New Features of Permutation Entropy in Ordinal Patterns Complexity Plane

Rasika Dilhani

Supervisor: Alejandro C. Frery

Submitted in partial fullfilment of the requirements
for Doctor of Philosophy
in Data Science.

Victoria University of Wellington
2025

Abstract

This proposal provides a brief introduction to ordinal patterns and a literature review related to the Bandt & Pompe research article. After reviewing the literature on ordinal patterns and permutation entropy, we can conclude that ordinal patterns play an important role in conducting research. Bandt & Pompe proposed analyzing time series from the perspective of ordinal patterns to develop fast and automated methods for extracting qualitative information from nonlinear time series. They introduced the concept of permutation entropy to measure the complexity of a system underlying a time series while taking into account the ordering patterns that represent variations in the data. In this report I show how to use data to extract all ordinal patterns. First, I use the sample dataset to calculate their ordinal patterns with embedding dimension 3; I use letters to convert time series data into patterns. Secondly, I get the histogram on the data. And third, calculate the permutation entropy and statistical complexity. Next, determine the confidence interval of H without and with time dependence and finally the confidence interval of C.

Bandt and Pompe's ordinal pattern methodology has been widely used to investigate the latent dynamics of time series through their entropy, also known as permutation entropy. Nevertheless, there are no theoretical findings regarding the distribution of the permutation entropy, which needs to take the correlation effect between patterns into account. Considering that our approximation ignores the series dependence between symbols and that the asymptotic distribution of the permutation entropy is Normal We compare this result with the Multinomial sample entropy with embedding dimension 3, which assumes independence, and find that the expression

of the asymptotic variance becomes more complex as the embedding dimension increases. After that, a hypothesis test is developed and used to differentiate between the bearing fault diagnoses for four distinct rotating machines.

Contents

Chapter 1

Introduction

This proposal presents the use of permutation entropy and ordinal patterns in time series analysis, including the calculation of pattern histograms, entropy, and complexity to better understand their statistical properties. Additionally, the confidence intervals for entropy and complexity under the multinormal distribution and features for time series clustering are discussed. The concept of ordinal patterns in time series was introduced by Bandt and Pompe [1]. This study focuses on features derived from Bandt & Pompe symbolization, specifically Shannon entropy. Future work will extend to other measures, such as Rényi entropy, Fisher information, and the confidence intervals for their entropy and complexity under the multinormal distribution. Features for time series clustering within ordinal patterns for time series analysis will be further studied.

Time series analysis is widely applied across various fields, including engineering, economics, physical sciences, and more. A time series is defined as a collection of observations x_t, each representing a realized value of a particular random variable X_t, where time can be either discrete or continuous.

Examples of time series applications include finance (e.g., analyzing exchange rate movements or commodity prices), biology (e.g., modeling the growth and decline of bacterial populations), medicine (e.g., tracking the spread of diseases like COVID-19 or influenza), and geoscience (e.g., predicting wet or dry days based on past weather conditions).

The primary goal of time series analysis is to understand the nature of the phenomenon represented by the observed sequence. Time domain and frequency domain methods are the two primary approaches used in time series analysis. The temporal approach relies on concepts such as auto-correlation and regressions, where a time series' present value is analyzed in relation to its own past values or the past values of other series. This method represents time series directly as a function of time. On the other hand, the spectral approach represents time series through spectral expansions, such as wavelets or Fourier modes [**?**].

However, these methods often require assumptions such as large sample sizes or normally distributed observations that are rarely met in real-world empirical data. For many statistical techniques to be valid, these assumptions must hold, but in practice, they are frequently violated. For example, traditional approaches to time series analysis, such as time domain and frequency domain methods, rely on assumptions that are not always valid in real-world data. The time domain approach, which uses techniques like auto-correlation and regression, assumes stationarity and often struggles with nonlinear or nonstationary data. Similarly, the frequency domain approach, which represents time series through spectral expansions such as wavelets or Fourier modes, may require assumptions about periodicity and may not effectively capture short-term fluctuations.

Many statistical methods in these approaches depend on specific conditions, such as large sample sizes or normally distributed observations. However, these assumptions are often unrealistic, leading to inaccurate or biased results. When such conditions are not met, alternative methods must be considered.

As a result, alternative methods, often referred to as non-parametric techniques, must be considered. These methods rely on the rank R_t of the observations x_t rather than their actual values, making them robust and applicable to a wide range of data sets. Since non-parametric tests do not assume a normal distribution, they are highly reliable. For example, the Kruskal-Wallis H test and the Wilcoxon test are effective tools for comparing two or more population probability distributions from independent random samples. However, these techniques are not always suitable for time series data, which often require specialized methods tailored to their unique characteristics.

To address these challenges, ordinal pattern methods provide a robust alternative. Instead of analyzing the absolute values of a time series, these methods focus on the order relationships between consecutive data points. This approach effectively captures the underlying dynamics of complex systems and offers several advantages.

The ordinal pattern-based method has become a widely used tool for characterizing complex time series. Since its introduction nearly twenty-three years ago by Bandt and Pompe in their foundational paper [1], it has been successfully applied across various scientific fields, including biomedical signal processing, optical chaos, hydrology, geophysics, econophysics, engineering, and biometrics. It has also been used in the characterization of pseudo-random number generators.

The method proposed by Bandt and Pompe has been highly successful in analyzing time series. They computed Shannon entropy from the histogram of causal patterns by transforming the time series into ordinal patterns while preserving the original data structure. The histogram is then constructed based on these patterns.

This approach offers several advantages. The resulting distribution is less sensitive to outliers, and the histogram does not depend on any predefined model. These properties make ordinal patterns a valuable and practical tool for time series analysis, especially when traditional methods prove inadequate. Additionally, this method can effectively identify chaotic components within a sequence of words.

Later, Rosso [**?**] introduced an additional dimension to this analysis —-

Statistical Complexity—derived from the same histogram of causal patterns.

Introduction to Ordinal Pattern Analysis

Ordinal patterns are derived from non-parametric time series data rather than relying on their actual values. Ordinal patterns are transformations that encode the sorting characteristics of values in $\mathbb{R}^D$ into $D!$ symbols. One of the possible encoding is the set of indexes that sort the D values in non-decreasing order, where D is called the "Embedding Dimension" and usually ranges between three to six.

To illustrate this idea, let $X = \{x_1, x_2, \dots, x_{(n+D-1)}\}$ be a real valued time series of length $n + D - 1$ without ties. As stated by Bandt & Pompe, if the $\{x_t\}_{t=1}^{n+D-1}$ takes infinitely many values, it is common to replace them with a symbol sequence $\{\pi_j\}$ consisting of finitely many symbols and then compute the entropy from this sequence. The corresponding symbol sequence naturally emerges from the time series without requiring any model assumptions. We compute $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ symbols from subsequences of embedding dimension D. There are $D!$ possible symbols: $\pi_j \in \pi = (\pi^1, \pi^2, \dots, \pi^{D!})$. The histogram of proportions $h = (h_1, h_2, \dots, h_{D!})$ in which the bin h_l is the proportion of symbols of type π^l of the total number of symbols. For convenience, we will model those symbols as a k dimensional random vector where $k = D!$.

Consider a series of n independent trials in which only one of k mutually exclusive events $\pi^1, \pi^2, \dots, \pi^k$ is observed with probability $p_1, p_2, \dots, p_k$, respectively, where $p_l \geq 0$ and $\sum_{l=1}^{k} p_l = 1$. Suppose that $N = (N_1, N_2, \dots, N_k)$ is the vector of random variables that, with $\sum_{l=1}^{k} N_l = n$, counts how many times the events $\pi^1, \pi^2, \dots, \pi^k$ occur in the n trials. Then, the joint probability distribution of N is

$$\Pr\big(N = (n_1, n_2, \dots, n_k)\big) = n! \prod_{\ell=1}^{k} \frac{p_\ell^{n_\ell}}{n_\ell!}, \tag{1.1}$$

where $n_l \geq 0$ and $\sum_{l=1}^{k} n_l = n$.

Problem Statement

To illustrate this concept, imagine tracking the mean monthly humidity in Wellington. We want to analyze how humidity changes throughout the year. By examining this data, you can uncover interesting patterns that highlight the variations in humidity across different months.

Table 1.1: Mean monthly humidity variations in Wellington throughout the year

Month	Mean of relative humidity
January	77.3
February	81
March	82.4
April	81.7
May	83.6
June	85.6
July	84.4
August	83.1
September	78.8
October	79.6
November	78.2
December	78.8

We can convert this actual data into ordinal patterns. To do this, for each month, we determine the order of the humidity values rather than their actual magnitudes. Each three-time-point sequence (which can be adjusted based on preference) is converted into an ordinal pattern. The sequence length can vary from 3 to 6 or more.

Table 1.2: Ordinal Patterns

# of patterns	Mean Humidity sequence	Ordinal Patterns
1	(77.3,81,82.4)	(123) $= \pi_1$
2	(81,82.4,81.7)	(132) $= \pi_2$
3	(82.4,81.7,83.6)	(213) $= \pi_3$
4	(81.7,83.6,85.6)	(123) $= \pi_4$
5	(83.6,85.6,84.4)	(132) $= \pi_5$
6	(85.6,84.4,83.1)	(321) $= \pi_6$
7	(84.4,83.1,78.8)	(321) $= \pi_7$
8	(83.1,78.8,79.6)	(312) $= \pi_8$
9	(78.8,79.6,78.2)	(231) $= \pi_9$
10	(79.6,78.2,78.8)	(312) $= \pi_{10}$

In this example, we have 10 mutually exclusive events such that $\pi_j \in \pi = (\pi^1, \pi^2, \ldots, \pi^{D!}) = (123), (132), (213), (231), (312), (321)$. The probability distribution of the mean humidity can be calculated accordingly.

Table 1.3: Probability function

Notation	Probability
$p(\pi^1)$	2/10
$p(\pi^2)$	2/10
$p(\pi^3)$	1/10
$p(\pi^4)$	1/10
$p(\pi^5)$	2/10
$p(\pi^6)$	2/10

We construct the histogram of proportions $h = (h_1, h_2, h_3, h_4, h_5, h_6)$, where each bin h_l represents the proportion of symbols of type π^l out of the total six symbols. The histogram graph is shown below.

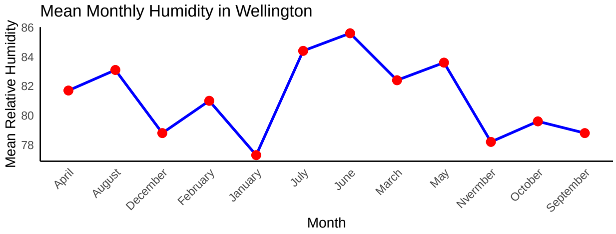

Figure 1.1: Mean Monthly humidity in Wellington

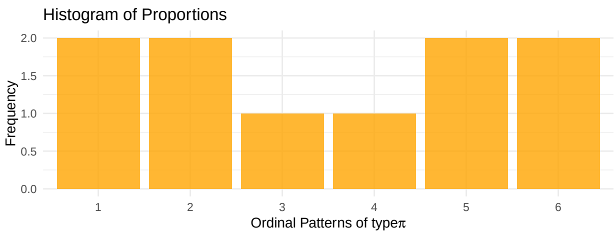

Figure 1.2: Histogram

Chapter 2

Literature Review

More about the methodology

In this concept first, the time series data is convert into order patterns and then calculate the sequence of probabilities from the patterns. Third, estimate the normalized Shannon entropy from the probability sequence. After that ordinal patterns complexity plane is use as a visual representation of the relationship between permutation entropy and the complexity measures.

The outline of this proposal is as follows: After the introduction, in Chapter 2 we have literature review, followed by more explanation of the details of this project in Chapter 3. Finally, we describe future plans for this project in Chapter 4.

2.1 Preliminaries

sssss

Chapter 3

The Research Project

In this chapter, we will explain the main ideas and the objectives of this research project.

3.1 Complexity Plane

Chapter 4

Future Works

As a short summary, we have completed the necessary preliminaries studies on various topics such as:

- mmmmmmmmmmmmmmmmmm
- mmmmmmmmmmmmmmmmmm
- mmmmmmmmmmmmmmmmm

and also we have examined the relevant research articles by Bandt & Pompe [1], Rey et.al [2, 3].

Bibliography

[1] BANDT, C., AND POMPE, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett. 88* (Apr 2002), 174102.

[2] REY, A. A., FRERY, A. C., GAMBINI, J., AND LUCINI, M. M. The asymptotic distribution of the permutation entropy. *Chaos: An Interdisciplinary Journal of Nonlinear Science 33*, 11 (11 2023), 113108.

[3] REY, A. A., FRERY, A. C., LUCINI, M., GAMBINI, J., CHAGAS, E. T. C., AND RAMOS, H. S. Asymptotic distribution of certain types of entropy under the multinomial law. *Entropy 25*, 5 (2023).