Symbolic transfer entropy rate is equal to transfer entropy rate for bivariate finite-alphabet stationary ergodic Markov processes

Taichi Haruna, Kohei Nakajima

Abstract—Transfer entropy is a measure of the magnitude and the direction of information flow between jointly distributed stochastic processes. In recent years, its permutation versions are considered in the literature to estimate the transfer entropy by counting the number of occurrence of orderings between values, not the values themselves. Here, we introduce the transfer entropy rate and its permutation version, the symbolic transfer entropy rate, and show that they are equal to each other for any bivariate finite-alphabet stationary ergodic Markov process. Our proof is based on the duality between values and orderings, which is introduced in [T. Haruna and K. Nakajima, Physica D 240, 1370 (2011)] and may give a coherent basis for the relationship between information theoretical quantities and their permutation versions defined on finite-alphabet stationary stochastic processes. We also discuss the relationship among the transfer entropy rate, the time-delayed mutual information rate and their permutation versions.

Index Terms—Permutation entropy, transfer entropy, symbolic transfer entropy, transfer entropy on rank vectors, Markov process

I. INTRODUCTION

Quantifying networks of information flows is critical to understand working of complex systems such as living, social and technological systems. Schreiber [1] introduced the notion of transfer entropy to measure the magnitude and the direction of information flow from one element to another element emitting stationary signals in a given system. It has been used to analyze information flows in real time series data from neuroscience [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11], and many other fields [12, 13, 14, 15, 16, 17, 18, 19].

The notion of permutation entropy introduced by Bandt and Pompe [20] has been proved that much of information contained in stationary time series can be captured by counting occurrences of orderings between values, not those of values themselves [21, 22, 23, 24, 25]. In particular, it is known that the entropy rate [26], which is one of the most fundamental quantity of stationary stochastic processes, is equal to the permutation entropy rate for any finite-alphabet stationary stochastic process [27]. The method of permutation has been applied across many disciplines [28].

The symbolic transfer entropy [29] is a permutation version of the transfer entropy and has been used as an efficient and conceptually simple way of quantifying information flows in real time series data [29, 30, 31, 32]. Another permutation version of the transfer entropy called transfer entropy on rank vectors has been introduced to improve the performance of the symbolic transfer entropy [33]. So far, most of the work on permutation versions of the transfer entropy are in application side. Although some theoretical results are known for the original transfer entropy [34, 35], so far no theoretical consideration has been given for its permutation versions to the best of the authors’ knowledge. The aim of this paper is to shed light on the relationship between the transfer entropy and their permutation versions from a theoretical perspective. In particular, we will consider the notions of transfer entropy rate [36] and symbolic transfer entropy rate and show that they are equal for any bivariate finite-alphabet stationary ergodic Markov process.

Our approach is based on the duality between values and orderings introduced by the authors [37]. In [37], the excess entropy [38], [39], [40], [41], [42], [43], [44], [45], which is an effective measure of complexity of stationary stochastic processes, and its permutation version is shown to be equal for any finite-alphabet stationary ergodic Markov process. Our result in this paper demonstrates the broad applicability of the duality between values and orderings for discussing the relationship between information theoretic measures and their permutation versions defined on finite-alphabet stationary stochastic processes.

This paper is organized as follows. In Section II we introduce the transfer entropy rate and the symbolic transfer entropy rate. We also discuss some combinatorial facts used in later sections. In Section III we give a proof of the equality between the transfer entropy rate and the symbolic transfer entropy rate which holds for bivariate finite-alphabet stationary ergodic Markov processes. In Section IV we discuss the relationship among the transfer entropy rate, the time-delayed mutual information rate and their permutation versions. Finally, in Section V we give concluding remarks.

II. DEFINITIONS AND PRELIMINARIES

Let \( A_n = \{1, 2, \cdots, n\} \) be a finite alphabet consisting of natural numbers from 1 to \( n \). In the following discussion, \( X = \)
\( \{X_1, X_2, \ldots \} \) and \( Y \equiv \{Y_1, Y_2, \ldots \} \) are jointly distributed finite-alphabet stationary stochastic processes, or equivalently, \((X, Y)\) is a bivariate finite-alphabet stationary stochastic process \( \{(X_1, Y_1), (X_2, Y_2), \ldots \}, \) where stochastic variables \( X_i \) and \( Y_j \) take their values in the alphabet \( A_n \) and \( A_m \), respectively. We use the notation \( X_i^L \equiv (X_1, X_2, \ldots, X_L) \) for simplicity. We write \( p(x_i^L, y_i^M) \) for the probability of the occurrence of words \( x_i^L \in A_n^L \) and \( y_i^M \in A_m^M \).

Originally, the notion of transfer entropy was introduced as a generalization of the entropy rate to bivariate processes \([1]\). Along this original motivation, here, we do not consider the transfer entropy but the transfer entropy rate \([56]\) from \( Y \) to \( X \) which is defined by

\[
t(X|Y) \equiv h(X) - h(X|Y),
\]

where \( h(X) \equiv \lim_{L \to \infty} H(X_i^L)/L \) is the entropy rate of \( X \), \( H(X_i^L) \equiv -\sum_{x_i^L \in A_n^L} p(x_i^L) \log_2 p(x_i^L) \) is the Shannon entropy of the occurrence of words of length \( L \) in \( X \) and \( h(X|Y) \) is the conditional entropy rate of \( X \) given \( Y \) defined by

\[
h(X|Y) \equiv \lim_{L \to \infty} H(X_{i+1}|X_i^L, Y_i^M),
\]

which always converges. \( t(X|Y) \) has the properties that (i) \( 0 \leq t(X|Y) \leq h(X) \) and (ii) \( t(X|Y) = 0 \) if \( X_i^L \) is independent of \( Y_i^M \) for all \( L \geq 1 \).

In order to introduce the notion of symbolic transfer entropy rate, we define a total order on the alphabet \( A_n \) by the usual “less-than-or-equal-to” relationship. Let \( S_L \) be the set of all permutations of length \( L \geq 1 \). We consider each permutation \( \pi \) of length \( L \) as a bijection on the set \( \{1, 2, \ldots, L\} \). Thus, each permutation \( \pi \in S_L \) can be identified with a sequence \( \pi(1) \cdots \pi(L) \). The permutation type \( \pi \in S_L \) of a given word \( x_i^L \in A_n^L \) is defined by re-ordering \( x_1, \ldots, x_L \) in ascending order, namely, \( x_i^L \) is of type \( \pi \) if we have \( x_{\pi(i)} \leq x_{\pi(i+1)} \) and \( \pi(i) < \pi(i+1) \) when \( x_{\pi(i)} = x_{\pi(i+1)} \) for \( i = 1, 2, \ldots, L - 1 \). For example, \( \pi(1) \pi(2) \pi(3) \pi(4) \pi(5) = 41352 \) for \( x_5^5 = 24213 \in A_5^5 \) because \( x_1 x_2 x_3 x_4 x_5 = \) 12234. The map \( \phi_n : A_n^L \rightarrow S_L \) sends each word \( x_i^L \) to its unique permutation type \( \pi = \phi_n(x_i^L) \).

We will use the notions of rank sequences and rank variables \([22]\) in some situations. The rank sequences of length \( L \) are words \( r_i^L \in A_n^L \) satisfying \( 1 \leq r_i \leq i \) for \( i = 1, \ldots, L \). The set of all rank sequences of length \( L \) is denoted by \( R_L \). It is clear that \( |R_L| = L! = |S_L| \). Each word \( x_i^L \in A_n^L \) can be mapped to a rank sequence \( r_i^L \) by defining \( r_i \equiv \sum_{j=1}^i \delta(x_j \leq x_i) \) for \( i = 1, \ldots, L \), where \( \delta(P) = 1 \) if the proposition \( P \) is true, otherwise \( \delta(P) = 0 \). We denote this map from \( A_n^L \) to \( R_L \) by \( \varphi_n \). It can be shown that the map \( \varphi_n : A_n^L \rightarrow R_L \) is compatible with the map \( \phi_n : A_n^L \rightarrow S_L \) in the following sense: there exists a bijection \( \iota : R_L \rightarrow S_L \) such that \( \iota \circ \varphi_n = \phi_n \). \([22]\) The rank variables associated with \( X \) are defined by \( R_i \equiv \sum_{j=1}^i \delta(X_j \leq X_i) \) for \( i = 1, \ldots, L \). In general, \( R \equiv \{R_1, R_2, \ldots\} \) is a non-stationary stochastic process.

The symbolic transfer entropy rate from \( Y \) to \( X \) is defined by

\[
t^*(X|Y) \equiv h^*(X) - h^*(X|Y),
\]

where \( h^*(X) \equiv \lim_{L \to \infty} H^*(X_i^L)/L \) is the permutation entropy rate which is known to exists and is equal to \( h(X) \) \([27]\), \( H^*(X_i^L) \equiv -\sum_{x_i^L \in A_n^L} p(x_i^L) \log_2 p(x_i^L) \) is the Shannon entropy of the occurrence of permutations of length \( L \) in \( X \), \( p(\pi) = \sum_{x_i^L=\pi} p(x_i^L) \) and \( h^*(X|Y) \) is given by

\[
h^*(X|Y) \equiv \lim_{L \to \infty} \left( H^*(X_{i+1}^L + Y_i^L) - H^*(X_i^L + Y_i^L) \right)
\]

if the limit in the right hand side exists. Here, \( H^*(X_i^L + Y_i^M) \) is defined by \( H^*(X_i^L + Y_i^M) \equiv -\sum_{\pi \in S_L, \pi' \in S_M} p(\pi, \pi') \log_2 p(\pi, \pi') \) where \( p(\pi, \pi') = \sum_{x_i^L=\pi, y_i^M=\pi'} p(x_i^L, y_i^M) \). Let \( R \) and \( S \) be rank variables associated with \( X \) and \( Y \), respectively. By the compatibility between \( \phi_k \) and \( \varphi_k \) for \( k = m, n \), we have \( H(R_i, S_i^M) = H^*(X_i^L + Y_i^M) \). Thus, \( h^*(X|Y) \) can be written as \( h^*(X|Y) = \lim_{L \to \infty} H(R_{i+1}, S_i^M) \) if \( h^*(X|Y) \) exists.

Note that the above definition of the symbolic transfer entropy rate \([3]\) is the rate version of the transfer entropy on rank vectors \([33]\) which is an improved version of the symbolic transfer entropy \([29]\). Of course, we can directly consider the rate version of the symbolic transfer entropy. However, its definition is more complicated than the rate version of the transfer entropy on rank vectors. Moreover, we can show that the rate version of the symbolic transfer entropy is equal to the rate version of the transfer entropy on rank vectors for bivariate finite-alphabet ergodic Markov processes, the class of stochastic processes considered in this paper.\(^1\) Hence, we here take the formula \([5]\) as the definition of the symbolic transfer entropy rate.

### III. Main Result

In this section, we give a proof of the following claim: for any bivariate finite-alphabet stationary ergodic Markov process \((X, Y)\),

\[
t(X|Y) = t^*(X|Y).
\]

In order to prove this claim, we introduce a map \( \mu : S_L \rightarrow \mathbb{N}^L \), where \( \mathbb{N} = \{1, 2, \ldots\} \) is the set of all natural numbers ordered by usual “less-than-or-equal-to” relationship, by the following procedure: First, given a permutation \( \pi \in S_L \), we decompose the sequence \( \pi(1) \cdots \pi(L) \) into maximal ascending subsequences. A subsequence \( i_1 \cdots i_k \) of a sequence \( i_1 \cdots i_L \) is called a maximal ascending subsequence if it is ascending, namely, \( i_j \leq i_{j+1} \leq \cdots \leq i_k \), and neither \( i_{j-1} \cdots i_j \cdots i_{j+k} \) nor \( i_j \cdots i_{j+k} \cdots i_{j+k+1} \) is a subsequence. Second, if \( \pi(1) \cdots \pi(\ell_1), \pi(\ell_1 + 1) \cdots \pi(\ell_2), \cdots, \pi(\ell_k - 1 + 1) \cdots \pi(L) \) is a decomposition of \( \pi(1) \cdots \pi(L) \) into maximal ascending subsequences, then we define a word \( x_i^L \in \mathbb{N}^L \) by \( x_1(1) = \cdots = x_{\pi(1)}(1) = 1, x_{\pi(1)+1}(1) = \cdots = x_{\pi(\ell_1)}(1) = 2, \cdots, x_{\pi(\ell_k-1)+1}(1) = \cdots = x_{\pi(L)}(1) = k \). We define \( \mu(\pi) = x_i^L \). For example, a decomposition of 25341 \( S_L \) into maximal ascending subsequences is 25, 34, 1. We obtain \( \mu(\pi) = x_1 x_2 x_3 x_4 x_5 = 31221 \)

\(^1\)The proof for this result will be presented elsewhere because it requires generalization to hidden Markov processes which is beyond the scope of this paper.
by putting \( x_{22}x_{33}x_{44}x_{1} = 11223 \). By construction, we have \( \phi_n \circ \mu(\pi) = \pi \) when \( \mu(\pi) \in A_n^L \) for any \( \pi \in S_L \).

The map \( \mu \) can be seen as a dual to the map \( \phi_n \) (or \( \varphi_n \)) in the following sense (Theorem 9 in [27]): Let us put

\[
B_{n,L} = \{ x^L_n \in A_n^L | \phi_n^-(\pi) = \{ x^L_i \} \text{ for some } \pi \in S_L \},
\]

\[
C_{n,L} = \{ \pi \in S_L | ||\phi_n^-(\pi)|| = 1 \}.
\]

Then, \( \phi_n \) restricted on \( B_{n,L} \) is a map into \( C_{n,L} \), \( \mu \) restricted on \( C_{n,L} \) is a map into \( B_{n,L} \), and they form a pair of mutually inverse maps. Furthermore, we have the following structural characterization of words in the set \( B_{n,L} \) defined by (6).

Since \( h(X) = h^*(X) \) holds for any finite-alphabet stationary process, proving (5) is equivalent to showing that the equality

\[
\lim_{L \to \infty} H(R_{L+1}|R^L_1, S^L_1) = \lim_{L \to \infty} H(X_{L+1}|X^L_1, Y^L_1)
\]

holds for any bivariate finite-alphabet stationary ergodic Markov process \( (X, Y) \). Without loss of generality, we can assume that each \( (x, y) \in A_n \times A_n \) appears with a positive probability \( p(x, y) > 0 \). By the assumed ergodicity, for any \( \epsilon > 0 \) if we take \( L \) sufficiently large, then

\[
\sum_{x_i^L, y_i^L} p(x_i^L, y_i^L) > 1 - \epsilon,
\]

where \( (\ast) \) is the condition that for any \( x \in A_n \) there exist \( 1 \leq i \leq \lfloor L/2 \rfloor < j \leq L \) such that \( x \neq x_i = x_j \), \( (\ast \ast) \) is the condition that for any \( y \in A_m \) there exist \( 1 \leq i' \leq \lfloor L/2 \rfloor < j' \leq L \) such that \( y = y_i = y_{j'} \) and \( |r| \) is the largest integer not greater than \( r \) for a real number \( r \).

We put \( D_{n,m,L} = \{(x_i^L, y_i^L) | (\ast) \text{ and } (\ast \ast) \text{ hold} \} \) and \( E_{n,m,L} = \{(r_i^L, s_i^L) | \varphi_n(x_i^L) = r_i^L, \varphi_m(y_i^L) = s_i^L \text{ for some } (x_i^L, y_i^L) \in D_{n,m,L} \} \). By (6), we have \( x_i^L \in B_{n,L} \) and \( y_i^L \in B_{m,L} \) for \( (x_i^L, y_i^L) \in D_{n,m,L} \). Thus, the map \( (x_i^L, y_i^L) \mapsto (\varphi_n(x_i^L), \varphi_m(y_i^L)) \) is a bijection from \( D_{n,m,L} \) to \( E_{n,m,L} \) due to the duality between \( \phi_n \) and \( \mu \) for \( k = m, n \).

In particular, we have \( p(x_i^L, y_i^L) = p(r_i^L, s_i^L) \) and \( p(r_{i+1}, s_{i+1}) = p(r_{i+1}, x_i^L, y_i^L) \) for \( (x_i^L, y_i^L) \in D_{n,m,L} \) and \( r_i^L = \varphi_n(x_i^L), s_i^L = \varphi_m(y_i^L) \).

Given any \( \epsilon > 0 \), let us take \( L \) large enough so that the inequality (8) holds. We shall evaluate each term in the right-hand side of (9). The second term in (9) is bounded by \( c \log_2 n \) which can be arbitrary small. In order to show the third term also converges to 0 as \( L \to \infty \), we use the Markov property: If \( (X, Y) \) is ergodic Markov, then by the similar discussion as in Section 4 of [27] we can show that

\[
\sum_{(r_i^L, s_i^L)} p(r_1^L, s_1^L) < C_{n,m,L} L^2 \log_2 (L + 1)
\]

for some \( C > 0, 0 \leq \lambda \leq 1 \) and positive integer \( a \). Thus, the absolute value of the third term is bounded by the quantity \( C_{n,m,L} L^2 \log_2 (L + 1) \) which goes to 0 as \( L \to \infty \). Finally, the first term is shown to 0 by the same discussion as in the proof of Lemma 1 of (27).

Thus, given \( (x_i^L, y_i^L) \in D_{n,m,L} \), the probability distribution

\[
p(r_{i+1} | x_i^L, y_i^L)
\]

is just a re-indexing of \( p(x_{L+1} | x_i^L, y_i^L) \), which implies that the first term is exactly equal to 0. This completes the proof of our claim. From the proof, we can also see that 

\[
t^X(X, Y) \leq t^X(X, Y)
\]

holds for any bivariate finite-alphabet stationary ergodic process \( (X, Y) \) if \( h^*(X|Y) \) exists for the process.

IV. ON THE RELATIONSHIP WITH THE TIME-DELAYED MUTUAL INFORMATION RATE

Apart from permutation, it is natural to ask whether the equality for the conditional entropy rate

\[
l_{\to \infty} H(X_{L+1} | X^L_1, Y^L_1) = \frac{1}{L} H(X_{L+1} | Y^L_1) \quad (10)
\]

holds or not, which is parallel to the equality for the entropy rate

\[
l_{\to \infty} H(X_{L+1} | X^L_1) = \frac{1}{L} H(X_{L+1} | Y^L_1) \quad (11)
\]

for any finite-alphabet stationary stochastic process \( X, Y \).

In this section, we will see that this question has an intimate relationship with the relationship between the transfer entropy rate and the time-delayed mutual information rate.

In general, (11) does not hold. For example, if \( X = Y \), then we have \( l_{\to \infty} H(X_{L+1} | X^L_1, Y^L_1) = h(X) \), while \( l_{\to \infty} \frac{1}{L} H(X_{L+1} | Y^L_1) = 0 \). However, note that the inequality

\[
l_{\to \infty} H(X_{L+1} | X^L_1, Y^L_1) \geq \frac{1}{L} H(X_{L+1} | Y^L_1) \quad (11)
\]

holds for any bivariate finite-alphabet stationary stochastic process \( (X, Y) \). Indeed, we have

\[
l_{\to \infty} H(X_{L+1} | X^L_1, Y^L_1) = \frac{1}{L} \sum_{i=1}^{L+1} H(X_i | X_{i-1}^L, Y_{i-1})
\]

for any \( i \), the first equality is due to the Cesàro mean theorem (if \( \lim_{L \to \infty} b_L = b \) then \( \lim_{L \to \infty} \frac{1}{L} \sum_{i=1}^{L+1} b_i = b \) and the last equality follows from the chain rule for the Shannon entropy. In the following, we give a sufficient condition for (10).

If there exists \( N > 0 \) such that if \( i > N \) then \( X_i \) is independent of \( Y_{i+j} \) given \( X_1^i, Y_1^i \) for any \( j \geq 0 \), that is, \( Pr(X_i = x_i, Y_{i+j} = y_{i+j} | X_1^i, Y_1^i) = Pr(X_i = x_i | X_1^i, Y_1^i) Pr(Y_{i+j} = y_{i+j} | X_1^i, Y_1^i) \) for any \( j \geq 0 \). Then, \( X_i \) is a decreasing sequence of \( L \). By the chain rule for the Shannon entropy, \( H(X_{L+1} | Y^L_1) = H(X_{L+1} | X^L_1, Y^L_1) + H(X_L | X_{L-1}^L, Y^L_1) + \cdots + H(X_1) = a_{L+L} + a_{L-1} + \cdots + a_{0} \).

Since the
for any Markov property:

\[ I(x_{i+1} \mid x_i) \leq \log_2 \sum_{y_1} p(x_{i+1} \mid x_i, y_1) \log_2 p(x_{i+1} \mid x_i, y_1) \]

by the inequality (11).

If the former sum is finite, by the Cesáro mean theorem, we obtain

\[ \lim_{L \to \infty} \frac{1}{L} H(X_{L+1} \mid Y_L) = \lim_{L \to \infty} \frac{1}{L} \sum_{i=N}^{L} a_{i,i} = \lim_{L \to \infty} a_{L,L} = \lim_{L \to \infty} H(X_{L+1} \mid Y_L^L). \]

Note that if the assumption holds, then it holds for \( N = 1 \) by stationarity. If \((X, Y)\) is Markov, then we can show that the assumption is equivalent to the following simpler condition by using the Markov property:

\[ p(x_2, y_2 \mid x_1, y_1) = p(x_2 \mid x_1) p(y_2 \mid y_1) \quad (12) \]

for any \( x_1, x_2 \in A_n \) and \( y_1, y_2 \in A_m \).

If (10) holds, then we obtain

\[ t(X \mid Y) = \lim_{L \to \infty} \frac{1}{L} I(X_{L+1}^L \mid Y_L^L), \]

where \( I(A; B) \) is the mutual information between stochastic variables \( A \) and \( B \). We call the quantity at the right hand side of (13) \textit{time-delayed mutual information rate} and denote it by \( i_{s+1}(X; Y) \). Note that we have \( t(X \mid Y) \leq i_{s+1}(X; Y) \) for any bivariate finite-alphabet stationary stochastic process \((X, Y)\) by the inequality (11).

For any bivariate finite-alphabet stationary stochastic process \((X, Y)\), one can see that \( i_{s+1}(X; Y) = h(X) + h(Y) - h(X \mid Y) \) holds. Hence, we have \( i_{s+1}(X; Y) = i_{s+1}(Y; X) = i(X; Y) \). Here, \( i(X; Y) = \lim_{L \to \infty} \frac{1}{L} I(X_{L+1}; Y_L) \) is the mutual information rate between \( X \) and \( Y \). Thus, when we consider the rate for mutual information between two jointly distributed finite-alphabet stationary stochastic processes \( X \) and \( Y \), which is defined in the limit \( L \to \infty \), time delay has no significance in contrast to the time-delayed mutual information which has been used in time series embedding \( [46] \) and detection of nonlinear interdependence between two time series at different time points \( [47], [48] \).

The result on the relationship between \( t(X \mid Y) \) and \( i_{s+1}(X \mid Y) \) when \((X, Y)\) is Markov can be summarized as follows: We have (13) if the condition (12) holds for any \( x_1, x_2 \in A_n \) and \( y_1, y_2 \in A_m \), when \((X, Y)\) is Markov.

Another interesting case is when both \((X, Y)\) and \( Y \) are Markov. In this case, a necessary and sufficient condition for (13) can be derived easily: If both \((X, Y)\) and \( Y \) are Markov, then we have (13) if and only if the condition

\[ p(x_2, y_2 \mid x_1, y_1) = p(x_2 \mid x_1) p(y_2 \mid y_1) \quad (14) \]

holds for any \( x_1, x_2 \in A_n \) and \( y_1, y_2 \in A_m \). Proving this claim is equivalent to showing \( h(X \mid Y) = h(X, Y) - h(Y) \).

By using the Markov property, we obtain \( h(X \mid Y) = h(X, Y) + h(Y) = H(X_{L+1} \mid Y_L) - H(X_{L+1} \mid Y_L, Y_2) \geq 0 \). In the last inequality, we have the equality if and only if \( Y_2 \) is independent of \( X_2^2 \) given \( Y_1 \), that is,

\[ p(x_1, x_2 \mid y_1) = p(x_1, x_2) p(y_1) \quad (15) \]

for any \( x_1, x_2 \in A_n \) and \( y_1, y_2 \in A_m \), which is equivalent to the condition in the claim.

Let us introduce the symbolic time-delayed mutual information rate by

\[ i_{s+1}^s(X; Y) \equiv \lim_{L \to \infty} \frac{1}{L} I^s(X_{L+1}; Y_L^L), \]

where \( I^s(X_{L+1}; Y_L^L) \equiv H^s(X_{L+1}) + H^s(Y_L^L) - H^s(X_{L+1}, Y_L^L) \) and discuss the relationship with the transfer entropy rate, the symbolic transfer entropy rate and the (time-delayed) mutual information rate. \( i_{s+1}^s(X; Y) \) exists for any bivariate finite-alphabet stationary stochastic process as we will see below.

Similar properties with the time-delayed mutual information rate hold for the symbolic time-delayed mutual information rate: First, we note that \( t^s(X \mid Y) \leq i_{s+1}^s(X \mid Y) \) holds for any bivariate finite-alphabet stationary stochastic process \((X, Y)\) such that \( t^s(X \mid Y) \) exists because the similar inequality with (11) holds for permutation versions of corresponding quantities. Second, the symbolic time-delayed mutual information rate also admits the following expression:

\[ i_{s+1}^s(X; Y) = h^s(X) + h^s(Y) - h^s(X, Y), \]

where \( h^s(X, Y) = \lim_{L \to \infty} H(R_{L+1}^s, S_L^s) / L \). Thus, if we introduce the symbolic mutual information rate between \( X \) and \( Y \) by \( i^s(X; Y) = \lim_{L \to \infty} \frac{1}{L} I^s(X_{L+1}; Y_L^L) \), then we have \( i_{s+1}^s(X; Y) = i_{s+1}^s(Y; X) = i^s(X; Y) = i^s(Y; X) \).

Since the symbolic time-delayed mutual information rate is a sum of permutation entropy rates, it holds that for any bivariate finite-alphabet stationary stochastic process \((X, Y)\),

\[ i_{s+1}^s(X; Y) = i_{s+1}^s(Y; X) \]

Hence, we obtain the following two claims: (1) If \((X, Y)\) is ergodic Markov and (12) holds for any \( x_1, x_2 \in A_n \) and \( y_1, y_2 \in A_m \), then we have...
t^i(X|Y) = t(X|Y) = i_{t+1}(X|Y) = i_{t+1}(X; Y), \quad (ii)

If (X, Y) is ergodic Markov, Y is Markov and $\begin{bmatrix} t \\ X \\ Y \end{bmatrix}$ holds for any $x_1, x_2 \in A_n$ and $y_1, y_2 \in A_m$, then we have $t^i(X|Y) = t(X|Y) = i_{t+1}(X; Y) = i_{t+1}(X; Y)$.

V. CONCLUDING REMARKS

In this paper, we proved that the equality between the transfer entropy rate and the symbolic transfer entropy rate holds for any bivariate finite-alphabet stationary ergodic Markov process, which is the first theoretical result on permutation versions of the transfer entropy. We also discussed the relationship between these quantities and the time-delayed mutual information rate and its permutation version.

Next natural question is how we can weaken the condition for [5]. At present, the authors are aware that the equality [5] can be at least extended to any finite-state finite-alphabet hidden Markov process whose state transition matrix is irreducible by almost the same discussion as in the ergodic Markov case. Research results along this line will be presented elsewhere.

We hope that our proof technique based on the duality between $\partial_n$ and $\mu$, which is called the duality between values and orderings in [37], opens up a systematic study on the relationship between the information theoretical quantities and their permutation versions.

ACKNOWLEDGMENT

The authors would like to thank D. Kugiumtzis for his useful comments and discussion. TH was supported by JST PRESTO program.

REFERENCES

[1] T. Schreiber, “Measuring information transfer,” Phys. Rev. Lett., vol. 85, pp. 461–464, 2000.
[2] M. Besserve, B. Schölkopf, N. K. Logothetis, and S. Panzeri, “Causal relationships between frequency bands of extracellular signals in visual cortex revealed by an information theoretic analysis,” J. Comput. Neurosci., vol. 29, pp. 547–566, 2010.
[3] A. Buehlmann and G. Deco, “Optimal information transfer in the cortex through synchronization,” PLoS Comput. Biol., vol. 6, p. e1000934, 2010.
[4] M. Garofalo, T. Nieus, P. Massobrio, and S. Martinoia, “Evaluation of interrelations between spontaneous low-frequency oscillations in cerebral hemodynamics and systemic cardiovascular dynamics,” NeuroImage, vol. 31, pp. 1592–1600, 2006.
[5] C. J. Honey, R. Koetter, M. Breakspear, and O. Sporns, “Network structure of cerebral cortex shapes functional connectivity on multiple time scales,” Proc. Natl. Acad. Sci. U.S.A., vol. 104, pp. 10240–10245, 2007.
[6] T. Katura, N. Tanaka, A. Obata, H. Sato, and A. Maki, “Quantitative evaluation of interrelations between spontaneous low-frequency oscillations in cerebral hemodynamics and systemic cardiovascular dynamics,” NeuroImage, vol. 31, pp. 1592–1600, 2006.
[7] N. Luedtke, N. K. Logothetis, and S. Panzeri, “Testing methodologies for the nonlinear analysis of causal relationships in neurovascular coupling,” Magn. Reson. Imaging, vol. 28, pp. 1113–1119, 2010.
[8] M. Lungarella and O. Sporns, “Mappoing information flow in sensorimotor networks,” PLoS Comput. Biol., vol. 2, p. e144, 2006.
[9] S. A. Neymotin, K. M. Jacobs, A. A. Fenton, and W. Lytton, “Synaptic information transfer in computer models of neocortical columns,” J. Comput. Neurosci., vol. 30, pp. 69–84, 2011.
[10] S. Sabesan, L. B. Good, K. S. Tsakalis, A.Spanias, D. M. Treiman, and L. D. Iasemidis, “Information flow and application to epiphenomenic focus localization from intracranial eeg,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 17, pp. 244–253, 2009.
[40] D. P. Feldman, C. S. McTague, and J. P. Crutchfield, “The organization of intrinsic computation: complexity-entropy diagrams and the diversity of natural information processing,” Chaos, vol. 18, p. 043106, 2008.
[41] R. Shaw, The Dripping Faucet as a Model Chaotic System. Aerial Press, Santa Cruz, California, 1984.
[42] P. Grassberger, “Toward a quantitative theory of self-generated complexity,” Int. J. Theor. Phys., vol. 25, pp. 907–938, 1986.
[43] W. Bialek, I. Nemenman, and N. Tishby, “Predictability, complexity, and learning,” Neural Computation, vol. 13, pp. 2409–2463, 2001.
[44] W. Li, “On the relationship between complexity and entropy for markov chains and regular languages,” Complex Systems, vol. 5, pp. 381–399, 1991.
[45] D. V. Arnold, “Information-theoretic analysis of phase transitions,” Complex Systems, vol. 10, pp. 143–155, 1996.
[46] A. M. Fraser and H. L. Swinney, “Independent coordinates for strange attractors from mutual information,” Phys. Rev. A, vol. 33, pp. 1134–1140, 1986.
[47] K. Kaneko, “Lyapunov analysis and information flow in coupled map lattices,” Physica D, vol. 23, pp. 436–447, 1986.
[48] J. A. Vastano and H. L. Swinney, “Information transport in spatiotemporal systems,” Phys. Rev. Lett., vol. 60, pp. 1773–1776, 1988.