2016 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2016, Yugawara, Japan, November 27th-30th, 2016

A new approach to non-uniformed sampled time series using ordinal partitions

Thomas Stemler[†‡⋆], Konstantinos Sakellariou[‡], Michael McCullough[‡] and Ben Luo[‡]

†Complex Data Modelling Group, The University of Western Australia
35 Stirling Hwy, Crawley WA, 6009, Australia
‡School of Mathematics and Statistics, The University of Western Australia
35 Stirling Hwy, Crawley WA, 6009, Australia
⋆ Postam Institut für Klimafolgenabschätzung (PIK), RD IV Transdisciplinary Concepts and Methods
Potsdam, Germany
Email: thomas.stemler@uwa.edu.au, konstantinos.sakellariou@research.uwa.edu.au

Abstract—The ordinal patterns symbolic methodology may be used to discriminate between deterministic and stochastic time series. Deterministic systems are characterised by forbidden patterns which would not occur even in infinitely long time series, whereas random systems may be exhibit any possible pattern. In this study, we examine the effects of highly irregularly-sampled time series, such those encountered in paleoclimate measurements, on the estimator of the proportion of forbidden patterns manifested by synthetic time series. We investigate three different sampling regimes and observe a high sensitivity in the results. The degree of asymmetry of the sampling distribution is shown to be a useful heuristic indicator for the reliability of the forbidden patterns estimator and the potential for detecting determinism. Exponentially distributed lead to gross underestimation of the total number of forbidden patterns. On the other hand, sampling in the presence of large chronological gaps can lead to relatively accurate estimates as long as the time series contains sufficiently many densely-sampled areas as well.

1. Introduction

Detecting a deterministic component in noisy data is an important problem in nonlinear time series analysis. The assumption of determinism underlies a large class of techniques which focus on the theory of dynamical systems. Symbolic dynamics tools have recently shown potential towards this goal. Ordinal patterns, in particular, comprise symbols obtained by a segmentation of a time series into elements of equal length. Patterns which cannot occur for a specified system are termed forbidden. Analysing the statisti- cal properties of the resulting sequence can shed light on the underlying dynamics. Deterministic time series are thought to always be characterised by forbidden patterns, in contrast to random systems whereby any possible pattern may be realised. Thus, the relative proportion of forbidden patterns can be used to detect determinism. We investigate the effects of highly irregular sampling, such as from paleoclimate or geological data, on the reliability of this statistic as estimated from time series data. Real-world data sets often appear in the form of non-uniformly sampled time series. This may be due to device failure, weather conditions, human error, the nature of the system (e.g. financial transactions data) or the measurement method (e.g. geological data) and other causes. For this study we are motivated by geoscientific and paleoclimate time series which are characterised by miss- ing entries and large chronological gaps. Although there exist several types of irregular sampling, which can vary from rather mildly to highly unevenly spaced data, the majority of established techniques in time series analysis assume regular sampling. Consequently, there is an increasing need to extend the applicability of existing techniques and create more sophisticated tools to reliably analyse irregularly sampled time series. For linear systems, the Lomb-Scargle periodogram (also known as Vaniceks least-square method) is an example towards this direction. For nonlinear systems, there exist a few notable (but very recent) examples such as the similarity estimators method (K. Rehfeld and J. Kurths, Climate of the Past 10, 107 (2014)), the distance metric for marked point processes by Suzuki, Hirata, and Aihara (S. Suzuki, Y. Hirata, and K. Aihara, Int. J. Bifurcation Chaos 20, 3699 (2010)) and the transformation-cost time series by Eroglu *et al.*(D. Eroglu, F. McRobie, I. Ozken, T. Stemler, K.-H. Wyrwoll, S. F. M. Breitenbach, N. Marwan, and J. Kurths, Nature Communications 7, 12929 (2016)).

In this paper, we further extend upon previous results by exploring cases of more severe irregular sampling, far removed from the regular sampling grid archetype. The types of unevenly sampled synthetic data we examine herein are more characteristic of geoscientific measure- ments such as geological or paleoclimate data, and pose a formidable challenge for the ensuing time series analysis under any methodological approach. The question we ask is about the reliability of the estimator of forbidden patterns as a criterion for the existence of determinism in highly irregularly sampled time series. The sampling schemes we examine here include a Poisson sampling process (exponentially distributed in- tervals), Pareto sampling (example of a power-law distri- bution) and Γ-distributed sampling periods for various levels of skewness.