.

This paper is organised as follows. In Sect. 2, basic notions of partially ordered sets are reviewed and partially ordered permutation entropies are defined. The main results are also presented. In Sect. 3 and Sect. 4, results on square partially ordered permutation entropies and triangular partially ordered permutation entropies are proved, respectively. In Sect. 4, concluding remarks are given.

2 Definitions and Main Results

2.1 Stationary Stochastic Processes, Entropy Rate and Excess Entropy

We consider discrete-time finite-alphabet stationary stochastic processes with a partial order on the alphabet. Let $\mathbf{X} = \{X_1, X_2, \dots\}$ be a discrete-time stationary stochastic process over a finite alphabet A (in short, SSP $\mathbf{X}$ over A), where X_i is an A-valued stochastic variable $X_i : \Omega \to A$ on a common probability space Ω for all $i = 1, 2, \dots$. For any word $x_{1:L} := x_1 x_2 \dots x_L := (x_1, x_2, \dots, x_L) \in A^L$ of length L, the following equality holds for any $k \geq 1$, because of the assumed stationarity:

$$\begin{aligned} &\mathrm{Prob}\{X_1 = x_1, X_2 = x_2, \dots, X_L = x_L\} \\ &= \mathrm{Prob}\{X_k = x_1, X_{k+1} = x_2, \dots, X_{L+k-1} = x_L\}. \end{aligned} \quad (1)$$

Consequently, we can write

$$p(x_{1:L}) = \Pr\{X_1 = x_1, X_2 = x_2, \dots, X_L = x_L\} \quad (2)$$

for the probability of occurrence of a word $x_{1:L}$ of length L.

In this paper, we consider two information-theoretic quantities for SSPs. One is entropy rate and the other is excess entropy. Recall that the *entropy rate* of an SSP $\mathbf{X}$ over A is defined as the average uncertainty of $\mathbf{X}$ per unit time:

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

where

$$H(X_{1:L}) = -\sum_{x_{1:L}\in A^L} p(x_{1:L}) \log_2 p(x_{1:L}) \quad (4)$$

is the joint Shannon entropy of $(X_1, \dots, X_L)$. It is well-known that the limit on the right-hand-side of Eq. (3) always exists [23].

The *excess entropy* $\mathbf{E}(\mathbf{X})$ of $\mathbf{X}$ quantifies the subextensive part of its entropy [9]:

$$\begin{aligned} \mathbf{E}(\mathbf{X}) &= \lim_{L\to\infty} \left(H(X_{1:L}) - h(\mathbf{X})L\right) \\ &= \sum_{L=1}^{\infty} \left(H(X_L|X_{1:L-1}) - h(\mathbf{X})\right). \end{aligned} \quad (5)$$

Since the conditional entropy $H(X_L|X_{1:L-1})$ monotonically approaches $h(\mathbf{X})$ from above, $\mathbf{E}(\mathbf{X})$ exists or otherwise diverges. It is known that $\mathbf{E}(\mathbf{X})$ can be written as mutual information between the left and right semi-infinite sequences of the stochastic variables [9]. Thus, $\mathbf{E}(\mathbf{X})$ measures the degree of global correlation in $\mathbf{X}$. In the literature, excess entropy is also called *effective measure of complexity* [24], *stored information* [25], *predictive information* [26], or simply *complexity* [27,28].

2.2 Partially Ordered Sets

In this paper, we assume that the finite alphabet A is equipped with a partial order $\leq$ and call it the *partially ordered finite alphabet*. Here, we recall basic definitions and terminologies of partially ordered sets that are used in this paper. A *partial order* on a set A is a binary relation $\leq$ on A satisfying the following three conditions [29]: (i) $a \leq a$ (reflexivity); (ii) if $a \leq b$ and $b \leq a$ then $a = b$ (antisymmetry); (iii) $a \leq b$ and $b \leq c$ imply $a \leq c$ (transitivity) for all $a, b, c \in A$. A set A equipped with a partial order $\leq$ is called a *partially ordered set* (*poset*) and is denoted by $(A, \leq)$. In the following, a poset is denoted by its underlying set unless otherwise required. For every set, the equality relation $=$ is a partial order on it called the *discrete order*. If a partial order $\leq$ on A satisfies either $a \leq b$ or $b \leq a$ for all $a, b \in A$, then it is called a *total order*. Clearly, the number of all total orders on a given set A with n members is $n!$.

Let $(A, \leq)$ be a poset. A self-map $f : A \to A$ is called an *automorphism* of A if it is order-preserving ($a \leq b$ implies $f(a) \leq f(b)$ for all $a, b \in A$) and has an order-preserving inverse; that is, there exists an order-preserving map $g : A \to A$ such that $f \circ g = g \circ f = \mathrm{id}_A$ holds, where id_A is the identity map of A ($\mathrm{id}_A(a) = a$ for all $a \in A$). When A is a finite poset, an order-preserving self-map $f : A \to A$ is an automorphism if and only if it is injective ($f(a) = f(b)$ implies $a = b$ for all $a, b \in A$) [29]. Obviously, id_A is an automorphism of A. We denote the set of all automorphisms of A by $\mathrm{Aut}(A)$. The size of $\mathrm{Aut}(A)$ measures the degree of symmetry of A, which is said to be *asymmetric* when $\mathrm{Aut}(A) = \{\mathrm{id}_A\}$. For example, every finite totally ordered set is asymmetric. A direct union of a finite number of finite totally ordered sets with different lengths is also asymmetric.

2.3 Permutation Entropies

In this subsection, we assume that $(A, \leq)$ is a finite totally ordered set with n elements. For example, $A = \{1, 2, \ldots, n\}$, with $\leq$ indicating the usual 'less-than-or-equal-to' relation between natural numbers. Let $\mathcal{S}_L$ be the set of permutations of length L. Each $\pi \in \mathcal{S}_L$ is a bijective map from the set $\{1, 2, \ldots, L\}$ to itself. For each $x_{1:L} \in A^L$, its *permutation type* is defined as a permutation $\pi \in \mathcal{S}_L$ satisfying $x_{\pi(i)} \leq x_{\pi(i+1)}$ for $i = 1, 2, \ldots, L-1$. When $x_{\pi(i)} = x_{\pi(i+1)}$, we require that π satisfies $\pi(i) < \pi(i+1)$. By this condition, the permutation type of $x_{1:L}$ is uniquely determined.

We define a map $\phi_{n,L} : A^L \to \mathcal{S}_L$ by sending each $x_{1:L} \in A^L$ to its permutation type. $\phi_{n,L}$ induces a partition of A^L in the following way: Two words $x_{1:L}, y_{1:L} \in A^L$ are contained in the same block of the partition if $\phi_{n,L}(x_{1:L}) = \phi_{n,L}(y_{1:L})$. Let $\mathbf{X}$ be an SSP over a totally ordered finite alphabet A. The probability of occurrence of $\pi \in \mathcal{S}_L$ is given by

$$p_*(\pi) = \sum_{x_{1:L} \in \phi_{n,L}^{-1}(\pi)} p(x_{1:L}). \tag{6}$$

The *permutation entropy rate* of $\mathbf{X}$ [1,2] is defined by

$$h_*(\mathbf{X}) = \lim_{L \to \infty} H_*(X_{1:L})/L, \tag{7}$$

where

$$H_*(X_{1:L}) = -\sum_{\pi \in \mathcal{S}_L} p_*(\pi) \log_2 p_*(\pi). \tag{8}$$

It is known that $h(\mathbf{X}) = h_*(\mathbf{X})$ for any SSP $\mathbf{X}$ over a totally ordered finite alphabet A [7,8,12].

The *permutation excess entropy* of $\mathbf{X}$ [12,15] is defined by

$$\mathbf{E}_*(\mathbf{X}) = \limsup_{L \to \infty} \left(H_*(X_{1:L}) - h_*(\mathbf{X}) L \right). \tag{9}$$

Unlike entropy rates, the excess entropy and the permutation excess entropy for an arbitrary SSP do not generally coincide [12]. However, when $\mathbf{X}$ is the output process of a hidden Markov model with an ergodic internal process, it is known that the equality $\mathbf{E}(\mathbf{X}) = \mathbf{E}_*(\mathbf{X})$ holds [14].

If we associate information on the equality between occurrences of symbols to the map $\phi_{n,L}$, then we obtain a finer partition of the set of words of length L. We can define the modified permutation entropy rate and the modified permutation excess entropy by making use of this partition [15]. We have shown that the modified permutation entropy rate is equal to the entropy rate for any SSP $\mathbf{X}$ over A, and the modified permutation excess entropy is equal to the excess entropy if $\mathbf{X}$ is ergodic [15].

These previous works indicate that permutations can be used to calculate the entropy rate of SSPs over A without any restrictions. However, we need certain conditions to capture more detailed information of SSPs, such as the excess entropy.

2.4 Partially Ordered Permutation Entropies

For a partially ordered finite alphabet A and $L \geq 1$, we define a map

$$\phi^s_{A,L} : A^L \to \mathcal{M}_L(\{0,1\}) \tag{10}$$

by sending each word $x_{1:L}$ of length L to a square matrix M whose (i,j)-entry m_{ij} is $m_{ij} = 1$ if $x_i \leq x_j$ and $m_{ij} = 0$ otherwise, where $\mathcal{M}_L(\{0,1\})$ is the set of all $\{0,1\}$-valued square matrices of order L. Note that $\phi^s_{A,L}(x_{1:L})$, regarded as a binary relation on the set $\{1, 2, \ldots, L\}$, defines a preorder (a binary relation satisfying reflexivity and transitivity) for each $x_{1:L} \in A^L$.

We also define another map

$$\phi^t_{A,L} : A^L \to \mathcal{T}_L(\{0,1\}) \tag{11}$$

by sending each word $x_{1:L}$ of length L to an upper-triangular matrix T whose (i,j)-entry t_{ij} is $t_{ij} = 1$ if $x_i \leq x_j$ and $t_{ij} = 0$ otherwise for $i \leq j$, and $t_{ij} = 0$ for $i > j$, where $\mathcal{T}_L(\{0,1\})$ is the set of all $\{0,1\}$-valued upper-triangular matrices of order L. $\phi^t_{A,L}(x_{1:L})$ defines a partial order on the set $\{1, 2, \ldots, L\}$. Our motivation for introducing $\phi^t_{A,L}$ is to drop information on the ties (equality $x_i = x_j$ for $i \neq j$) from $\phi^s_{A,L}$. However, $\phi^t_{A,L}$ disregards information beyond ties since $t_{ij} = 0$ for $i > j$ regardless of whether $x_i = x_j$. We could define a map from A^L to $\mathcal{M}_L(\{0,1\})$ that exactly disregards ties. For such a map, our main theorems (Theorems 1 and 3) hold because the claims we will prove in the following are stronger results.

Maps $\phi^s_{A,L}$ and $\phi^t_{A,L}$ both induce a partition of A^L as follows: $x_{1:L}$ and $y_{1:L}$ are contained in the same block of the partition if $\phi^u_{A,L}(x_{1:L}) = \phi^u_{A,L}(y_{1:L})$ where $u = t$ or $u = s$. When A is a finite totally ordered set, the partition of A^L by $\phi^t_{A,L}$ is identical to that induced by the permutation type of words [12], whereas the partition of A^L by $\phi^s_{A,L}$ is that induced by the permutation type and arrangement of equalities [15]. Hence, when A is a totally ordered finite alphabet, the square partially ordered permutation entropies and the triangular partially ordered permutation entropies introduced below are reduced to the modified permutation entropies and the permutation entropies, respectively.

Let $\mathbf{X}$ be an SSP over a partially ordered finite alphabet A. For any $M \in \mathcal{M}_L(\{0,1\})$ and $T \in \mathcal{T}_L(\{0,1\})$, their probabilities of occurrence in $\mathbf{X}$ are given by

$$p_s(M) = \sum_{x_{1:L} \in \left(\phi^s_{A,L}\right)^{-1}(M)} p(x_{1:L}) \tag{12}$$

and

$$p_t(T) = \sum_{x_{1:L} \in \left(\phi^t_{A,L}\right)^{-1}(T)} p(x_{1:L}), \tag{13}$$

respectively. The *square partially ordered permutation entropy rate* of $\mathbf{X}$ is defined by

$$h_s(\mathbf{X}) = \lim_{L\to\infty} H_s(X_{1:L})/L, \tag{14}$$

where

$$H_s(X_{1:L}) = -\sum_{M \in \mathcal{M}_L(\{0,1\})} p_s(M) \log_2 p_s(M). \tag{15}$$

Similarly, we define the *triangular partially ordered permutation entropy rate* of $\mathbf{X}$ as

$$h_t(\mathbf{X}) = \lim_{L\to\infty} H_t(X_{1:L})/L, \tag{16}$$

where

$$H_t(X_{1:L}) = -\sum_{T \in \mathcal{T}_L(\{0,1\})} p_t(T) \log_2 p_t(T). \tag{17}$$

In the following, we will show that both limits for the right-hand-sides of Eqs. (14) and (16) exist and are equal to the entropy rate of $\mathbf{X}$. That is,

Theorem 1

$$h(\mathbf{X}) = h_s(\mathbf{X}) = h_t(\mathbf{X}) \tag{18}$$

holds for every SSP $\mathbf{X}$ *over a partially ordered finite alphabet.*

The partially ordered permutation excess entropies are also defined by replacing H and h in Eq. (5) by H_s and h_s or by H_t and h_t: We define the *square partially ordered permutation excess entropy* by

$$\mathbf{E}_s(\mathbf{X}) = \limsup_{L\to\infty} \left(H_s(X_{1:L}) - h_s(\mathbf{X})L\right) \tag{19}$$

and the *triangular partially ordered permutation excess entropy* by

$$\mathbf{E}_t(\mathbf{X}) = \limsup_{L\to\infty} \left(H_t(X_{1:L}) - h_t(\mathbf{X})L\right). \tag{20}$$

Unlike entropy rates, partially ordered permutation excess entropies are not necessarily equal to the excess entropy for every SSP over a partially ordered finite alphabet. The order structure of the alphabet is relevant.

Theorem 2 *Let* A *be a partially ordered finite alphabet.*

$$\mathbf{E}(\mathbf{X}) = \mathbf{E}_s(\mathbf{X}) \tag{21}$$

for every ergodic SSP $\mathbf{X}$ *over* A *if and only if* A *is asymmetric.*

Theorem 3 *Let* A *be a partially ordered finite alphabet.*

$$\mathbf{E}(\mathbf{X}) = \mathbf{E}_t(\mathbf{X}) \tag{22}$$

for every SSP $\mathbf{X}$ *over* A*, which is the output process of a hidden Markov model with an ergodic internal process if and only if* A *is asymmetric.*

Since every finite totally ordered set is asymmetric, Theorems 2 and 3 are extensions of our previous results in Refs. 12, 14, 15.

3 Square Partially Ordered Permutation Entropies

In this section, we prove the results on the square partially ordered permutation entropies. The following Lemma is straightforward.

Lemma 1 *Let $(A, \leq)$ be a finite poset. For any pair of words $x_{1:L}$ and $y_{1:L}$ of length $L \geq 1$, if $\phi^s_{(A,\leq),L}(x_{1:L}) = \phi^s_{(A,\leq),L}(y_{1:L})$, then $\phi^s_{(A,=),L}(x_{1:L}) = \phi^s_{(A,=),L}(y_{1:L})$.*

Lemma 1 says that the partition of A^L by $\phi^s_{(A,\leq),L}$ is a refinement of the one by $\phi^s_{(A,=),L}$. For the partition of A^L induced by $\phi^s_{(A,=),L}$, we have $S(n,k)$ blocks with size $(n)_k = n(n-1)\dots(n-k+1)$ for each $k = 1, 2, \dots, n$, where $n = |A|$ is the cardinality of A and $S(n,k)$ is a Stirling number of the second kind. Thus, $|\left(\phi^s_{(A,\leq),L}\right)^{-1}(M)| \leq n!$ for any $M \in \mathcal{M}_L(\{0,1\})$. This implies the following upper bound on the difference between $H(X_{1:L})$ and $H_s(X_{1:L})$.

Lemma 2 *Let $\mathbf{X}$ be an SSP over a partially ordered finite alphabet A. It holds that*

$$0 \leq H(X_{1:L}) - H_s(X_{1:L}) \leq \alpha^s_{\mathbf{X},A,L} \log_2(|A|!), \qquad (23)$$

where

$$\alpha^s_{\mathbf{X},A,L} = \sum_{\substack{M \in \mathcal{M}_L(\{0,1\}),\\ |\left(\phi^s_{A,L}\right)^{-1}(M)| > 1}} p_s(M). \qquad (24)$$

It is straightforward to obtain the equality of entropy rates $h(\mathbf{X}) = h_s(\mathbf{X})$ in Theorem 1 from Lemma 2.

An SSP $\mathbf{X}$ over a finite alphabet A is called *ergodic* if the relative frequency of each word $x_{1:L}$ converges to $p(x_{1:L})$ in probability [30]. That is, for any word $x_{1:k}$ of length $k \geq 1$, any $\epsilon > 0$ and any $\delta > 0$, there exists a natural number L_0 such that if $L > L_0$, then

$$\mathrm{Prob}\{|F_{x_{1:k},L} - p(x_{1:k})| < \delta\} > 1 - \epsilon, \qquad (25)$$

where $F_{x_{1:k},L}$ is the number of occurrences of $x_{1:k}$ in the sequence $X_1, X_2, \dots, X_L$ divided by $L - k + 1$.

Lemma 3 *Let A be a finite poset. If A is not asymmetric, then there exists an ergodic SSP $\mathbf{X}$ over A such that $\mathbf{E}(\mathbf{X}) > \mathbf{E}_s(\mathbf{X})$.*

Proof. Let A be a non-asymmetric finite poset. There exists an automorphism f of A such that $f \neq \mathrm{id}_A$. Choose $x \in A$ such that $x \neq f(x)$. Suppose that x appears in a word $x_{1:L}$ and set $y_{1:L} = f(x_1)f(x_2)\dots f(x_L)$. Since f is an isomorphism, we have $\phi^s_{A,L}(x_{1:L}) = \phi^s_{A,L}(y_{1:L})$. On the other hand, $x_{1:L} \neq y_{1:L}$ because $x \neq f(x)$. Thus, $\left(\phi^s_{A,L}\right)^{-1}(M) \geq 2$ for $M = \phi^s_{A,L}(x_{1:L})$.

Consider an i.i.d. process $\mathbf{X}$ over A such that every symbol occurs with the same probability. We have

$$\begin{aligned}
&H(X_{1:L}) - H_s(X_{1:L}) \\
=&\sum_{\substack{M \in \mathcal{M}_L(\{0,1\}),\\ |\left(\phi^s_{A,L}\right)^{-1}(M)| > 1}} p_s(M) \log_2 |\left(\phi^s_{A,L}\right)^{-1}(M)| \\
\geq&\sum_{\substack{x_{1:L} \in A^L,\\ \exists j \text{ s.t. } x_j = x}} p(x_{1:L}) \\
=&1 - \sum_{\substack{x_{1:L} \in A^L,\\ \forall j \ x_j \neq x}} p(x_{1:L}) \\
=&1 - (1 - p(x))^L \to 1 \text{ as } L \to \infty.
\end{aligned}$$

Since $h(\mathbf{X}) = h_s(\mathbf{X})$ by Theorem 1, we obtain the strict inequality $\mathbf{E}(\mathbf{X}) > \mathbf{E}_s(\mathbf{X})$. □

Lemma 4 *Let A be a finite poset. If A is asymmetric, then $\alpha^s_{\mathbf{X},A,L} \to 0$ as $L \to \infty$ for every ergodic SSP $\mathbf{X}$ over A.*

Proof. Assume that A is an asymmetric finite poset. Let $\mathbf{X}$ be an ergodic SSP over A. To prove that $\alpha^s_{\mathbf{X},A,L} \to 0$ as

$L \to \infty$, it is sufficient to show that $|\left(\phi^s_{A,L}\right)^{-1}(M)| = 1$ for all $M = \phi^s_{A,L}(x_{1:L})$, where $x_{1:L}$ satisfies the following condition: for all $x \in A$ there exists $1 \le j \le L$ such that $x = x_j$. Indeed, if this is true, then

$$\begin{aligned}\alpha^s_{\mathbf{X},A,L} =& \sum_{\substack{M \in \mathcal{M}_L(\{0,1\}),\\ |\left(\phi^s_{A,L}\right)^{-1}(M)|>1}} p_s(M)\\ =& 1 - \sum_{\substack{M \in \mathcal{M}_L(\{0,1\}),\\ |\left(\phi^s_{A,L}\right)^{-1}(M)|=1}} p_s(M)\\ \le& 1 - \sum_{x_{1:L} \text{ satisfying } (*)} p(x_{1:L}),\end{aligned}$$

where $(*)$ is the following condition: for all $x \in A$ with $p(x) > 0$ there exists j such that $x_j = x$. However, the sum over $x_{1:L}$ satisfying $(*)$ approaches 1 as $L \to \infty$, since $\mathbf{X}$ is ergodic.

For any word $x_{1:L} \in A^L$, let us assume that for all $x \in A$ there exists $1 \le j \le L$ such that $x = x_j$. Let $M = \phi^s_{A,L}(x_{1:L})$. Suppose there exists $y_{1:L} \in \left(\phi^s_{A,L}\right)^{-1}(M)$ such that $y_{1:L} \ne x_{1:L}$. Since $\phi^s_{A,L}(x_{1:L}) = \phi^s_{A,L}(y_{1:L})$, we have

$$x_i \le x_j \Leftrightarrow y_i \le y_j \text{ for all } 1 \le i,j \le L. \quad (26)$$

This implies

$$x_i = x_j \Leftrightarrow y_i = y_j \text{ for all } 1 \le i,j \le L. \quad (27)$$

Since every $x \in A$ appears in $x_{1:L}$, we can define a map $f : A \to A$ by $f(x_i) = y_i$ for $i = 1,2,\dots,L$. Note that f is well-defined due to (27), which also implies that f is injective. Finally, $x \le y \Leftrightarrow f(x) \le f(y)$ for all $x, y \in A$ by (26). Thus, f is an injective order-preserving self-map on the finite poset A, which implies that f is an automorphism of A. Now, f cannot be an identity map, because $x_{1:L} \ne y_{1:L}$. This contradicts the assumption that A is asymmetric. Hence, it holds that $|\left(\phi^s_{A,L}\right)^{-1}(M)| = 1$ for $M = \phi^s_{A,L}(x_{1:L})$, where $x_{1:L}$ is such that every $x \in A$ appears at least once in it. □

It is clear that Theorem 2 follows immediately from Theorem 1 and Lemmas 2, 3, and 4.

4 Triangular Partially Ordered Permutation Entropies

There is no analogue of Lemma 1 for $\phi^t_{A,L}$. However, we have the following size estimate for the inverse images of upper-triangular matrices by $\phi^t_{A,L}$.

Lemma 5 *Let A be a finite poset and $L \ge 1$. It holds that*

$$|\left(\phi^t_{A,L}\right)^{-1}(T)| \le O(L^{|A|-1}) \quad (28)$$

for every $T \in \mathcal{T}_L(\{0,1\})$.

Proof. We appeal to an induction on $|A|$. The first step $|A| = 1$ is trivial. Assume that the claim holds when $|A| \le n$, and consider the case $|A| = n + 1$. Since A is a finite poset, we can take a maximal element z of A. Let $T \in \mathcal{T}_L(\{0,1\})$. The total number of $x_{1:L}$ such that $\phi^t_{A,L}(x_{1:L}) = T$ and $x_i \ne z$ for all $1 \le i \le L$ is bounded by $O(L^{n-1})$ by the induction hypothesis. To bound the total number of those containing at least one z, let us divide them by the smallest subscript i such that $x_i = z$, which is denoted by $i_{x_{1:L}}$. If $i_{x_{1:L}} = i_{y_{1:L}}$ and $\phi^t_{A,L}(x_{1:L}) = \phi^t_{A,L}(y_{1:L}) = T$ for $x_{1:L}, y_{1:L} \in A^L$, then every occurrence of z in $x_{1:L}$ has the same subscript as that in $y_{1:L}$ due to the maximality of z. Thus, the total number of $x_{1:L}$ such that $\phi^t_{A,L}(x_{1:L}) = T$ and $i_{x_{1:L}} = j$ for a given j is the number of words of length $L-k$ that are mapped to the upper-triangular matrix T', which is obtained by removing all rows and columns corresponding to subscripts i such that $x_i = z$, which, in turn, is at most $O(L^{n-1})$ by the induction hypothesis, where k is the number of occurrences of z in $x_{1:L}$. Hence, the total number of words $x_{1:L}$ containing at least one z is at most $L \times O(L^{n-1}) = O(L^n)$. This completes the inductive step and the claim follows by mathematical induction. □

Thus, we obtain an analogue of Lemma 2.

Lemma 6 *Let* $\mathbf{X}$ *be an SSP over a partially ordered finite alphabet* A*. It holds that*

$$0 \leq H(X_{1:L}) - H_t(X_{1:L}) \leq \alpha^t_{\mathbf{X},A,L} \log_2 O(L^{|A|-1}), \quad (29)$$

where

$$\alpha^t_{\mathbf{X},A,L} = \sum_{\substack{T \in \mathcal{T}_L(\{0,1\}), \\ |(\phi^t_{A,L})^{-1}(M)|>1}} p_t(T). \quad (30)$$

The second equality of entropy rates $h(\mathbf{X}) = h_t(\mathbf{X})$ in Theorem 1 immediately follows from Lemma 6.

A quadruple $(\Sigma, A, \{U^{(x)}\}_{x\in A}, \mu)$ satisfying the following three conditions is called a *hidden Markov model* (HMM) [31], where Σ and A are finite sets, $U^{(x)}$ is a $|\Sigma| \times |\Sigma|$ matrix for each $x \in A$, and μ is a probability distribution on Σ.

(i) $U^{(x)}_{ss'} \geq 0$ for any $s, s' \in \Sigma$ and any $x \in A$;

(ii) $\sum_{s',x} U^{(x)}_{ss'} = 1$ for any $s \in \Sigma$;

(iii) and $\mu(s') = \sum_{s,x} \mu(s) U^{(x)}_{ss'}$ for any $s' \in \Sigma$.

If we introduce a $|\Sigma| \times |\Sigma|$ matrix U by $U = \sum_{a\in A} U^{(a)}$, then the triple (Σ, U, μ) defines the *underlying Markov chain*, with Σ a state set, U a state transition matrix, and μ a stationary probability distribution. The stationary process arising from the underlying Markov chain is called an *internal process*. The internal process is ergodic if and only if U is *irreducible* [32]: for any $s, s' \in \Sigma$, there exists $k > 0$, such that $(U^k)_{ss'} > 0$. An *output process* of an HMM is the SSP $\mathbf{X} = \{X_1, X_2, \ldots\}$ over A with the joint probability distributions:

$$\begin{aligned} &\mathrm{Prob}\{X_1 = x_1, X_2 = x_2, \ldots, X_L = x_L\} \\ &= \sum_{s,s'} \mu(s) \left(U^{(x_1)} \ldots U^{(x_L)}\right)_{ss'} \end{aligned} \quad (31)$$

for all $x_1, \ldots, x_L \in A$ and $L \geq 1$.

Lemma 7 *Let* A *be a finite poset. If* A *is not asymmetric, then there exists an SSP* $\mathbf{X}$ *over* A*, which is the output process of an HMM with an ergodic internal process such that* $\mathbf{E}(\mathbf{X}) > \mathbf{E}_t(\mathbf{X})$.

Proof. Let A be a non-asymmetric finite poset. The inequality $\mathbf{E}(\mathbf{X}) > \mathbf{E}_t(\mathbf{X})$ holds for the same i.i.d. process $\mathbf{X}$ over A given in the proof of Lemma 3. This can be shown in the same way as in Lemma 3. □

Lemma 8 *Let* A *be a finite poset. If* A *is asymmetric, then* $\alpha^t_{\mathbf{X},A,L} \to 0$ *exponentially fast as* $L \to \infty$ *for every SSP* $\mathbf{X}$ *over* A*, which is the output process of an HMM with an ergodic internal process.*

Proof. Let A be an asymmetric finite poset and $\mathbf{X}$ the output process of an HMM $(\Sigma, A, \{U^{(x)}\}_{x\in A}, \mu)$ with an ergodic internal process. We shall show that $|\left(\phi^t_{A,L}\right)^{-1}(T)| =$

1 for any $T = \phi^t_{A,L}(x_{1:L})$ where $x_{1:L}$ is such that every $x \in A$ occurs at least once both in $x_{1:\lfloor L/2 \rfloor}$ and $x_{\lfloor L/2 \rfloor+1:L}$ ($\lfloor r \rfloor$ is the largest integer less than or equal to a real number r). This implies $\alpha^t_{\mathbf{X},A,L} \to 0$ exponentially fast as $L \to \infty$. Indeed, we have

$$\begin{aligned}\alpha^t_{\mathbf{X},A,L} =& \sum_{\substack{T \in \mathcal{T}_L(\{0,1\}),\\ |(\phi^t_{A,L})^{-1}(T)|>1}} p_t(T) \\ =& 1 - \sum_{\substack{T \in \mathcal{T}_L(\{0,1\}),\\ |(\phi^t_{A,L})^{-1}(T)|=1}} p_t(T) \\ \leq& 1 - \sum_{x_{1:L} \text{ satisfying } (**)} p(x_{1:L}),\end{aligned}$$

where $(**)$ is the following condition: for all $x \in A$ with $p(x) > 0$ there exist $1 \leq j \leq \lfloor L/2 \rfloor < k \leq L$ such that $x_j = x_k = x$. However, the sum over $x_{1:L}$ satisfying $(**)$ approaches 1 exponentially fast as $L \to \infty$ by the ergodicity of the internal process of HMM $(\Sigma, A, \{U^{(x)}\}_{x\in A}, \mu)$ (See Lemma 3 in Ref. 14 and the proof of Lemma 12 in Ref. 12).

Let $x_{1:L}$ satisfy the following condition: for any $x \in A$ there exist j, k such that $1 \leq j \leq \lfloor L/2 \rfloor < k \leq L$ and $x_j = x_k = x$. Suppose that there exists $y_{1:L} \in (\phi^t_{A,L})^{-1}(T)$ such that $y_{1:L} \neq x_{1:L}$ for $T = \phi^t_{A,L}(x_{1:L})$. Since $\phi^t_{A,L}(x_{1:L}) = \phi^t_{A,L}(y_{1:L})$, we have

$$x_i \leq x_j \Leftrightarrow y_i \leq y_j \text{ for all } 1 \leq i \leq j \leq L. \tag{32}$$

Let us define a map f from A to itself by sending each $x \in A$ to y_{i_x}, where i_x is the minimum subscript i such that $x_i = x$. We have $i_x \leq \lfloor L/2 \rfloor$ by the choice of $x_{1:L}$. Let $x \leq x'$ in A. If $i_x \leq i_{x'}$, then $f(x) = y_{i_x} \leq y_{i_{x'}} = f(x')$ by (32). When $i_{x'} < i_x$, take a subscript $k > \lfloor L/2 \rfloor$ such that $y_k = y_{i_{x'}}$. By (32), we have $x_k \geq x_{i_{x'}} = x' \geq x = x_i$. Again by (32), we have $f(x) = y_i \leq y_k = f(x')$. Thus, f is order-preserving. Let $f(x) = f(x')$ for $x, x' \in A$. Take subscripts $k, l > \lfloor L/2 \rfloor$ such that $x_k = x$ and $x_l = x'$. By (32), we have $y_k \geq y_{i_x} = y_{i_{x'}}$. Again by (32), $x = x_k \geq x_{i_{x'}} = x'$. Similarly, we obtain $x' \geq x$ from $y_l \geq y_{i_{x'}} = y_{i_x}$. This implies that f is injective. Thus, f is shown to be an injective order-preserving self-map on the finite poset A; that is, it is an automorphism of A.

Next, we define another self-map g on A by $g(x) = y_{j_x}$ where $j_x > \lfloor L/2 \rfloor$ is the maximum subscript i such that $x_i = x$. We can show that g is also an automorphism of A in the same way as in the case of f.

Now, suppose that both f and g are the identity map of A. For a given $x \in A$, consider any subscript k such that $x_k = x$. Since $i_x \leq k \leq j_x$ and $x_{i_x} = x_k = x_{j_x}$, we have $x = f(x) = y_{i_x} \leq y_k \leq y_{j_x} \leq g(x) = x$ by (32), implying that $y_k = x = x_k$. However, this contradicts $x_{1:L} \neq y_{1:L}$. Therefore, at least one of f and g must be a non-identity automorphism of A. This in turn contradicts the assumption that A is asymmetric. Thus, it must hold that $|(\phi^t_{A,L})^{-1}(T)| = 1$ for $T = \phi^t_{A,L}(x_{1:L})$ where $x_{1:L}$ is such that every $x \in A$ occurs at least once in both $x_{1:\lfloor L/2 \rfloor}$ and $x_{\lfloor L/2 \rfloor+1:L}$. This completes the proof of the claim. □

One can see that Theorem 3 follows immediately from Theorem 1 and Lemmas 6, 7, and 8.

5 Concluding Remarks

One main message of this paper is that asymmetry of the alphabet is key to obtaining equalities between entropies and their partial-order analogues. Since every finite totally ordered set is asymmetric, the results in this paper demonstrate one reason why the original permutation entropies can capture various aspects of complexity of time series so well.

Let us go back to the motivating example mentioned in Sect. 1 to discuss implications of the results obtained in this paper. Let $(A_1, \leq_1), \ldots, (A_N, \leq_N)$ be finite totally ordered sets, and let $\mathbf{X}$ be an SSP over $A_1 \times \ldots A_N$ with $N \geq 2$. On one hand, $\mathbf{X}$ can be considered as an N-variate process where the i-th component takes its value in the totally ordered alphabet A_i. On the other hand, if we introduce the pointwise order $\leq$ into $A_1 \times \cdots \times A_N$ ($(x_1, \ldots, x_N) \leq (y_1, \ldots, y_N)$ for $(x_1, \ldots, x_N), (y_1, \ldots, y_N) \in A_1 \times \cdots \times A_N$ if $x_i \leq_i y_i$ for all $1 \leq i \leq N$), then $\mathbf{X}$ can be regarded as a univariate process over the partially ordered finite alphabet $A_1 \times \cdots \times A_N$. In Ref. 14, we took the former viewpoint to show that the entropy rate of $\mathbf{X}$ can be calculated as the multivariate permutation entropy rate by introducing the joint probability of N-tuples of permutations. We took the latter viewpoint in this paper, and Theorem 1 shows that the entropy rate of $\mathbf{X}$ can be also calculated as the triangular and square partially ordered permutation entropy rates. The partition of the set of words of length L over $A_1 \times \cdots \times A_N$ used in the latter approach is coarser than that used in the former approach. This means that we need less information than that assumed in previous work in order to obtain the entropy rate of $\mathbf{X}$. However, Theorems 2 and 3 indicate that the latter approach is limited in calculating information-theoretic quantities for $\mathbf{X}$ beyond the entropy rate, even under the ergodicity assumptions, because the pointwise order $\leq$ on $A_1 \times \cdots \times A_N$ is not asymmetric in general.

The equalities between the excess entropy and the partially ordered excess entropies requires convergence of the difference between the Shannon entropy of the words of length L and the Shannon entropy of the $\{0, 1\}$-valued matrices of order L. The proofs of Theorems 2 and 3 reveal that asymmetry of the alphabet is a necessary and sufficient condition for convergence under the ergodicity assumptions. On the other hand, equalities between the entropy rate and the partially ordered permutation entropy rates require a weaker condition, namely, that the difference of the Shannon entropies grow slower than the length of the words. The proof of Theorem 1 shows that this condition holds, regardless of the degree of symmetry of the alphabet.

We can obtain equalities between several other entropies and their partial-order analogues under the same assumption as in Theorem 2 and Theorem 3. For example, the transfer entropy rate is equal to the square partially ordered symbolic transfer entropy rate (which can be defined in the obvious manner by extending the definition of the rate of symbolic transfer entropy in Ref. 14) for every ergodic SSP over a finite partially ordered alphabet A if A is asymmetric. A similar statement holds for the triangular partially ordered symbolic transfer entropy rate. The same can also be done for the rate of transfer entropy on rank vectors [33] and momentary information transfer [34]. The proofs of these claims can be straightforwardly given by extending Lemmas 2, 4, 6 and 8 to multivariate versions.

References

1. C. Bandt and B. Pompe. Permutation entropy: a natural complexity measure for time series. *Phys. Rev. Lett.*, 88:174102, 2002.
2. J. M. Amigó. *Permutation Complexity in Dynamical Systems*. Springer-Verlag Berlin Heidelberg, 2010.
3. J. M. Amigó and K. Keller. Permutation entropy: One concpet, two approaches. *Eur. Phys. J. Special Topics*, 222:263–273, 2013.
4. M. Zanin, L. Zunino, O. Rosso, and D. Papo. Permutation entropy and its main biomedical and econophysics applications: A review. *Entropy*, 14:1553–1577, 2012.
5. J. M. Amigó, K. Keller, and V. A. Unakafova. Ordinal symbolic analysis and its application to biomedical recordings. *Phil. Trans. R. Soc. A*, 373:20140091, 2015.
6. K. Keller, T. Mangold, I. Stolz, and J. Werner. Permutation entropy: New ideas and challenges. *Entropy*, 19:134, 2017.
7. J. M. Amigó, M. B. Kennel, and L. Kocarev. The permutation entropy rate equals the metric entropy rate for ergodic information sources and ergodic dynamical systems. *Physica D*, 210:77–95, 2005.
8. J. M. Amigó. The equality of Kolmogorov-Sinai entropy and metric permutation entropy generalized. *Physica D*, 241:789–793, 2012.
9. J. P. Crutchfield and D. P. Feldman. Regularities unseen, randomness observed: Levels of entropy convergence. *Chaos*, 15:25–54, 2003.
10. T. Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, 85:461–464, 2000.
11. M. Staniek and K. Lehnertz. Symbolic transfer entropy. *Phys. Rev. Lett.*, 100:158101, 2008.
12. T. Haruna and K. Nakajima. Permutation complexity via duality between values and orderings. *Physica D*, 240:1370–1377, 2011.
13. T. Haruna and K. Nakajima. Symbolic transfer entropy rate is equal to transfer entropy rate for bivariate finite-alphabet stationary ergodic markov processes. *Eur. Phys. J. B*, 86:230, 2013.
14. T. Haruna and K. Nakajima. Permutation complexity and coupling measures in hidden markov models. *Entropy*, 15:3910–3930, 2013.
15. T. Haruna and K. Nakajima. Permutation approach to finite-alphabet stationary stochastic processes based on the duality between values and orderings. *Eur. Phys. J. Special Topics*, 222:383–399, 2013.
16. T. Haruna and K. Nakajima. Permutation excess entropy and mutual information between the past and future. *Int. J. Comput. Ant. Sys.*, 26:197–209, 2014.
17. R. Monetti, W. Bunk, T. Aschenbernner, and F. Jamitzky. Characterizing synchronization in time series using information measures extracted from symbolic representations. *Phys. Rev. E*, 79:046207, 2009.
18. J. M. Amigó, R. Monetti, T. Aschenbernner, and W. Bunk. Transcripts: An algebraic approach to coupled time series. *CHAOS*, 22:013105, 2012.
19. R. Monetti, W. Bunk, T. Aschenbernner, S. Springer, and J. M. Amigó. Information directionality in coupled time series using transcripts. *Phys. Rev. E*, 88:022911, 2013.

20. J. M. Amigó, T. Aschenbernner, W. Bunk, and R. Monetti. Dimensional reduction of conditional algebraic multi-information via transcripts. *Information Sciences*, 278:298–310, 2014.

21. C. Bian, C. Qin, Q. D. Y. Ma, and Q. Shen. Modified permutation-entropy analysis of heartbeat dynamics. *Phys. Rev. E*, 85:021906, 2012.

22. T. Haruna. Partially ordered permutation complexity of coupled time series. *Physica D*, 2018. https://doi.org/10.1016/j.physd.2018.09.002.

23. T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley & Sons, Inc., 1991.

24. P. Grassberger. Toward a quantitative theory of self-generated complexity. *Int. J. Theor. Phys.*, 25:907–938, 1986.

25. R. Shaw. *The Dripping Faucet as a Model Chaotic System.* Aerial Press, Santa Cruz, California, 1984.

26. W. Bialek, I. Nemenman, and N. Tishby. Predictability, complexity, and learning. *Neural Computation*, 13:2409–2463, 2001.

27. W. Li. On the relationship between complexity and entropy for markov chains and regular languages. *Complex Systems*, 5:381–399, 1991.

28. D. V. Arnold. Information-theoretic analysis of phase transitions. *Complex Systems*, 10:143–155, 1996.

29. B. S. W. Schröder. *Ordered Sets: An Introduction.* Springer Science+Business Media, New York, 2003.

30. R. B. Ash. *Information Theory.* Interscience Publishers, 1965.

31. Brian D. O. Anderson. The realization problem for hidden markov models. *Math. Control Signals Systems*, 12:80–120, 1999.

32. P. Walters. *An Introduction to Ergodic Theory.* Springer-Verlag, New York, 1982.

33. D. Kugiumtzis. Transfer entropy on rank vectors. *J. Nonlin. Sys. Appl.*, 3:73–81, 2012.

34. B. Pompe and J. Runge. Momentary information transfer as a coupling measure of time series. *Phys. Rev. E*, 83:051122, 2011.