Permutation Complexity via Duality between Values and Orderings

Taichi Haruna[*,1,2], Kohei Nakajima [3]

[1] Department of Earth & Planetary Sciences, Graduate School of Science, Kobe University, 1-1, Rokkodaicho, Nada, Kobe, 657-8501, JAPAN

[2] PRESTO, Japan Science and Technology Agency (JST), 4-1-8 Honcho Kawaguchi, Saitama 332-0012, Japan

[3] Artificial Intelligence Laboratory, Department of Informatics, University of Zurich, Andreasstrasse 15, 8050 Zurich, Switzerland

E-mail: tharuna@penguin.kobe-u.ac.jp (T. Haruna)

Tel & Fax: +81-78-803-5739

Abstract

We study the permutation complexity of finite-state stationary stochastic processes based on a duality between values and orderings between values. First, we establish a duality between the set of all words of a fixed length and the set of all permutations of the same length. Second, on this basis, we give an elementary alternative proof of the equality between the permutation entropy rate and the entropy rate for a finite-state stationary stochastic processes first proved in [Amigó, J.M., Kennel, M. B., Kocarev, L., 2005. *Physica D* 210, 77-95]. Third, we show that further information on the relationship between the structure of values and the structure of orderings for finite-state stationary stochastic processes beyond the entropy rate can be obtained from the established duality. In particular, we prove that the permutation excess entropy is equal to the excess entropy, which is a measure of global correlation present in a stationary stochastic process, for finite-state stationary ergodic Markov processes.

Keywords: Permutation entropy; Excess entropy; Duality; Stationary stochastic processes; Ergodic Markov processes

[*]Corresponding author

1 Introduction

One of the most intriguing recent findings in the science of complexity is that much of the information contained in stationary time series can be captured by orderings between values [1]. Bandt and Pompe [2] first introduced the notion of permutation entropy which quantifies the average uncertainty of orderings between values per time unit, in contrast to the entropy rate for stationary stochastic processes or the Kolmogorov-Sinai entropy for dynamical systems, both of which quantify the average uncertainty of values per time unit. Bandt et al. [3] proved that the permutation entropy is equal to the Kolmogorov-Sinai entropy for piecewise monotone maps on one-dimensional intervals. Amigó et al. [4] showed that the permutation entropy rate is equal to the entropy rate for any finite-state stationary stochastic process [1]. They also generalized the results of [3] to ergodic maps on intervals of arbitrary dimensions by considering the limits of finite-state stationary stochastic processes. Keller and Sinn [5] took a different approach from that of [4] to generalize the results of [3]. The topological permutation entropy was also studied by Bandt et al. [3], Misiurewicz [6] and Amigó and Kennel [7].

In this paper, we study the permutation complexity of finite-state stationary stochastic processes based on a duality between values and orderings between values. Orderings between values induce a coarse-graining of the set of all words of a fixed length. Namely, two words are mapped to the same ordering (permutation) if order-relationships between values in both words are the same. In the case of shift maps on the unit interval, Elizalde [8] performed enumerations associated with such a coarse-graining. In our case, the enumeration is similar, but much simpler than that of [8]. However, we emphasize a dual structure existing between the set of all words of a fixed length and the set of all permutations of the same length. Indeed, we show that there is a kind of minimal realization map from the latter to the former. We can make the pair of the coarse-graining map and the minimal realization map form a Galois connection [9], which is a categorical adjunction [10] between partially ordered sets, by introducing suitable partial orders on the sets at both sides. We present an elementary alternative proof for the equality between the permutation entropy rate and the entropy

[1]Amigó et al. stated that the equality holds for finite-state stationary ergodic processes in Theorem 2 and an inequality holds for the non-ergodic case in Theorem 6 in [4]. However, one can see that they actually proved the equality for any finite-state stationary stochastic process if he or she examine their proof carefully. This point is corrected in Amigó's recent book [1].

rate based on the duality between values and orderings.

We can study the further relationship between the structure of values and the structure of orderings for finite-state stationary stochastic processes beyond the entropy rate equality if we make use of the duality between values and orderings in more depth. Here, we consider the excess entropy which is a measure of global correlation present in finite-state stationary stochastic processes. The excess entropy has an old history in complex systems study [11, 12, 13]. However, it is still of recent research interest. For example, Feldman et al. [14] proposed the entropy-complexity diagrams based on the entropy rate and the excess entropy to analyze various types of natural information processing. We define the permutation excess entropy and show that the permutation excess entropy is equal to the excess entropy for finite-state stationary ergodic Markov processes. We also present a simple non-ergodic counter-example with a strict inequality.

Let us give a rough sketch of our proof strategy for the main results. Let ϕ be the coarse-graining map sending each word of length $L(\geq 1)$ from a finite alphabet to its associated permutation of length L. Given a finite-state stationary stochastic process, only permutations π such that the size of $\phi^{-1}(\pi)$ is greater than 1 may contribute to the difference between the entropy rate and the permutation entropy rate of the process. If we denote the probability that those permutations occur by q_L, then we can show that the difference (≥ 0) before the normalization (division by L) and taking the limit of $L \to \infty$ is bounded from above by the probability q_L multiplied by a function of L whose growth rate is $\log L$ by using the fact that the size of $\phi^{-1}(\pi)$ is given by a binomial coefficient depending on L for any permutation π of length L (Lemma 10). The equality between the entropy rate and the permutation entropy rate is immediate from this bound (Theorem 11). Furthermore, if the process is ergodic Markov, then we can show that q_L diminishes exponentially fast as $L \to \infty$ by using a characterization of words s_1^L such that $\phi^{-1}(\pi) = \{s_1^L\}$ for some π and the irreducibility of the associated transition matrix. This leads to the equality between the excess entropy and the permutation excess entropy (Theorem 14). We note that those words s_1^L such that $\phi^{-1}(\pi) = \{s_1^L\}$ for some π can be seen as a special type of "stable objects" under the duality between the coarse-graining map ϕ and the minimal realization map (Theorem 9 (iii)).

This paper is organized as follows. In Section 2, we establish the duality between values and orderings. In Section 3, we give a proof of the equality between the permutation entropy rate and the entropy rate for finite-state stationary stochastic processes based on the duality. In Section 4, we prove the equality between the permutation excess entropy and the excess entropy for finite-state stationary ergodic Markov processes and give a non-ergodic counter-example with a strict inequality.

2 Duality between Values and Orderings

In this section we establish the duality between values and orderings.

2.1 Permutations and Rank Sequences

Let A be an alphabet. We consider the case that the cardinality $|A|$ of A is finite or countably infinite. If $|A| = n$ $(n = 1, 2, \cdots)$, then we write $A = A_n = \{1, 2, \cdots, n\}$. If $n = \infty$, then $A = A_\infty$ is identified with the set of all natural numbers $\mathbb{N} = \{1, 2, 3, \cdots\}$. We consider A_n $(n = 1, 2, \cdots, \infty)$ is not just a set, but a totally ordered set ordered by the 'less-than-or-equal-to' relationship $\leq$ between natural numbers. In the following discussion, if we write just A, then A can be either A_n or $A_\infty = \mathbb{N}$.

Let $A^L = \underbrace{A \times \cdots \times A}_{L}$ for $L \geq 1$. Each element $w \in A^L$ is called a *word* of length L. If $w = (s_1, \cdots, s_L) \in A^L$, then we write $w = s_1 \cdots s_L = s_1^L$.

Let $\mathcal{S}_L$ be the set of all *permutations* of length L, namely, $\mathcal{S}_L$ is the set of all bijections on the set $\{1, 2, \cdots, L\}$. For $s_1^L \in A^L$ and $\pi \in \mathcal{S}_L$, we say that s_1^L is of type π if we have $s_{\pi(i)} \leq s_{\pi(i+1)}$ and $\pi(i) < \pi(i+1)$ when $s_{\pi(i)} = s_{\pi(i+1)}$ for $i = 1, 2, \cdots, L-1$. For example, $\pi(1)\pi(2)\pi(3)\pi(4)\pi(5) = 24315$ for $s_1^5 = 31213$ because $s_2 s_4 s_3 s_1 s_5 = 11233$.

Each word $s_1^L \in A^L$ has a unique permutation type $\pi \in \mathcal{S}_L$. Hence, the correspondence $s_1^L \mapsto \pi$ defines a many-to-one (in general) map $\phi : A^L \to \mathcal{S}_L$, which coarse-grains the set A^L of words of length L by their permutation types.

We make use of the notion of *rank sequence* introduced in [4]. In some situations, discussions might become facilitated if we use rank sequences instead of permutations. However, as far as the authors are aware, their compatibility with the map ϕ sending words to associated permutations has not been presented explicitly so far. Hence, it may not be worthless to study them here.

A word $r_1^L \in \mathbb{N}^L$ is called a *rank sequence* of length L if it satisfies $1 \leq r_i \leq i$ for $i = 1, \cdots, L$. We denote the set of all rank sequences of length L by $\mathcal{R}_L$. Note that there exists a bijection between $\mathcal{S}_L$ and $\mathcal{R}_L$ because $|\mathcal{S}_L| = L! = |\mathcal{R}_L|$.

Each word $s_1^L \in A^L$ gives rise to a rank sequence $r_1^L \in \mathcal{R}_L$ in the following way:

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

where $\delta(X) = 1$ if the proposition X is true, otherwise $\delta(X) = 0$. Namely, r_i is the number of indices j $(1 \leq j \leq i)$ such that $s_j \leq s_i$. This correspondence $s_1^L \mapsto r_1^L$ defines a map $\varphi : A^L \to \mathcal{R}_L$.

In the following discussion, we will show that there is a bijection $\iota : \mathcal{R}_L \to \mathcal{S}_L$ such that $\iota \circ \varphi = \phi$, namely, the following diagram commutes:

$$\begin{tikzcd} A^L \arrow[r, "\phi"] \arrow[dr, "\varphi"'] & \mathcal{S}_L \\ & \mathcal{R}_L. \arrow[u, "\iota"'] \end{tikzcd}$$

Given a rank sequence $r_1^L \in \mathcal{R}_L$, we define a permutation $\iota(r_1^L) = \pi \in \mathcal{S}_L$ inductively as follows: first, we define $\pi(1) = \max\{i | r_i = 1,\ 1 \leq i \leq L\}$. $\pi(1)$ is well-defined because we have $r_1 = 1$. Second, we define

$$\pi(2) = \max\{i | r_i^{(1)} = \min_{j \neq \pi(1)} \{r_j^{(1)}\}\}$$

where $r_1^{(1)} \cdots r_L^{(1)}$ is a rank sequence defined by

$$r_i^{(1)} = \begin{cases} r_i - 1 & \text{if } i > \pi(1) \\ r_i & \text{otherwise.} \end{cases}$$

In general, we define

$$\pi(k) = \max\{i | r_i^{(k-1)} = \min\{r_j^{(k-1)} | j \neq \pi(1), \cdots, \pi(k-1)\}\}$$

for $k = 2, \cdots, L$, where $r_1^{(k-1)} \cdots r_L^{(k-1)}$ is a rank sequence defined by

$$r_i^{(k-1)} = \begin{cases} r_i^{(k-2)} - 1 & \text{if } i > \pi(k-1) \text{ and } i \neq \pi(1), \cdots, \pi(k-2) \\ r_i^{(k-2)} & \text{otherwise,} \end{cases}$$

and $r_i^{(0)} = r_i$. By construction, this procedure defines a unique permutation $\iota(r_1^L) = \pi \in \mathcal{S}_L$.

For example, consider $r_1^5 = 11342 \in \mathcal{R}_5$. $\pi = \iota(11342) \in \mathcal{S}_5$ is obtained by the following calculations:

$$\begin{aligned}
\pi(1) &= \max\{i|r_i = 1\} = 2,\ r_1^{(1)5} = 11231, \\
\pi(2) &= \max\{i|r_i^{(1)} = \min_{j\neq 2}\{r_j^{(1)}\}\} = 5,\ r_1^{(2)5} = 11231, \\
\pi(3) &= \max\{i|r_i^{(2)} = \min_{j\neq 2,5}\{r_j^{(2)}\}\} = 1,\ r_1^{(3)5} = 11121, \\
\pi(4) &= \max\{i|r_i^{(3)} = \min_{j\neq 1,2,5}\{r_j^{(3)}\}\} = 3,\ r_1^{(4)5} = 11111, \\
\pi(5) &= \max\{i|r_i^{(4)} = \min_{j\neq 1,2,3,5}\{r_j^{(4)}\}\} = 4.
\end{aligned}$$

Lemma 1

$$r_{\pi(k)}^{(k-1)} = 1$$

for $k = 1, 2, \cdots, L$.

Proof. It is sufficient to show that $r_j^{(k-1)} = 1$ for some $j \neq \pi(1), \cdots, \pi(k-2)$. Consider the minimum index j such that $j \notin \{\pi(1), \cdots, \pi(k-2)\}$. Then, we have $r_j^{(k-1)} = r_j - (j-1)$ because $\{1, \cdots, j-1\} \subseteq \{\pi(1), \cdots, \pi(k-2)\}$. However, $1 \leq r_j \leq j$ and $r_j^{(k-1)} \geq 1$ by construction. Hence, $r_j = j$ and we obtain $r_j^{(k-1)} = 1$.

$\square$

Proposition 2 *The map* $\iota : \mathcal{R}_L \to \mathcal{S}_L$ *is a bijection.*

Proof. It is sufficient to show that ι is injective because $|\mathcal{R}_L| = |\mathcal{S}_L| = L! < \infty$. Assume that $\iota(r_1^L) = \iota(\tilde{r}_1^L) = \pi$ for $r_1^L, \tilde{r}_1^L \in \mathcal{R}_L$, $\pi \in \mathcal{S}_L$. We have $r_i^{(L-1)} = \tilde{r}_i^{(L-1)} = 1$ for $i = 1, \cdots, L$ by Lemma 1 because $r_{\pi(k)}^{(k-1)} = r_{\pi(k)}^{(L-1)}$ for $k = 1, \cdots, L$. We can reconstruct both r_1^L and $\tilde{r}_1^L$ from $\overline{r}_1^{(L-1)L} := r_1^{(L-1)L} = \tilde{r}_1^{(L-1)L} = \underbrace{11\cdots 1}_{L}$ by the following procedure: first, we add 1 to the $\pi(L)$-th 1 in $\overline{r}_1^{(L-1)L}$ if $\pi(L) > \pi(L-1)$, and do nothing otherwise. The obtained sequence $\overline{r}_1^{(L-2)L}$ is identical to both $r_1^{(L-2)L}$ and $\tilde{r}_1^{(L-2)L}$ because $\iota(r_1^L) = \iota(\tilde{r}_1^L) = \pi$. Second, we add 1 to $\overline{r}_{\pi(L)}^{(L-2)}$ if $\pi(L) > \pi(L-2)$, and do nothing otherwise, and add 1 to $\overline{r}_{\pi(L-1)}^{(L-2)}$ if $\pi(L-1) > \pi(L-2)$, and do nothing otherwise. If we call the obtained sequence $\overline{r}_1^{(L-3)L}$, then we have $\overline{r}_1^{(L-3)L} = r_1^{(L-3)L} = \tilde{r}_1^{(L-3)L}$. In general, if we define

$$\overline{r}_i^{(L-k)} = \begin{cases} \overline{r}_i^{(L-(k-1))} + 1 & \text{if } i \in \{\pi(L-(k-1)), \cdots, \pi(L)\} \text{ and } i > \pi(L-k) \\ \overline{r}_i^{(L-(k-1))} & \text{otherwise} \end{cases}$$

for $k = 2, \cdots, L$, then we have $\overline{r}_1^{(L-k)L} = r_1^{(L-k)L} = \tilde{r}_1^{(L-k)L}$. In particular, we obtain $\overline{r}_1^{(0)L} = r_1^L = \tilde{r}_1^L$ for $k = L$.

□

Proposition 3 *$\iota \circ \varphi = \phi$.*

Proof. We have to show that $\iota(\varphi(s_1^L)) = \phi(s_1^L)$ for any word $s_1^L \in A^L$. Put $\pi = \phi(s_1^L)$, $\tilde{\pi} = \iota(\varphi(s_1^L))$ and $r_1^L = \varphi(s_1^L)$. We shall show that $\pi(k) = \tilde{\pi}(k)$ for $k = 1, \cdots, L$ inductively. First, we show that $\pi(1) = \tilde{\pi}(1)$. By the definition of ϕ and ι, $\pi(1)$ is the index i of the minimum-leftmost s_i and $\tilde{\pi}(1)$ is the maximum index i such that $r_i = 1$. We have

$$r_i = 1 \Leftrightarrow s_j > s_i \text{ for } j = 1, \cdots, i-1$$

by the definition of rank sequences. Hence, $s_j > s_{\tilde{\pi}(1)}$ for $j = 1, \cdots, \tilde{\pi}(1)-1$. On the other hand, we have $s_{\tilde{\pi}(1)} \leq s_j$ for $j = \tilde{\pi}(1), \cdots, L$. Indeed, if there exists $j > \tilde{\pi}(1)$ such that $s_{\tilde{\pi}(1)} > s_j$, then $r_j > 1$ must hold because $\tilde{\pi}(1)$ is the maximum index i such that $r_i = 1$. Hence, there exists $j_1 < j$ such that $s_{j_1} \leq s_j$. If $j_1 \leq \tilde{\pi}(1)$, then this contradicts $s_k > s_{\tilde{\pi}(1)}$ for $k = 1, \cdots, \tilde{\pi}(1)-1$. So, we have $\tilde{\pi}(1) < j_1 < j$. Since $s_{\tilde{\pi}(1)} > s_j \geq s_{j_1}$, the same argument can be applied to j_1 instead of j. Thus, we obtain a strictly decreasing infinite sequence of indices $j_1 j_2 \cdots$ such that $\tilde{\pi}(1) < \cdots < j_2 < j_1 < j$. However, this is impossible because the number of indices between $\tilde{\pi}(1)$ and j is finite. Therefore, $s_{\tilde{\pi}(1)}$ is the minimum-leftmost value in s_1^L, which implies that $\tilde{\pi}(1) = \pi(1)$.

Now, suppose that $\tilde{\pi}(1) = \pi(1), \cdots, \tilde{\pi}(k) = \pi(k)$, where $1 \leq k \leq L-1$. We would like to show that $\tilde{\pi}(k+1) = \pi(k+1)$. By the definition of ϕ and ι, $\pi(k+1)$ is the index i of the minimum-leftmost s_i except for $s_{\pi(1)}, \cdots, s_{\pi(k)}$ and $\tilde{\pi}(k+1)$ is the maximum index i such that $r_i^{(k)} = 1$ except for $\tilde{\pi}(1), \cdots, \tilde{\pi}(k)$, where we have $\tilde{\pi}(1) = \pi(1), \cdots, \tilde{\pi}(k) = \pi(k)$ by the assumption of the mathematical induction. For an appropriate permutation $(i_1, \cdots, i_k)$ of $(1, \cdots, k)$, we have

$$\pi(i_1) < \cdots < \pi(i_m) < \tilde{\pi}(k+1) < \pi(i_{m+1}) < \cdots < \pi(i_k).$$

It must hold that $r_{\tilde{\pi}(k+1)} = m+1$ because $r^{(k)}_{\tilde{\pi}(k+1)} = 1$. The number of indices j for $j = 1, \cdots, \tilde{\pi}(k+1)-1$ such that $s_j \leq s_{\tilde{\pi}(k+1)}$ is m by the definition of $r_{\tilde{\pi}(k+1)}$. On the other hand, we have $s_{\pi(i_1)}, \cdots, s_{\pi(i_m)} \leq s_{\tilde{\pi}(k+1)}$ by the definition of π. Hence, the equality

$$\{j | s_j \leq s_{\tilde{\pi}(k+1)} \text{ and } 1 \leq j < \tilde{\pi}(k+1)\} = \{\pi(i_1), \cdots, \pi(i_m)\}$$

holds. Thus, if $j \neq \pi(i_1), \cdots, \pi(i_m)$ and $1 \leq j < \tilde{\pi}(k+1)$, then we have $s_j > s_{\tilde{\pi}(k+1)}$. This implies that $\tilde{\pi}(k+1) \leq \pi(k+1)$ because if $\pi(k+1) < \tilde{\pi}(k+1)$, then $s_{\pi(k+1)} > s_{\tilde{\pi}(k+1)}$, which contradicts the assumption that $s_{\pi(k+1)}$ takes the minimum value except for $s_{\pi(1)}, \cdots, s_{\pi(k)}$. For the other inequality, assume that $\pi(i_{m'}) < \pi(k+1) < \pi(i_{m'+1})$. We have $s_j > s_{\pi(k+1)}$ for $j \neq \pi(i_1), \cdots, \pi(i_{m'})$ because $s_{\pi(k+1)}$ takes the minimum-leftmost value except for $s_{\pi(1)}, \cdots, s_{\pi(k)}$. On the other hand, it follows that $s_{\pi(i_1)}, \cdots, s_{\pi(i_{m'})} \leq s_{\pi(k+1)}$ by the definition of π. Hence, we have $r_{\pi(k+1)} = \sum_{j=1}^{\pi(k+1)} \delta(s_j \leq s_{\pi(k+1)}) = m'+1$, which implies that $r^{(k)}_{\pi(k+1)} = 1$. Thus, we obtain $\pi(k+1) \leq \tilde{\pi}(k+1)$ because $\tilde{\pi}(k+1)$ is the maximum index i such that $r^{(k)}_i = 1$ except for $\pi(1), \cdots, \pi(k)$.

□

Corollary 4 *For $s_1^L, t_1^L \in A^L$, the following statements are equivalent:*

(i) $\phi(s_1^L) = \phi(t_1^L)$.

(ii) For all $1 \leq j \leq k \leq L$, $s_k \leq s_j \Leftrightarrow t_k \leq t_j$.

Proof. Assume that $\phi(s_1^L) = \phi(t_1^L) = \pi \in \mathcal{S}_L$. Then, we have

$$\begin{aligned} s_{\pi(1)} \leq s_{\pi(2)} \leq \cdots \leq s_{\pi(L)} \text{ and} \\ t_{\pi(1)} \leq t_{\pi(2)} \leq \cdots \leq t_{\pi(L)}. \end{aligned}$$

Hence, (ii) holds. For the reverse direction, assume that (ii) holds. Then, we have $\sum_{k=1}^{j} \delta(s_k \leq s_j) = \sum_{k=1}^{j} \delta(t_k \leq t_j)$ for any $1 \leq j \leq L$, which implies $\varphi(s_1^L) = \varphi(t_1^L)$. Hence, we have $\phi(s_1^L) = \iota \circ \varphi(s_1^L) = \iota \circ \varphi(t_1^L) = \phi(t_1^L)$.

□

2.2 The Coarse-Graining Map ϕ

Now, we are ready to study properties of the coarse-graining map $\phi : A^L \to \mathcal{S}_L$ in detail.

Lemma 5 *Let $\pi \in \mathcal{S}_L$. Assume that there is no $s_1^L \in A_{i-1}^L$ such that $\phi(s_1^L) = \pi$, but there exists $s_1^L \in A_i^L$ such that $\phi(s_1^L) = \pi$ for some $i \geq 1$ (when $i = 1$ we define $A_{i-1} = A_0 = \emptyset$).*

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

(ii) $|\phi^{-1}(\pi)| = \binom{L+n-i}{n-i}$, where $n \geq i$ and the domain of ϕ is set to A_n^L.

Proof. (i): First, we prove the uniqueness. If $i = 1$, then we have nothing to do. So, we assume that $i \geq 2$. Suppose that $\phi(s_1^L) = \phi(t_1^L) = \pi$ and $s_1^L, t_1^L \in A_i^L$. If $s_1^L \neq t_1^L$, then there exists j such that $s_{\pi(j)} \neq t_{\pi(j)}$. We can assume that $s_{\pi(j)} < t_{\pi(j)}$ without loss of generality. Let us define a word u_1^L by

$$u_{\pi(k)} = \begin{cases} s_{\pi(k)} & k = 1, \cdots, j-1, \\ t_{\pi(k)} - 1 & k = j, \cdots, L. \end{cases}$$

We claim that $\phi(u_1^L) = \pi$. Indeed, it is clear that we have $u_{\pi(k-1)} \leq u_{\pi(k)}$ and $\pi(k-1) < \pi(k)$ when $u_{\pi(k-1)} = u_{\pi(k)}$, for $k \neq j$. When $k = j$, we have $s_{\pi(j-1)} \leq s_{\pi(j)} \leq t_{\pi(j)} - 1$ by the assumption. Suppose that $s_{\pi(j-1)} = t_{\pi(j)} - 1$. It follows that $s_{\pi(j-1)} = s_{\pi(j)}$, which implies that $\pi(j-1) < \pi(j)$. Thus, we have $\phi(u_1^L) = \pi$. However, this contradicts the assumption that there is no $s_1^L \in A_{i-1}^L$ such that $\phi(s_1^L) = \pi$ because $u_1^L \in A_{i-1}^L$.

Next, suppose that $\phi(t_1^L) = \pi$ for $t_1^L \in A_n^L$, $n \geq i$. Let us show that $s_{\pi(k)} \leq t_{\pi(k)}$ for $k = 1, \cdots, L$. If $i = 1$, then we have nothing to do because $s_{\pi(k)} = 1$ for all k. So, we assume that $i \geq 2$. If there exists j such that $s_{\pi(j)} > t_{\pi(j)}$, then a word u_1^L defined by

$$u_{\pi(k)} = \begin{cases} t_{\pi(k)} & k = 1, \cdots, j-1, \\ s_{\pi(k)} - 1 & k = j, \cdots, L. \end{cases}$$

satisfies $\phi(u_1^L) = \pi$ and $u_1^L \in A_{i-1}^L$ by the same reason in the proof of the uniqueness, which violates the assumption that there is no $s_1^L \in A_{i-1}^L$ such that $\phi(s_1^L) = \pi$. Hence, if we define $c_k = t_k - s_k$ for $k = 1, \cdots, L$, then $c_k \geq 0$ and $c_{\pi(L)} \leq n - i$ because $t_{\pi(L)} \leq n$ and $s_{\pi(L)} = i$. The remaining task for us is to show that $c_{\pi(k)} \leq c_{\pi(k+1)}$ for $k = 1, \cdots, L-1$. If $i = 1$, then $c_{\pi(k)} = t_{\pi(k)} - 1 \leq t_{\pi(k+1)} - 1 = c_{\pi(k+1)}$ for $k = 1, \cdots, L-1$. Suppose that $i \geq 2$ and $c_{\pi(j+1)} < c_{\pi(j)}$ for some j. Then, we have

$$\begin{aligned} c_{\pi(j+1)} < c_{\pi(j)} \quad &\Leftrightarrow \quad t_{\pi(j+1)} - s_{\pi(j+1)} < t_{\pi(j)} - s_{\pi(j)} \\ &\Leftrightarrow \quad s_{\pi(j)} + \left(t_{\pi(j+1)} - t_{\pi(j)}\right) < s_{\pi(j+1)}. \end{aligned}$$

This implies that

$$s_{\pi(j)} \leq s_{\pi(j)} + \left(t_{\pi(j+1)} - t_{\pi(j)}\right) \leq s_{\pi(j+1)} - 1 \tag{1}$$

because $t_{\pi(j+1)} \geq t_{\pi(j)}$. Let us introduce a word u_1^L by

$$u_{\pi(k)} = \begin{cases} s_{\pi(k)} & k = 1, \cdots, j, \\ s_{\pi(k)} - 1 & k = j+1, \cdots, L. \end{cases}$$

We claim that $\phi(u_1^L) = \pi$ and $u_1^L \in A_{i-1}^L$, which contradicts the assumption that there is no $s_1^L \in A_{i-1}^L$ such that $\phi(s_1^L) = \pi$. We only need to show that $\pi(j) < \pi(j+1)$ when $u_{\pi(j)} = u_{\pi(j+1)}$. However, by (1), if $s_{\pi(j)} = s_{\pi(j+1)} - 1$, then we have $t_{\pi(j)} = t_{\pi(j+1)}$, which implies that $\pi(j) < \pi(j+1)$.

(ii): The number of sequences $c_1 \cdots c_L$ satisfying $0 \leq c_{\pi(1)} \leq \cdots \leq c_{\pi(L)} \leq n - i$ is given by a binomial coefficient $\binom{L+n-i}{n-i}$. Hence, we have $|\phi^{-1}(\pi)| \leq \binom{L+n-i}{n-i}$ by (i). On the other hand, given a sequence $c_1 \cdots c_L$ such that $0 \leq c_{\pi(1)} \leq c_{\pi(2)} \leq \cdots \leq c_{\pi(L)} \leq n - i$, $t_1^L \in A_n^L$ defined by $t_k = s_k + c_k$ for $k = 1, \cdots, L$ clearly satisfies $\phi(t_1^L) = \pi$. Hence, we have $|\phi^{-1}(\pi)| \geq \binom{L+n-i}{n-i}$.

□

If there is no word $s_1^L \in A_{i-1}^L$ such that $\phi(s_1^L) = \pi$, but there exists a (unique) word $s_1^L \in A_i^L$ such that $\phi(s_1^L) = \pi$ for $i \geq 1$, then we say that π *appears for the first time at* i. We denote the number of permutations $\pi \in \mathcal{S}_L$ that appear for the first time at i by $\nu(i, L)$. By Lemma 5, we have $\nu(1, L) = 1$ and

$$\nu(n, L) = n^L - \sum_{i=1}^{n-1} \binom{L+n-i}{n-i} \nu(i, L) \tag{2}$$

for $n \geq 2$.

The following Proposition 6 and the subsequent paragraph in this subsection are only for the record. They will not be used in later sections. So, readers who are interested in only the main results of this paper can skip them.

Proposition 6 *A closed-form expression for $\nu(n,L)$ is given by the following formula:*

$$\nu(n,L)=\sum_{i=0}^{n-1}(-1)^i\binom{L+1}{i}(n-i)^L. \tag{3}$$

Proof. We prove the formula by mathematical induction on n. if $n=1$, then we have $\nu(1,L)=1$. Assume that the formula holds for natural numbers $1,2,\cdots,n$. Then, we have

$$\begin{aligned}
\nu(n+1,L) &= (n+1)^L-\sum_{i=1}^{n}\binom{L+n+1-i}{n+1-i}\nu(i,L)\\
&= (n+1)^L-\sum_{i=1}^{n}\binom{L+n+1-i}{n+1-i}\sum_{k=0}^{i-1}(-1)^k\binom{L+1}{k}(i-k)^L\\
&= (n+1)^L+\sum_{j=1}^{n}\left(\sum_{\substack{i-k=j,\\ 1\le i\le n}}(-1)^{k+1}\binom{L+n+1-i}{n+1-i}\binom{L+1}{k}\right)j^L.
\end{aligned}$$

It is enough to show that

$$\sum_{\substack{i-k=j,\\ 1\le i\le n}}(-1)^{k+1}\binom{L+n+1-i}{n+1-i}\binom{L+1}{k}=(-1)^{n+1-j}\binom{L+1}{n+1-j}$$

for $j=1,\cdots,n$. If we put $l=n-j$, then this is equivalent to showing that

$$\sum_{k=0}^{l}(-1)^{l-k}\binom{L+1+l-k}{L}\binom{L+1}{k}=\binom{L+1}{l+1} \tag{4}$$

for $l=0,1,\cdots,n-1$. Consider the equality

$$(1+x)^{-(L+1)}(1+x)^{L+1}=1 \tag{5}$$

which holds for $|x| < 1$. The left-hand side of (5) can be written as

$$\left(\sum_{p=0}^{\infty}(-1)^p\binom{L+p}{L}x^p\right)\left(\sum_{q=0}^{L+1}\binom{L+1}{q}x^q\right).$$

If we compare the coefficient of x^{l+1} for $l = 0, 1, \cdots$ in both sides of the equality (5), then we obtain

$$\sum_{p+q=l+1}(-1)^p\binom{L+p}{L}\binom{L+1}{q} = 0.$$

After a few algebras, we can derive the desired equality (4).

□

Note that (3) is identical to a closed-form expression for the *Eulerian number* $\left\langle {L \atop n-1} \right\rangle$ [15], where the Eulerian number $\left\langle {a \atop b} \right\rangle$ is the number of permutations π of $\{1, \cdots, a\}$ that have exactly b ascents, namely, b places with $\pi(j) < \pi(j+1)$. The equality (2) is equivalent to the so-called *Worpitzky's identity*:

$$n^L = \sum_{k=L-n}^{L-1}\left\langle {L \atop k} \right\rangle\binom{n+k}{L}. \tag{6}$$

Indeed, one can obtain the Worpitzky's identity (6) from (2) by a few algebras using the symmetry law $\left\langle {L \atop i-1} \right\rangle = \left\langle {L \atop L-i} \right\rangle$.

2.3 The Minimal Realization Map μ

For any $\pi \in \mathcal{S}_L$, we can construct a word $s_1^L \in \mathbb{N}^L$ such that $\phi(s_1^L) = \pi$ in the following procedure: first, we decompose the sequence $\pi(1)\cdots\pi(L)$ into *maximal ascending sequences*. A subsequence $i_j \cdots i_{j+k}$ of a sequence $i_1 \cdots i_L$ is called a *maximal ascending sequence* if it is ascending, namely, $i_j \leq i_{j+1} \leq \cdots \leq i_{j+k}$, and neither $i_{j-1}i_j \cdots i_{j+k}$ nor $i_j \cdots i_{j+k}i_{j+k+1}$ is ascending. Suppose $\pi(1)\cdots\pi(i_1)$, $\pi(i_1+1)\cdots\pi(i_2), \cdots, \pi(i_{k-1}+1)\cdots\pi(L)$ is a decomposition of $\pi(1)\cdots\pi(L)$ into maximal ascending sequences. If we define a word $s_1^L \in \mathbb{N}^L$ by

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

then we have $\phi(s_1^L) = \pi$ by construction. Thus, π appears for the first time at most k. We denote the word s_1^L by $\mu(\pi)$. μ defines a map $\mu : \mathcal{S}_L \to \mathbb{N}^L$ such that $\phi \circ \mu(\pi) = \pi$ for any $\pi \in \mathcal{S}_L$.

For example, if $\pi \in \mathcal{S}_5$ is given by $\pi(1)\pi(2)\pi(3)\pi(4)\pi(5) = 24315$, then its decomposition into maximal ascending sequences is $24, 3, 15$. If we put $s_2s_4s_3s_1s_5 = 11233$, then we obtain $\mu(\pi) = s_1s_2s_3s_4s_5 = 31213$.

Let $\pi \in \mathcal{S}_L$ appear for the first time at n. By Lemma 5, there exists a unique word $s_1^L \in A_n^L$ such that $\phi(s_1^L) = \pi$. We say that s_1^L is a *minimal realization* of π. In the following, we shall show that $\mu(\pi)$ is a minimal realization of π.

Proposition 7 *The following statements are equivalent:*

(i) $s_1^L \in A_n^L$ *is a minimal realization of some permutation* $\pi \in \mathcal{S}_L$ *that appears for the first time at* n.

(ii) For any $1 \leq i \leq n-1$, *there exists* $1 \leq j < k \leq L$ *such that* $s_j = i+1,\ s_k = i$.

Proof. When $n = 1$, the equivalence is trivial. So, we assume that $n \geq 2$ in the following discussion.

(i)$\Rightarrow$(ii): Let $s_1^L \in A_n^L$ be a minimal realization of $\pi \in \mathcal{S}_L$ that appears for the first time at n. Suppose that statement (ii) does not hold. Then, there exists $1 \leq i \leq n-1$ such that, for any $1 \leq j, k \leq n$, if $s_k = i$ and $s_j = i+1$, then $k < j$. Let us define a word t_1^L by

$$t_j = \begin{cases} s_j - 1 & \text{if } s_j = i+1, \\ s_j & \text{otherwise.} \end{cases}$$

We claim that $\phi(t_1^L) = \pi$. By Corollary 4, it is sufficient to show that $s_k \leq s_j \Leftrightarrow t_k \leq t_j$ for all $1 \leq k \leq j \leq L$. Fix $1 \leq k \leq j \leq L$. Assume that $s_k \leq s_j$. If $s_j = i+1$, then we have $t_j = s_j - 1 = i$. If we also have $s_k = i+1$, then $t_k = s_k - 1 = i = t_j$. Otherwise, we have $s_k \neq i+1$. Thus, we obtain $s_k \leq i$ because $s_k \leq s_j = i+1$. Then, $t_k = s_k \leq i = t_j$. On the other hand, if $s_j \neq i+1$, then we have $t_j = s_j$. Thus, we obtain $t_k \leq s_k \leq s_j = t_j$. To show the reverse direction, let us assume that $t_k \leq t_j$. If $s_j = i+1$, then $t_j = s_j - 1 = i$. If we also have $s_k = i+1$, then $s_k = s_j$. Otherwise, we have $s_k \neq i+1$, then $t_k = s_k$ so that $s_k = t_k \leq t_j = i < s_j$. On the other hand, if $s_j \neq i+1$, then we have $t_j = s_j$. If we also have $s_k \neq i+1$, then $s_k = t_k \leq t_j = s_j$. Otherwise, we have $s_k = i+1$, then $t_k = s_k - 1$. Suppose that $s_k > s_j$. Then, $s_j < s_k = i+1$ and $i = s_k - 1 = t_k \leq t_j = s_j$. Hence, $s_j = i$. Since we have assumed that (ii) does not hold, we obtain $j < k$. However, this contradicts our other assumption that $k \leq j$. Hence, we have $s_k \leq s_j$.

Suppose that there exists j such that $s_j = i+1$. Then, we have $t_1^L \neq s_1^L$. This contradicts the uniqueness of minimal realization of π because both s_1^L and t_1^L are contained in A_n^L. Suppose that there exists no j such that $s_j = i+1$. Since π appears for the first time at n and s_1^L is its minimal realization, we have $s_{\pi(L)} = n$. Hence, $i+1 < n$ should hold. Let us take the least j such that $i+1 < s_{\pi(j)}$ and put it as j_0. If we define a word t_1^L by

$$t_{\pi(j)} = \begin{cases} s_{\pi(j)} & \text{if } j < j_0, \\ s_{\pi(j)} - 1 & \text{if } j \geq j_0, \end{cases}$$

then we have $\phi(t_1^L) = \pi$. Indeed, $t_{\pi(j_0-1)} = s_{\pi(j_0-1)} < i+1 \leq s_{\pi(j_0)} - 1 = t_{\pi(j_0)}$ because $i+1 < s_{\pi(j_0)}$. On the other hand, we have $t_1^L \in A_{n-1}^L$, which is a contradiction.

(ii)$\Rightarrow$(i): Assume that $s_1^L \in A_n^L$ satisfies (ii). Let $t_1^L \in A_{n-i}^L$ be a minimal realization of $\pi = \phi(s_1^L)$. We shall show that $i = 0$. By Lemma 5, we have $t_{\pi(k)} \leq s_{\pi(k)}$ for $k = 1, \cdots, L$ and $0 \leq c_{\pi(1)} \leq \cdots \leq c_{\pi(L)} = n - (n-i) = i$ for $c_k = s_k - t_k$. Suppose there exists j such that $1 \leq c_{\pi(j)}$. Take the least j such that $1 \leq c_{\pi(j)}$ and put it j_0. Now, consider the least k such that $s_{\pi(k)} = s_{\pi(j_0)}$ and the largest k' such that $s_{\pi(k')} = s_{\pi(j_0)}$ and put them as k_0 and k_1, respectively. Then, we have $t_{\pi(k_0)} = t_{\pi(k_1)}$. Indeed, $t_{\pi(k_0)} = s_{\pi(k_0)} - c_{\pi(k_0)} = s_{\pi(k_1)} - c_{\pi(k_0)} \geq s_{\pi(k_1)} - c_{\pi(k_1)} = t_{\pi(k_1)}$. On the other hand, $t_{\pi(k_0)} \leq t_{\pi(k_1)}$ because $k_0 \leq k_1$. Thus, we obtain $t_{\pi(k_0)} = t_{\pi(k_1)}$. This means that $c_{\pi(k_0)} = c_{\pi(k_1)}$, which, in turn, implies $c_{\pi(k)} = c_{\pi(j_0)}$ for all $k_0 \leq k \leq k_1$. (Thus, $j_0 = k_0$.) If we define a word u_1^L by

$$u_{\pi(k)} = \begin{cases} s_{\pi(k)} - 1 & \text{if } k_0 \leq k \leq k_1, \\ s_{\pi(k)} & \text{otherwise}, \end{cases}$$

then we have $\phi(u_1^L) = \pi$. To show this, we should care for only $k = k_0 - 1, k_0$ and $k = k_1, k_1 + 1$. First, let us consider the former. By the definition of u_1^L and k_0, we have $u_{\pi(k_0-1)} = s_{\pi(k_0-1)} = t_{\pi(k_0-1)}$. We also have $u_{\pi(k_0)} = s_{\pi(k_0)} - 1 \geq t_{\pi(k_0)}$ because $s_{\pi(k_0)} - t_{\pi(k_0)} = c_{\pi(k_0)} \geq 1$. Hence, $u_{\pi(k_0-1)} = t_{\pi(k_0-1)} \leq t_{\pi(k_0)} \leq u_{\pi(k_0)}$. If $u_{\pi(k_0-1)} = u_{\pi(k_0)}$, then $t_{\pi(k_0-1)} = t_{\pi(k_0)}$, which implies that $\pi(k_0 - 1) < \pi(k_0)$. The latter is obvious because $u_{\pi(k_1)} = s_{\pi(k_1)} - 1 < s_{\pi(k_1+1)} = u_{\pi(k_1+1)}$.

Now, if we put $s_{\pi(j_0)} = a(\geq 2)$, then there exist $j_1 < j_2$ such that $s_{j_1} = a$ and $s_{j_2} = a-1$ by (ii). By the construction of u_1^L, we have $u_{j_1} = a-1 = u_{j_2}$. This implies that u_1^L and s_1^L have different rank sequences because $\varphi(u_1^L)_{j_2} > \varphi(s_1^L)_{j_2}$. Thus, we have $\phi(u_1^L) = \iota \circ \varphi(u_1^L) \neq \iota \circ \varphi(s_1^L) = \phi(s_1^L)$, which is a contradiction.

$\square$

Corollary 8 *For $\pi \in \mathcal{S}_L$, $\mu(\pi)$ is a minimal realization of π.*

Proof. Let

$$\pi(1)\cdots\pi(j_1),\ \pi(j_1+1)\cdots\pi(j_2),\cdots,\pi(j_{k-1}+1)\cdots\pi(L)$$

be a decomposition of $\pi(1)\cdots\pi(L)$ into maximal ascending sequences. If $s_1^L = \mu(\pi)$, then

$$\begin{aligned} s_{\pi(1)} = \cdots = s_{\pi(j_1)} = 1, s_{\pi(j_1+1)} = \cdots = s_{\pi(j_2)} = 2, \\ \cdots, s_{\pi(j_{k-1})+1} = \cdots = s_{\pi(L)} = k \end{aligned}$$

by the definition of μ. For each $1 \leq i \leq k-1$, we have $s_{\pi(j_i)} = i$, $s_{\pi(j_i+1)} = i+1$ and $\pi(j_i) > \pi(j_i+1)$. Hence, condition (ii) of Proposition 7 is satisfied by s_1^L. Since $\phi(s_1^L) = \pi$, s_1^L is a minimal realization of π.

$\square$

2.4 The Duality

We can make the pair of maps

$$\mathbb{N}^L \underset{\mu}{\overset{\phi}{\rightleftarrows}} \mathcal{S}_L$$

form a Galois connection [9] in the following way: we consider the set $\mathcal{S}_L$ as an ordered set with the discrete order, namely, we define an order relation $\leq_{\mathcal{S}_L}$ on $\mathcal{S}_L$ by $\pi \leq_{\mathcal{S}_L} \pi' :\Leftrightarrow \pi = \pi'$. On the other hand, we introduce an order relation $\leq_{\mathbb{N}^L}$ on $\mathbb{N}^L$ by $s_1^L \leq_{\mathbb{N}^L} t_1^L :\Leftrightarrow \phi(s_1^L) = \phi(t_1^L) =: \pi$ and there exist $0 \leq c_{\pi(1)} \leq \cdots \leq c_{\pi(L)}$ such that $s_k = c_k + t_k$. By Corollary 8, we have

$$\phi(s_1^L) \leq_{\mathcal{S}_L} \pi \Leftrightarrow s_1^L \leq_{\mathbb{N}^L} \mu(\pi)$$

for $s_1^L \in \mathbb{N}^L$ and $\pi \in \mathcal{S}_L$.

If we restrict the domain of the map ϕ to A_n^L, we obtain the following form of the duality stated in Theorem 9 (iv) bellow. Theorem 9 summarizes the main results of this section.

Theorem 9 *Let us set the domain of the coarse-graining map ϕ to A_n^L.*

(i) For $\pi \in \mathcal{S}_L$, if $\phi^{-1}(\pi) \neq \emptyset$, then the value of $|\phi^{-1}(\pi)|$ takes a binomial coefficient $\binom{L+n-i}{n-i}$ for some $1 \leq i \leq n$.

(ii) For $\pi \in \mathcal{S}_L$, the following two statements are equivalent:

(a) $|\phi^{-1}(\pi)| = 1$.

(b) π appears for the first time at n.

(iii) For $s_1^L \in A_n^L$, the following three statements are equivalent:

(c) $\phi^{-1}(\pi) = \{s_1^L\}$ for some $\pi \in \mathcal{S}_L$.

(d) For any $1 \leq i \leq n-1$ there exists $1 \leq j < k \leq L$ such that $s_j = i+1, s_k = i$.

(e) $s_1^L \notin A_{n-1}^L$ and $s_1^L = \mu \circ \phi(s_1^L)$.

(iv) If we restrict ϕ on the subset of A_n^L consisting of words satisfying one of the three equivalent conditions in (iii), then ϕ gives a one-to-one correspondence between these words and permutations of length L satisfying one of the two equivalent conditions in (ii) with its inverse μ.

Proof. (i) If π appears for the first time at $i \leq n$, then $|\phi^{-1}(\pi)| = \binom{L+n-i}{n-i}$ by Lemma 5 (ii).

(ii) (a)$\Rightarrow$(b): Suppose $|\phi^{-1}(\pi)| = 1$ and π appears for the first time at $i \leq n$. By (i), $\binom{L+n-i}{n-i} = 1$ holds. This happens if and only if $i = n$.

(b)$\Rightarrow$(a): If π appears for the first time at n, then there exists a unique $s_1^L \in A_n^L$ such that $\phi(s_1^L) = \pi$. Hence, $\phi^{-1}(\pi) = \{s_1^L\}$.

(iii) (c)$\Rightarrow$(d),(e): Assume $\phi^{-1}(\pi) = \{s_1^L\}$ for some $\pi \in \mathcal{S}_L$. By (ii), π appears for the first time at n. Hence, s_1^L is a minimal realization of π. Hence, (d) holds for s_1^L by Proposition 7. To see (e) holds, first observe that s_1^L cannot be contained in A_{n-1}^L. We also have $s_1^L = \mu(\pi) = \mu(\phi(s_1^L))$ because $\mu(\pi)$ is a minimal realization of π by Corollary 8.

(d)$\Rightarrow$(c): If (d) holds for s_1^L, then s_1^L is a minimal realization of some $\pi \in \mathcal{S}_L$ that appears for the first time at n by Proposition 7. Hence, we have $\phi^{-1}(\pi) = \{s_1^L\}$ by the uniqueness of minimal realization.

(e)⇒(c): Assume $s_1^L \notin A_{n-1}^L$ and $s_1^L = \mu(\phi(s_1^L))$. s_1^L is a minimal realization of $\phi(s_1^L)$ by Corollary 8. $\phi(s_1^L)$ appears for the first time at n since $s_1^L \notin A_{n-1}^L$. Hence, $\phi^{-1}(\phi(s_1^L)) = \{s_1^L\}$ holds by (ii).

(iv) Let us put $X = \{s_1^L \in A_n^L | s_1^L \notin A_{n-1}^L,\ s_1^L = \mu \circ \phi(s_1^L)\}$ and $Y = \{\pi \in \mathcal{S}_L || \phi^{-1}(\pi)| = 1\}$. If $s_1^L \in X$, then $\phi^{-1}(\phi(s_1^L)) = \{s_1^L\}$. Hence, ϕ restricted on X is a map from X into Y. On the other hand, μ restricted on Y is a map from Y into X. Indeed, π appears for the first time at n by (ii). Since $\mu(\pi)$ is a minimal realization of π by Corollary 8, it must hold that $\phi^{-1}(\pi) = \{\mu(\pi)\}$. Thus, we have $\mu(\pi) \notin A_{n-1}^L$ and $\mu(\pi) \in A_n^L$. We also have $\mu(\pi) = \mu \circ \phi \circ \mu(\pi)$ because $\phi \circ \mu$ is an identity on $\mathcal{S}_L$. Now, μ restricted on Y is a left inverse of ϕ restricted on X by the definition of X. It is also a right inverse because $\phi \circ \mu$ is an identity on $\mathcal{S}_L$. □

3 Permutation Entropy Rate Revisited

Let $\mathbf{S} = \{S_1, S_2, \cdots\}$ be a finite-state stationary stochastic process, where stochastic variables S_i take their values in A_n. Stationarity means that

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

for any $k, L \geq 1$ and $s_1, \cdots, s_L \in A_n$. For simplicity, we write $p(s_1^L) = p(s_1 \cdots s_L)$ instead of $\Pr\{S_1 = s_1, \cdots, S_L = s_L\}$. In the following discussion, we set the domain of the map ϕ introduced in Section 2 to A_n^L.

The *entropy rate* $h(\mathbf{S})$ of a finite-state stationary stochastic process $\mathbf{S} = \{S_1, S_2, \cdots\}$ is defined by

$$h(\mathbf{S}) = \lim_{L \to \infty} \frac{1}{L} H(S_1^L), \tag{7}$$

where $H(S_1^L) = H(S_1, \cdots, S_L) = -\sum_{s_1^L \in A_n^L} p(s_1^L) \log p(s_1^L)$. Here, we take the base of the logarithm as 2. It is well-known that the limit exists for any finite-state stationary stochastic process [16].

The *permutation entropy rate* $h^*(\mathbf{S})$ of a finite-state stationary stochastic process $\mathbf{S} = \{S_1, S_2, \cdots\}$ is defined by

$$h^*(\mathbf{S}) = \lim_{L \to \infty} \frac{1}{L} H^*(S_1^L), \tag{8}$$

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

$\pi \in \mathcal{S}_L$. Amigó et al. proved that the limit exists for all finite-state stationary stochastic processes and is equal to $h(\mathbf{S})$ [1, 4]. They first showed the equality with the assumption of the ergodicity. Then, they proceeded to the general case by appealing to the ergodic decomposition theorem of the entropy rate.

If we make use of *rank variables* $R_i = \sum_{j=1}^{n} \delta(S_j \le S_i)$ for $i = 1, 2, \cdots$ introduced in [4], then the permutation entropy rate has the following alternative expression by Proposition 2 and Proposition 3:

$$h^*(\mathbf{S}) = \lim_{L \to \infty} \frac{1}{L} H(R_1^L).$$

Intuitively, the entropy rate quantifies the uncertainty of values per unit symbol on the one hand, while the permutation entropy rate quantifies the uncertainty of orderings between values per unit symbol on the other hand.

In the following discussion, we give an elementary alternative proof of $h(\mathbf{S}) = h^*(\mathbf{S})$ for a finite-state stationary stochastic process $\mathbf{S} = \{S_1, S_2, \cdots\}$ based on the duality between values and orderings established in Section 2.

Lemma 10

$$0 \le H(S_1^L) - H^*(S_1^L) \le \left(\sum_{\substack{\pi \in \mathcal{S}_L, \\ |\phi^{-1}(\pi)| > 1}} p(\pi) \right) n \log(L + n). \tag{9}$$

Proof.

$$\begin{aligned} H(S_1^L) - H^*(S_1^L) &= - \sum_{s_1^L \in A_n^L} p(s_1^L) \log p(s_1^L) + \sum_{\pi \in \mathcal{S}_L} p(\pi) \log p(\pi) \\ &= \sum_{\pi \in \mathcal{S}_L} \left(- \sum_{s_1^L \in \phi^{-1}(\pi)} p(s_1^L) \log p(s_1^L) + \left(\sum_{s_1^L \in \phi^{-1}(\pi)} p(s_1^L) \right) \log p(\pi) \right) \\ &= \sum_{\substack{\pi \in \mathcal{S}_L, \\ p(\pi) > 0}} \left(- \sum_{s_1^L \in \phi^{-1}(\pi)} p(s_1^L) \log \frac{p(s_1^L)}{p(\pi)} \right) \\ &= \sum_{\substack{\pi \in \mathcal{S}_L, \\ p(\pi) > 0}} p(\pi) \left(- \sum_{s_1^L \in \phi^{-1}(\pi)} \frac{p(s_1^L)}{p(\pi)} \log \frac{p(s_1^L)}{p(\pi)} \right). \end{aligned}$$

Now, we have

$$0 \leq -\sum_{s_1^L \in \phi^{-1}(\pi)} \frac{p(s_1^L)}{p(\pi)} \log \frac{p(s_1^L)}{p(\pi)} \leq n \log(L+n)$$

for $\pi \in \mathcal{S}_L$ such that $\phi^{-1}(\pi) \neq \emptyset$ and $p(\pi) > 0$ because the value of $|\phi^{-1}(\pi)|$ takes a binomial coefficient $\binom{L+n-i}{n-i}$ for some $1 \leq i \leq n$ by Theorem 9 (i). Note that if $i = n$, then $|\phi^{-1}(\pi)| = 1$, which implies

$$-\sum_{s_1^L \in \phi^{-1}(\pi)} \frac{p(s_1^L)}{p(\pi)} \log \frac{p(s_1^L)}{p(\pi)} = 0.$$

□

Theorem 11 *For any finite-state stationary stochastic process* $\mathbf{S} = \{S_1, S_2, \cdots\}$, $h(\mathbf{S}) = h^*(\mathbf{S})$.

Proof. Since we have

$$\sum_{\substack{\pi \in \mathcal{S}_L, \\ |\phi^{-1}(\pi)| > 1}} p(\pi) \leq 1 \text{ and } \frac{\log(L+n)}{L} \underset{L\to\infty}{\to} 0,$$

we obtain

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

by Lemma 10.

□

4 Permutation Excess Entropy

The *excess entropy* [17] $\mathbf{E}(\mathbf{S})$ of a finite-state stationary stochastic process $\mathbf{S} = \{S_1, S_2, \cdots\}$ is defined by

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

if the limit on the right-hand side exists. The excess entropy $\mathbf{E}(\mathbf{S})$ is a measure of global correlation present in a finite-state stationary stochastic process $\mathbf{S} = \{S_1, S_2, \cdots\}$. If $\mathbf{E}(\mathbf{S})$ exists, then we can write [17]

$$\mathbf{E}(\mathbf{S}) = \sum_{L=1}^{\infty} \Big(H(S_L | S_1^{L-1}) - h(\mathbf{S}) \Big) = \lim_{L\to\infty} I(S_1^L; S_{L+1}^{2L}), \tag{11}$$

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

The *permutation excess entropy* $\mathbf{E}^*(\mathbf{S})$ of a finite-state stationary stochastic process $\mathbf{S} = \{S_1, S_2, \cdots\}$ is defined by

$$\mathbf{E}^*(\mathbf{S}) = \lim_{L\to\infty} \left(H^*(S_1^L) - h^*(\mathbf{S})L \right), \tag{12}$$

if the limit on the right-hand side exists.

It is straightforward to obtain a similar alternative expression for the permutation excess entropy $\mathbf{E}^*(\mathbf{S})$ to that for the excess entropy (11), when $\mathbf{E}^*(\mathbf{S})$ exists:

$$\mathbf{E}^*(\mathbf{S}) = \sum_{L=1}^{\infty} \Big(H(R_L | R_1^{L-1}) - h^*(\mathbf{S}) \Big). \tag{13}$$

Note that we also have the equality $h^*(\mathbf{S}) = \lim_{L\to\infty} H(R_L | R_1^{L-1})$ which is an analog to the alternative expression for the entropy rate $h(\mathbf{S}) = \lim_{L\to\infty} H(S_L | S_1^{L-1})$ because the right-hand side expression in (13) converges. We can prove that the permutation excess entropy $\mathbf{E}^*(\mathbf{S})$ also admits a mutual information expression if the process $\mathbf{S}$ is ergodic Markov, which will be presented elsewhere [18].

We would like to know whether $\mathbf{E}(\mathbf{S}) = \mathbf{E}^*(\mathbf{S})$ holds or not for a given finite-state stationary stochastic process $\mathbf{S}$. In the rest of the paper, we give a partial answer to this problem. In particular, we will show that $\mathbf{E}(\mathbf{S}) = \mathbf{E}^*(\mathbf{S})$ for any finite-state stationary ergodic Markov process.

Note that we always have $\mathbf{E}^*(\mathbf{S}) \leq \mathbf{E}(\mathbf{S})$ if the limits on both sides exist because $H^*(S_1^L) \leq H(S_1^L)$ and $h^*(\mathbf{S}) = h(\mathbf{S})$ by Lemma 10 and Theorem 11, respectively. To show $\mathbf{E}^*(\mathbf{S}) = \mathbf{E}(\mathbf{S})$, it is sufficient to show that

$$\left(\sum_{\substack{\pi \in \mathcal{S}_L, \\ |\phi^{-1}(\pi)| > 1}} p(\pi) \right) \log L \underset{L\to\infty}{\longrightarrow} 0$$

if $\mathbf{E}(\mathbf{S})$ exists. Let us put

$$q_L := \sum_{\substack{\pi \in \mathcal{S}_L, \\ |\phi^{-1}(\pi)|>1}} p(\pi).$$

Lemma 12 *Let ϵ be a positive real number and L be a positive integer. Assume that for any $s \in A_n$,*

$$\Pr\{s_1^{\lfloor L/2 \rfloor} | s_j \neq s \text{ for any } 1 \leq j \leq \lfloor L/2 \rfloor\} \leq \epsilon$$

holds, where $\lfloor x \rfloor$ is the largest integer not greater than x. Then, we have $q_L \leq 2n\epsilon$.

Proof. We shall prove

$$\sum_{\substack{\pi \in \mathcal{S}_L, \\ |\phi^{-1}(\pi)|=1}} p(\pi) \geq 1 - 2n\epsilon.$$

Let us consider a word $s_1^L \in A_n^L$ satisfying the following two conditions:

(i) Each symbol $s \in A_n$ appears in $s_1^{\lfloor L/2 \rfloor}$ at least once.

(ii) Each symbol $s \in A_n$ appears in $s_{\lfloor L/2 \rfloor+1}^L$ at least once.

By the assumption of the lemma, we have

$$\Pr\{s_1^{\lfloor L/2 \rfloor} | \text{(i) holds}\} \geq 1 - n\epsilon,$$

because

$$\Pr\{s_1^{\lfloor L/2 \rfloor} | \text{(i) holds}\} + \sum_{s=1}^{n} \Pr\{s_1^{\lfloor L/2 \rfloor} | s_j \neq s \text{ for any } 1 \leq j \leq \lfloor L/2 \rfloor\} \geq 1.$$

Similarly,

$$\Pr\{s_{\lfloor L/2 \rfloor+1}^L | \text{(ii) holds}\} \geq 1 - n\epsilon$$

holds because of the stationarity. Hence, we have both

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

which imply

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

It is clear that a word $s_1^L \in A_n^L$ satisfying both (i) and (ii) fulfills condition (d) in Theorem 9 (iii). Hence, by Theorem 9 (iv), we obtain

$$\sum_{\substack{\pi \in \mathcal{S}_L, \\ |\phi^{-1}(\pi)|=1}} p(\pi) = \sum\nolimits^* p(s_1^L) \geq \Pr\{s_1^L | \text{both (i) and (ii) hold}\} \geq 1 - 2n\epsilon,$$

where $\sum^*$ is the sum over all words s_1^L satisfying the condition (d) in Theorem 9 (iii).

□

As a first simple application of Lemma 12, let us consider a stochastic process $\mathbf{S} = \{S_1, S_2, \cdots\}$ such that the stochastic variables S_i are independent and identically distributed, namely, each symbol $s \in A_n$ appears at a probability $p(s) > 0$ independently. If we put $0 < \alpha := \min_{s \in A_n}\{p(s)\} < 1$, then we have

$$\Pr\{s_1^{\lfloor L/2 \rfloor} | s_j \neq s \text{ for any } 1 \leq j \leq \lfloor L/2 \rfloor\} = (1 - p(s))^{\lfloor L/2 \rfloor} \leq \left\{(1-\alpha)^{\frac{1}{2}}\right\}^L.$$

Thus, by Lemma 12, we have

$$H(S_1^L) - H^*(S_1^L) \leq 2n^2 \left\{(1-\alpha)^{\frac{1}{2}}\right\}^L \log(L+n) \underset{L \to \infty}{\longrightarrow} 0.$$

However, in this case, $\mathbf{E}^*(\mathbf{S}) = \mathbf{E}(\mathbf{S})$ is obvious from $\mathbf{E}^*(\mathbf{S}) \leq \mathbf{E}(\mathbf{S})$ because $\mathbf{E}(\mathbf{S}) = 0$.

Let $\mathbf{S} = \{S_1, S_2, \cdots\}$ be a *finite-state stationary ergodic Markov process* with a set of states A_n and a transition matrix $P = (p_{ij})$, where $p_{ij} \geq 0$ for all $1 \leq i, j \leq n$ and $\sum_{j=1}^n p_{ij} = 1$ for all $1 \leq i \leq n$. It is known that a finite-state stationary Markov process is ergodic if and only if its transition matrix P is irreducible [19]: a matrix P is *irreducible* if for all $1 \leq i, j \leq n$ there exists $l > 0$ such that $p_{ij}^{(l)} > 0$, where $p_{ij}^{(l)}$ is the (i,j)-th element of P^l. By the Perron-Frobenius theorem for irreducible non-negative matrices, there exists a unique *stationary distribution* $\mathbf{p} = (p_1, \cdots, p_n)$ such that $p_i > 0$ for all $1 \leq i \leq n$, $\sum_{i=1}^n p_i = 1$ and $\sum_{i=1}^n p_i p_{ij} = p_j$ for all $1 \leq j \leq n$, namely, ${}^tP\mathbf{p} = \mathbf{p}$, where ${}^tP$ is the transpose of the matrix P. Then, we have $p(s_1^L) = p_{s_1} p_{s_1 s_2} \cdots p_{s_{L-1} s_L}$ for $s_1^L \in A^L$. The entropy rate $h(\mathbf{S})$ and the excess entropy $\mathbf{E}(\mathbf{S})$ of a finite-state stationary Markov process $\mathbf{S} = \{S_1, S_2, \cdots\}$ are given by $h(\mathbf{S}) = -\sum_{i,j=1}^n p_i p_{ij} \log p_{ij}$ and $\mathbf{E}(\mathbf{S}) = -\sum_{i=1}^n p_i \log p_i + \sum_{i,j=1}^n p_i p_{ij} \log p_{ij}$, respectively.

Let L be a positive integer. Let us put $N := \lfloor L/2 \rfloor$. Given a symbol $s \in A_n$, we would like to evaluate

$$\begin{aligned} \beta_s &:= \Pr\{s_1^N | s_j \neq s \text{ for any } 1 \leq j \leq N\} \\ &= \sum_{\substack{s_j \neq s, \\ 1 \leq j \leq N}} p(s_1 \cdots p_N) = \sum_{\substack{s_j \neq s, \\ 1 \leq j \leq N}} p_{s_1} p_{s_1 s_2} \cdots p_{s_{N-1} s_N}. \end{aligned}$$

If $n = 1$ then $\beta_1 = 0$. So, this case is trivial. Hence, we assume $n \geq 2$ in the following discussion. If we introduce a matrix P_s whose (i,j)-th elements are defined by

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

then we can write

$$\beta_s = \langle (P_s)^{N-1} \mathbf{u}_s, \mathbf{p} \rangle,$$

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

Since P_s is a non-negative matrix, the following statements hold by the Perron-Frobenius theorem for non-negative matrices:

(i) There exists a non-negative eigenvalue λ such that any other eigenvalue of P_s has absolute value not greater than λ.

(ii) $\lambda \leq \max_i \{\sum_{j=1}^n (P_s)_{ij}\} = 1$.

(iii) There exists a non-negative right eigenvector $\mathbf{v}$ corresponding to the eigenvalue λ.

Lemma 13 $\lambda < 1$.

Proof. Suppose that $\lambda = 1$. Then, we have $P_s \mathbf{v} = \mathbf{v}$. For any positive integer l, we have

$$\langle \mathbf{v}, \mathbf{p} \rangle = \langle P_s^l \mathbf{v}, \mathbf{p} \rangle \leq \langle P^l \mathbf{v}, \mathbf{p} \rangle = \langle \mathbf{v}, \left({}^t P\right)^l \mathbf{p} \rangle = \langle \mathbf{v}, \mathbf{p} \rangle,$$

since $P_s \leq P$. Thus, we obtain $\langle \left(P^l - P_s^l\right) \mathbf{v}, \mathbf{p} \rangle = 0$, which implies that $\left(P^l - P_s^l\right) \mathbf{v} = \mathbf{0}$ because $\mathbf{p}$ is a positive vector and $\left(P^l - P_s^l\right) \mathbf{v}$ is a non-negative vector. Now, let us fix any $1 \leq j \leq n$. There exists l such that

$p_{sj}^{(l)} > 0$ because P is irreducible. Since the elements in the s-th row of the matrix P_s^l are all 0, we have $\sum_{k=1}^{n} p_{sk}^{(l)} v_k = 0$, where we put $\mathbf{v} = (v_1, \cdots, v_n)$. Thus, we obtain $v_j = 0$ because $p_{sj}^{(l)} > 0$, $p_{sk}^{(l)} \geq 0$ and $v_k \geq 0$ for all $1 \leq k \leq n$. Since $1 \leq j \leq n$ is arbitrary, $\mathbf{v} = \mathbf{0}$ must hold. However, this contradicts $\mathbf{v} \neq \mathbf{0}$ because $\mathbf{v}$ is an eigenvector.

□

Now, let $P_s = S + T$ be a Jordan-Chevalley decomposition of the matrix P_s, where S is a diagonalizable matrix and T is a nilpotent matrix. Let A be an invertible matrix such that $A^{-1}SA = D$, where D is a diagonal matrix. Since T is nilpotent, there exists a positive integer k such that T^k is a zero matrix. We also have $ST = TS$. If we put $E := A^{-1}TA$ then E^k is a zero matrix and $DE = ED$. Thus, for sufficiently large N,

$$\begin{aligned} P_s^{N-1} &= A(D+E)^N A^{-1} \\ &= A \left(\sum_{i=0}^{k-1} \binom{N-1}{i} D^{N-1-i} E^i \right) A^{-1} = \lambda^{N-k} O(N^{k-1}), \end{aligned}$$

where the big-O notation $O(N^{k-1})$ for a matrix means that each element of the matrix is $O(N^{k-1})$. Hence, we obtain $\beta_s = \lambda^{N-k} O(N^{k-1})$. Since $0 \leq \lambda < 1$ by Lemma 13, we get the following theorem by combining Lemma 10 and Lemma 12:

Theorem 14 *Let $\mathbf{S} = \{S_1, S_2, \cdots\}$ is a finite-state stationary ergodic Markov process. Then, the permutation excess entropy $\mathbf{E}^*(\mathbf{S})$ exists and $\mathbf{E}^*(\mathbf{S}) = \mathbf{E}(\mathbf{S})$.*

We can construct a finite-state stationary non-ergodic Markov process such that $\mathbf{E}(\mathbf{S}) \neq \mathbf{E}^*(\mathbf{S})$ immediately. For example, let $n = 2$ and

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We choose a stationary distribution $\mathbf{p} = (p_1, p_2) = (\frac{1}{2}, \frac{1}{2})$. Then we have $p(\underbrace{00\cdots0}_{L}) = p(\underbrace{11\cdots1}_{L}) = \frac{1}{2}$. Hence we have $h(\mathbf{S}) = h^*(\mathbf{S}) = 0$ and $\mathbf{E}(\mathbf{S}) = -p_1 \log p_1 - p_2 \log p_2 = 1$. On the other hand, we have $\mathbf{E}^*(\mathbf{S}) = 0$ because $\phi(\underbrace{00\cdots0}_{L}) = \phi(\underbrace{11\cdots1}_{L}) \in \mathcal{S}_L$.

Acknowledgments

TH was supported by the JST PRESTO program.

References

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

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

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

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

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

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

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

[8] S. Elizalde, The number of permutations realized by a shift, SIAM Journal of Discrete Mathematics 23 (2009) 765-786.

[9] B. A. Davey, H. A. Priestley, Introduction to Lattices and Order, second ed., Cambridge Univ. Press, Cambridge, 2002.

[10] S. MacLane, Categories for the Working Mathematician, second ed., Springer-Verlag, New York, 1998.

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

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

[13] R. Shaw, The Dripping Faucet as a Model Chaotic System, Aerial Press, Santa Cruz, California, 1984.

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

[15] R. L. Graham, D. E. Knuth, O. Patashnik, Concrete Mathematics, second ed., Addison-Wesley Publishing Company, Inc, 1994.

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

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

[18] T. Haruna, K. Nakajima, Permutation excess entropy and mutual information between the past and future, submitted to 10th International Conference on Computing Anticipatory Systems.

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