VICTORIA UNIVERSITY OF
WELLINGTON
TE HERENGA WAKA

1897

New Features of Permutation Entropy in Ordinal Patterns Complexity Plane

Rasika Dilhani

Supervisor: Alejandro C. Frery

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 introduces the use of permutation entropy within the ordinal patterns complexity plane for time series analysis. The concept of order patterns from time series is introduced by Bandt and Pompe [1].

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 time domain 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 frequency domain 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.

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.

However, traditional methods often encounter difficulties when dealing with nonlinear or nonstationary data, which are prevalent in real-world scenarios such as machine diagnostics. To overcome these challenges, ordinal pattern methods offer a robust alternative. Instead of analyzing the absolute values of a time series, ordinal patterns focus on the order relationships between consecutive data points. This approach effectively captures the underlying dynamics of complex systems and provides several advantages.

In the context of machine diagnostics, ordinal pattern methods enable efficient detection of anomalies or changes in operational conditions, facilitating more accurate identification of malfunctioning machines. This makes ordinal patterns a valuable and practical addition to the toolkit for time series analysis, particularly when traditional methods fall short.

We propose using ordinal patterns derived from non-parametric time series data instead of relying on their actual values. This approach in-

volves comparing the values of the series at n equally spaced time points $t, t+\tau, ..., t+(n-1)\tau$ to determine the corresponding order patterns. Here τ, serves as a scale parameter, allowing us to analyze the series at different scales.

To illustrate this concept, consider a set of observations $\{x_t\}_{t=1}^T$. Instead of focusing on the magnitude of the values, we utilize only the order relationships, such as $x_t < x_s$ or $x_t > x_s$ assuming that $x_t \in \mathbb{R}$. For simplicity, cases where $x_t = x_s$, for $t \neq s$ are treated separately or excluded.

Problem Statement

We consider food preferences among individuals as an example of time series data. Meal choices vary widely among people, reflecting their unique tastes and priorities. For instance, when selecting main meals such as beef, pork, chicken, fish, and vegetables, each person has distinct preferences. By analyzing these choices, we can uncover intriguing patterns that highlight the diversity in individual selections.

Imagine two individuals with different meal preferences. Person 1 enjoys pork the most, followed by beef, fish, chicken, and vegetables. In contrast, Person 2 prefers chicken first, followed by fish, vegetables, beef, and pork. A simple graph illustrating these preferences would show that their choices do not overlap significantly, emphasizing their unique tastes.

Expanding this analysis to a larger group allows us to observe even more diverse and complex patterns. Each individual's preference forms a unique data point, and collectively, they create a rich tapestry of variation. This exploration provides valuable insights into how food choices differ across individuals and groups, making the study both meaningful and captivating.

The concept of ordinal patterns in time series can be effectively demonstrated through a real-world example. Traditionally, numerous algorithms, techniques, and heuristics have been utilized to estimate complexity mea-
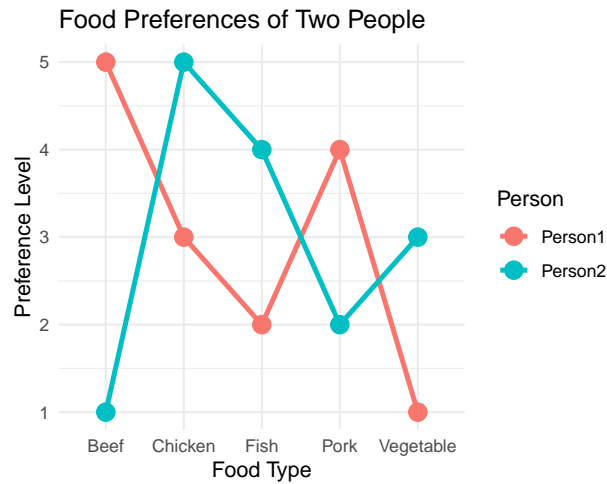
Figure 1.1: Food preference for two people

sures from real-world data. However, these methods often prove effective only for low-dimensional dynamical systems and struggle to perform well when noise is introduced into the data.

The Bandt and Pompe method overcomes this limitation by offering a robust approach that remains reliable even in noisy environments. In time series analysis, key complexity measures—such as entropy, fractal dimension, and Lyapunov exponents—play a crucial role. These measures are instrumental in comparing neighboring values, helping to uncover the underlying structure and dynamics of the data.

The advantages of Bandt & Pompe methods:

- Simplicity

- Extremely fast calculation

- Robutness

- Invariance with respect to nonlinear monotonous transformation

One disadvantage of this method is that, although it is extremely fast and robust, it is best suited for large datasets and scenarios where there is

limited time for preprocessing or fine-tuning parameters.

However, despite the widespread use of ordinal patterns to study the latent dynamics of time series through permutation entropy, there are no established theoretical results concerning the distribution of permutation entropy that account for the correlation effects between patterns. Nonetheless, Rey et al. [2] demonstrated that the asymptotic distribution of permutation entropy is normal. They compared their findings with those of multinomial sample entropy, which assumes independence between patterns. Notably, the expression for the asymptotic variance becomes increasingly complex as the embedding dimension increases.

This proposal has three objectives in order to continue this research work.

- Define a data base of time series for clustering, i.e., finding similar time series.

- Extract all the features we know from their Bandt & Pompe symbolization (Shannon, Tsallis and Renyi entropies, Fisher information measure, complexities, and the available confidence intervals)

- Use those features for time series clustering

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:

- mmmmmmmmmmmmmmmmm

- mmmmmmmmmmmmmmmmmm

- mmmmmmmmmmmmmmmm

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).