Preface

Huge amount of information is available as time series in many scientific fields: geophysics, astronomy, biophysics, quantitative finance, Internet traffic, etc. Processing so many time series is possible only by means of automatic algorithms usually designed in data mining. One of the critical tasks which has to be achieved by these algorithms is the automatic estimation of the trend contained in an arbitrary noisy time series. The aim of our book is to provide several automatic algorithms for nonmonotonic trend estimation. We do not intend to review the existing automatic trend estimation algorithms, but to present a thorough analysis for those presented in this book.

Obviously, an automatic algorithm is not able to work for all imaginable time series. By its automatic feature we mean that, without any subjective intervention, it efficiently processes time series of a well-defined type. The greater the diversity of the time series types, the more "automatic" the algorithm is. Therefore in designing a trend estimation algorithm an essential component is the method to evaluate its accuracy for a large diversity of time series. However, the algorithms are very often tested under unrealistic conditions and on too small number of time series. One reason for this situation is that the time series theory is dominated by stationary stochastic processes. The theoretical results for nonstationary time series containing a trend hold only under restrictive conditions, seldom satisfied by the real time series.

When the statistical theory is not applicable, Monte Carlo experiments can be used to evaluate the accuracy of the automatic algorithms. Even then the results are useful only if the members of the statistical ensemble have a diversity comparable with that of the real time series. The main difficulty is to generate realistic nonmonotonic trends. Usually, Monte Carlo simulations are performed on artificial time series much simpler than those encountered in practice, with monotonic (linear, power-law, exponential, and logarithmic) or periodic (sinusoidal) trends. The approach based on numerical Monte Carlo experiments in our book is much more general and the trends generated by our original algorithm are meaningful for real time series.

Chapter 1 contains fundamentals in probability theory, statistics, and time series theory which are used in the rest of the book. We analyze the autoregressive noise of order one denoted AR(1), which is a simple model depending only on two parameters: the variance and the constant of the serial correlation. Even for more complex noises an AR(1) model is a zero order approximation capturing their most important features. The noise serial correlation essentially influences the accuracy of the estimated trend because when it increases, the large-scale fluctuations of the noise cannot be distinguished from the trend variations.

In Chap. 2 we construct the statistical ensemble on which the Monte Carlo experiments are performed. There is no rigorous mathematical method to demonstrate that the variability of the obtained artificial time series is rich enough to simulate the variability of the real time series. In fact we construct an independent "numerical reality" on which we perform numerical experiments. Therefore, our approach is more typical to computational physics than to data mining or mathematical statistics. As examples of Monte Carlo experiments we evaluate the confidence interval for a method to estimate the serial correlation parameter of an AR(1) noise and we present a numerical method for testing if a time series is uncorrelated.

In Chaps. 3 and 4 we analyze in detail the accuracy of the classical algorithms of polynomial fitting and moving average in the case of arbitrary nonmonotonic trends. The quality of the estimated trend depends mainly on three parameters: the number of the time series values, the ratio between the amplitudes of the trend variations and the noise fluctuations, and the serial correlation of the noise. Our analysis shows that even in the case of the simplest trend estimation algorithms, due to the many parameters on which the artificial time series depend, a realistic evaluation of their performances is difficult and laborious.

In the last three chapters we present our original automatic algorithms for processing nonstationary time series containing a stationary noise superposed over a nonmonotonic trend. Their performances are tested by means of numerical experiments of the same type as those used in the previous chapters. The algorithms are designed to work on any time series, even if it has only a few values. Obviously, the best results are obtained for an AR(1) noise superposed over a deterministic trend with at least several hundreds of values. For other types of time series the outcomes of the algorithms have to be statistically analyzed by Monte Carlo experiments.

In Chap. 5 we design an automatic algorithm, called the averaged conditional displacement (ACD), to estimate a monotonic trend as a piecewise linear curve. The Monte Carlo experiments indicate that its accuracy is comparable with that of the classical methods, but it has the advantage to be automatic and to describe a much richer set of monotonic trend shapes. Applied to a time series with an arbitrary nonmonotonic trend, the ACD algorithm extracts one of the possible monotonic components which can be associated with the given trend. The probability that the estimated monotonic component is real can be estimated by a method based on surrogate time series.

In Chap. 6 we define the timescale of a local extremum of a time series such that it allows a classification of the local extrema with respect to their importance for the global shape of the time series. The local extrema with scales greater than a given value provide a partition of a noisy time series in segments which approximate the monotonic parts of the trend from a time series. The quality of this approximation is improved by first applying a moving average to the noisy time series. We use the monotonic component estimated by the ACD algorithm as a reference to measure the magnitude of the nonmonotonic variations of a time series. In this way we can build a criterion to stop the partition of a time series when the resulting segments may be considered monotonic.

In the last chapter we give an automatic form to the repeated central moving average (RCMA) analyzed in Chap. 4. In order to adjust the parameters of the RCMA algorithm to the characteristics of the processed time series, we have designed two simple statistical methods to estimate the noise serial correlation and the ratio between the amplitudes of the trend variations and of the noise fluctuations. The partitioning algorithm presented in Chap. 6 is used now to determine the local extrema of the estimated trend which corresponds to the real trend and not to the smoothed noise.

We illustrate the functioning of the analyzed algorithms by processing time series from astrophysics, finance, biophysics, and paleoclimatology. The examples of real time series are typical to the complex situations encountered in practice: data missing from the time series, superposition of several types of noises, long time series with tens of thousands of values, non-Gaussian probability distributions with fat tails, repeated values of the time series, additional conditions imposed on time series by the physical laws governing the studied phenomenon.

Our analysis is restricted to AR(1) noises superposed over nonmonotonic trends, but our methods can be applied to study other noise models. Such new applications could be: autoregressive noise of higher orders, long-range correlated noises, unevenly sampled time series, asymmetric probability distribution of the time series values. Obviously, the number of parameters could increase and the analysis of the accuracy of the estimated trend would become more burdensome.

We have limited our analysis to four methods of trend estimation: two classical (polynomial fitting and moving average) and two original and automatic (one for monotonic trends and the other for arbitrary nonmonotonic trends). Other trend estimation methods can be analyzed using the same type of Monte Carlo experiments. In order to obtain significant results, it is essential to use a statistical ensemble of artificial time series with a variety of trend shapes at least as rich as that generated by our algorithm presented in Chap. 2.

Even if the main definitions and theorems used in the book are briefly presented, nevertheless it is recommended that the reader has the knowledge of basic notions in probability, mathematical statistics, and time series theory. This book is of interest for researchers who need to process nonstationary time series. Detailed descriptions of all the numerical methods presented in the book allow the reader to reproduce the original automatic algorithms for trend estimation and time series partitioning. In addition, the source codes in MATLAB of the computer programs implementing them are freely available on the web so that the researchers who merely apply trend estimation algorithms could successfully use them.

Cluj-Napoca, April 2012

Călin Vamoş
Maria Crăciun

http://www.springer.com/978-94-007-4824-8

Automatic trend estimation
Vamos¸, C.; Cr˘aciun, M.
2013, X, 131 p. 77 illus., Softcover
ISBN: 978-94-007-4824-8