Time Series Interpolation Algorithms

Melissa Van Bussel

MATH4800H Project Proposal, Winter 2019

Background

A *time series* is a set $\{X_t\}$ of observations, one taken at each time t. They are regularly used in a variety of disciplines, including (but not limited to) economics, sociology, and physics. *Time Series Analysis* refers to techniques used to draw inferences from temporal data and to gain meaningful information about the process from which the time series originates. Some applications of time series analysis include prediction of natural disasters, weather forecasting, and stock market predictions.

A time series $\{X_t\}$ with index set $\mathbb{Z}$ is said to be *stationary* if:

1. $\mathbf{E}\left[|X_t|^2\right] < \infty \quad \forall t \in \mathbb{Z}$ (that is, it has finite variance)
2. $\mathbf{E}\left[X_t\right] = m \quad \forall t \in \mathbb{Z}$ (that is, the expected value is the same for each observation in the time series)
3. $\gamma_x(r, s) = \gamma_x(r + A, s + A) \quad \forall A, r, s, \in \mathbb{Z}$ (that is, the autocovariance function is time-independent)

Time Series Analysis is a very active field of Statistics, and there are a variety of sophisticated and powerful methods to analyze stationary time series. However, in practice, we often have data that is problematic – for example, there may be missing values, or observations may have been recorded at irregular time intervals [1]. To account for this problem, *interpolation* must be used. *Interpolation* refers to the construction of new data points in a set of known data points. When working with time series, we wish to interpolate observations whenever there are gaps in our data, so that we may obtain a time series which has an observation for each $t \in \mathbb{Z}$, where $\mathbb{Z}$ is the index set, and each t is spaced by the same amount.

Proposal

This Honours Project will be focused on researching a number of time series interpolation algorithms in the scientific literature, implementing them in code, and testing and benchmarking them. Algorithms will be implemented using R, by creating functions, documenting them, and creating an R package [2] to be published online for others to use. Once these functions have been implemented, they will be tested on real-world time series data, in order to compare the performance of various interpolation methods. They will be evaluated on a number of criteria such as estimation variance, mean-squared error (MSE) and other metrics regularly used in academic papers of this nature (such as those suggested in Table 1 of *Interpolation in Time Series: An Introductive Overview of Existing Methods, Their Performance Criteria and Uncertainty* [3].

According to a recent paper [3] (October 2017) published by Lepot et. al., there is a gap in the literature concerning the performance comparison of various time series interpolation algorithms. If significant progress is made on this Honours Project, the results should be publishable in an academic journal.

References

[1] Wesley S. Burr. *Air Pollution and Health: Time Series Tools and Analysis.* Queen's University, PhD thesis. 2012.

[2] Wesley S. Burr (2012). *tsinterp: A Time Series Interpolation Package for R.* R Package.

[3] Mathieu Lepot, Jean-Baptiste Aubin, and François H.L.R. Clemens. *Interpolation in Time Series: An Introductive Overview of Existing Methods, Their Performance and Uncertainty Assessment.* Water 2017, 9(10), 796.

[4] Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods.* Springer. 2006.