PERMUTONS AND ENTROPY

SÁRA SZEPESSY

Supervisor: Balázs Maga

Abstract. Permutons are continuous limit objects for sequences of large permutations. This project explores their complexity through the lens of permutation entropy. We construct a permuton that exhibits oscillating entropy behavior, where the entropy rate fluctuates between zero and infinity. This construction is motivated by the conjecture that such non-convergence may be the generic case for permutons associated with continuous measure-preserving functions.

1. Introduction

A permuton is a probability measure on the unit square $[0,1]^2$ with uniform marginals. They arise as the natural limit objects of permutation sequences when $n \to \infty$. This theory has received significant attention in recent years as a central framework in the study of combinatorial limits. One powerful approach to characterizing the structural complexity of these objects is through the study of their various entropies.

Shannon entropy. We briefly recall the definition of Shannon entropy and a fundamental inequality regarding the entropy of mixed distributions, which will be essential for our later estimates.

Definition 1. *(entropy) The Shannon entropy or just entropy of a discrete random variable X is*

$$H[X] = -\sum_x \mathbb{P}(X = x) \log \mathbb{P}(X = x) = -\mathbb{E}[\log \mathbb{P}(X)]$$

when the sum exists.

Note that if X is a discrete random variable taking values in a set of size n, then $H[X] \le \log n$, with equality if and only if X is uniformly distributed.

We establish the following fundamental properties of Shannon entropy to be used in our entropy rate estimations:

Lemma 1 (Refinement)**.** *Let $\mathcal{P} = (p_1, \dots, p_k)$ be a probability distribution. Let $\mathcal{Q}$ be a refinement of $\mathcal{P}$ given by $\mathcal{Q} = (q_{1,1}, \dots, q_{1,l_1}; \dots; q_{k,1}, \dots, q_{k,l_k})$ such that $\sum_{j=1}^{l_i} q_{i,j} = p_i$ for each $i \in \{1, \dots, k\}$. Then:*

$$H(\mathcal{Q}) \ge H(\mathcal{P}).$$

Lemma 2. *Let $(\nu_k)_{k=1}^m$ be probability measures on the same finite ground set E and $(\alpha_k)_{k=1}^m$ a nonnegative vector with ℓ_1 norm $\|\alpha\|_1$. Then*

$$\sum_{k=1}^m \alpha_k H(\nu_k) \le \|\alpha\|_1 H\left(\frac{\sum_{k=1}^m \alpha_k \nu_k}{\|\alpha\|_1}\right) \le \sum_{k=1}^m \alpha_k H(\nu_k) + \|\alpha\|_1 H\left(\frac{(\alpha_k)_{k=1}^m}{\|\alpha\|_1}\right).$$

Sampling entropy of permutons. Given a permuton μ, we obtain a pattern τ by sampling k points according to the measure μ and recording the relative order of their second coordinates from left to right. This produces the random permutation τ_μ. We let $t(\tau, \mu) = \mathbb{P}(\tau_\mu = \tau)$ represent the probability that this sampling procedure yields the pattern τ and denote the distribution of this random permutation by $\mu^{(n)}$.

It is also natural to quantify the information captured by this sampling procedure, namely by determining the Shannon entropy of $\mu^{(n)}$, gaining the pattern entropy sequence $(H(\mu^{(n)}))_{n\in\mathbb{N}}$. For simplicity, we will denote it by $H(\mu^{(n)})$ from now on.

An important class of permutons is associated with measure-preserving functions. Throughout this discussion, we consider functions $f : [0,1] \to [0,1]$ that preserve the one-dimensional Lebesgue measure.

For any such f, the associated permuton μ_f is defined as the uniform probability measure supported on the graph of f.

Although calculating the sampling entropies for a given permuton is often difficult, the following result of Balázs Maga [1] provides a surprisingly elegant limit for the asymptotic entropy of μ_f, when f satisfies certain regularity conditions:

Theorem 1. *Suppose $f : [0,1] \to [0,1]$ is measure-preserving and piecewise continuously differentiable with finitely many pieces. Then* $\frac{H(\mu^{(n)})}{n} \to \int_{[0,1]} \log |f'|$.

Note that this is identical to Rokhlin's formula [4] about the Kolmogorov–Sinai entropy of the underlying measure-preserving function ($h_{KS}(f)$), when $|f'| > K$ for some $K > 1$.

Permutation entropy for interval maps. We now consider a classical definition of permutation entropy for interval maps. While this framework does not require the map to be measure-preserving, we shall restrict our discussion to the measure-preserving setting to maintain consistency with the permuton constructions discussed previously.

Let $f : I \to I$ be an interval map. For any initial point x, we consider its first n iterates $(x, f(x), \ldots, f^{n-1}(x))$. Given a permutation $\pi = (k_1, \ldots, k_n)$ of $\{0, \ldots, n-1\}$, we define P_π as the set of points x whose orbit is ordered according to π:

$$P_\pi = \{x \in I \mid f^{k_1}(x) < f^{k_2}(x) < \cdots < f^{k_n}(x)\}$$

For $n \geq 2$, we let the partition $\mathcal{P}_n^*$ be the collection of all such sets P_π that are non-empty. In this construction, we disregard points where $f^i(x) = f^j(x)$ for some $i \neq j$; accordingly, we generally assume that the set of such points has (Lebesgue) measure zero. We define the order n permutation entropy of the map f with respect to the probability measure ν as the Shannon entropy of the distribution induced by the partition $\mathcal{P}_n^*$:

$$H_n^{\mathrm{perm}}(\nu, f) = -\sum_{P_\pi \in \mathcal{P}_n^*} \nu(P_\pi) \log(\nu(P_\pi)).$$

If ν is clear from the context, we simply write $H_n^{\mathrm{perm}}(f)$. Given a permuton μ associated to a measure-preserving function, we denote the distribution induced by the partition $\mathcal{P}_n^*$ by μ^n. Using this notation, we have $H_n^{\mathrm{perm}}(f) = H(\mu_f^n)$.

In parallel to the sampling entropy result for permutons (Theorem 1), the permutation entropy also recovers the Kolmogorov-Sinai entropy of the map as n grows. Specifically, for a wide class of maps, the following result holds [2]:

Theorem 2. *For any piecewise monotone interval map $f : [0,1] \to [0,1]$ with finitely many pieces, and for each f-invariant probability measure:*

$$\lim_{n\to\infty} \frac{1}{n} H_n^{\mathrm{perm}}(\nu, f) = h_{KS}(\nu, f).$$

As a direct consequence of this theorem, we observe that maps with zero Kolmogorov-Sinai entropy, such as rotations, have a vanishing permutation entropy rate [3, Prop 4.7].

Proposition 1. *Let $f : [0,1] \to [0,1]$ be an irrational rotation, i.e., $f(x) = x + \alpha \pmod 1$ with $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Then $h_{KS}(f) = 0$. Consequently,* $\frac{H_n^{\mathrm{perm}}(f)}{n} \to 0$ *as $n \to \infty$.*

2. Motivation

The parallels between Theorem 1 and Theorem 2 suggest a deep connection between the sampling and permutation entropies of permutons. For measure-preserving maps, both quantities converge to the Kolmogorov-Sinai entropy under mild regularity conditions. This might suggest that the entropy rate is typically well-behaved. However, a recent result of Balázs Maga shows that essentially the opposite is true: for a generic continuous measure-preserving function $f \in \mathcal{C}(\lambda)$ (in the sense of Baire category), the entropy rate satisfies[1]:

$$\liminf_{n\to\infty} \frac{H(\mu_f^{(n)})}{n} = 0 \quad \text{and} \quad \limsup_{n\to\infty} \frac{H(\mu_f^{(n)})}{n \log n} = 1.$$

Given the close relationship between these two notions of complexity in the regular setting, it is natural to ask whether permutation entropy also exhibits such erratic behavior. This motivates our construction of a permuton where the normalized permutation entropy $H(\mu^n)/n$ oscillates between zero and infinity along different subsequences. While the example we provide is not yet continuous, it serves as an existence result, and the techniques intended for its continuous extension represent a likely building block for proving that such behavior is generic in $\mathcal{C}(\lambda)$.

3. Permuton with Infinite Entropy Rate

We now provide a construction for a permuton μ exhibiting the asymptotic behavior $\frac{H_n^{\text{perm}}(\mu)}{n} \to \infty$. To begin, we fix n and construct a permuton μ_n that maximizes the order n permutation entropy. Given that $H_n^{\text{perm}} = H(\mathcal{P}_n^*) \leq \log(|\mathcal{P}_n^*|) \leq \log(n!)$ with equality if and only if $\mathcal{P}_n^*$ consists of $n!$ sets of equal measure, the entropy is maximized when each permutation $\pi \in S_n$ occurs with uniform probability.

To proceed, we define the cyclic variants of a permutation $\pi = (k_1, \ldots, k_n) \in S_n$ as the set

$$C(\pi) = \{(k_1, \ldots, k_n), (k_2, \ldots, k_n, k_1), \ldots, (k_n, k_1, \ldots, k_{n-1})\}.$$

Since each such set contains exactly n distinct permutations and $C(\pi) \neq C(\pi')$ implies $C(\pi) \cap C(\pi') = \emptyset$, S_n can be decomposed into $(n-1)!$ disjoint cyclic classes:

$$S_n = \bigsqcup_{i=1}^{(n-1)!} C(\pi_i)$$

for a properly chosen set of representative permutations $\{\pi_1, \ldots, \pi_{(n-1)!}\}$.

We now choose a set of representative permutations $\{\pi_1, \ldots, \pi_{(n-1)!}\}$ and construct the permuton $\mu_{C(\pi_i)}$ for each $i = 1, \ldots, (n-1)!$ as following. For a given class $C(\pi_i)$, let $(k_0, k_1, \ldots, k_{n-1}) \in C(\pi_i)$ be a representative starting with $k_0 = 0$. We distribute a total Lebesgue measure of $\frac{1}{n}$ into each of the n subsquares $[\frac{k_j}{n}, \frac{k_j+1}{n}] \times [\frac{k_{j+1}}{n}, \frac{k_{j+1}+1}{n}]$, $j = 0, \ldots, n-1$, where we define $k_n = k_0 = 0$; see Figure1. To ensure H_n^{perm} is well-defined on our permuton, we distribute the measure among the n subsquares by placing identical, scaled copies of the graph of a measure-preserving function f_n into each square. Specifically, the measure within each of the choosen subsquare is supported on a copy of the graph of f_n that has been scaled by a factor of $1/n$ and shifted to the appropriate grid position. We will define our exact choice of f_n later. Let f be the resulting measure-preserving function on the unit interval. By this construction, if a point x belongs to the interval $[\frac{k_j}{n}, \frac{k_j+1}{n}]$, its iterates satisfy $f^l(x) \in [\frac{k_{j+l}}{n}, \frac{k_{j+l}+1}{n}]$. Consequently, we obtain $x \in P_{\pi^{-1}}$ with $\pi = (k_j, k_{j+1}, \ldots, k_{j-1})$. Thus, choosing x uniformly randomly, we obtain each π^{-1} with $\pi \in C(\pi_i)$ with equal probability.

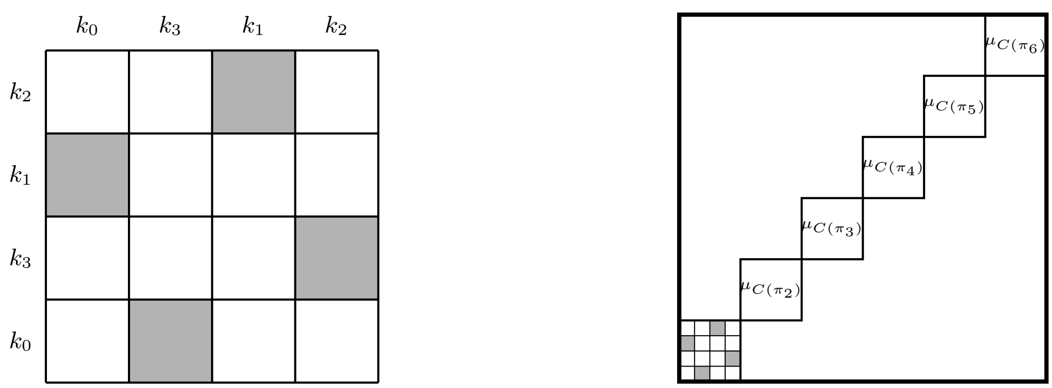

Figure 1. Left: Grid representation of the permuton $\mu_{C((2,3,1,0))}$. Right: The permuton μ_4.

Finally, by arranging the $(n-1)!$ component permutons $\mu_{C(\pi_i)}$ in a diagonal structure (see Figure 1) and assigning each a weight of $\frac{1}{(n-1)!}$, we obtain the permuton μ_n, where each permutaion $\pi \in S_n$ is obtained with uniform probability, as required.

We are now ready to construct the permuton μ. Let $(c_k)_{k=1}^{\infty}$ be a non-decreasing sequence of positive integers. We partition the unit square $[0,1]^2$ by placing a sequence of squares S_k of side length 2^{-k} along the diagonal, starting from the bottom-left corner for $k = 1$. On each square S_k, we embed the permuton μ_{c_k} with a weight of 2^{-k}. To ensure that the order-n permutation entropy H_n^{perm} is well-defined for all n across each component μ_{c_k}, we define $f_{c_k}(x) = x + \alpha_{c_k}$ $(mod\ 1)$ with $\alpha_{c_k} \in \mathbb{Q}^*$ arbitrary.

For each n, we define the probability distributions μ_1^n and μ_2^n as the normalized sums of the components $\mu_{c_k}^n$ for indices $c_k \geq n$ and $c_k < n$, respectively, that is $\mu_1^n = a_n \cdot \sum_{k:c_k \geq n} \frac{1}{2^k} \mu_{c_k}^n$ and $\mu_2^n = b_n \cdot \sum_{k:c_k<n} \frac{1}{2^k} \mu_{c_k}^n$. This way we get:

$$\mu^n = \left(\sum_{k:c_k \geq n} \frac{1}{2^k} \right) \mu_1^n + \left(\sum_{k:c_k<n} \frac{1}{2^k} \right) \mu_2^n$$

Using Prop.2:

$$\frac{H^{\text{perm}}(\mu^n)}{n} \geq \sum_{k:c_k \geq n} \frac{1}{2^k} \cdot \frac{H^{\text{perm}}(\mu_1^n)}{n} + \sum_{k:c_k<n} \frac{1}{2^k} \cdot \frac{H^{\text{perm}}(\mu_2^n)}{n} \geq \sum_{k:c_k \geq n} \frac{1}{2^k} \cdot \frac{\log(n!)}{n} + 0 \tag{1}$$

Here we used the fact that for each $c_k \geq n$, we obtain all permutation of S_n in μ_{c_k} with the same probability. Let us choose

$$c_k = 2^{2^{2^k}},$$

then $c_k \geq n \iff k \geq \log_2 \log_2 \log_2(n)$. Then the following lower bound holds:

$$\sum_{k:c_k \geq n} \frac{1}{2^k} \geq 2^{-\log_2 \log_2 \log_2(n)} = \frac{1}{\log_2 \log_2(n)}.$$

Substituting back to (1) we get:

$$\frac{H^{\text{perm}}(\mu^n)}{n} \geq \frac{\log(n!)}{n \cdot \log\log(n)} \geq \frac{\frac{n}{2}\log(n)}{n \cdot \log\log(n)} = \frac{1}{2} \frac{\log(n)}{\log\log(n)} \to \infty, \text{ as } n \to \infty. \tag{2}$$

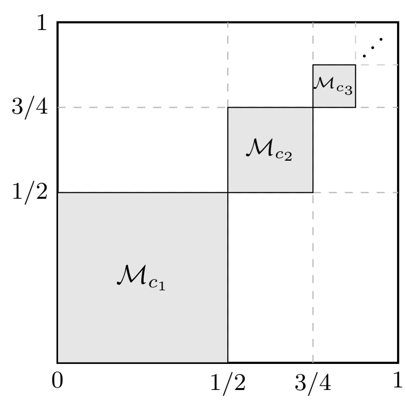

FIGURE 2. The dyadic-diagonal construction of the permuton μ. The components $\mathcal{M}_{c_k}$ are embedded in squares along the diagonal with side lengths 2^{-k} and weights 2^{-k}.

4. A Permuton with Oscillating Entropy Rate

Let us define the permuton μ_β as the associated permuton of the measure-preserving function $g_\beta(x) = x + \beta$ $(mod\ 1)$. For each $k \in \mathbb{N}$, choose $\beta_k \leq \frac{1}{k}$, so that the first k iterates of g_{β_k} do not "overlap" for any starting point x. More precisely, if we fix k, for any $n \leq k$ we have $\{\pi \in S_n : P_\pi \in \mathcal{P}_n^*\} = C(id)$, therefore $|\mathcal{P}_n^*| = n$ for all $n \leq k$.

Let $f_{c_k}(x) = x + \beta_k = g_{\beta_k}$ be the local map used to construct μ_{c_k}. This choice defines the micro-structure at level c_k. For $n \geq c_k$, a point's orbit follows the macro-pattern of length c_k determined by the chosen cyclic permutation $\pi \in S_{c_k}$. However, the precise ordering of iterates is further refined by the local behavior of f_{c_k} inside each subsquare. To understand the structure of μ_{c_k}, consider the orbit of a point under this construction. Globally, the point 'jumps' between subsquares according to the cyclic permutation $\pi \in S_{c_k}$ at each time step. Locally, however, the point's relative position within each subsquare is governed by the map f_{c_k}. Because each subsquare contains an identical, scaled copy of f_{c_k}, the point effectively 'climbs' the graph of the function while being shifted across the grid. Specifically, let $\mathcal{P}$ be a partition formed by $c_k!$ scaled copies of the partition associated with f_{c_k}. For any point starting in an element $I \in \mathcal{P}$, the sequence of subsquares it visits is fixed by the cycle, and its relative vertical rank within those squares is fixed by the local dynamics of f_{c_k}. Thus, the resulting order-n permutation is uniquely determined for all $x \in I$. This shows that $\mathcal{P}$ is a refinement of $\mathcal{P}_n^*(\mu_{c_k})$, allowing us to bound the entropy: $H_n^{\text{perm}}(\mu_{c_k}) \leq H_n^{\text{perm}}(\mathcal{P}) \leq H_n^{\text{perm}}(f_{c_k}) + \log(c_k!)$. By applying the mixing inequality for Shannon entropy (Proposition 2), we can bound the total entropy of the order-n permutation distribution: $H_n^{\text{perm}}(\mu_{c_k}) \leq H_n^{\text{perm}}(\mathcal{P})\mu_{c_k}^n \leq H_n^{\text{perm}}(f_{c_k}) + \log(c_k!)$. Thus, using Prop.1, $\frac{H_n^{\text{perm}}(\mu_{c_k})}{n} \to 0$, as $n \to \infty$.

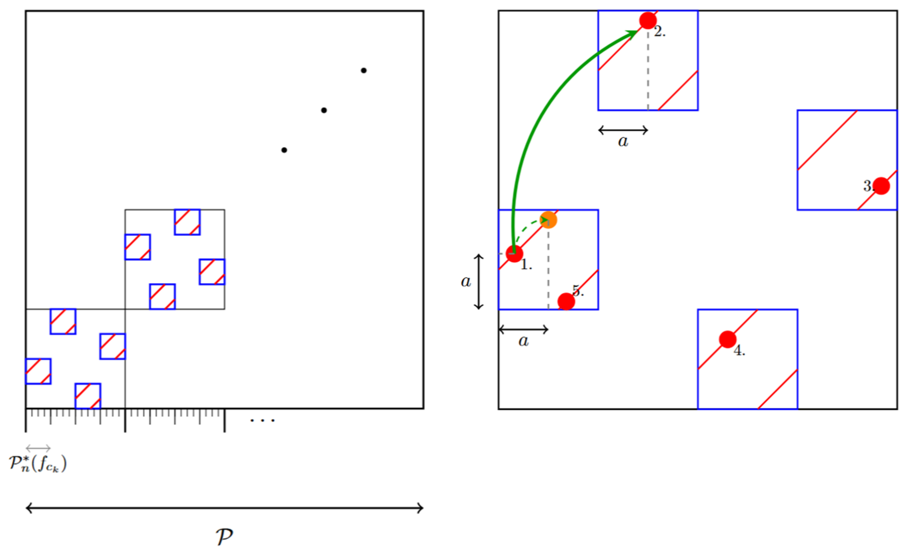

FIGURE 3. Left: The permuton μ_4 with the rotation map embedded and a schematic representation of the partition $\mathcal{P}$; Right: First 5 iterates in the subsquare associated to $C(1,3,2,0)$.

Recursive Construction of the Permuton. Finally, we are ready to construct our permuton μ with oscillating entropy behavior. We define μ using the same dyadic-diagonal construction described previously. To each j-th subsquare along the diagonal, we assign a permuton with weight 2^{-j}. The specific assignment is determined by two sequences: a sequence of indices $n_k \to \infty$ and a sequence of indices $m_k \to \infty$, chosen such that $n_{k-1} < m_k < n_k$.

For each $k \in \mathbb{N}$, the assignments are as follows:

- To the index $j = n_k$, we assign the high-entropy permuton $\mu_{c_{n_k}}$.
- To all other indices j in the interval $m_k \leq j < m_{k+1}$, we assign the rotational permuton $\mu_{\beta_{m_k}}$.

Our goal is to show that by carefully choosing the growth rates of n_k and m_k, the resulting entropy rate along the sequence m_k satisfies:

$$\lim_{k\to\infty} \frac{H_{m_k}^{\text{perm}}(\mu)}{m_k} = 0.$$

Estimating the Entropy of the Preceding Indices. We define our sequences inductively. For $k > 1$, assume that $n_{k'}$ and $m_{k'}$ have been defined for all $k' < k$. For $k = 1$, we simply assume $m_1 = 1$ and $n_1 = 2$. Suppose we have fixed m_k. We seek an upper bound for the entropy contributed by the indices $j < m_k$. To this end, let γ_k be a normalizing factor such that the weighted sum of the measures forms a probability distribution $\nu_k^{m_k}$:

$$\nu_k^{m_k} = \gamma_k \left[\sum_{i=1}^{k-1} \left(\sum_{\substack{m_i \le j < m_{i+1} \\ j \ne n_i}} 2^{-j} \right) \mu_{\beta_{m_i}}^{m_k} + \sum_{i=1}^{k-1} 2^{-n_i} \mu_{c_{n_i}}^{m_k} \right].$$

Applying the mixing inequality for Shannon entropy and noting that the number of distinct measure components is at most $2k$, we obtain:

$$H(\nu_k^{m_k}) \le \log(2k) + \sum_{i=1}^{k-1} H_{m_k}^{\text{perm}}(\mu_{\beta_{m_i}}) + \sum_{i=1}^{k-1} H_{m_k}^{\text{perm}}(\mu_{c_{n_i}}).$$

Since k is fixed, each term in the summands represents the permutation entropy of a fixed permuton. Because these entropies vanish as the sampling order $m_k \to \infty$, we can choose N sufficiently large such that $H(\nu_k^{m_k}) < \frac{1}{3k}$ for all $m_k > N$.

Estimating the Entropy of the Tail. For the indices $j \ge m_k$, we construct the probability measure $\rho_k^{m_k,n_k}$ in a similar fashion:

$$\rho_k^{m_k,n_k} = \delta_{m_k,n_k} \left(a_{m_k,n_k} \cdot B_1(m_k,n_k) + b_{m_k,n_k} \cdot B_2(m_k,n_k) \right)$$

where:

- δ_{m_k,n_k} is the normalization factor.
- B_1 represents the component of the permuton associated with future indices $j = n_{k'}$ for all $k \le k'$.
- B_2 represents the component associated with the remaining indices $j \ge m_k$ (where $j \ne n_{k'}$).
- a_{m_k,n_k} and b_{m_k,n_k} are the respective total weights of these components within μ.

Note that while the specific values of $n_{k'}$ (for $k' \ge k$) are not yet determined, we can nonetheless establish uniform upper bounds for the entropy. These bounds will depend exclusively on the value of n_k and m_k.

To establish an upper bound, we recall that the rotation parameters $\beta_{k'}$ are chosen such that the permutation partition satisfies $\mathcal{P}_{m_k}^*(B_2) = C(\text{id})$. This ensures that the iterates within the B_2 component exhibit minimal combinatorial complexity. For the B_1 component, we choose n_k sufficiently large so that the weight a_{m_k,n_k} effectively suppresses the high-entropy terms. Specifically, we have:

$$\frac{H_{m_k}^{\text{perm}}(\rho_k^{m_k,n_k})}{m_k} \le \frac{\log 2}{m_k} + \frac{\log(m_k!)}{m_k \cdot 2^{n_k-1}} + \frac{\log m_k}{m_k} \le \frac{\log 2}{m_k} + \frac{\log(m_k)}{2^{n_k-1}} + \frac{\log m_k}{m_k}.$$

The first and third terms depend only on m_k; thus, we can choose $N < m_k$ such that their sum is less than $\frac{1}{3k}$. Finally, by selecting $n_k > m_k$ sufficiently large, the middle term also falls below $\frac{1}{3k}$, yielding the desired bound.

Combining the estimates we obtain the global bound for the order-m_k permutation entropy of μ

$$\frac{H_{m_k}^{\text{perm}}(\mu)}{m_k} \le \frac{1}{k} + \frac{\log 2}{m_k} \to 0, \text{ as } k \to \infty.$$

Note that from the previous example we already have $\frac{H_{n_k}^{\text{perm}}(\mu)}{n_k} \to \infty$ (the only difference is a constant factor of $1/2$, which does not affect the divergence of the entropy rate). This confirms that the permuton μ exhibits oscillating entropy behavior, where the entropy rate vanishes along the sequence m_k but diverges along n_k.

References

[1] B. Maga (2025), *On the sampling entropy of permutons* https://arxiv.org/pdf/2503.18518.pdf

[2] C. Bandt, Gerhard Keller, and Bernd Pompe (2002). "Entropy of interval maps via permutations". In: Nonlinearity 15 , p. 1595.

[3] O. Sarig (2023), *Lecture Notes on Ergodic Theory*, Weizmann Institute of Science, Available: `https://www.weizmann.ac.il/math/sarigo/sites/math.sarigo/files/uploads/ergodicnotes.pdf`

[4] Y. Coudène (2016), "Ergodic Theory and Dynamical Systems", Universitext, Springer-Verlag London. Available: `https://www.ime.usp.br/~sylvain/RokhlinFormulaEntropy.pdf`

Declaration of AI usage

The AI language model Gemini was used to rephrase and polish portions of the text to improve clarity and wording, as well as to generate LaTeX code for the figures.