Proceedings of the 5th International Conference on Statistics: Theory and Applications (ICSTA'23)
Brunel University, London, United Kingdom – August 03-05, 2023
Paper No. 110
DOI: 10.11159/icsta23.110

Fuzzy Clustering Of Ordinal Time Series Based On Two Novel Distances

Ángel López-Oriona[1], Christian H. Weiss[2], José A. Vilar[1]
[1]Research Group MODES, Research Center for Information and Communication Technologies (CITIC), University of A Coruña, Spain
a.oriona@udc.es; jose.vilarf@udc.es
[2]Department of Mathematics and Statistics, Helmut Schmidt University, Hamburg, Germany
weissc@hsu-hh.de

***Abstract* -** Clustering of time series is a central machine learning task with applications in many fields. While most procedures focus on real-valued time series, very few works consider series with alternative ranges. In this paper, the problem of clustering ordinal time series is addressed. To this aim, two novel distances between ordinal series are introduced and used as input for the fuzzy *C*-medoids algorithm. Both metrics are based on estimated cumulative probabilities, thus automatically taking advantage of the underlying ordering existing in the series range. The corresponding clustering algorithms are able to group series generated from similar underlying stochastic processes, achieve accurate results with series coming from a wide variety of models and are computationally efficient. Moreover, the consideration of the fuzzy approach allows the techniques to properly handle time series showing an uncertain behaviour. An extensive simulation study shows that the proposed methods outperform several alternative procedures.

***Keywords*:** ordinal time series, clustering, serial measures, cumulative probabilities

1. Introduction

Time series clustering concerns the problem of splitting a set of unlabelled time series into homogeneous groups in such a way that similar series are placed together in the same group and dissimilar series are located in different groups. This unsupervised classification tool is often used to characterise different dynamic patterns without the need to analyse and model each single time series, which is computationally intensive when dealing with datasets of large size. Several time series clustering algorithms have been proposed in the 21st century, which has witnessed a growing interest in this challenging field. An extensive overview including current advances, future prospects, interesting references and specific application areas is provided by [1].

The majority of clustering methods focus on real-valued time series while the treatment of series with categorical range, so-called categorical time series (CTS), has received much less attention. Additionally, to the best of our knowledge, all the proposed clustering methods for CTS are designed for the general case in which the series take nominal values, i.e., when no underlying ordering exists in the categorical range. Clearly, these methods are still valid when such inherent ordering exists, that is, when the set that is subject to clustering contains ordinal time series (OTS). However, when applying these procedures in OTS datasets, one completely ignores the natural ordering, which could be an important factor when identifying the underlying clustering partition. Moreover, OTS datasets appear rather naturally in several application domains including finance [2], environmental sciences [3] or medicine [4], among others. Based on previous considerations, one can state that there is a clear need for the construction of clustering algorithms specifically designed to deal with OTS.

The main goal of this paper is to introduce fuzzy clustering algorithms for OTS capable of: (i) grouping together ordinal sequences generated from similar stochastic processes, (ii) achieving accurate results with series coming from a broad variety of ordinal models, and (iii) performing the clustering task in an efficient manner. To this aim, we propose two dissimilarity measures between OTS which are based on features quantifying marginal properties and serial dependence patterns. Both metrics are used as input for the standard fuzzy *C*-medoids algorithm, which allows for the assignment of gradual memberships of the OTS to clusters. Specifically, the consideration of the fuzzy approach enables the techniques to properly handle time series showing an uncertain behaviour. Assessment of the clustering approaches is carried out by means of a comprehensive simulation study including different ordinal processes commonly used in the literature.

The rest of the paper is organised as follows. The two proposed distances between OTS are presented in Section 2 along with an illustrative example. In Section 3, fuzzy clustering algorithms based on both dissimilarities are introduced. The methods are evaluated in Section 4 through some simulations. Some concluding remarks are summarised in Section 5. One proposition is given in the paper, but its proof is not provided due to space limitations.

2. Two distances between OTS

In this section, we present two novel dissimilarities between ordinal series.

2.1. Some background on ordinal processes

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a strictly stationary stochastic process having the ordered categorical range $\mathcal{S} = \{s_0, \ldots, s_n\}$ with$s_0 < s_1 < \cdots < s_n$. Process $\{X_t\}_{t\in\mathbb{Z}}$ is often referred to as ordinal *process*, while categories in $\mathcal{S}$ are frequently called *states*. Let $\{C_t\}_{t\in\mathbb{Z}}$ be the count process with range $\{0, \ldots, n\}$ generating the ordinal process $\{X_t\}_{t\in\mathbb{Z}}$, i.e., $X_t = s_{C_t}$. It is well known that the distributional properties of $\{C_t\}_{t\in\mathbb{Z}}$ (e.g., stationarity) are properly inherited by $\{X_t\}_{t\in\mathbb{Z}}$ [5]. In particular, the marginal probabilities, p_i, and the bivariate probabilities at lag $l \in \mathbb{Z}$, $p_{ij}(l)$, can be expressed as $p_i = P(X_t = s_i) = P(C_t = i)$ and $p_{ij}(l) = P(X_t = s_j, X_{t-l} = s_i) = P(C_t = j, C_{t-l} = i)$, $i, j = 0, \ldots, n$. Note that both the marginal and the bivariate probabilities are still well defined in the general case of a stationary stochastic process with nominal range, i.e., when no underlying ordering exists in the range $\mathcal{S}$. In addition, in an ordinal process, one can consider the corresponding cumulative probabilities, which are defined as $f_i = P(X_t \leq s_i) = P(C_t \leq i)$ and $f_{ij}(l) = P(X_t \leq s_i, X_{t-l} \leq s_j) = P(C_t \leq i, C_{t-l} \leq s_j)$, $i, j = 0, \ldots, n-1$, $l \in \mathbb{Z}$, for the marginal and the bivariate case, respectively.

In practice, quantities $p_i, p_{ij}(l), f_i$ and $f_{ij}(l)$ must be estimated from a T-length realisation of the ordinal process, $\overline{X}_t = \{\overline{X}_1, \ldots, \overline{X}_T\}$, usually referred to as *ordinal time series* (OTS). Natural estimates of these probabilities are given by $p_i = \frac{\sum_{k=1}^{T} I(\overline{X}_k = s_i)}{T}$, $p_{ij}(l) = \frac{\sum_{k=1}^{T-l} I(\overline{X}_k = s_i) I(\overline{X}_{k+l} = s_j)}{T-l}$, $f_i = \frac{\sum_{k=1}^{T} I(\overline{X}_k \leq s_i)}{T}$, $f_{ij}(l) = \frac{\sum_{k=1}^{T-l} I(\overline{X}_k \leq s_i) I(\overline{X}_{k+l} \leq s_j)}{T-l}$, where $I(\cdot)$ denotes the indicator function.

Probabilities $p_i, p_{ij}(l), f_i$ and $f_{ij}(l)$ can be used to represent the process $\{X_t\}_{t\in\mathbb{Z}}$ in terms of marginal and serial dependence patterns. An alternative way of describing a given ordinal process is by means of features measuring classical statistical properties (e.g., centrality, dispersion...) in the ordinal setting. A practical approach to define these quantities consists of considering expected values of some well-known distances between ordinal categories [6]. Table 1 presents some of the features defined by [6] in the particular case of the underlying ordinal distance being the so-called block distance, denoted by $d_{o,1}$, which is defined as $d_{o,1}(s_i, s_j) = |i - j|$ for a pair of states s_i and s_j. Note that $d_{o,1}$ provides a natural way of assessing dissimilarity between ordinal categories, since it treats the ordinal data as if assigning equidistant numbers (scores) to the different states.

Median ($\text{loc}_{d_{o,1}}$)	Dispersion ($\text{disp}_{d_{o,1}}$)	Asymmetry ($\text{asym}_{d_{o,1}}$)	Skewness ($\text{skew}_{d_{o,1}}$)	Ordinal Cohen's κ ($\kappa_{d_{o,1}}(l)$)
$\sum_{i=0}^{n-1}(i+1)(f_{i+1} - f_i)$	$2\sum_{i=0}^{n-1} f_i(1 - f_i)$	$\sum_{i=0}^{n-1}(1 - f_i - f_{n-i-1})^2$	$2\sum_{i=0}^{n-1} f_i - 1$	$\frac{\sum_{i=0}^{n-1}(f_{ii}(l) - f_i^2)}{\sum_{i=0}^{n-1} f_i(1-f_i)}$

Table 1. Some features of an ordinal process.

The first four measures in Table 1 summarise the marginal behaviour of the process, while the ordinal Cohen's κ, $\kappa_{d_{o,1}}(l)$, evaluates the degree of serial dependence exhibited by $\{X_t\}_{t\in\mathbb{Z}}$ at a given lag $l \in \mathbb{Z}$. When dealing with the realisation $\{X_t\}_{t\in\mathbb{Z}}$, these features can be estimated by considering the quantities $\hat{f}_i$ and $\hat{f}_{ij}(l)$. The corresponding estimates are denoted as $\widehat{loc}_{d_{o,1}}$, $\widehat{disp}_{d_{o,1}}$, $\widehat{asym}_{d_{o,1}}$, $\widehat{skew}_{d_{o,1}}$ and $\hat{\kappa}_{d_{o,1}}(l)$. A detailed analysis of the asymptotic properties of these estimates is provided in [6].

2.2. Two novel dissimilarities between ordinal time series

Suppose we have two stationary ordinal processes $\{X_t^{(1)}\}_{t\in\mathbb{Z}}$ and $\{X_t^{(2)}\}_{t\in\mathbb{Z}}$. A simply dissimilarity criterion between both processes can be established by measuring discrepancy between their corresponding representations in terms of cumulative probabilities. In this way, for a given collection of L lags, $\mathcal{L} = \{l_1, \dots, l_L\}$, we define a distance d_1 as $d_1(X_t^{(1)}, X_t^{(2)}) = d_{1,M}(X_t^{(1)}, X_t^{(2)}) + d_{1,B}(X_t^{(1)}, X_t^{(2)})$, with $d_{1,M}(X_t^{(1)}, X_t^{(2)}) = \sum_{i=0}^{n-1}(f_i^{(1)} - f_i^{(2)})^2$ and $d_{1,B}(X_t^{(1)}, X_t^{(2)}) = \sum_{k=1}^{L}\sum_{i=0}^{n-1}\sum_{j=0}^{n-1}(f_{ij}^{(1)}(l_k) - f_{ij}^{(2)}(l_k))^2$, where the superscripts (1) and (2) are used to indicate that the corresponding probabilities refer to processes $\{X_t^{(1)}\}_{t\in\mathbb{Z}}$ and $\{X_t^{(2)}\}_{t\in\mathbb{Z}}$, respectively.

Note that the terms $d_{1,M}$ and $d_{1,B}$ in metric d_1 assess dissimilarity between marginal and lagged bivariate probabilities, respectively. Moreover, the latter term involves the set $\mathcal{L}$, which must be fixed in advance according to the lags at which one wishes to evaluate serial dependence. It is worth remarking that, by considering the cumulative probabilities in the definition of d_1 , we obtain a dissimilarity measure which takes into account the underlying ordering existing in both processes.

An alternative dissimilarity measure considering features based on the block distance $d_{o,1}$ is defined as $d_2(X_t^{(1)}, X_t^{(2)}) = d_{2,M}(X_t^{(1)}, X_t^{(2)}) + d_{2,B}(X_t^{(1)}, X_t^{(2)})$, with $d_{2,M}(X_t^{(1)}, X_t^{(2)}) = ||(\frac{loc_{d_{o,1}}^{(1)}}{n}, \frac{2disp_{d_{o,1}}^{(1)}}{n}, \frac{asym_{d_{o,1}}^{(1)}}{n}, \frac{skew_{d_{o,1}}^{(1)}}{n}) - (\frac{loc_{d_{o,1}}^{(2)}}{n}, \frac{2disp_{d_{o,1}}^{(2)}}{n}, \frac{asym_{d_{o,1}}^{(2)}}{n}, \frac{skew_{d_{o,1}}^{(2)}}{n})||^2$ and $d_{2,B}(X_t^{(1)}, X_t^{(2)}) = \sum_{k=1}^{L}(\kappa_{d_{o,1}}^{(1)}(l_k) - \kappa_{d_{o,1}}^{(2)}(l_k))^2$.

In the same way as d_1, dissimilarity d_2 is formed by the terms $d_{2,M}$ and $d_{2,B}$, which assess discrepancy between marginal and serial behaviour of both ordinal processes, respectively. Specifically, the marginal component contains the normalised versions of the quantities in Table 1 [6]. Thus, each one of the features is expected to exhibit approximately the same weight in the computation of $d_{2,B}$.

Since, in practice, we only have finite-length realisations of both ordinal processes, the values of d_1 and d_2 are unknown and must be properly estimated. The corresponding estimates take the form $\hat{d}_p(\overline{X}_t^{(1)}, \overline{X}_t^{(2)}) = \hat{d}_{p,M}(X_t^{(1)}, X_t^{(2)}) + \hat{d}_{p,B}(X_t^{(1)}, X_t^{(2)})$, $p = 1,2$, where $\overline{X}_t^{(1)}$ and $\overline{X}_t^{(2)}$ are realisations (not necessarily with the same length) from processes $\{X_t^{(1)}\}_{t\in\mathbb{Z}}$ and $\{X_t^{(2)}\}_{t\in\mathbb{Z}}$, respectively, and $\hat{d}_{p,M}$ and $\hat{d}_{p,B}$ are proper estimates of $d_{p,M}$ and $d_{p,B}$ computed by considering the estimates $\hat{f}_i^{(h)}$, $\hat{f}_{ij}^{(h)}(l_k)$, $p, h = 1,2$, respectively, for each one of the realisations.

2.3. Motivating example

This section illustrates the advantages of using cumulative probabilities to differentiate between ordinal processes. This is shown by means of a toy example involving synthetic data. For the sake of simplicity, we focus on the marginal case. Let us consider three stationary processes having the ordinal range $\mathcal{S} = \{s_0, s_1, s_2, s_3\}$, denoted by $X_t^{(1)}$, $X_t^{(2)}$ and $X_t^{(3)}$, with marginal distributions given by the vectors of probabilities $\boldsymbol{p}_i = (P(X_t^{(i)} = s_0), \dots, P(X_t^{(i)} = s_3))$, $i = 1,2,3$, respectively, such that $\boldsymbol{p}_1 = (0.4,0.1,0.1,0.4)$, $\boldsymbol{p}_2 = (0.1,0.4,0.1,0.4)$, $\boldsymbol{p}_3 = (0.1,0.1,0.4,0.4)$. Let us define a metric d^* measuring dissimilarity between two processes by means of the squared Euclidean distance between the corresponding vectors of marginal probabilities. In this way, pairwise distances for the set of processes $\{X_t^{(1)}, X_t^{(2)}, X_t^{(3)}\}$ are given by $d^*(X_t^{(1)}, X_t^{(2)}) = d^*(X_t^{(1)}, X_t^{(3)}) = d^*(X_t^{(2)}, X_t^{(3)}) = 0.18$. According to metric d^*, each pair of processes exhibit the same amount of dissimilarity. However, the underlying ordering in the set $\mathcal{S}$ suggests that process $\{X_t^{(1)}\}_{t\in\mathbb{Z}}$ should be closer to $\{X_t^{(2)}\}_{t\in\mathbb{Z}}$ than to $\{X_t^{(3)}\}_{t\in\mathbb{Z}}$, since category s_1 is closer to s_0 than category s_2. Therefore, one could state that distance d^* is not appropriate to evaluate dissimilarity between the marginal distributions of two ordinal processes.

Let us consider now the metric $d_{1,M}$ (see Section 2.2), which is defined as the squared Euclidean distance between the corresponding vectors of cumulative probabilities. These vectors take the form $\boldsymbol{f}_1 = (0.4,0.5,0.6,1)$, $\boldsymbol{f}_2 = (0.1,0.5,0.6,1)$ and $\boldsymbol{f}_3 = (0.1,0.2,0.6,1)$ for processes $X_t^{(1)}$, $X_t^{(2)}$ and $X_t^{(3)}$, respectively. Pairwise distances based on $d_{1,M}$ are given by $d_{1,M}(X_t^{(1)}, X_t^{(2)}) = 0.09, d_{1,M}(X_t^{(1)}, X_t^{(3)}) = 0.18, d_{1,M}(X_t^{(2)}, X_t^{(3)}) = 0.09$. According to dissimilarity $d_{1,M}$, the pair $(X_t^{(1)}, X_t^{(2)})$ is closer than the pair $(X_t^{(1)}, X_t^{(3)})$. Moreover, process $X_t^{(2)}$ is located at the same distance from $X_t^{(1)}$ and $X_t^{(3)}$. This is reasonable, since the marginal distribution of both $X_t^{(1)}$ and $X_t^{(3)}$ can be obtained from the distribution of $X_t^{(2)}$ by transferring the same amount of probability either one step backward (s_0) or upward (s_2) from category s_1, respectively. In essence, cumulative probabilities allow us to better differentiate between ordinal distributions because they implicitly take into account the underlying count processes (see Section 2.1). Specifically, the amount of dissimilarity is lower when the differences between marginal distributions happen at closer categories. Therefore, metric $d_{1,M}$ assigns distance values which are consistent with the ordering existing in the range $\mathcal{S}$.

The above example highlights the importance of considering cumulative probabilities to properly measure dissimilarity between ordinal processes. Previous considerations can be justified by means of the following proposition, which expresses metric $d_{1,M}$ in terms of discrepancies between marginal probabilities.

Proposition 1. Given two stationary ordinal processes $\{X_t\}_{t\in\mathbb{Z}}$ and $\{Y_t\}_{t\in\mathbb{Z}}$ with range $\mathcal{S} = \{s_0, s_1, \dots, s_n\}$ and marginal distributions $(p_0, p_1, \dots, p_n)$ and $(q_0, q_1, \dots, q_n)$, respectively, with $n > 0$, the distance $d_{1,M}$ between processes $\{X_t\}_{t\in\mathbb{Z}}$ and $\{Y_t\}_{t\in\mathbb{Z}}$ can be written as $d_{1,M}(X_t, Y_t) = \sum_{i=0}^{n-1}(n-i)(p_i - q_i)^2 + 2\sum_{j=0}^{n-2}\sum_{k=j+1}^{n-1}(n-k)(p_j - q_j)(p_k - q_k)$.

According to Proposition 1, distance $d_{1,M}$ can be expressed as the sum of two terms. The first term contains the squared differences, $(p_i - q_i)^2$, while the second term includes the cross products $(p_j - q_j)(p_k - q_k)$. In both cases, specific weights are given to the corresponding differences. In particular, the weights are higher when marginal probabilities at lower categories are considered, which means that discrepancies in earlier states exhibit a larger influence in the computation of $d_{1,M}$ than discrepancies in later states. Note that this is consistent with the distance computations above, where the three processes are equidistant when considering marginal probabilities (metric d^*), but they exhibit a different, more reasonable configuration when considering cumulative probabilities (metric $d_{1,M}$).

Previous analyses illustrate the advantages of using cumulative probabilities when measuring dissimilarity between the marginal distributions of two ordinal processes. An analogous argument could be provided when assessing dissimilarity between lagged bivariate distributions. The theoretical considerations for the bivariate case are not shown in this version of the manuscript for the sake of simplicity.

3. Fuzzy clustering algorithms for ordinal time series

This section is devoted to introduce fuzzy clustering algorithms for ordinal series which are based on the proposed dissimilarities d_1 and d_2.

Consider a set of s ordinal time series, $\mathbb{S} = \{\overline{X}_t^{(1)}, \dots, \overline{X}_t^{(s)}\}$, not necessarily having the same length. We wish to perform fuzzy clustering on the elements of $\mathbb{S}$ in such a way that the series generated from similar underlying stochastic processes are grouped together. To this aim, we propose to use a fuzzy C-medoids clustering model based on the distances introduced in Section 2.2, which tries to find the subset of $\mathbb{S}$ of size C, $\tilde{\mathbb{S}} = \{\tilde{X}_t^{(1)}, \dots, \tilde{X}_t^{(C)}\}$, whose elements are usually referred to as medoids, and the $s \times C$ matrix of fuzzy coefficients, $\boldsymbol{U} = (u_{ic})$, $i = 1, \dots, s$, $c = 1, \dots, C$, which define the solution of the minimisation problem

$$\min_{\tilde{\mathbb{S}}, \boldsymbol{U}} \sum_{i=1}^{s} \sum_{c=1}^{C} u_{ic}^{m} d_p(i,c) \, with \, respect \, to \sum_{c=1}^{C} u_{ic} = 1, u_{ic} \geq 0, \qquad (1)$$

where $d_p(i,c) = d_p(\overline{X}_t^{(i)}, \tilde{X}_t^{(c)})$, $p = 1,2$, $u_{ic} \in [0,1]$ represents the membership degree of the ith CTS in the cth cluster and $m > 1$ is a real number usually referred to as fuzziness parameter, which regulates the fuzziness of the partition. For $m = 1$, the crisp version of the algorithm is obtained, so the solution takes the form $u_{ic} = 1$ if the ith series pertains to cluster c and $u_{ic} = 0$ otherwise. As the value of m increases, the boundaries between clusters get softer and the resulting partition is fuzzier. The constrained optimization problem in (1) can be solved by means of the Lagrangian multipliers method, which provides an iterative algorithm that alternately optimises the membership degrees and the medoids. Specifically, the iterative solutions for the membership degrees are given by $u_{ic} = [\sum_{c'=1}^{C} (\frac{d_p(i,c)}{d_p(i,c')})^{\frac{1}{m-1}}]^{-1}$, for $p = 1,2$, $i = 1, \dots, s$, $c = 1, \dots, C$. Once the membership degrees are obtained through (1), the C series minimising the objective function in (1) are selected as new medoids. Specifically, for each $c \in \{1, \dots, C\}$, the index j_c satisfying $j_c = argmin_{1 \leq j \leq s} \sum_{i=1}^{s} u_{ic}^{m} d_p(X_t^{(i)}, X_t^{(j)})$, $p = 1,2$. This two-step procedure is repeated until there is no change in the medoids or a maximum number of iterations is reached. An outline of the corresponding clustering algorithm is given in Algorithm 1.

Fix C, m, $max.iter$ and $p \in \{1,2\}$
Set $iter = 0$
Pick the initial medoids $\widetilde{\mathcal{S}} = \{\widetilde{X}_t^{(1)}, \dots, \widetilde{X}_t^{(C)}\}$
repeat
    Set $\widetilde{\mathcal{S}}_{\mathrm{OLD}} = \widetilde{\mathcal{S}}$ {Store the current medoids}
    Compute u_{ic}, $i = 1, \dots, s$, $c = 1, \dots, C$, using (18)
    For each $c \in \{1, \dots, C\}$, determine the index $j_c \in \{1, \dots, s\}$ using (19)
    return $\widetilde{X}_t^{(c)} = X_t^{(j_c)}$, for $c = 1, \dots, C$ {Update the medoids}
    $iter \leftarrow iter + 1$
until $\widetilde{\mathcal{S}}_{\mathrm{OLD}} = \widetilde{\mathcal{S}}$ or $iter = max.iter$
return The final fuzzy partition and set of medoids

Algorithm 1.The fuzzy C-medoids algorithm based on the proposed distances.

4. Simulation study

In this section, we carry out a set of simulations with the aim of evaluating the behaviour of both d_1 and d_2 in different scenarios of OTS clustering. First we describe some alternative dissimilarities that we consider for comparison purposes. Next we explain how the performance of the metrics is measured along with the corresponding simulation mechanism and results.

4.1. Alternative metrics

To shed light on the performance of the proposed fuzzy clustering algorithm, it was compared with some other models based on alternative dissimilarities. The corresponding approaches are described below.

- *A procedure employing the probability mass function*. This method considers a dissimilarity defined in the same way as d_O, but replacing the quantities $f_i^{(k)}$ and $f_{ij}^{(k)}(l)$ by the estimates $p_i^{(k)}$ and $p_{ij}^{(k)}(l)$, respectively. The corresponding metric is called d_{PMF}. Note that distance d_{PMF} is still well defined when treating with nominal time series, thus ignoring the underlying ordering. Therefore, performance of d_{PMF} is a straightforward benchmark for the proposed metric d_1, which is specifically designed to deal with ordinal series.

- *Autocorrelation-based clustering*. [7] proposed a distance measure between time series based on the autocorrelation function. Specifically, each time series is described by means of a vector $(\rho(l_1), \dots, \rho(l_L))$ whose components are the estimated autocorrelations for a given set of lags. The metric is defined as the squared Euclidean distance between the corresponding vectors. We denote this dissimilarity as d_{ACF}. Note that d_{ACF} is only defined for numerical time series, but the distance can be easily computed in the ordinal case by considering the associated count time series (see Section 2.1).

• *Quantile-based clustering*. [8] introduced a clustering method using a dissimilarity based on quantile dependence. Specifically, each series is replaced by a feature vector containing estimates of the so-called quantile autocovariance function for several pairs of probability levels $(\tau, \tau') \in [0,1]^2$ and a fixed set of lags. The proposed metric, denoted by d_{QAF}, is defined as the squared Euclidean distance between two vector representations. As in the case of d_{ACF}, computation of distance d_{QAF} must be carried out by considering the corresponding count time series. Several time series clustering procedures employing quantile-based features have been proposed in the literature [9, 10, 11]. This type of methods usually show a great performance when the clusters are characterised by different nonlinear structures.

4.2. Experimental design and results

In this section, the performance of metrics d_1 and d_2 is analysed in a simulation study. We consider three simple scenarios consisting of four clusters represented by the same type of generating processes, denoted by $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ and $\mathcal{C}_4$. Each one of the groups contains 5 OTS, which gives rise to a set of 20 OTS to be subject to clustering. The generating models concerning the count process $\{C_t\}_{t\in\mathbb{Z}}$ in each group are given below for each one of the scenarios.

• **Scenario 1**. *Fuzzy clustering of OTS based on binomial AR(p) models*. Let $\pi \in (0,1)$, $\rho \in (max\{\frac{-\pi}{1-\pi}, \frac{1-\pi}{\pi}\},1)$, $\beta = \pi(1-\rho)$, $\alpha = \beta + \rho$. Let the count process $\{C_t\}_{t\in\mathbb{Z}}$ be defined by the recursion $C_t = \sum_{i=1}^{p} D_{t,i}(\alpha \odot_t C_{t-i} + \beta \odot_t (n - C_{t-i}))$, where the $(D_{t,1}, \dots, D_{t,p})$ are independent variables which are distributed according to MULT$(1; \phi_1, \dots, \phi_p)$ with $\phi_1 + \dots + \phi_p = 1$ and $\odot_t$ denotes the binomial thinning operator performed at a given time t. Specifically, for a random variable Y with range $\{0, \dots, n'\}$ and $\alpha' \in (0,1)$, the binomial thinning operator, denoted by $\odot$, is defined as $\alpha' \odot Y = \sum_{i=1}^{Y} Z_i$, where the Z_i are independent random variables following a Bernoulli distribution with parameter α'. The processes considered in this scenario are two binomial AR(1) models and two binomial AR(2) models with vectors of coefficients given by $\mathcal{C}_1$: $(\alpha, \beta) = (0.32,0.17)$, $\mathcal{C}_2$: $(\alpha, \beta) = (0.38,0.21)$, $\mathcal{C}_3$: $(\alpha, \beta, \phi_1, \phi_2) = (0.20,0.09,0.1,0.9)$ and $\mathcal{C}_4$: $(\alpha, \beta, \phi_1, \phi_2) = (0.26,0.13,0.5,0.5)$, respectively.

• **Scenario 2**. *Fuzzy clustering of OTS based on binomial INARCH(p) models.* Let $\beta, \alpha_1, \dots, \alpha_p$ be real numbers such that $\beta, \beta + \sum_{i=1}^{p} \alpha_i \in (0,1)$, and assume that the count process $\{C_t\}_{t\in\mathbb{Z}}$ verifies $C_t | C_{t-1}, C_{t-2}, \dots \sim Bin(n, \beta + \frac{1}{p}\sum_{i=1}^{p} \alpha_i C_{t-i})$. The considered processes are two binomial INARCH(1) models and two binomial INARCH(2) models with vectors of coefficients given by $\mathcal{C}_1$: $(\alpha_1, \beta) = (0.30,0.35)$, $\mathcal{C}_2$: $(\alpha_1, \beta) = (0.30,0.40)$, $\mathcal{C}_3$: $(\alpha_1, \alpha_2, \beta) = (0.1,0.1,0.2)$ and $\mathcal{C}_4$: $(\alpha_1, \alpha_2, \beta) = (0.1,0.1,0.4)$, respectively.

• **Scenario 3**. *Fuzzy clustering of OTS based on ordinal logit AR(1) models*. Let $\{C_t\}_{t\in\mathbb{Z}}$ be a count process and denote by $\{\boldsymbol{Y}_t = (Y_{t,0}, \dots, Y_{t,n})\}_{t\in\mathbb{Z}}$ its binarization (i.e., $C_t = k$, if and only if $Y_{t,k} = 1$ and $Y_{t,k'} = 0$, $k' \neq k$) and by $\{\boldsymbol{Y}_t^* = (Y_{t,0}, \dots, Y_{t,n-1})\}_{t\in\mathbb{Z}}$ its reduced binarization. Let the process $\{Q_t\}_{t\in\mathbb{Z}}$ be formed by independent variables following a standard logistic distribution and assume that $C_t = s_j \, if and \, only \, if Q_t - \boldsymbol{Y}_t^* \boldsymbol{\alpha}^\top = [\eta_{j-1}, \eta_j)$, where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ and $-\infty = \eta_{-1} < \eta_0 < \dots < \eta_{n-1} < \eta_n = +\infty$ are threshold parameters which can be represented by means of the vector $\boldsymbol{\eta} = (\eta_0, \dots, \eta_{n-1})$. The considered processes are four 6-state ordinal logit AR(1) models with vector of coefficients $\boldsymbol{\alpha} = (-2,-1,0,1,2)$ and vector of thresholds given by $\mathcal{C}_1$: $\boldsymbol{\eta} = (0.4,0.8,1.2,1.6,2)$, $\mathcal{C}_2$: $\boldsymbol{\eta} = (0.6,1.2,1.8,2.4,3)$, $\mathcal{C}_3$: $\boldsymbol{\eta} = (0.8,1.6,2.4,3.2,4)$ and $\mathcal{C}_4$: $\boldsymbol{\eta} = (1,2,3,4,5)$.

	Scenario 1					Scenario 2					Scenario 3				
	$\widehat{d}_1$	$\widehat{d}_2$	$\widehat{d}_{PMF}$	$\widehat{d}_{ACF}$	$\widehat{d}_{QAF}$	$\widehat{d}_1$	$\widehat{d}_2$	$\widehat{d}_{PMF}$	$\widehat{d}_{ACF}$	$\widehat{d}_{QAF}$	$\widehat{d}_1$	$\widehat{d}_2$	$\widehat{d}_{PMF}$	$\widehat{d}_{ACF}$	$\widehat{d}_{QAF}$
$m = 1.2$	**0.98**	0.95	0.93	0.27	0.96	0.69	**0.71**	0.63	0.30	0.58	**0.91**	0.78	0.90	0.30	0.70
$m = 1.4$	**0.91**	0.82	0.79	0.23	0.87	0.64	**0.67**	0.53	0.26	0.51	**0.78**	0.72	0.69	0.28	0.62
$m = 1.6$	**0.81**	0.68	0.65	0.19	0.76	**0.59**	**0.59**	0.45	0.23	0.45	**0.64**	**0.64**	0.52	0.25	0.53
$m = 1.8$	**0.70**	0.57	0.53	0.15	0.66	**0.52**	0.51	0.37	0.18	0.37	0.52	**0.55**	0.40	0.23	0.45
$m = 2.0$	**0.60**	0.47	0.43	0.12	0.56	**0.46**	0.44	0.31	0.16	0.33	0.43	**0.47**	0.33	0.20	0.38

Table 2. Average values of ARIF obtained by the fuzzy C-medoids algorithm for several metrics.

The simulation study was carried out as follows. For each scenario, 5 OTS of length $T = 600$ were generated from each process. In all cases, the range of the count process $\{C_t\}_{t\in\mathbb{Z}}$ was set to $\{0,1,\dots,5\}$, giving rise to ordinal realisations with range $\{s_0, s_1, \dots, s_5\}$. Several values of the fuzziness parameter m were considered, namely $m \in \{1.2,1.4,1.6,1.8,2\}$. Given a scenario and a value for m, 200 simulations were carried out. In each trial, the fuzzy C-medoids algorithm based on d_1, d_2, d_{PMF}, d_{ACF} and d_{QAF} was applied by using the corresponding value of m as input. The number of clusters was set to $C = 4$. The collection of lags $\mathcal{L}$ was set to $\mathcal{L} = \{1,2\}$ in Scenarios 1 and 2 and to $\mathcal{L} = \{1\}$ in Scenario 3, thus considering the maximum number of defining lags existing in each scenario. The effectiveness of each clustering procedure was assessed by means of the fuzzy extension of the Adjusted Rand Index (ARIF). Computation of dissimilarity d_{QAF} requires fixing a set of probability levels. In this regard, we executed the corresponding fuzzy C-medoids algorithm by considering the sets $\mathcal{T}_1 = \{0.1,0.5,0.9\}$, $\mathcal{T}_2 = \{0.3,0.5,0.7\}$ and $\mathcal{T}_3 = \{0.4,0.8\}$ independently. Throughout the paper, only the results corresponding to the set associated with the highest average value for ARIF are presented.

Average values of ARIF are given in Table 2. Overall, the proposed metrics outperform the alternative dissimilarities by a considerable margin. The distance based on cumulative probabilities d_1 achieves the best scores in Scenario 1, while no significant differences are observed between d_1 and d_2 in Scenarios 2 and 3. In sum, results in Table 2 highlight the importance of considering dissimilarities specifically designed to deal with OTS (i.e., taking into account the underlying ordering existing in the range of the series) when performing clustering of ordinal series. It is worth highlighting that similar conclusions are reached by considering different values for T and alternative clustering quality indexes.

5. Conclusions

In this paper, we have proposed two novel distances be- tween OTS which automatically take advantage of the underlying ordering existing in the series range. Both dissimilarities are used as input to the classical fuzzy C-medoids algorithm, which allows for the assignment of gradual memberships of the OTS to the different groups. This is particularly useful when dealing with time series datasets, where large amounts of uncertainty are frequent due to regime shifts. To assess the performance of the clustering algorithms, a simulation study including different types of ordinal processes was considered. The methods were compared with clustering algorithms based on alternative dissimilarities. Overall, the proposed clustering algorithms showed the best performance. Specifically, they outperformed some techniques specifically designed to deal with real-valued and with nominal time series.

References

[1] T. W. Liao, "Clustering of time series data: A survey". Pattern Recognition, vol. 38, no. 11, pp. 1857–1874, 2005.

[2] C. Pamminger and S. Frühwirth-Schnatter. "Model-based clustering of categorical time series". Bayesian Analysis, vol. 5, no. 2, pp. 345-368, 2010.

[3] C. H. Weiß, "Regime-switching discrete arma models for categorical time series". Entropy, vol. 22, no. 4, pp. 458,

2020.
[4] J. Koss, S. Tinaz and H. D. Tagare, “Hierarchical denoising of ordinal time series of clinical scores”. IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 7, pp. 3507-3516, 2022.
[5] C. H. Weiß, “An introduction to discrete-valued time series”. John Wiley & Sons, 2018.
[6] C. H. Weiß, “Distance-based analysis of ordinal data and ordinal time series”. Journal of the American Statistical Association, vol. 115, no. 531, pp. 1189-1200, 2019.
[7] P. D’Urso and E. A. Maharaj. “Autocorrelation-based fuzzy clustering of time series”. Fuzzy Sets and Systems, vol. 160, no. 24, pp. 3565-3589, 2009.
[8] B. Lafuente-Rego and J. A. Vilar. “Clustering of time series using quantile autocovariances”. Advances in Data Analysis and classification, vol. 10, no. 3, pp. 391-415, 2016.
[9] A. M. Alonso, F. J. Nogales and C. Ruiz, “Hierarchical clustering for smart meter electricity loads based on quantile autocovariances”. IEEE Transactions on Smart Grid, vol. 11, no. 5, pp. 4522-4530, 2020.
[10] A. López-Oriona and J. A. Vilar, “Quantile cross-spectral density: A novel and effective tool for clustering multivariate time series”. Expert Systems with Applications, vol. 185, pp. 115677, 2021.
[11] A. López-Oriona, J. A. Vilar and P. D’Urso, “Quantile-based fuzzy clustering of multivariate time series in the frequency domain”. Fuzzy Sets and Systems, vol. 443, pp. 115-154, 2022.