The Asymptotic Distribution of the Permutation Entropy

A. A. Rey,[1] A. C. Frery,[2] J. Gambini,[3] and M. M. Lucini[4]

[1)] *CIDIA - Universidad Nacional de Hurlingham, Villa Santos Tesei 1688, República Argentina; and CPSI - Universidad Tecnológica Nacional Regional Buenos Aires, Ciudad Autónoma de Buenos Aires 1179, República Argentina*

[2)] *School of Mathematics and Statistics - Victoria University of Wellington, Wellington 6140, New Zealand*

[3)] *CIDIA - Universidad Nacional de Hurlingham, Villa Santos Tesei 1688, República Argentina; CPSI - Universidad Tecnológica Nacional Regional Buenos Aires, Ciudad Autónoma de Buenos Aires 1179, República Argentina; and Dpto. Ing. Computación - Universidad Nacional de Tres de Febrero, Sáenz Peña 1674, República Argentina*

[4)] *Facultad de Ciencias Exactas, Naturales y Agrimensura - Universidad Nacional del Nordeste and CONICET, Corrientes 3400, República Argentina*

(*Electronic mail: andrea.rey@unahur.edu.ar)

Ordinal Patterns serve as a robust symbolic transformation technique, enabling the unveiling of latent dynamics within time series data. This methodology involves constructing histograms of patterns, followed by the calculation of both entropy and statistical complexity — an avenue yet to be fully understood in terms of its statistical properties. While asymptotic results can be derived by assuming a Multinomial distribution for histogram proportions, the challenge emerges from the non-independence present in the sequence of ordinal patterns. Consequently, the direct application of the Multinomial assumption is questionable. This study focuses on the computation of the asymptotic distribution of permutation entropy, considering the inherent patterns' correlation structure. Furthermore, the research delves into a comparative analysis, pitting this distribution against the entropy derived from a Multinomial law. We present simulation algorithms for sampling time series with prescribed histograms of patterns and transition probabilities between them. Through this analysis, we better understand the intricacies of ordinal patterns and their statistical attributes.

The methodology of ordinal patterns introduced by Bandt and Pompe has been extensively applied to study the latent dynamics of time series via their entropy, named permutation entropy. However, there are no theoretical results about the permutation entropy's distribution, which must consider the correlation effect between patterns. In this work, we prove that the asymptotic distribution of the permutation entropy is Normal. Since the expression of the asymptotic variance is more complex as the embedding dimension increases, we compare this result with the Multinomial sample entropy, which assumes independence. A hypothesis test is then derived, and it is applied to distinguish meteorological time series of distinct locations, as well as to differentiate biological signals such as ECGs.

I. INTRODUCTION

Signal analysis through ordinal patterns has been widely used since they were presented in Bandt and Pompe[2]. In this seminal article, the authors proposed a way of analyzing time series using ordinal patterns rather than the actual values. Their approach consists of transforming typically small and overlapping windows of m observations (m is called "embedding dimension," and the corresponding values are called "words") into their ordinal patterns, i.e., the set of indexes that sort the values in the window. The time series $\boldsymbol{x} = (x_1, x_2, \dots, x_{n+m-1})$ is transformed into the sequence of patterns $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_n)$, with $\pi_j \in \Pi^m = \{\pi^{(1)}, \pi^{(2)}, \dots, \pi^{(m!)}\}$ where $\pi^{(i)}$, $i = 1, \dots, m!$, is the set of possible patterns provided the values in $\boldsymbol{x}$ have no ties. The analysis is then carried on $\boldsymbol{\pi}$ rather than on the original values $\boldsymbol{x}$.

This approach has several interesting features, among them:

1. The sequence of patterns $\boldsymbol{\pi}$ is invariant to monotonically increasing transformations of $\boldsymbol{x}$.

2. The ordinal patterns are less sensitive to outliers than the original values.

Such robustness is one of the reasons why the Bandt & Pompe symbolization approach has become a trendy and successful way of analyzing time series[1].

The marginal[3] analysis of ordinal patterns proceeds by computing the histogram of proportions $\widehat{\boldsymbol{q}}$ of the patterns observed in $\boldsymbol{\pi}$:

$$\widehat{q}_i = \frac{\#\{\pi_j \in \boldsymbol{\pi} : \pi_j = \pi^{(i)}\}}{n}, \text{ for every } 1 \leq i \leq m!.$$

Time series with different underlying dynamics often exhibit different histograms. Extreme cases are strictly monotonic series that produce a single pattern and white noise that produces histograms with approximately equal proportions.

The entropy, in general, and the Shannon entropy are very sensitive to these differences. It also provides insightful information about the predictability of the series. For these reasons, the Shannon entropy of the histogram of ordinal patterns, called "Shannon Permutation Entropy" (or "Permutation Entropy," for short), plays a central role in this kind of analysis[2]. Given $\boldsymbol{q} = (q_1, q_2, \dots, q_{m!})$ the probability vector of the ordinal patterns, the Permutation Entropy is defined by

$$S(\boldsymbol{q}) = -\sum_{i=1}^{m!} q_i \ln q_i, \tag{1}$$

and its normalized version is

$$H(\boldsymbol{q}) = \frac{S(\boldsymbol{q})}{\ln m!}. \tag{2}$$

The Permutation Entropy, along with the Statistical Complexity, define the "Entropy-Complexity Plane," a closed manifold in which time series become points whose position reveals the type of underlying dynamics.

However, the full potential of the marginal analysis of Ordinal Patterns has not been fully exploited. Leyva et al.[9] state that one of the longstanding problems is attaching confidence regions to points in the Entropy-Complexity plane.

Chagas et al.[5] obtained empirical confidence regions for the points that white noise produces in the Entropy-Complexity plane. The findings were derived from simulations, and to extend the same approach to different time series, users need to generate additional simulations under the underlying model. Rey et al.[11] obtained the asymptotic distribution of certain types of entropies of histograms under the Multinomial law and developed a hypothesis test for comparing histograms of different numbers of bins. In order to use these results with the Permutation Entropy, the authors introduced a simplification: the patterns in $\boldsymbol{\pi}$ are independent. However, Elsinger[6] showed that ordinal patterns are not independent.

In this work, we obtain the asymptotic distribution of the permutation entropy considering ordinal patterns dependence. This distribution is a Normal law whose variance is larger than the variance of the asymptotic model under the independence simplification, i.e., with bins that obey a Multinomial distribution. When comparing meteorological time series, we assess the impact of using the latter instead of the better-adjusted former model. We present simulation algorithms for sampling time series with prescribed histograms of patterns and transition probabilities between them.

Elsinger[6] computed the transition probabilities from state π_i at time t to state π_j at time $t+1$ We use these expressions to calculate the Permutation Entropy asymptotic distribution. Then, we assess the committed error when simplifying that the ordinal patterns are independent.

This paper unfolds as follows. Section II recalls the main properties of ordinal patterns distribution, the asymptotic entropy distribution, and the dependency structure for the embedding dimension $m = 3$. In Section III, we compare asymptotic variances for actual and simplified models. In Sections IV and V, the results of applying the computed asymptotic variances to simulated and actual time series are shown. We conclude the article in Section VI. In Appendix A, the conditional probabilities of ordinal patterns transitions are computed.

II. ORDINAL PATTERNS DISTRIBUTION

Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_{n+m-1})$ be a real-valued time series without ties of length $T = n + m - 1$, and transform it into the series of symbols $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_n)$ of patterns observed over words of size m (the embedding dimension). Let $\boldsymbol{x}_m(t) = (x_t, x_{t+1}, \ldots, x_{t+m-1})$ be a part of $\boldsymbol{x}$, for $t = 1, 2, \ldots, n$. The subsequence $\boldsymbol{x}_m(t)$ is π_i-type if

$$\begin{cases} x_{t+i_1} \leq x_{t+i_2} \leq \cdots \leq x_{t+i_m} \text{ and} \\ i_{s-1} \leq i_s \text{ if } x_{t+i_{s-1}} = x_{t+i_s}, \end{cases} \tag{3}$$

where $(i_1, i_2, \ldots, i_m)$ is a permutation of the numbers $0, 1, \ldots, m-1$, and $s \in \{1, 2, \ldots, m\}$. Since each sequence $\boldsymbol{x}_m(t)$ is associated with a symbol, we obtain a sequence of ordinal patterns from the time series $\boldsymbol{x}$ given by $(\psi_1, \psi_2, \ldots, \psi_n)$. This sequence can be modeled as a realization of the ordinal pattern process whose possible states belong to Π^m. Notice that, for $i = 1, 2, \ldots, m!$, $\boldsymbol{x}_m(t)$ is π_i-type if and only if $\psi_t = \pi_i$.

The indicator function $\mathbb{1}_{\pi_i}(t)$ is defined by

$$\mathbb{1}_{\pi_i}(t) = \begin{cases} 1 & \text{if } \boldsymbol{x}_m(t) \text{ is } \pi_i\text{-type}, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Let $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_{m!})$ be the vector of random variables that count the number of occurrences of π_i in n trials, i.e., for $i = 1, 2, \ldots, m!$:

$$Z_i = \sum_{t=1}^{n} \mathbb{1}_{\pi_i}(t). \tag{5}$$

Due to the overlapping between $\boldsymbol{x}_m(t)$ and $\boldsymbol{x}_m(t+h)$ for $h = 1, 2, \ldots, m-1$, the ordinal patterns ψ_t and ψ_{t+h} are dependent for all $t = 1, 2, \ldots, n$ and $h = 1, 2, \ldots, m-1$. On the contrary, ψ_t and ψ_{t+h} are independent if $h \geq m$ and the stationary ordinal pattern process is $(m-1)$-dependent. For $i = 1, 2, \ldots, m!$, let q_i be the probability of observing the state π_i, denote the vector of probabilities as $\boldsymbol{q} = (q_1, q_2, \ldots, q_{m!})$, and express as $\mathbf{D}_{\boldsymbol{q}} = \mathrm{Diag}(q_1, q_2, \ldots, q_{m!})$ the associated diagonal matrix. The transition probability of reaching state π_j at time $t + \ell$ from state π_i at time t, for $\ell = 1, 2, \ldots, m-1$, is denoted by $q_{ij}^{(\ell)}$. We collect these transition probabilities in the matrix $\mathbf{Q}^{(\ell)}$ whose elements are $q_{ij}^{(\ell)} = \Pr(\psi_t = \pi_i \wedge \psi_{t+\ell} = \pi_j)$.

From Yamashita Rios de Sousa and Hlinka[12], we know that the expected value of Z_i is given by:

$$\mathrm{E}(Z_i) = nq_i,\ i \in \{1, \dots, m!\}, \tag{6}$$

and that the covariance between Z_i and Z_j is given by:

$$\mathrm{Cov}(Z_i, Z_j) = n\delta_{ij}q_j - n^2 q_i q_j + \sum_{\ell=1}^{m-1} (n-\ell) q_{ij}^{(\ell)} q_{ji}^{(\ell)},\ i, j \in \{1, \dots, m!\}, \tag{7}$$

where δ_{ij} is the Kronecker delta function.

A. Asymptotic Distribution of the Permutation Entropy

Let $\boldsymbol{X}_n = (X_{1,n}, X_{2,n}, \dots, X_{K,n})$, with $n \in \mathbb{N}$, be a sequence of independent and identically distributed K-variate vectors of random variables. Suppose that as n tends to infinity, $\sqrt{n}(X_{1,n} - \theta_1, X_{2,n} - \theta_2, \dots, X_{K,n} - \theta_K)$ converges in distribution to the multivariate Normal law $\mathcal{N}(\mathbf{0}, S_{\boldsymbol{X}})$ where $S_{\boldsymbol{X}}$ is the covariance matrix. Consider $h_1, h_2, \dots, h_K$ real-functions continuously differentiable in a neighborhood of the parameter point $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_K)$, such that the matrix of partial derivatives M defined by $M_{ij} = \partial h_i / \partial \theta_j$ for $1 \le i, j \le K$ is non-singular in this neighborhood. The multivariate version of the Delta Method, cf. Lehmann and Casella[8], states that

$$\sqrt{n}\big[h_1(\boldsymbol{X}_n) - h_1(\boldsymbol{\theta}), h_2(\boldsymbol{X}_n) - h_2(\boldsymbol{\theta}), \dots, h_K(\boldsymbol{X}_n) - h_K(\boldsymbol{\theta})\big] \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\big(\mathbf{0}, M S_{\boldsymbol{X}} M^{\mathrm{T}}\big). \tag{8}$$

The maximum likelihood (ML) estimator of q_i is the relative frequency $\widehat{q}_i = Z_i/n$, $1 \le i \le m!$. We will apply the Delta method to the sequence:

$$\boldsymbol{X}_n = (\hat{q}_{1,n}, \hat{q}_{2,n}, \dots, \hat{q}_{m!,n}) = \left(\frac{Z_1}{n}, \frac{Z_2}{n}, \dots, \frac{Z_{m!}}{n}\right). \tag{9}$$

By Yamashita Rios de Sousa and Hlinka[12], Eq. (30), if $\boldsymbol{W}_n = \frac{1}{\sqrt{n}}(\boldsymbol{Z} - n\boldsymbol{q})$ then

$$\boldsymbol{W}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\big(\mathbf{0}, \boldsymbol{\Sigma}\big), \tag{10}$$

where

$$\boldsymbol{\Sigma} = \mathbf{D}_{\boldsymbol{q}} - (2m-1)\boldsymbol{q}\boldsymbol{q}^{\mathrm{T}} + \sum_{\ell=1}^{m-1} \left(\mathbf{Q}^{(\ell)} + {\mathbf{Q}^{(\ell)}}^{\mathrm{T}}\right). \tag{11}$$

The diagonal elements are

$$\boldsymbol{\Sigma}_{ii} = (\mathbf{D}_{\boldsymbol{q}})_{ii} - (2m-1)(\boldsymbol{q}\boldsymbol{q}^{\mathrm{T}})_{ii} + \sum_{\ell=1}^{m-1} \left(\mathbf{Q}^{(\ell)} + {\mathbf{Q}^{(\ell)}}^{\mathrm{T}}\right)_{ii} = q_i - (2m-1)q_i^2 + 2\sum_{\ell=1}^{m-1} \mathbf{Q}_{ii}^{(\ell)}, \tag{12}$$

and, if $i \neq j$, then

$$\mathbf{\Sigma}_{ij} = (\mathbf{D}_{\boldsymbol{q}})_{ij} - (2m-1)(\boldsymbol{q}\boldsymbol{q}^{\mathrm{T}})_{ij} + \sum_{\ell=1}^{m-1} \left(\mathbf{Q}^{(\ell)} + \mathbf{Q}^{(\ell)\mathrm{T}}\right)_{ij} = -(2m-1)q_i q_j + \sum_{\ell=1}^{m-1} \left(\mathbf{Q}_{ij}^{(\ell)} + \mathbf{Q}_{ji}^{(\ell)}\right). \tag{13}$$

To verify the hypothesis of the Delta Method, notice that,

$$\boldsymbol{W}_n = \frac{1}{\sqrt{n}}(\boldsymbol{Z} - n\boldsymbol{q}) = \frac{n}{\sqrt{n}}\left(\frac{\boldsymbol{Z}}{n} - \boldsymbol{q}\right) = \sqrt{n}(\widehat{\boldsymbol{q}} - \boldsymbol{q}) = \sqrt{n}(\boldsymbol{X}_n - \boldsymbol{q}). \tag{14}$$

Let $\Omega_i = \{(x_1, x_2, \dots, x_{m!}) \in \mathbb{R}^{m!} : x_i > 0 \wedge x_s \geq 0 \text{ for } s = 1, 2, \dots, m! \text{ and } s \neq i\}$, for $i = 1, 2, \dots, m!$. Finally, we define the functions involved in the Delta Method for $i = 1, 2, \dots, m!$, as $h_i : \Omega_i \to \mathbb{R}$:

$$h_i(q_1, q_2, \dots, q_{m!}) = q_i \ln q_i, \tag{15}$$

which verify, for $1 \leq i, j \leq m!$, that

$$\frac{\partial h_i}{\partial q_j} = \begin{cases} \ln q_i + 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

If $q_i = 0$, the function h_i is omitted. Let B be the matrix of partial derivatives, which is diagonal. Using the result given by Eq. (8), we obtain the limit joint distribution:

$$\sqrt{n}\left[h_1(\widehat{q}_1) - h_1(q_1), h_2(\widehat{q}_2) - h_2(q_2), \dots, h_{m!}(\widehat{q}_{m!}) - h_{m!}(q_{m!})\right] \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\boldsymbol{q}}), \tag{17}$$

where $\mathbf{\Sigma}_{\boldsymbol{q}} = B\mathbf{\Sigma}B^{\mathrm{T}}$. The matrix of partial derivatives is diagonal, for $1 \leq i, j \leq m!$, then it holds that

$$\begin{aligned}(\mathbf{\Sigma}_{\boldsymbol{q}})_{ij} &= \sum_{r=1}^{m!}(B\mathbf{\Sigma})_{ir}B_{rj}^{\mathrm{T}} = \sum_{r=1}^{m!}\sum_{s=1}^{m!} B_{is}(\mathbf{\Sigma})_{sr}B_{rj}^{\mathrm{T}} = B_{ii}(\mathbf{\Sigma})_{ij}B_{jj}^{\mathrm{T}} \\ &= \begin{cases} (\ln q_i + 1)^2 \mathbf{\Sigma}_{ii} & \text{if } i = j, \\ (\ln q_i + 1)(\ln q_j + 1)\mathbf{\Sigma}_{ij} & \text{if } i \neq j. \end{cases}\end{aligned} \tag{18}$$

Since the Shannon entropy is a linear combination of the functions $\{h_1, h_2, \dots, h_{m!}\}$ with all scalars equal to -1, following Lehmann and Casella[8], we have that:

$$\sqrt{n}[S(\widehat{\boldsymbol{q}}) - S(\boldsymbol{q})] \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\left(0, \sigma_{\boldsymbol{q}}^2\right), \tag{19}$$

where

$$\begin{aligned}\sigma_{\boldsymbol{q}}^2 = \sum_{i=1}^{m!}(\mathbf{\Sigma}_{\boldsymbol{q}})_{ii} + 2\sum_{i=1}^{m!-1}\sum_{j=i+1}^{m!}(\mathbf{\Sigma}_{\boldsymbol{q}})_{ij} &= \sum_{i=1}^{m!}(\ln q_i + 1)^2\left[q_i - (2m-1)q_i^2 + 2\sum_{\ell=1}^{m-1}\mathbf{Q}_{ii}^{(\ell)}\right] \\ &- 2\sum_{i=1}^{m!-1}\sum_{j=i+1}^{m!}(\ln q_i + 1)(\ln q_j + 1)\left[(2m-1)q_i q_j - \sum_{\ell=1}^{m-1}\left(\mathbf{Q}_{ij}^{(\ell)} + \mathbf{Q}_{ji}^{(\ell)}\right)\right].\end{aligned} \tag{20}$$

Eqs. (19) and (20) are our core result: the asymptotic variance of the Shannon entropy of ordinal patterns considering their correlation structure. It is worth noticing that, for practical purposes, given n sufficiently large, we can use that $S(\boldsymbol{q})$ has a Normal distribution with mean equal to $S(\widehat{\boldsymbol{q}})$ and variance equal to $\sigma^2_{\boldsymbol{q}}/n$.

In addition, for $\alpha \in (0,1)$ and n sufficiently large, the $(1-\alpha)100\%$ confidence interval of $S(\boldsymbol{q})$ is given by

$$S(\widehat{\boldsymbol{q}}) \pm z_{\frac{\alpha}{2}} \sigma_{\boldsymbol{q}}/\sqrt{n}. \tag{21}$$

where $z_{\frac{\alpha}{2}}$ is the $\alpha/2$-quantile of a standard normal random variable.

In the case of the normalized Permutation Entropy, the result given in Eq. (19) yields to

$$\sqrt{n}[H(\widehat{\boldsymbol{q}}) - H(\boldsymbol{q})] \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\big(0, \sigma^2_{\boldsymbol{q}}/(\ln m!)^2\big), \tag{22}$$

B. Embedding Dimension $m = 3$

In this section, we focus on the case $m = 3$. The possible ordinal patterns are shown in Figure 1.

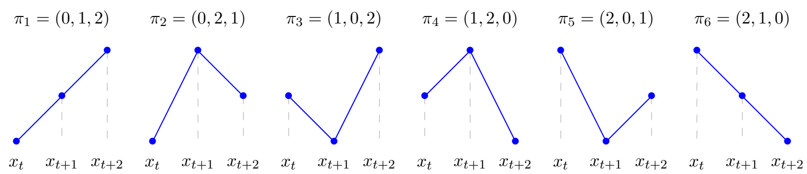


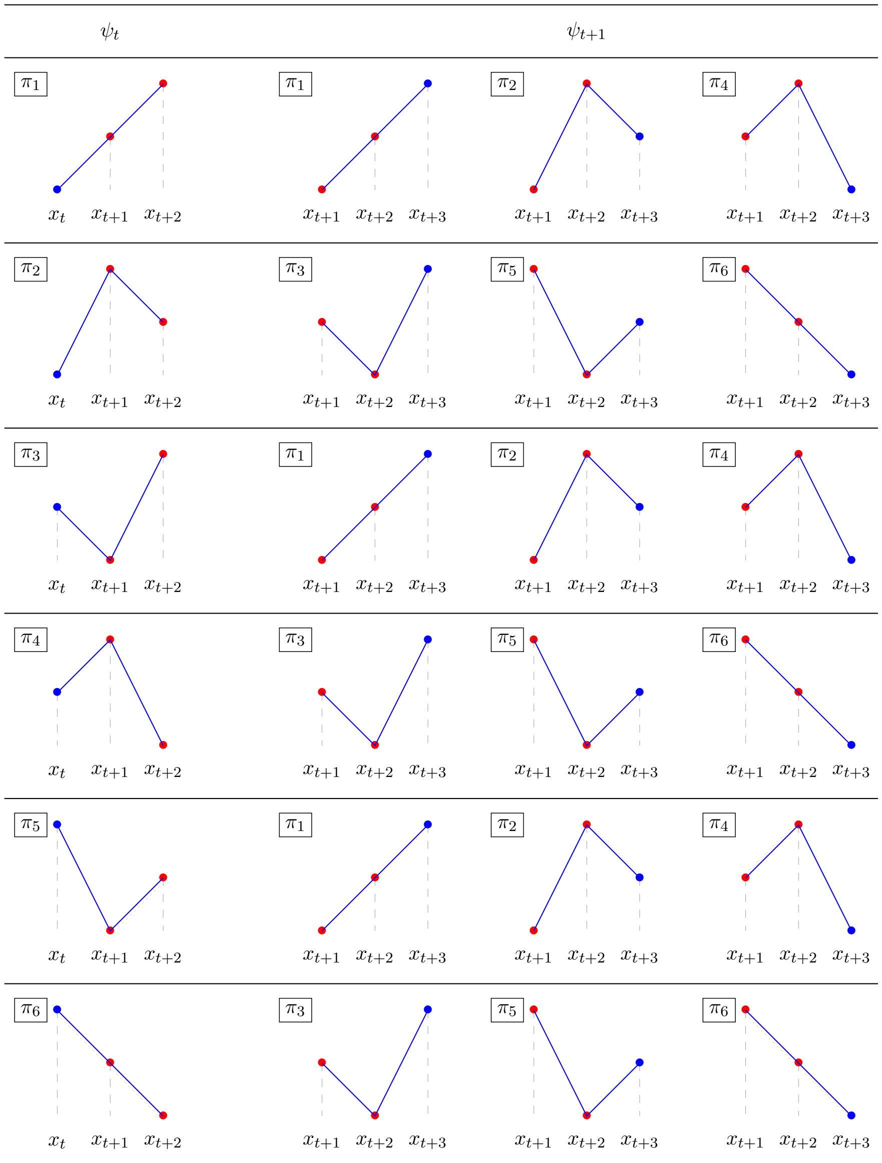
FIG. 1: Ordinal patterns for $m = 3$.

For $\ell = 1$, if $\boldsymbol{x}_3(t) = (x_t, x_{t+1}, x_{t+2})$ is π_i-type then $\boldsymbol{x}_3(t+1) = (x_{t+1}, x_{t+2}, x_{t+3})$ could be π_j-type for only three possible values of j depending on the position of x_{t+3} in the sorted sequence. Table I shows all the possible one-step transitions between symbols.

As a consequence of the forbidden transitions, $q^{(1)}_{ij} = 0$ for $(i,j) \in (I_1 \times I_2) \cup (I_3 \times I_4)$, where $I_1 = \{1,3,5\}$, $I_2 = \{3,5,6\}$, $I_3 = \{2,4,6\}$ and $I_4 = \{1,2,4\}$. On the remaining cases, notice that $q^{(1)}_{ij} = \Pr(\psi_t = \pi_i \wedge \psi_{t+1} = \pi_j) = \Pr(\psi_t = \pi_i)\Pr(\psi_{t+1} = \pi_j \mid \psi_t = \pi_i) = q_i \Pr(\psi_{t+1} = \pi_j \mid \psi_t = \pi_i)$. We compute the conditional probabilities (see Appendix A for details):

$$\Pr(\psi_{t+1} = \pi_j \mid \psi_t = \pi_i) = \begin{cases} 0.50 & \text{if } (i,j) \in \{(1,5),(2,3),(3,6),(4,1),(5,4),(6,2)\}, \\ 0.25 & \text{if } (i,j) \in \{(1,1),(1,3),(2,1),(2,5),(3,2),(3,4), \\ & (4,3),(4,5),(5,2),(5,6),(6,4),(6,6)\}. \end{cases} \tag{23}$$

TABLE I: Possible ordinal pattern transitions for $m = 3$ and $\ell = 1$. Red dots represent common observations in two subsequent patterns; blue dots are the possible next values.



Thus, we conclude that:

$$Q^{(1)} = \begin{pmatrix} 0.25q_1 & 0.25q_1 & 0 & 0.50q_1 & 0 & 0 \\ 0 & 0 & 0.25q_2 & 0 & 0.25q_2 & 0.50q_2 \\ 0.25q_3 & 0.50q_3 & 0 & 0.25q_3 & 0 & 0 \\ 0 & 0 & 0.25q_4 & 0 & 0.50q_4 & 0.25q_4 \\ 0.50q_5 & 0.25q_5 & 0 & 0.25q_5 & 0 & 0 \\ 0 & 0 & 0.50q_6 & 0 & 0.25q_6 & 0.25q_6 \end{pmatrix}. \tag{24}$$

Let us now compute the transitions for $\ell = 2$ (see Appendix A for details), i.e.

$q_{ij}^{(2)} = \Pr(\psi_t = \pi_i \wedge \psi_{t+2} = \pi_j) = \Pr(\psi_t = \pi_i)\Pr(\psi_{t+2} = \pi_j \mid \psi_t = \pi_i) = q_i \Pr(\psi_{t+2} = \pi_j \mid \psi_t = \pi_i)$. The conditional probabilities are:

$$\Pr(\psi_{t+2} = \pi_j \mid \psi_t = \pi_i) = \begin{cases} 0.30 & \text{if } (i,j) \in \{(1,4),(1,6),(2,4),(2,6),(5,1),(5,3),(6,1),(6,3)\}, \\ 0.20 & \text{if } (i,j) \in \{(3,2),(3,4),(4,2),(4,5)\}, \\ 0.15 & \text{if } (i,j) \in \{(1,2),(1,5),(2,2),(2,5),(3,1),(3,3),(3,4),(3,6), \\ & (4,1),(4,3),(4,4),(4,6),(5,1),(5,5),(6,1),(6,5)\}, \\ 0.05 & \text{if } (i,j) \in \{(1,1),(1,3),(2,1),(2,3),(5,4),(5,6),(6,4),(6,6)\}. \end{cases} \tag{25}$$

And then

$$Q^{(2)} = \begin{pmatrix} 0.05q_1 & 0.05q_1 & 0.15q_1 & 0.15q_1 & 0.30q_1 & 0.30q_1 \\ 0.15q_2 & 0.15q_2 & 0.20q_2 & 0.20q_2 & 0.15q_2 & 0.15q_2 \\ 0.05q_3 & 0.05q_3 & 0.15q_3 & 0.15q_3 & 0.30q_3 & 0.30q_3 \\ 0.30q_4 & 0.30q_4 & 0.15q_4 & 0.15q_4 & 0.05q_4 & 0.05q_4 \\ 0.15q_5 & 0.15q_5 & 0.20q_5 & 0.20q_5 & 0.15q_5 & 0.15q_5 \\ 0.30q_6 & 0.30q_6 & 0.15q_6 & 0.15q_6 & 0.05q_6 & 0.05q_6 \end{pmatrix}. \tag{26}$$

These are the only transition matrices required to obtain the asymptotic variance of the Permutation Entropy.

III. COMPARISON WITH THE MULTINOMIAL MODEL

Consider a series of n independent trials in which only one of the K mutually exclusive events is observed with probability $p_1, p_2, \dots, p_K$, respectively, such that $p_i \geq 0$ and $\sum_{i=1}^K p_i = 1$. Let $\boldsymbol{N} = (N_1, N_2, \dots, N_K)$ be the vector of random variables that count the number of occurrences of these events in the n trials, with $\sum_{i=1}^K N_i = n$. Then, if the events are independent, the joint distribution of $\boldsymbol{N}$ is

$$\Pr\big(\boldsymbol{N} = (n_1, n_2, \dots, n_K)\big) = n! \prod_{i=1}^{K} \frac{p_i^{n_i}}{n_i!}, \tag{27}$$

where $n_i \geq 0$ and $\sum_{i=1}^K n_i = n$. This model is the Multinomial distribution with n trials and probability vector $\boldsymbol{p} = (p_1, p_2, \dots, p_K)$.

Chagas et al.[4], using the multivariate Delta method, proved that

$$\sqrt{n}[S(\widehat{\boldsymbol{p}}) - S(\boldsymbol{p})] \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}\big(0, \nu_{\boldsymbol{p}}^2\big), \tag{28}$$

where

$$\nu_{\boldsymbol{p}}^2 = \sum_{i=1}^{K} p_i(1-p_i)(\ln p_i + 1)^2 - 2\sum_{j=1}^{K-1}\sum_{i=j+1}^{K} p_j p_i (\ln p_j + 1)(\ln p_i + 1). \tag{29}$$

The vector of permutation probabilities $\boldsymbol{q} = (q_1, q_2, \dots, q_{m!})$ stems from dependent states and, thus, Eqs. (28) and (29) are the result of a simplification of the problem. However, an interesting question is assessing the independence assumption's cost, which implies a Multinomial distribution. Considering $K = m!$, the asymptotic means given in Eq. (19) and (28) are both zero, but the variances are different. After some calculus, it can be observed the following relation between the asymptotic variances:

$$\sigma_{\boldsymbol{q}}^2 = \nu_{\boldsymbol{q}}^2 - \sum_{i=1}^{m!} (\ln q_i + 1)^2 \left[(2m-2) q_i^2 - 2 \sum_{\ell=1}^{m-1} \mathbf{Q}_{ii}^{(\ell)} \right] - \\ 2 \sum_{i=1}^{m!-1} \sum_{j=i+1}^{m!} (\ln q_i + 1)(\ln q_j + 1) \left[(2m-2) q_i q_j - \sum_{\ell=1}^{m-1} \left(\mathbf{Q}_{ij}^{(\ell)} + \mathbf{Q}_{ji}^{(\ell)} \right) \right]. \quad (30)$$

In the case of white noise where all the q_i are equal, $\nu_{\boldsymbol{q}}^2$ vanishes, but $\sigma_{\boldsymbol{q}}^2$ does not.

IV. EXPERIMENTS WITH SIMULATED DATA

Throughout this section, we consider $m = 3$ as the embedding dimension and $\boldsymbol{x}$ a time series of length $T = n + 2$. Let $\boldsymbol{N} = (N_1, N_2, \dots, N_6)$ be the vector of symbols' frequency, where N_i is the number of symbols of type $\pi^{(i)}$. Note that $\boldsymbol{N}$ is a particular case of the vector of random variables $\boldsymbol{Z}$, defined in Eq. (5). Thus, the total number of symbols is $n = \sum_{i=1}^{6} N_i$. In this case, the probability function of the permutations is $\boldsymbol{q} = (q_1, q_2, \dots, q_6)$ where $q_i = N_i/n$ for $i = 1, 2, \dots, 6$.

In the following, we consider n as a multiple of 6. If $\boldsymbol{x}$ is a white noise time series, then $N_i = n/6$ and $q_i = 1/6$ for all $i = 1, 2, \dots, 6$.

We propose variations of the vector of frequencies given by the following models.

1. One is One (OiO): there exist $j, j' \in \{1, 2, \dots, 6\}$, $j \neq j'$ such that $N_j = 1$, $N_{j'} = n/3 - 1$, and $N_i = n/6$ for $i \neq j, j'$.

2. Half and Half (HaH): there exist three components of the vector $\boldsymbol{N}$ equal to $N_i = n/6 + a$ and the other three are equal to $N_i = n/6 - a$, where $a \in \{1, 2, \dots, n/6 - 1\}$.

3. Linear (Lin): let n be a multiple of 21; i.e., $n = 21b$, the components of the vector $\boldsymbol{N}$ are $b, 2b, 3b, 4b, 5b, 6b$, not necessarily in this order.

Algorithms 1, 2, and 3 can be used to generate time series with models OiO, HaH, and Lin, respectively, depending on two parameters, $\alpha \in \mathbb{R}$ and $\epsilon > 0$. Figure 2 shows time series of length 128 generated by these algorithms using the following parameters:

- Model OiO: $n = 126$, $\alpha = 0$, $\epsilon = 0.01$;

- Model HaH: $n = 126$, $\alpha = 0$, $a = 7$;

- Model Lin: $b = 6$, $\alpha = 20$, $\epsilon = 0.01$.

Algorithm 1: Time series generation under the OiO model.

Input: n multiple of 6, $\alpha \in \mathbb{R}$, $\epsilon > 0$

Output: Time series $\boldsymbol{x}$ length $T = n + 2$

$x_1 \leftarrow \alpha + 0.5$;

$x_2 \leftarrow \alpha$;

$x_3 \leftarrow \alpha + 0.75$;

$x_4 \leftarrow \alpha + 1$;

$x_5 \leftarrow \alpha + 0.25$;

for $i = 6, 7, \ldots, 2n/3 + 2$ **do**

  switch i **do**

    case $i \equiv 0(4)$ **do**

      $x_i \leftarrow \alpha + 0.75$;

    case $i \equiv 1(4)$ **do**

      $x_i \leftarrow \alpha + 0.25$;

    case $i \equiv 2(4)$ **do**

      $x_i \leftarrow \alpha + 0.5$;

    case $i \equiv 3(4)$ **do**

      $x_i \leftarrow \alpha$;

for $j = 0, 1, \ldots, n/3 - 1$ **do**

  $x_{2T/3+3+j} \leftarrow \alpha + 0.375 - j\epsilon$;

Algorithm 2: Time series generation under the HaH model.

Input: n multiple of 6, $\alpha \in \mathbb{R}$, $a \in \{1, 2, \ldots, n/6 - 1\}$

Output: Time series $\boldsymbol{x}$ length $T = n + 2$

$x_1 \leftarrow \alpha$;

$x_2 \leftarrow \alpha + 0.5$;

$\beta \leftarrow \alpha$;

$L = n/6 + a$;

for $i = 1, 2, \ldots, L$ **do**

  $x_{3i} \leftarrow \beta + 1$;

  $x_{3i+1} \leftarrow \beta + 0.75$;

  $x_{3i+2} \leftarrow \beta + 1.25$;

  $\beta \leftarrow x_{3i+2}$;

$\gamma \leftarrow x_{3L+1} - 0.25$;

for $i = 1, 2, \ldots, L - 2a$ **do**

  $x_{3L+3i} \leftarrow \gamma$;

  $x_{3L+3i+1} \leftarrow \gamma - 0.75$;

  $x_{3L+3i+2} \leftarrow \gamma - 0.5$;

  $\gamma \leftarrow x_{3L+3i+1} - 0.25$;

Algorithm 3: Time series generation under the Lin model.

Input: b positive integer, $\alpha \in \mathbb{R}$, $\epsilon > 0$

Output: Time series $\boldsymbol{x}$ of length $T = 21b + 2$

$x_1 \leftarrow \alpha$;

$x_2 \leftarrow \alpha - 2$;

$\beta \leftarrow (x_1 + x_2)/2$;

for $i = 0, 1, \ldots, b-1$ **do**

- $x_{21i+3} \leftarrow \beta$;
- $x_{21i+4} \leftarrow (x_{21i+2} + x_{21i+3})/2$;
- $x_{21i+5} \leftarrow x_{21i+3} + 0.5$;
- $x_{21i+6} \leftarrow x_{21i+4} - 0.5$;
- $x_{21i+7} \leftarrow x_{21i+5} + 0.5$;
- $x_{21i+8} \leftarrow x_{21i+6} - 0.5$;
- $x_{21i+9} \leftarrow (x_{21i+7} + x_{21i+8})/2$;
- $x_{21i+10} \leftarrow x_{21i+8} - 0.5$;
- **for** $j = 1, 2, \ldots, 7$ **do**
  - $x_{21i+10+j} \leftarrow x_{21i+10} + j\epsilon$;
- $x_{21i+18} \leftarrow x_{21i+16} - 0.5$;
- **for** $\ell = 1, 2, \ldots, 5$ **do**
  - $x_{21i+18+\ell} \leftarrow x_{21i+18} + \ell\epsilon$;
- $\beta \leftarrow (x_{21i+22} + x_{21i+23})/2$;

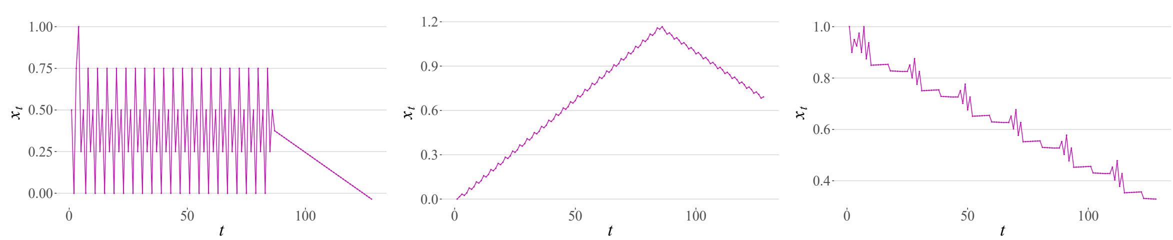


FIG. 2: Examples of time series generated by Algorithms 1 (left), 2 (middle), and 3 (right).

Table II shows the resulting vectors $\boldsymbol{N}$ and $\boldsymbol{q}$, corresponding to the simulated time series. These results are close to the theoretical values.

The asymptotic Permutation Entropy distribution is Normal, whose mean only depends on the vector of permutation probabilities. Then, the mean value is the same for the actual distribution of the ordinal patterns or the simplified distribution. However, the asymptotic variance is different. As we can notice by Eqs. (20) and (29), the expression of the asymptotic variance under the actual distribution of the ordinal patterns is computationally more involved. Figure 3 shows the values of the asymptotic standard

TABLE II: Vectors of frequencies and probabilities of the time series shown in Figure 2.

Model	N_1	N_2	N_3	N_4	N_5	N_6	q_1	q_2	q_3	q_4	q_5	q_6
OiO	1	21	21	21	21	41	0.0079	0.1667	0.1667	0.1667	0.1667	0.3254
HaH	28	28	28	14	14	14	0.2222	0.2222	0.2222	0.1111	0.1111	0.1111
Lin	36	6	12	24	18	30	0.0476	0.1905	0.0952	0.2381	0.1429	0.2857

deviations for the simulated data in function of the value of n for the actual distribution of the ordinal patterns and the Multinomial distribution. The difference between them is smaller as n increases. Moreover, the difference between $\sigma^2_{\boldsymbol{q}}/n$ and $\nu^2_{\boldsymbol{q}}/n$ is less than 10^{-3} for $n \geq 3000$ and the model OiO, and for $n \geq 1000$ and the models HaH and Lin. Thus, the ordinal pattern transitions dependence might be sacrificed for the benefit of the computational cost.

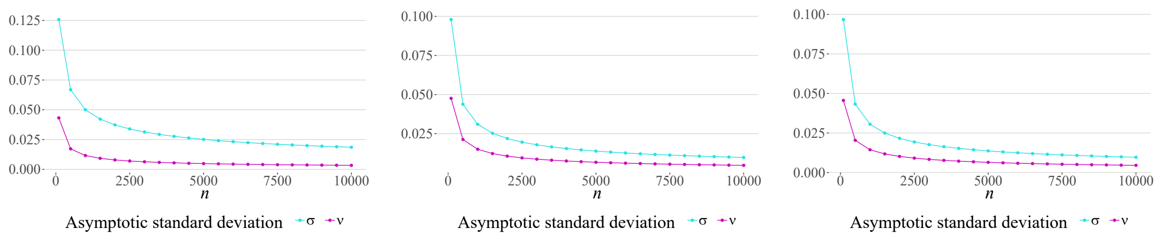


FIG. 3: Asymptotic standard deviations $\sigma_{\boldsymbol{q}}/n$ and $\nu_{\boldsymbol{q}}/n$ for the models OiO (left), HaH (middle), and Lin (right).

V. EXPERIMENTS WITH ACTUAL DATA

This section shows the practical application of asymptotic variance as a valuable tool for distinguishing various dynamics within a set of time series data. Building on the technical insights introduced in Section II A, we illustrate its utility by employing a hypothesis test, initially proposed in Chagas et al.[4], to contrast different time series dynamics while assuming the Multinomial model. Leveraging our primary result concerning the asymptotic distribution as defined in (19), we improve that test to encompass the actual ordinal pattern distribution, incorporating correlation. We then apply this approach to analyzing meteorological data and biological signals. The results suggest that this method is promising for classification across diverse research fields grounded in time-series analysis.

A. Meteorological time-series

In the present study, we used observations of the following meteorological characteristics: minimum and maximum daily temperatures (in Fahrenheit degrees) and daily precipitation. The values were measured in Dublin Phoenix Park (Ireland), Edinburgh Royal Botanic Garden (Scotland), and Miami International Airport (United States of America). These datasets are available at the Climate Data Online website (https://www.ncei.noaa.gov/cdo-web/), which is supported by the National Oceanic and Atmospheric Administration (NOAA). We consider the period from 8 August 1992 to 30 December 2019. These daily observations and the frequencies of their ordinal patterns are shown in Fig. 4. It is worth mentioning that tied data present in these time series are treated using the sequential order. It means that, for example, the sequences $x_t = x_{t+1} < x_{t+2}$ and $x_t < x_{t+1} = x_{t+2}$ are π_1-type. The confidence intervals of these series using the actual distribution of the ordinal patterns and the Multinomial model are shown in Figure 5, and their lengths are presented in Table III. Since the asymptotic variances are more significant if we consider the ordinal pattern correlation, these confidence intervals are wider than the ones obtained assuming independence. Notice that the confidence intervals under the Multinomial model are roughly half the size of the intervals considering the correlation structure. Such difference may impact the comparison of time series, leading to the rejection of the hypothesis that there is no difference when, in fact, there is not enough evidence for that.

TABLE III: Confidence intervals under the true and the Multinomial models of the minimum (top) and maximum (middle) daily temperatures and daily precipitations (bottom) in Dublin, Edinburgh, and Miami, from 8 August 1992 until 30 December 2019.

Feature	City	Mean H	Confidence interval length: True	Confidence interval length: Multinomial	Length difference
Minimum temperature	Dublin	0.97680	0.00023	0.00012	0.00011
	Edinburgh	0.97946	0.00022	0.00011	0.00011
	Miami	0.95747	0.00030	0.00016	0.00014
Maximum temperature	Dublin	0.96755	0.00027	0.00014	0.00013
	Edinburgh	0.97159	0.00025	0.00013	0.00012
	Miami	0.91806	0.00040	0.00022	0.00018
Daily precipitation	Dublin	0.92396	0.00040	0.00021	0.00018
	Edinburgh	0.90310	0.00044	0.00024	0.00020
	Miami	0.78393	0.00058	0.00034	0.00025

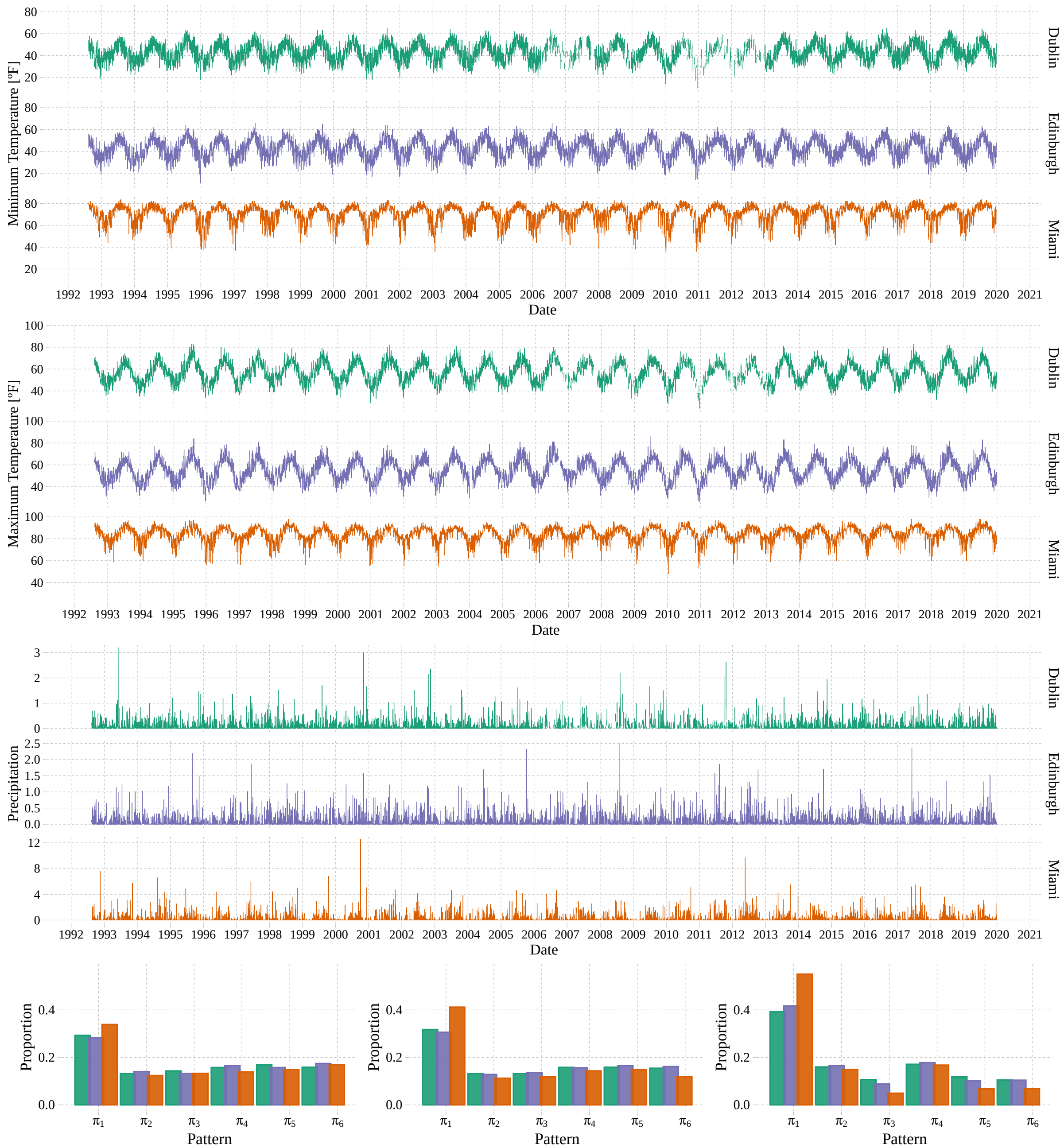


FIG. 4: Minimum (first row) and maximum (second row) daily temperatures, and daily precipitations (third row) in Dublin, Edinburgh, and Miami, from 8 August 1992 until 30 December 2019. Histograms of the ordinal patterns (last row) of the minimum and maximum temperatures and precipitations, from left to right.

Dublin and Edinburgh have similar temperate oceanic climates with similar precipitation and temperature regimes. Conversely, the weather conditions in Miami, a tropical monsoon climate, differ significantly. Table IV shows the p-values of the test that state the same underlying process as the null hypothesis, assuming the actual distribution of the ordinal patterns (with correlation) and the Multinomial model (with independence).

Asymptotic Distribution PE

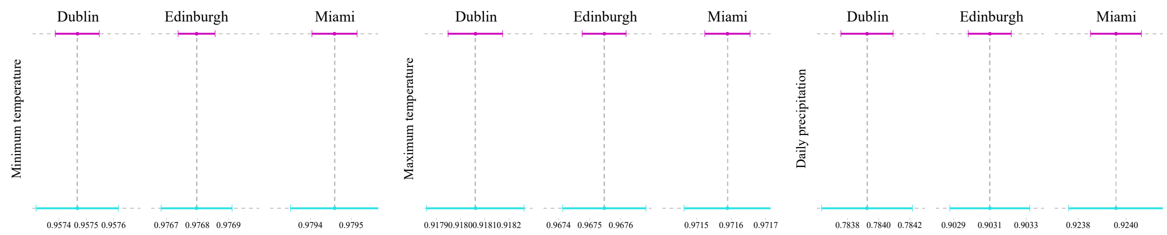


FIG. 5: Confidence intervals under the true (cyan) and the Multinomial (violet) models of the minimum (left) and maximum (middle) daily temperatures, and daily precipitations (right) in Dublin, Edinburgh, and Miami, from 8 August 1992 until 30 December 2019.

It can be seen that the decisions of the tests are the same, assuming independence or not, except for the case of daily precipitation in Dublin and Edinburgh, where the null hypothesis is rejected at 5 % of significance under the Multinomial model. However, they are not rejected at the same level when considering the correlation among patterns.

TABLE IV: Results of p-values for testing the same dynamic of the meteorological time series.

Data	Model	Dublin / Edinburgh	Dublin / Miami	Edinburgh / Miami
Minimum	True	0.7379	4.41×10^{-2}	1.95×10^{-2}
Temperature	Multinomial	0.5192	1.51×10^{-4}	1.03×10^{-5}
Maximum	True	0.6627	4.99×10^{-5}	7.89×10^{-6}
Temperature	Multinomial	0.4098	2.67×10^{-13}	6.66×10^{-16}
Precipitation	True	0.1603	8.88×10^{-16}	4.22×10^{-11}
	Multinomial	0.0102	0.0000	0.0000

B. Biologogical signals

In order to assess the versatility of our approach across various contexts, we conducted an analysis using a dataset of electrocardiograms (ECGs) sourced from the PhysioNet platform of the Computational Physiology Laboratory at the Massachusetts Institute of Technology, available at https://physionet.org/. This dataset encompasses ECGs depicting normal sinus rhythm[7] and ECGs from patients exhibiting cardiac arrhythmias[10].

To illustrate some of the potential benefits of applying our theoretical findings, we randomly selected ten signals from this repository. Five ECGs corresponded to normal rhythm, while the remaining five were associated with arrhythmia. For clarity and brevity, we standardized the nomenclature of these signals, as presented in Table V.

The subsequent hypothesis test was conducted using the initial 10 000 observations from each signal, and the results are visualized in Figure 6.

TABLE V: Selected ECG signals.

ECG Group	PhysioNet id	Our id
Normal rhythm	16265	N1
	16273	N2
	16420	N3
	16786	N4
	19830	N5
Arrhythmia	109	A1
	116	A2
	208	A3
	212	A4
	222	A5

We applied two tests, one considering the actual model of the ordinal patterns and the other one under the Multinomial model that assumes independence. The null hypothesis of our tests states that two time series share the same dynamic. Each ECG record was compared with the rest of the signals. The decisions are shown in Table VI: "NR" denotes not a rejection, and "R" is a rejection. It can be noticed that when the actual model was used, all the ECG with normal rhythm were grouped in the same class. The same held for the ECG with arrhythmia. Moreover, the null hypothesis was rejected in most comparisons between a signal with regular rhythm and one with arrhythmia. On the other hand, when the test was applied assuming independence, the errors of Type I and Type II increased.

VI. CONCLUSIONS

In this work, we computed the Shannon entropy asymptotic distribution of the ordinal patterns obtained from a time series. The overlapping sequences that define the ordinal patterns induce their serial dependence. Thus, the asymptotic distribution expression requires a large amount of computational effort. We compared these results with the Shannon entropy asymptotic distribution of ordinal patterns under the Multinomial model, which assumes independence. Both asymptotic distributions are Normal with the same asymptotic mean. We studied the relationship between the asymptotic variances.

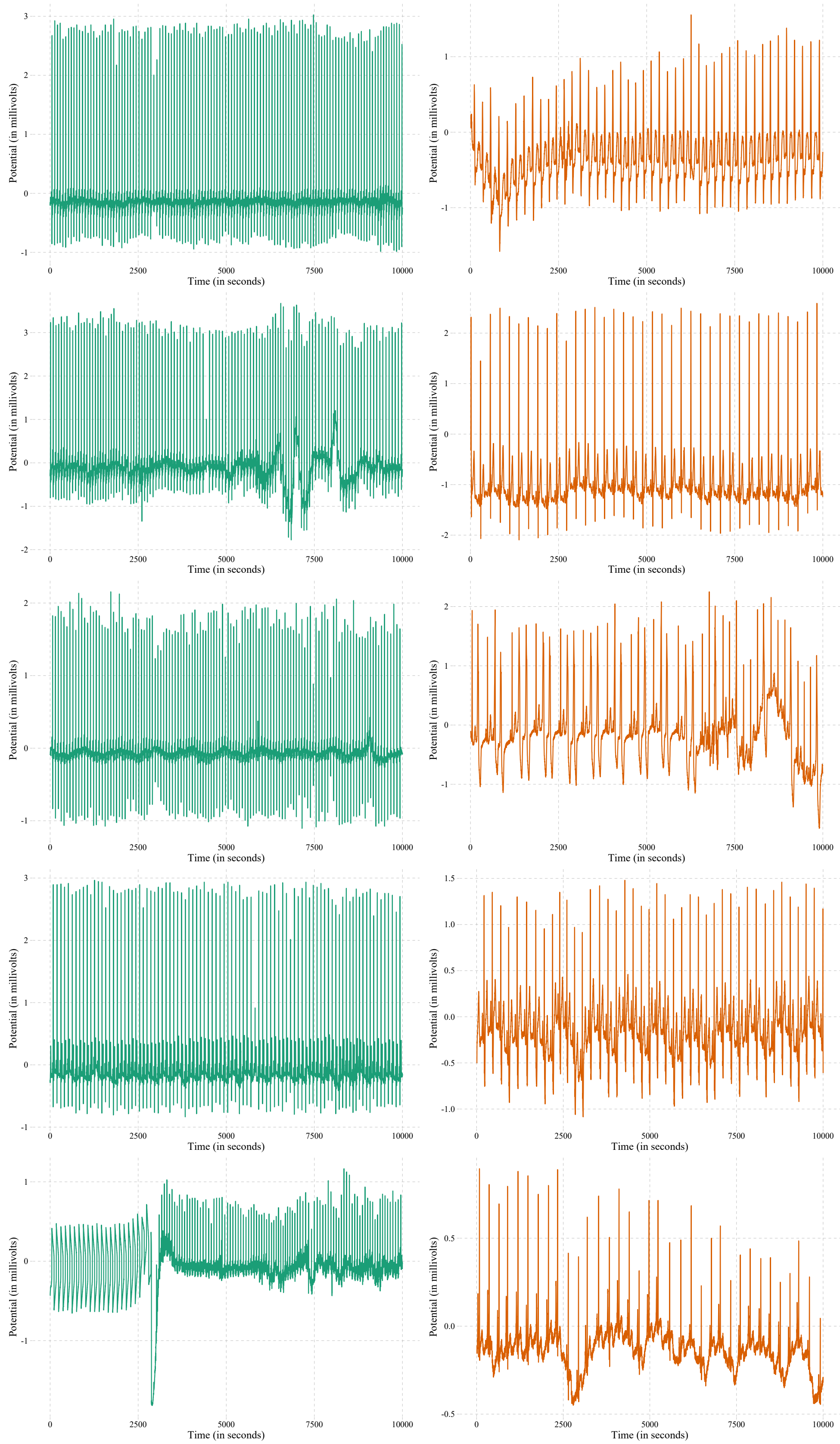


FIG. 6: Selected ECG from PhysioNet database. Normal rhythm: N1, N2, N3, N4, and N5 (left column from top to bottom). Arrhythmia: A1, A2, A3, A4, and A5 (right column from top to bottom).

TABLE VI: Decisions of applying the test under the true model (left) and the Multinomial model (right); "NR" denotes there is no significant statistical evidence to reject the null hypothesis, and "R" means that there is.

	True Model											**Multinomial Model**									
	N1	**N2**	**N3**	**N4**	**N5**	**A1**	**A2**	**A3**	**A4**	**A5**		**N1**	**N2**	**N3**	**N4**	**N5**	**A1**	**A2**	**A3**	**A4**	**A5**
N1	NR	NR	NR	NR	NR	R	R	NR	NR	R	**N1**	NR	R	R	R	R	R	R	R	R	NR
N2		NR	NR	NR	NR	R	R	R	R	R	**N2**		NR	NR	NR	NR	NR	NR	NR	NR	NR
N3			NR	NR	NR	R	R	NR	R	R	**N3**			NR	R	R	NR	NR	R	R	NR
N4				NR	NR	R	R	R	R	R	**N4**				NR	NR	NR	NR	NR	NR	R
N5					NR	R	R	R	R	NR	**N5**					NR	NR	NR	NR	NR	R
A1						NR	NR	NR	NR	NR	**A1**						NR	NR	NR	NR	R
A2							NR	NR	NR	NR	**A2**							NR	NR	NR	R
A3								NR	NR	NR	**A3**								NR	NR	R
A4									NR	NR	**A4**									NR	R
A5										NR	**A5**										NR

We found that: *(i)* the variance of the actual model is larger than the variance of the Multinomial model, and *(ii)* the two variances coincide for a suitable selection of the time series lengths.

For an embedding dimension equal to 3, we found the explicit expression of the asymptotic distribution. We quantified the difference between the asymptotic variances with serial correlation (the actual model) and under independence (the Multinomial model) We applied hypothesis tests with and without the independence assumption to meteorological time series from three locations, two with similar regimes and a third one with different climate conditions. Only in one case, concerning the daily precipitation, does the decision differ. Finally, we applied the same tests to ECG signals with regular rhythm and arrhythmia. In this case, the independence assumption yields more cases with a wrong decision on whether to reject the null hypothesis. Accordingly, our theoretical result is a suitable tool for classifying time series.

Appendix A CONDITIONAL PROBABILITIES

This section presents a detailed computation of the conditional probabilities used in Section II A. The results are obtained for $\ell = 1$ and $m = 3$.

- If $\psi_t = \pi_1$, $x_t \le x_{t+1} \le x_{t+2}$, then there are three possibilities:
  - $x_{t+3} \le x_{t+1} \le x_{t+2}$, i.e. $\psi_{t+1} = \pi_4$, provided that $x_{t+3} \le x_t \le x_{t+1} \le x_{t+2}$ or $x_t \le x_{t+3} \le x_{t+1} \le x_{t+2}$. Thus, $\Pr(\psi_{t+1} = \pi_4 \mid \psi_t = \pi_1) = 1/2$.

  - $x_{t+1} \leq x_{t+3} \leq x_{t+2}$, i.e. $\psi_{t+1} = \pi_2$, provided that $x_t \leq x_{t+1} \leq x_{t+3} \leq x_{t+2}$. Thus, $\Pr(\psi_{t+1} = \pi_2 \mid \psi_t = \pi_1) = 1/4$.
  - $x_{t+1} \leq x_{t+2} \leq x_{t+3}$, i.e. $\psi_{t+1} = \pi_1$, provided that $x_t \leq x_{t+1} \leq x_{t+2} \leq x_{t+3}$. Thus, $\Pr(\psi_{t+1} = \pi_1 \mid \psi_t = \pi_1) = 1/4$.

- If $\psi_t = \pi_2$, $x_t \leq x_{t+2} \leq x_{t+1}$, then there are three possibilities:
  - $x_{t+3} \leq x_{t+2} \leq x_{t+1}$, i.e. $\psi_{t+1} = \pi_6$, provided that $x_{t+3} \leq x_t \leq x_{t+2} \leq x_{t+1}$ or $x_t \leq x_{t+3} \leq x_{t+2} \leq x_{t+1}$. Thus, $\Pr(\psi_{t+1} = \pi_6 \mid \psi_t = \pi_2) = 1/2$.
  - $x_{t+2} \leq x_{t+3} \leq x_{t+1}$, i.e. $\psi_{t+1} = \pi_5$, provided that $x_t \leq x_{t+2} \leq x_{t+3} \leq x_{t+1}$. Thus, $\Pr(\psi_{t+1} = \pi_5 \mid \psi_t = \pi_2) = 1/4$.
  - $x_{t+2} \leq x_{t+1} \leq x_{t+3}$, i.e. $\psi_{t+1} = \pi_3$, provided that $x_t \leq x_{t+2} \leq x_{t+1} \leq x_{t+3}$. Thus, $\Pr(\psi_{t+1} = \pi_3 \mid \psi_t = \pi_2) = 1/4$.

- If $\psi_t = \pi_3$, $x_{t+1} \leq x_t \leq x_{t+2}$, then there are three possibilities:
  - $x_{t+3} \leq x_{t+1} \leq x_{t+2}$, i.e. $\psi_{t+1} = \pi_4$, provided that $x_{t+3} \leq x_{t+1} \leq x_t \leq x_{t+2}$. Thus, $\Pr(\psi_{t+1} = \pi_4 \mid \psi_t = \pi_3) = 1/4$.
  - $x_{t+1} \leq x_{t+3} \leq x_{t+2}$, i.e. $\psi_{t+1} = \pi_2$, provided that $x_{t+1} \leq x_t \leq x_{t+3} \leq x_{t+2}$ or $x_{t+1} \leq x_{t+3} \leq x_t \leq x_{t+2}$. Thus, $\Pr(\psi_{t+1} = \pi_2 \mid \psi_t = \pi_3) = 1/2$.
  - $x_{t+1} \leq x_{t+2} \leq x_{t+3}$, i.e. $\psi_{t+1} = \pi_1$, provided that $x_{t+1} \leq x_t \leq x_{t+2} \leq x_{t+3}$. Thus, $\Pr(\psi_{t+1} = \pi_1 \mid \psi_t = \pi_3) = 1/4$.

- If $\psi_t = \pi_4$, $x_{t+2} \leq x_t \leq x_{t+1}$, then there are three possibilities:
  - $x_{t+3} \leq x_{t+2} \leq x_{t+1}$, i.e. $\psi_{t+1} = \pi_6$, provided that $x_{t+3} \leq x_{t+2} \leq x_t \leq x_{t+1}$. Thus, $\Pr(\psi_{t+1} = \pi_6 \mid \psi_t = \pi_4) = 1/4$.
  - $x_{t+2} \leq x_{t+3} \leq x_{t+1}$, i.e. $\psi_{t+1} = \pi_5$, provided that $x_{t+2} \leq x_t \leq x_{t+3} \leq x_{t+1}$ or $x_{t+2} \leq x_{t+3} \leq x_t \leq x_{t+1}$. Thus, $\Pr(\psi_{t+1} = \pi_5 \mid \psi_t = \pi_4) = 1/2$.
  - $x_{t+2} \leq x_{t+1} \leq x_{t+3}$, i.e. $\psi_{t+1} = \pi_3$, provided that $x_{t+2} \leq x_t \leq x_{t+1} \leq x_{t+3}$. Thus, $\Pr(\psi_{t+1} = \pi_3 \mid \psi_t = \pi_4) = 1/4$.

- If $\psi_t = \pi_5$, $x_{t+1} \leq x_{t+2} \leq x_t$, then there are three possibilities:
  - $x_{t+3} \leq x_{t+1} \leq x_{t+2}$, i.e. $\psi_{t+1} = \pi_4$, provided that $x_{t+3} \leq x_{t+1} \leq x_{t+2} \leq x_t$. Thus, $\Pr(\psi_{t+1} = \pi_4 \mid \psi_t = \pi_5) = 1/4$.
  - $x_{t+1} \leq x_{t+3} \leq x_{t+2}$, i.e. $\psi_{t+1} = \pi_2$, provided that $x_{t+1} \leq x_{t+3} \leq x_{t+2} \leq x_t$. Thus, $\Pr(\psi_{t+1} = \pi_2 \mid \psi_t = \pi_5) = 1/4$.
  - $x_{t+1} \leq x_{t+2} \leq x_{t+3}$, i.e. $\psi_{t+1} = \pi_1$, provided that $x_{t+1} \leq x_{t+2} \leq x_t \leq x_{t+3}$ or $x_{t+1} \leq x_{t+2} \leq x_{t+3} \leq x_t$. Thus, $\Pr(\psi_{t+1} = \pi_1 \mid \psi_t = \pi_5) = 1/2$.

- If $\psi_t = \pi_6$, $x_{t+2} \leq x_{t+1} \leq x_t$, then there are three possibilities:
  - $x_{t+3} \leq x_{t+2} \leq x_{t+1}$, i.e. $\psi_{t+1} = \pi_6$, provided that $x_{t+3} \leq x_{t+2} \leq x_{t+1} \leq x_t$. Thus, $\Pr(\psi_{t+1} = \pi_6 \mid \psi_t = \pi_6) = 1/4$.
  - $x_{t+2} \leq x_{t+3} \leq x_{t+1}$, i.e. $\psi_{t+1} = \pi_5$, provided that $x_{t+2} \leq x_{t+3} \leq x_{t+1} \leq x_t$. Thus, $\Pr(\psi_{t+1} = \pi_5 \mid \psi_t = \pi_6) = 1/4$.
  - $x_{t+2} \leq x_{t+1} \leq x_{t+3}$, i.e. $\psi_{t+1} = \pi_3$, provided that $x_{t+2} \leq x_{t+1} \leq x_{t+3} \leq x_t$ or $x_{t+2} \leq x_{t+1} \leq x_t \leq x_{t+3}$. Thus, $\Pr(\psi_{t+1} = \pi_3 \mid \psi_t = \pi_6) = 1/2$.

Analogously to the previous computations, we proceed as follows for $\ell = 2$.

- If $\psi_t = \pi_1$, $x_t \leq x_{t+1} \leq x_{t+2}$, then there are six possibilities:
  - $x_{t+2} \leq x_{t+3} \leq x_{t+4}$, i.e. $\psi_{t+2} = \pi_1$, provided that $x_t \leq x_{t+1} \leq x_{t+2} \leq x_{t+3} \leq x_{t+4}$. Thus, $\Pr(\psi_{t+1} = \pi_1 \mid \psi_t = \pi_1) = (1/4) \cdot (1/5) = 1/20$.
  - $x_{t+2} \leq x_{t+4} \leq x_{t+3}$, i.e. $\psi_{t+2} = \pi_2$, provided that $x_t \leq x_{t+1} \leq x_{t+2} \leq x_{t+4} \leq x_{t+3}$. Thus, $\Pr(\psi_{t+1} = \pi_2 \mid \psi_t = \pi_1) = (1/4) \cdot (1/5) = 1/20$.
  - $x_{t+3} \leq x_{t+2} \leq x_{t+4}$, i.e. $\psi_{t+2} = \pi_3$, provided that $x_{t+3} \leq x_t \leq x_{t+1} \leq x_{t+2} \leq x_{t+4}$, $x_t \leq x_{t+3} \leq x_{t+1} \leq x_{t+2} \leq x_{t+4}$ or $\leq x_t \leq x_{t+1} \leq x_{t+3} \leq x_{t+2} \leq x_{t+4}$. Thus, $\Pr(\psi_{t+1} = \pi_3 \mid \psi_t = \pi_1) = (3/4) \cdot (1/5) = 0.15$.
  - $x_{t+4} \leq x_{t+2} \leq x_{t+3}$, i.e. $\psi_{t+2} = \pi_4$, provided that $x_{t+4} \leq x_t \leq x_{t+1} \leq x_{t+2} \leq x_{t+3}$, $x_t \leq x_{t+4} \leq x_{t+1} \leq x_{t+2} \leq x_{t+3}$ or $x_t \leq x_{t+1} \leq x_{t+4} \leq x_{t+2} \leq x_{t+3}$. Thus, $\Pr(\psi_{t+1} = \pi_4 \mid \psi_t = \pi_1) = (1/4) \cdot (3/5) = 0.15$.
  - $x_{t+3} \leq x_{t+4} \leq x_{t+2}$, i.e. $\psi_{t+2} = \pi_5$, provided that $x_{t+3} \leq x_{t+4} \leq x_t \leq x_{t+1} \leq x_{t+2}$, $x_{t+3} \leq x_t \leq x_{t+4} \leq x_{t+1} \leq x_{t+2}$, $x_{t+3} \leq x_t \leq x_{t+1} \leq x_{t+4} \leq x_{t+2}$, $x_t \leq x_{t+3} \leq x_{t+4} \leq x_{t+1} \leq x_{t+2}$, $x_t \leq x_{t+3} \leq x_{t+1} \leq x_{t+4} \leq x_{t+2}$, or $\leq x_t \leq x_{t+1} \leq x_{t+3} \leq x_{t+4} \leq x_{t+2}$. Thus, $\Pr(\psi_{t+1} = \pi_5 \mid \psi_t = \pi_1) = 3!/20$.
  - $x_{t+4} \leq x_{t+3} \leq x_{t+2}$, i.e. $\psi_{t+2} = \pi_6$, provided that $x_{t+4} \leq x_{t+3} \leq x_t \leq x_{t+1} \leq x_{t+2}$, $x_{t+4} \leq x_t \leq x_{t+3} \leq x_{t+1} \leq x_{t+2}$, $x_{t+4} \leq x_t \leq x_{t+1} \leq x_{t+3} \leq x_{t+2}$, $x_t \leq x_{t+4} \leq x_{t+3} \leq x_{t+1} \leq x_{t+2}$, $x_t \leq x_{t+4} \leq x_{t+1} \leq x_{t+3} \leq x_{t+2}$ or $\leq x_t \leq x_{t+1} \leq x_{t+4} \leq x_{t+3} \leq x_{t+2}$. Thus, $\Pr(\psi_{t+1} = \pi_6 \mid \psi_t = \pi_1) = 3!/20$.
- If $\psi_t = \pi_2$, $x_t \leq x_{t+2} \leq x_{t+1}$, then there are six possibilities:
  - $x_{t+2} \leq x_{t+3} \leq x_{t+4}$, i.e. $\psi_{t+2} = \pi_1$, there are 3 ways to locate $x_{t+3} \leq x_{t+4}$. Thus, $\Pr(\psi_{t+1} = \pi_1 \mid \psi_t = \pi_2) = 3/20$.
  - $x_{t+2} \leq x_{t+4} \leq x_{t+3}$, i.e. $\psi_{t+2} = \pi_2$, there are 3 ways to locate $x_{t+4} \leq x_{t+3}$. Thus, $\Pr(\psi_{t+1} = \pi_2 \mid \psi_t = \pi_2) = 3/20$.
  - $x_{t+3} \leq x_{t+2} \leq x_{t+4}$, i.e. $\psi_{t+2} = \pi_3$, there are 2 ways to locate x_{t+3} and 2 ways to locate x_{t+4} Thus, $\Pr(\psi_{t+1} = \pi_3 \mid \psi_t = \pi_2) = 1/5$.

- $x_{t+4} \leq x_{t+2} \leq x_{t+3}$, i.e. $\psi_{t+2} = \pi_4$, there are 2 ways to locate x_{t+3} and 2 ways to locate x_{t+4}. Thus, $\Pr(\psi_{t+1} = \pi_4 \mid \psi_t = \pi_2) = 1/5$.
- $x_{t+3} \leq x_{t+4} \leq x_{t+2}$, i.e. $\psi_{t+2} = \pi_5$, there are 3 ways to locate $x_{t+3} \leq x_{t+4}$. Thus, $\Pr(\psi_{t+1} = \pi_5 \mid \psi_t = \pi_1) = 3/20$.
- $x_{t+4} \leq x_{t+3} \leq x_{t+2}$, i.e. $\psi_{t+2} = \pi_6$, there are 3 ways to locate $x_{t+4} \leq x_{t+3}$. Thus, $\Pr(\psi_{t+1} = \pi_6 \mid \psi_t = \pi_1) = 3/20$.

- If $\psi_t = \pi_3$, $x_{t+1} \leq x_t \leq x_{t+2}$, then there are six possibilities:
  - $x_{t+2} \leq x_{t+3} \leq x_{t+4}$, i.e. $\psi_{t+2} = \pi_1$, there is only one way to locate $x_{t+3} \leq x_{t+4}$. Thus, $\Pr(\psi_{t+1} = \pi_1 \mid \psi_t = \pi_3) = 1/20 = 1/20$.
  - $x_{t+2} \leq x_{t+4} \leq x_{t+3}$, i.e. $\psi_{t+2} = \pi_2$, there is only one way to locate $x_{t+4} \leq x_{t+3}$. Thus, $\Pr(\psi_{t+1} = \pi_2 \mid \psi_t = \pi_3) = 1/20$.
  - $x_{t+3} \leq x_{t+2} \leq x_{t+4}$, i.e. $\psi_{t+2} = \pi_3$, there are 3 ways to locate x_{t+3} and only one way to locate x_{t+4} Thus, $\Pr(\psi_{t+1} = \pi_3 \mid \psi_t = \pi_3) = 3/20$.
  - $x_{t+4} \leq x_{t+2} \leq x_{t+3}$, i.e. $\psi_{t+2} = \pi_4$, there are 3 ways to locate x_{t+4} and only one way to locate x_{t+3}. Thus, $\Pr(\psi_{t+1} = \pi_4 \mid \psi_t = \pi_3) = 3/20$.
  - $x_{t+3} \leq x_{t+4} \leq x_{t+2}$, i.e. $\psi_{t+2} = \pi_5$, there are 3! ways to locate $x_{t+3} \leq x_{t+4}$. Thus, $\Pr(\psi_{t+1} = \pi_5 \mid \psi_t = \pi_3) = 3!/20$.
  - $x_{t+4} \leq x_{t+3} \leq x_{t+2}$, i.e. $\psi_{t+2} = \pi_6$, there are 3! ways to locate $x_{t+4} \leq x_{t+3}$. Thus, $\Pr(\psi_{t+1} = \pi_6 \mid \psi_t = \pi_3) = 3!/20$.

- If $\psi_t = \pi_4$, $x_{t+2} \leq x_t \leq x_{t+1}$, then there are six possibilities:
  - $x_{t+2} \leq x_{t+3} \leq x_{t+4}$, i.e. $\psi_{t+2} = \pi_1$, there are 3! ways to locate $x_{t+3} \leq x_{t+4}$. Thus, $\Pr(\psi_{t+1} = \pi_1 \mid \psi_t = \pi_4) = 3!/20$.
  - $x_{t+2} \leq x_{t+4} \leq x_{t+3}$, i.e. $\psi_{t+2} = \pi_2$, there are 3! ways to locate $x_{t+4} \leq x_{t+3}$. Thus, $\Pr(\psi_{t+1} = \pi_2 \mid \psi_t = \pi_4) = 3!/20$.
  - $x_{t+3} \leq x_{t+2} \leq x_{t+4}$, i.e. $\psi_{t+2} = \pi_3$, there is only one way to locate x_{t+3} and 3 ways to locate x_{t+4} Thus, $\Pr(\psi_{t+1} = \pi_3 \mid \psi_t = \pi_4) = 3/20$.
  - $x_{t+4} \leq x_{t+2} \leq x_{t+3}$, i.e. $\psi_{t+2} = \pi_4$, there 3 ways to locate x_{t+3} and only one way to locate x_{t+4}. Thus, $\Pr(\psi_{t+1} = \pi_4 \mid \psi_t = \pi_4) = 3/20$.
  - $x_{t+3} \leq x_{t+4} \leq x_{t+2}$, i.e. $\psi_{t+2} = \pi_5$, there is only one way to locate $x_{t+3} \leq x_{t+4}$. Thus, $\Pr(\psi_{t+1} = \pi_5 \mid \psi_t = \pi_4) = 1/20$.
  - $x_{t+4} \leq x_{t+3} \leq x_{t+2}$, i.e. $\psi_{t+2} = \pi_6$, there is only one way to locate $x_{t+4} \leq x_{t+3}$. Thus, $\Pr(\psi_{t+1} = \pi_6 \mid \psi_t = \pi_4) = 1/20$.

- If $\psi_t = \pi_5$, $x_{t+1} \leq x_{t+2} \leq x_t$, then there are six possibilities:

  - $x_{t+2} \leq x_{t+3} \leq x_{t+4}$, i.e. $\psi_{t+2} = \pi_1$, there are 3 ways to locate $x_{t+3} \leq x_{t+4}$. Thus, $\Pr(\psi_{t+1} = \pi_1 \mid \psi_t = \pi_5) = 3/20$.
  - $x_{t+2} \leq x_{t+4} \leq x_{t+3}$, i.e. $\psi_{t+2} = \pi_2$, there are 3 ways to locate $x_{t+4} \leq x_{t+3}$. Thus, $\Pr(\psi_{t+1} = \pi_2 \mid \psi_t = \pi_5) = 3/20$.
  - $x_{t+3} \leq x_{t+2} \leq x_{t+4}$, i.e. $\psi_{t+2} = \pi_3$, there are 2 ways to locate x_{t+3} and 2 ways to locate x_{t+4} Thus, $\Pr(\psi_{t+1} = \pi_3 \mid \psi_t = \pi_5) = 1/5$.
  - $x_{t+4} \leq x_{t+2} \leq x_{t+3}$, i.e. $\psi_{t+2} = \pi_4$, there are 2 ways to locate x_{t+3} and 2 ways to locate x_{t+4}. Thus, $\Pr(\psi_{t+1} = \pi_4 \mid \psi_t = \pi_5) = 1/5$.
  - $x_{t+3} \leq x_{t+4} \leq x_{t+2}$, i.e. $\psi_{t+2} = \pi_5$, there are 3 ways to locate $x_{t+3} \leq x_{t+4}$. Thus, $\Pr(\psi_{t+1} = \pi_5 \mid \psi_t = \pi_5) = 3/20$.
  - $x_{t+4} \leq x_{t+3} \leq x_{t+2}$, i.e. $\psi_{t+2} = \pi_6$, there are 3 ways to locate $x_{t+4} \leq x_{t+3}$. Thus, $\Pr(\psi_{t+1} = \pi_6 \mid \psi_t = \pi_5) = 3/20$.

- If $\psi_t = \pi_6$, $x_{t+2} \leq x_{t+1} \leq x_t$, then there are six possibilities:
  - $x_{t+2} \leq x_{t+3} \leq x_{t+4}$, i.e. $\psi_{t+2} = \pi_1$, there are 3! ways to locate $x_{t+3} \leq x_{t+4}$. Thus, $\Pr(\psi_{t+1} = \pi_1 \mid \psi_t = \pi_6) = 3!/20$.
  - $x_{t+2} \leq x_{t+4} \leq x_{t+3}$, i.e. $\psi_{t+2} = \pi_2$, there are 3! ways to locate $x_{t+4} \leq x_{t+3}$. Thus, $\Pr(\psi_{t+1} = \pi_2 \mid \psi_t = \pi_6) = 3!/20$.
  - $x_{t+3} \leq x_{t+2} \leq x_{t+4}$, i.e. $\psi_{t+2} = \pi_3$, there are 3 ways to locate x_{t+4} and only one way to locate x_{t+3} Thus, $\Pr(\psi_{t+1} = \pi_3 \mid \psi_t = \pi_6) = 3/20$.
  - $x_{t+4} \leq x_{t+2} \leq x_{t+3}$, i.e. $\psi_{t+2} = \pi_4$, there are 3 ways to locate x_{t+3} and only one way to locate x_{t+4}. Thus, $\Pr(\psi_{t+1} = \pi_4 \mid \psi_t = \pi_6) = 3/20$.
  - $x_{t+3} \leq x_{t+4} \leq x_{t+2}$, i.e. $\psi_{t+2} = \pi_5$, there is only one way to locate $x_{t+3} \leq x_{t+4}$. Thus, $\Pr(\psi_{t+1} = \pi_5 \mid \psi_t = \pi_6) = 1/20$.
  - $x_{t+4} \leq x_{t+3} \leq x_{t+2}$, i.e. $\psi_{t+2} = \pi_6$, there is only one way to locate $x_{t+4} \leq x_{t+3}$. Thus, $\Pr(\psi_{t+1} = \pi_6 \mid \psi_t = \pi_6) = 1/20$.

CONFLICT OF INTEREST STATEMENT

The authors have no conflicts of interest to disclose.

REPRODUCIBILITY AND REPLICABILITY

The public repository https://gitlab.ecs.vuw.ac.nz/freryal/comparison-asymptotic-models-ordinal-patterns contains all the artefacts (text, code, and data) related to this article.

REFERENCES

[1] J. M. Amigó and O. A. Rosso. Ordinal methods: Concepts, applications, new developments, and challenges—in memory of Karsten Keller (1961–2022). *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(8), aug 2023. doi:10.1063/5.0167263.

[2] Christoph Bandt and Bernd Pompe. Permutation entropy: a natural complexity measure for time series. *Physical Review Letters*, 88(17):174102, 2002. doi: 10.1103/PhysRevLett.88.174102.

[3] E. Chagas, A. C. Frery, O. A. Rosso, and H. S. Ramos. Analysis and classification of SAR textures using Information Theory. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:663–675, 2021. doi: 10.1109/JSTARS.2020.3031918.

[4] E. T. C. Chagas, A. C. Frery, J. Gambini, M. M. Lucini, H. S. Ramos, and A. A. Rey. Statistical properties of the entropy from ordinal patterns. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 32:113118, 2022. doi:10.1063/5.0118706.

[5] E. T. C. Chagas, M. Queiroz-Oliveira, O. A. Rosso, H. S. Ramos, C. G. S. Freitas, and A. C. Frery. White noise test from ordinal patterns in the Entropy-Complexity plane. *International Statistical Review*, 2022. doi:10.1111/insr.12487.

[6] H. Elsinger. Independence tests based on symbolic dynamics. Technical report, Working Paper, 2010.

[7] Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220, 2000. doi:10.1161/01.cir.101.23.e215.

[8] E. L Lehmann and G. Casella. *Theory of point estimation*. Springer Science & Business Media, 2006.

[9] I. Leyva, J. Martínez, C. Masoller, O. Rosso, and M. Zanin. 20 years of ordinal patterns: Perspectives and challenges. *Europhysics LettersEuropean Physical Society*, 138(3), 2022. doi:10.1209/0295-5075/ac6a72.

[10] George B. Moody and Roger G. Mark. The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3):45–50, 2001. doi:10.1109/51.932724.

[11] A. A. Rey, A. C. Frery, M. Lucini, J. Gambini, E. T. C. Chagas, and H. S. Ramos. Asymptotic distribution of certain types of entropy under the multinomial law. *Entropy*, 25(5), 2023. ISSN 1099-4300. doi:10.3390/e25050734.

[12] A. M. Yamashita Rios de Sousa and J. Hlinka. Assessing serial dependence in ordinal patterns processes using chi-squared tests with application to EEG data analysis. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 32(7):073126, 2022. doi: 10.1063/5.0096954.