SPRINGER BRIEFS IN PHYSICS

Călin Vamoş · Maria Crăciun

Automatic Trend Estimation

Springer

SpringerBriefs in Physics

For further volumes:
http://www.springer.com/series/8902

Călin Vamoş · Maria Crăciun

Automatic Trend Estimation

Springer

Călin Vamoş
"Tiberiu Popoviciu" Institute of Numerical Analysis
Romanian Academy
Cluj-Napoca
Romania

Maria Crăciun
"Tiberiu Popoviciu" Institute of Numerical Analysis
Romanian Academy
Cluj-Napoca
Romania

ISSN 2191-5423 ISSN 2191-5431 (electronic)
ISBN 978-94-007-4824-8 ISBN 978-94-007-4825-5 (eBook)
DOI 10.1007/978-94-007-4825-5
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2012942924

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

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

Contents

1 Introduction 1
1.1 Discrete Stochastic Processes and Time Series 1
1.2 Trend Definition and Estimation 4
1.3 AR(1) Stochastic Process 7
References 12

2 Monte Carlo Experiments 15
2.1 Monte Carlo Statistical Ensembles. 15
2.2 Numerical Generation of Trends 18
2.3 Numerical Generation of Noisy Time Series. 21
2.4 Statistical Hypothesis Testing 26
References 30

3 Polynomial Fitting 31
3.1 Polynomial Fitting. 31
3.2 Polynomial Fitting of Artificial Time Series 34
3.3 An Astrophysical Example 38
References 42

4 Noise Smoothing 43
4.1 Repeated Central Moving Average 43
4.2 Smoothing of Artificial Time Series 49
4.3 A Financial Example 53
References 58

5 Automatic Estimation of Monotonic Trends 61
5.1 Average Conditional Displacement (ACD) Algorithm 61
5.2 Automatic ACD Algorithm. 65
5.3 Evaluation of the ACD Algorithm. 70
5.4 A Climatological Example 74

5.5 Monotonic Components of Nonmonotonic Trends 76
References 80

6 Estimation of Monotonic Trend Segments from a Noisy Time Series 81
6.1 Time Scale of Local Extrema 81
6.2 Local Extrema of Noisy Time Series 85
6.3 Local Extrema of RCMA Trends. 89
6.4 Significant Local Extrema of a Real Time Series 92
References 97

7 Automatic Estimation of Arbitrary Trends 99
7.1 Automatic RCMA (AutRCMA). 99
7.2 Statistical Significance of the Local Extrema of the AutRCMA Trend 105
References 110

Appendix A: Statistical Properties of the Linear Regression 111

Appendix B: Spurious Serial Correlation Induced by MA. 113

Appendix C: Continuous Analogue of the ACD Algorithm 117

Appendix D: Standard Deviation of a Noise Superposed over a Monotonic Trend 121

Appendix E: Construction of a Partition of Scale Δn 127

Appendix F: Estimation of the Ratio Between the Trend and Noise Magnitudes 129

Chapter 1
Introduction

A complete presentation of the theory of stochastic processes can be found in any treatise on the probability theory, e.g., [18] and for time series theory one can use [4]. In this introductory chapter we briefly present some basic notions which are used in the rest of the book. The main methods to estimate trends from noisy time series are introduced in Sect. 1.2. In the last section we discuss the properties of the order one autoregressive stochastic process AR(1) which has the serial correlation described by a single parameter and which is a good first approximation for many noises encountered in real phenomena.

1.1 Discrete Stochastic Processes and Time Series

At the occurrence of an event ω the *random variable* X takes the value $X(\omega) = x$. We follow the practice of denoting by small letters the realizations of the random variable denoted by the corresponding capital letters. Throughout this book we consider only continuous random variables with real values. If the random variable is absolutely continuous, then it has a *probability density function* (pdf) denoted $p(x)$. The *cumulative distribution function* (cdf) $F(x) = P(X \leq x)$ is the probability that the random variable X takes on a value less than or equal to x. We denote the *mean* of the random variable by $\mu = \langle X \rangle$ and its *variance* by $\sigma^2 = \langle (X - \langle X \rangle)^2 \rangle$.

The evolution in time of a random phenomenon is modeled by a *stochastic process*, i.e., a family of random variables $\{X(t),\ t \in I \subset \mathbf{R}\}$ defined on the same probability space and indexed by a set of real numbers I. In this book we study only discrete stochastic processes for which I contains equidistant sampling moments. The observations are made at discrete time moments $t_n = t_0 + (n-1)\Delta t$, where $n = 1, 2, \ldots, N$, Δt is the sampling interval, and t_0 is the initial time. The observed values $x_n \equiv x(t_n)$ are realizations of the corresponding random variables $X_n \equiv X(t_n)$. Although the number of observations is always finite, we assume that there is an infinite stochastic process $\{X_n, n = 0, \pm 1, \pm 2, \ldots\}$ whose realizations for $n < 1$ and $n > N$ have not

C. Vamoş and M. Crăciun, *Automatic Trend Estimation*, SpringerBriefs in Physics,
DOI: 10.1007/978-94-007-4825-5_1, © The Author(s) 2012

been observed. To distinguish between the infinite stochastic process which models the time evolution of the natural phenomenon and its measurements, we call *time series* the finite sequence of real numbers $\{x_n, n = 1, 2, ..., N\}$.

The joint cdf of the random variables $X_{n_1}, X_{n_2}, \ldots, X_{n_m}$ is the probability that their values are smaller than some given values

$$F_{\mathbf{n}}(\mathbf{x}) = P(X_{n_1} \leq x_1, \ldots, X_{n_m} \leq x_m),$$

where $\mathbf{n} = (n_1, \ldots, n_m)$ and $\mathbf{x} = (x_1, \ldots, x_m)$. For absolutely continuous random variables there exists the joint pdf $p_{\mathbf{n}}(\mathbf{x})$. A stochastic process is (strictly) *stationary* if for every index vector $\mathbf{n}$ and integer d we have $F_{\mathbf{n}+d\mathbf{1}} = F_{\mathbf{n}}$ or $p_{\mathbf{n}+d\mathbf{1}} = p_{\mathbf{n}}$, where $\mathbf{1} = (1, 1, \ldots, 1)$, i.e., its joint probabilities do not change under temporal translations. From this definition, for $m = 1$ it follows that all the components of a stationary process have the same probability distribution $p_n(x) = p(x)$ for all integers n. Such a stochastic process is called *identically distributed*.

The *autocovariance function* of a stochastic process with finite variance for all its components ($\sigma_n^2 < \infty$) is defined as

$$\gamma(n, m) = \langle (X_n - \langle X_n \rangle)(X_m - \langle X_m \rangle) \rangle. \tag{1.1}$$

Obviously $\gamma(n, n) = \sigma_n^2$. If the stochastic process is stationary, then

$$\gamma(n + d, m + d) = \gamma(n, m), \tag{1.2}$$

for all n, m, d integers and the autocovariance function depends only on the lag $h = n - m$ so that $\gamma(h) \equiv \gamma(h, 0)$. It is easy to show that $\gamma(0) \geq 0$, $\gamma(h) = \gamma(-h)$, and $|\gamma(h)| \leq \gamma(0)$ for any h. The *autocorrelation function* of a stationary stochastic process is defined as $\rho(h) = \gamma(h)/\gamma(0)$ and then $\rho(0) = 1$.

Usually the observed time series do not satisfy the condition imposed to strictly stationary stochastic processes. Furthermore, the analysis of time series is often reduced only to the statistical moments of second order. Therefore one defines a subclass of the stationary process more suitable for modeling of real phenomena. A stochastic process is *weak-stationary* if $\langle |X_n^2| \rangle < \infty$, $\langle X_n \rangle = \mu$ for all integers n and satisfies Eq. (1.2). A special weak-stationary process is the *white noise*, for which the components are uncorrelated $\gamma(h) = \sigma^2 \delta_{h0}$, where δ_{nm} is the Kronecker delta. Such a stochastic process is denoted by $X_n \sim WN(\mu, \sigma^2)$.

Another subclass of stationary processes contains the *independent and identically distributed* (i.i.d.) stochastic processes. The components of an i.i.d. process are mutually independent $p_{\mathbf{n}}(\mathbf{x}) = p_{n_1}(x_1) p_{n_2}(x_2) \ldots p_{n_m}(x_m)$. They are also identically distributed $p_{n_i}(x_i) = p(x_i)$ and then $p_{\mathbf{n}+h\mathbf{1}}(\mathbf{x}) = p(x_1) p(x_2) \ldots p(x_m) = p_{\mathbf{n}}(\mathbf{x})$ so that, if the stochastic process is infinite, the stationarity condition (1.2) is satisfied.

If the properties of the components of a stochastic process vary in time, then the stochastic process is *nonstationary*. As an example of nonstationary stochastic process we consider the *random walk* $\{X_n, n = 0, 1, 2, \ldots\}$ defined as

$$X_n = X_{n-1} + Z_n \quad \text{for} \quad n > 0, \tag{1.3}$$

where $\{Z_n\}$ is an i.i.d. stochastic process with zero mean, variance σ^2, and $X_0 = Z_0$. Obviously $\langle X_n \rangle = 0$ and for $n \leq m$ the autocovariance function given by Eq. (1.1) becomes

$$\gamma(n, m) = \langle X_n X_m \rangle = \langle X_n (X_n + Z_{n+1} + \cdots + Z_m) \rangle = \langle X_n^2 \rangle$$

because X_n depends only on $Z_0, Z_1, \ldots, Z_n$ which are independent of $Z_{n+1}, \ldots, Z_m$. Because

$$\langle X_n^2 \rangle = \sum_{k=0}^{n} \langle Z_k^2 \rangle + \sum_{k \neq l}^{n} \langle Z_k Z_l \rangle = (n+1)\sigma^2$$

we have

$$\gamma(n, m) = (1 + \min\{n, m\})\sigma^2. \tag{1.4}$$

Hence the autocovariance function of the random walk is not invariant to temporal translations and $\{X_n\}$ is a nonstationary process.

In practice we do not have access to random variables or stochastic processes but only to their realizations and we have to use the methods of the mathematical statistics in order to estimate the parameters of the observed phenomena. Let us consider a random variable X and one of its realizations $x^{(s)}$.[1] The set formed by S independent realizations $\{x^{(1)}, x^{(2)}, \ldots, x^{(S)}\}$ is called *sample of volume* S and it allows the estimation of the parameters of X. For instance the mean $\mu = \langle X \rangle$ is approximated by the *sample mean*

$$\mu^{\text{est}} \equiv \widehat{\mu} = \frac{1}{S} \sum_{s=1}^{S} x^{(s)}. \tag{1.5}$$

We make the convention that the quantities computed by means of a sample are denoted by a hat or with the superscript 'est'. By means of the law of large numbers one proves under rather general conditions that $\widehat{\mu}$ tends to μ when S tends to infinity. In the same way we define the *sample variance*

$$\widehat{\sigma}^2 = \frac{1}{S} \sum_{s=1}^{S} \left(x^{(s)} - \widehat{\mu} \right)^2. \tag{1.6}$$

Analogous relations can be used for a stationary time series $\{x_1, x_2, \ldots, x_N\}$. Instead of the sample mean (1.5) we define the *temporal mean*

[1] We have changed the usual notation x_s in order to avoid the confusion with the terms of a time series.

$$\overline{x} = \frac{1}{N}\sum_{n=1}^{N} x_n. \tag{1.7}$$

Since the associated stochastic process $\{X_n\}$ is stationary, all the terms in the sum have identical pdfs and $\overline{x}$ tends to its theoretical mean μ when N tends to infinity. A similar analogy can be made for the sample variance (1.6). The serial correlation of a time series is characterized by the *sample autocovariance function*

$$\widehat{\gamma}(h) = \frac{1}{N}\sum_{n=1}^{N-h}(x_{n+h} - \overline{x})(x_n - \overline{x}), \quad 0 \le h < N. \tag{1.8}$$

For $-N < h \le 0$, we have $\widehat{\gamma}(h) = \widehat{\gamma}(-h)$. If X_n is a linear combination of the components of an i.i.d. stochastic process with finite fourth order moment and the sum of the absolute values of the linear combination coefficients is finite, then the estimator (1.8) is biased, but its asymptotic distribution has the mean equal to the theoretical autocovariance function ([4], Chap. 7). The *sample autocorrelation function* is given by

$$\widehat{\rho}(h) = \widehat{\gamma}(h)/\widehat{\gamma}(0), \quad |h| < N. \tag{1.9}$$

1.2 Trend Definition and Estimation

Stochastic processes model the random phenomena as opposed to the deterministic phenomena which are modeled by numerical functions of time. There are many situations when different random and deterministic phenomena overlap. In the simplest case, a deterministic and a random phenomenon, mutually independent, are superposed (for example the instrumental noise affecting a measured physical quantity). The most frequently used model is the stochastic process

$$X_n = f_n + Z_n, \tag{1.10}$$

where $\{Z_n\}$ is a stationary stochastic process with zero mean $\langle Z_n\rangle = 0$ named *additive noise* and $f_n = f(t_n)$ are the values at the sampling moments of the deterministic function named *trend*.

We denote by $p_Z(z)$ the pdf of Z_n and by $p_X(x,n)$ that of X_n. Because $\{Z_n\}$ is stationary, p_Z does not depend on n. According to Eq. (1.10), p_X is equal to p_Z translated by f_n

$$p_X(x,n) = p_Z(x - f_n). \tag{1.11}$$

The explicit dependence of p_X on the time index n indicates that $\{X_n\}$ is a nonstationary process with the mean varying in time

$$\langle X_n \rangle = f_n + \langle Z_n \rangle = f_n.$$

Using this relation in Eq. (1.1) we obtain

$$\gamma_X(n, m) = \langle Z_n Z_m \rangle = \gamma_Z(n - m),$$

hence its autocovariance function is identical with that of the stationary noise, the nonstationarity of $\{X_n\}$ being restricted only to its mean.

We cannot apply the usual statistical methods to a single realization of a nonstationary stochastic process given by Eq. (1.10)

$$x_n = f_n + z_n. \tag{1.12}$$

If more time series are available $\{x_n^{(s)}\}$, $s = 1, 2, \ldots, S$, then for each term of the stochastic process a sample is available and we can evaluate the trend using the simple average given by Eq. (1.5)

$$f_n = \frac{1}{S} \sum_{s=1}^{S} x_n^{(s)}. \tag{1.13}$$

In practice more time series obtained under exactly the same conditions are rarely available. Usually we have to analyze a single time series and then the trend must be estimated and removed from the given time series.

The estimation of the trend $\{f_n\}$ from the time series $\{x_n\}$ is more accurate if the properties of the trend are distinct from those of the noise. Usually the trend is qualitatively characterized as being the component of a time series that is "slowly changing in time" [3]. It is implicitly assumed that the noise does not have this property. In order to formulate more precisely this observation, we denote by τ_Z the time scale of the stochastic process $\{Z_n\}$. It is defined as the time interval for which $\{z_n\}$ becomes uncorrelated. For example, a white noise has $\tau_Z = 1$. The condition that the trend has a slow variation in comparison with the noise means that the trend values remain strongly correlated over intervals of the order τ_Z. If we denote by τ_f the time scale for which the trend values are correlated, then the property of slow variation of the trend corresponds to the condition $\tau_Z < \tau_f$. Usually the trend has only a few monotonic components on the whole time series, so τ_f is of the magnitude order of the time series length. Then the condition $\tau_Z < \tau_f$ represents the requirement that the values of the noise become uncorrelated on time intervals smaller than the length of the analyzed time series.

Obviously, there are situations when $\tau_Z \sim \tau_f$, for example, if the noise is strongly correlated or if the time series is too short. In such cases the slowest fluctuations of the noise are confounded with the variations due to the trend, and their separation is very difficult. In such cases the false trend is called *stochastic trend*. It is only partially possible to avoid such errors when we have more information on the specific properties of noise or trend.

Confusion is often made between the trend definition and trend estimation. The definition is based on the different nature of the two terms in Eq. (1.10): the trend is the deterministic part of a time series with additive noise. Then it can be determined by means of Eq. (1.13) if we dispose of a sufficient number of time series (1.12) generated in the same conditions by the same phenomenon. Due to its random nature, the averaging in Eq. (1.13) eliminates the noise from the averaged time series. Difficulties occur when we have to estimate the trend from a single time series. Therefore a statement as "a rigorous and satisfactory definition of either the trend of nonlinear nonstationary data or the corresponding detrending operation still is lacking" [17] really refers to trend estimation, not to trend definition.

There are many methods of trend estimation [1, 13], but most of them belong to two major classes, parametric and nonparametric. A parametric method chooses for $f(t)$ a functional form depending on several parameters and then computes their values by regression. As a typical parametric method we analyze in Chap. 3 the *polynomial fitting*. The estimation of the trend obtained with this method becomes worse as the form of the function $f(t)$ becomes more complicated and as the noise serial correlation increases.

Among the nonparametric methods to estimate the trend we study in Chap. 4 the noise smoothing by *moving average* (MA). The *central moving average* (CMA) is defined as

$$\vartheta_x(n) = \frac{1}{2K+1} \sum_{k=-K}^{K} x_{n+k}, \tag{1.14}$$

where K is the semi-length of the averaging window given by a positive integer. Applying the CMA to the time series (1.12) we obtain

$$\vartheta_x(n) = \vartheta_f(n) + \vartheta_z(n). \tag{1.15}$$

Since the noise is more fluctuant than the trend, it is more damped by the CMA and then $\vartheta_x(n) \approx \vartheta_f(n)$. If the trend is not too much distorted by smoothing, then the averaged time series is a good estimator of the trend $f_n \approx \vartheta_x(n)$.

Every method for trend estimation can be transformed into an automatic algorithm if the optimum values of its parameters for the analyzed time series are automatically determined. The difficulty consists in the great variability of the possible time series. For instance, the ratio between the amplitude of the trend variations and the amplitude of noise fluctuations has a significant influence on the trend estimation quality. If the time series is dominated by noise, then the degree of the estimated polynomial trend must be small, such that the estimated trend should not follow the noise fluctuations. If the CMA is used, then K must take large values such that the noise should be strongly smoothed. Inversely, when the time series is dominated by trend, then the optimum degree of the polynomial trend increases and K decreases. The accuracy of the estimated trend also depends on the serial correlation of the noise. Therefore, before choosing the correct values of the parameters of the automatic algorithm,

we have to estimate the main properties of the time series containing an arbitrary nonmonotonic trend.

Sometimes the time series analysis is performed not in order to estimate the trend, but to study the noise. In this case we say that a *detrending* method is applied. *Differencing* is such an algorithm. We define the difference operator ∇_d for a positive integer d as

$$\nabla_d X_n = X_{n+d} - X_n, \tag{1.16}$$

where $\{X_n\}$ is a stochastic process. This definition also holds for deterministic functions. Since the operator ∇_d is linear, its action on the stochastic process (1.10) is separated into two terms

$$\nabla_d X_n = \nabla_d f_n + \nabla_d Z_n. \tag{1.17}$$

If the step d is small enough, then the term due to the trend is much smaller than the term due to the noise because according to its definition the trend has a slow variation. Hence we have the approximation

$$\nabla_d X_n \approx \nabla_d Z_n. \tag{1.18}$$

It is obvious that it is not always possible to find such a value for d, for example, if the noise fluctuations have an amplitude of the same magnitude order as the trend variation over a sample interval.

1.3 AR(1) Stochastic Process

The autoregressive-moving average (ARMA) processes are the fundamental models in time series theory. A detailed presentation can be found in [4]. In this book we use only a particular form of ARMA processes. The process $\{Z_n, n = 0, \pm 1, \pm 2, \ldots\}$ is an AR(1) process if it is stationary and for every n

$$Z_n = \phi Z_{n-1} + W_n. \tag{1.19}$$

where $\{W_n\} \sim WN(0, \sigma^2)$. One can specify more exactly the properties of such stochastic processes if the noise characteristics are better defined. For example, if the white noise is Gaussian, then the AR(1) process is also Gaussian.

By iteratively applying Eq. (1.19) h times we obtain

$$Z_n = W_n + \phi W_{n-1} + \cdots + \phi^{h-1} W_{n-h+1} + \phi^h Z_{n-h}. \tag{1.20}$$

Because the process $\{Z_n\}$ is stationary, the random variables Z_n and Z_{n-h} have the same norm and then the norm of the last term in Eq. (1.20) is ϕ^h times the norm of the left term. If $|\phi| < 1$, then for any positive number $\varepsilon < 1$, there exists $h > \ln \varepsilon / \ln |\phi| > 0$ such that the last term from the right side can be neglected

$$Z_n = W_n + \phi W_{n-1} + \cdots + \phi^{h-1} W_{n-h+1} + \mathcal{O}(\varepsilon) \tag{1.21}$$

for $\varepsilon \to 0$. From Eq. (1.21) it follows that the influence of the white noise reduces as h increases and Z_n depends only on the past values of the noise. Such a stochastic process obtained for $|\phi| < 1$ is called *causal*.

In the following we analyze the basic properties of the causal AR(1) process. If we take the mean of Eq. (1.19) we obtain $\langle Z_n \rangle = 0$. Its square is

$$Z_n^2 = \phi^2 Z_{n-1}^2 + 2\phi Z_{n-1} W_n + W_n^2.$$

In accordance with Eq. (1.21) the random variables Z_{n-1} and W_n are independent, and then $\langle Z_{n-1} W_n \rangle = 0$. If we take the mean of the last equation, we obtain a relation between the variances of successive random variables

$$\sigma_n^2 = \phi^2 \sigma_{n-1}^2 + \sigma^2.$$

Because the AR(1) process is stationary, it follows that for every n we have the same value for the variance $\sigma_n \equiv \sigma_s$ and then

$$\sigma_s^2 = \frac{\sigma^2}{1 - \phi^2}. \tag{1.22}$$

In order to compute the autocovariance function we multiply Eq. (1.19) with Z_{n-h} and we take the mean. Because the mean of Z_n vanishes and $\langle Z_{n-h} W_n \rangle = 0$, we obtain

$$\gamma(h) \equiv \langle Z_n Z_{n-h} \rangle = \phi \gamma(h-1)\,.$$

By applying successively this relation and taking into account that $\gamma(0) = \sigma_s^2$, we have

$$\gamma(h) = \sigma_s^2 \phi^h. \tag{1.23}$$

The spectral density is the Fourier transform of the autocovariance function

$$g(\nu) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \mathrm{e}^{-2\pi i h \nu} \gamma(h),$$

where the positive number ν is the frequency. By direct calculation with $\gamma(h)$ given by Eq. (1.23) we obtain

$$g(\nu) = \frac{\sigma^2}{2\pi} \frac{1}{1 - 2\phi \cos 2\pi\nu + \phi^2}. \tag{1.24}$$

We will use the causal AR(1) process for the most of the stationary noise models. One of its advantages is that its serial correlation depends on the single parameter ϕ.

In addition, it has been used in many models of natural phenomena: the wind speed fluctuations [7], radar signals [5], climatic phenomena variability [11], electroencephalographic activity [10], heart interbeat time series [6], the daily temperature fluctuations [8], X-ray emission from the active galactic nuclei [14], sunspots variability [12], etc. As a result of their simple mathematical properties and their direct physical interpretation, the realizations of AR(1) processes have been used as artificial time series to analyze some numerical algorithms for monotonic trend removal [15], some surrogate data test for nonlinearity [9] or for renormalization group analysis [2].

In the following we briefly discuss the solution of Eq. (1.19) for other values of the parameter ϕ. In the nonstationary case $|\phi| = 1$, all the terms in Eq. (1.20) have unit coefficients and for any delay h none of them can be neglected. Every random variable Z_n is an infinite sum of terms with the same norm, hence its norm is infinite and Eq. (1.19) has no stationary solution. For $\phi = 1$ the AR(1) process has the same recursive formula as the random walk (1.3), but while the former has no initial term, the latter begins with Z_0.

When $|\phi| > 1$, Eq. (1.21) is not true any more because in this case ϕ^h increases when h increases. However, the same reasoning as in the case $|\phi| < 1$ can be repeated if we write Eq. (1.19) in the form

$$Z_{n-1} = \phi^{-1} Z_n - \phi^{-1} W_n.$$

In this way we obtain a stationary causal AR(1) process because $|\phi^{-1}| < 1$, but in the reverse temporal direction and with the variance of the noise equal to $\phi^{-1}\sigma$. It means that the formulas can be obtained from those for the causal ones by replacing ϕ with ϕ^{-1} and σ with $\phi^{-1}\sigma$. Such a stochastic process is called *acausal*.

The values of the parameter ϕ can be both positive and negative. In order to qualitatively characterize the difference between the two situations we use the fact that for $\phi = 0$, the AR(1) process reduces to a white noise with uncorrelated terms. When $\phi > 0$, from Eq. (1.19) it follows that the fluctuations due to the white noise are superposed over the term ϕZ_{n-1} which memorizes a part of the previous value of the time series. Hence, for large ϕ the successive values of the time series are close to each other and the fluctuations due to the white noise are small. Therefore, in comparison with a realization of a white noise, for $\phi > 0$ the graphical representation of an AR(1) process is less fluctuant and resembles to a deterministic trajectory disturbed by a random fluctuation (Fig. 1.1c).

When $\phi < 0$ the white noise is superposed over the term $-|\phi| X_{n-1}$ which has an opposite sign to the previous term of the time series. Consequently, the white noise fluctuations are enhanced and the series values fluctuate more rapidly than the white noise, as shown in Fig. 1.1a. The successive values of the autocovariance function (1.23) are of opposite signs and the time series is called anticorrelated.

In practice the time series have a finite length and usually they are considered realizations of a finite part of a stochastic process of infinite length. For an AR(1) process, the first term of the time series is correlated with the preceding term which has not been recorded. But the first term of a numerically generated time series

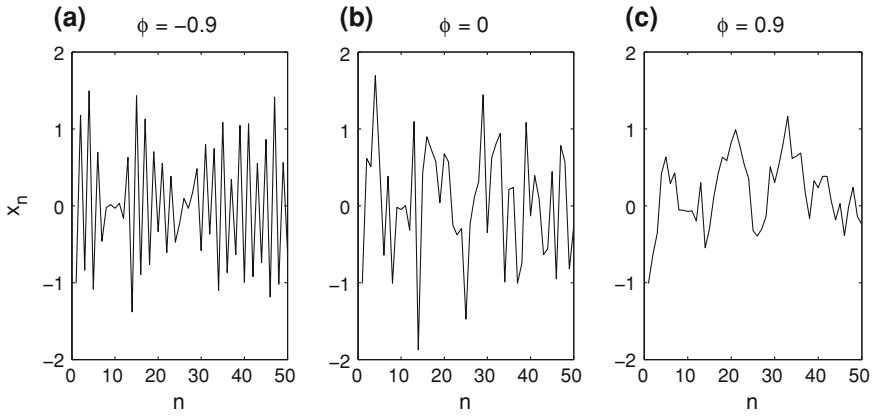

Fig. 1.1 Realizations of an anticorrelated (**a**), uncorrelated (**b**), and correlated (**c**) AR(1) process

cannot be related to realizations of other preceding random variables. Therefore, a numerically generated time series is never strictly a realization of a finite part of an ideal stationary stochastic process of infinite length. Since Eq. (1.19) defining the AR(1) process is recursive, the first term must be defined by an additional relation. As we show in the following, the manner in which this additional relation is chosen can essentially modify the properties of the stochastic process.

We call *finite AR(1) process* a stochastic process of finite length satisfying the recursive relation in Eq. (1.19). We consider only the case when the white noise $\{W_n\}$ is Gaussian, otherwise one can obtain only asymptotic distributions of the AR(1) process. Let us denote the finite AR(1) process by $\{\widehat{Z}_n\}$, $n = 1, 2, \ldots, N$. Because $\widehat{Z}_n$ satisfies Eq. (1.19) for $n > 1$, by successive applications of this relation we can express the terms of the stochastic process as a finite sum

$$\widehat{Z}_n = W_n + \phi W_{n-1} + \cdots + \phi^{n-2} W_2 + \phi^{n-1} \widehat{Z}_1. \tag{1.25}$$

In the following we consider only the causal AR(1) processes with $|\phi| < 1$. As shown above, the acausal process is equivalent with a causal one generated in reverse temporal order.

If $\widehat{Z}_1$ is a Gaussian random variable with variance $\widehat{\sigma}_1$ and zero mean, then from Eq. (1.25) it follows that $\widehat{Z}_n$ is the sum of n Gaussian random variables, hence it has also a Gaussian distribution with variance

$$\widehat{\sigma}_n^2 = \sigma^2 \left(1 + \phi^2 + \phi^4 + \cdots + \phi^{2(n-2)}\right) + \widehat{\sigma}_1^2 \phi^{2(n-1)}.$$

Applying the formula for the sum of a geometric series we have

$$\widehat{\sigma}_n^2 = \sigma_s^2 + \left(\widehat{\sigma}_1^2 - \sigma_s^2\right) \phi^{2(n-1)}, \tag{1.26}$$

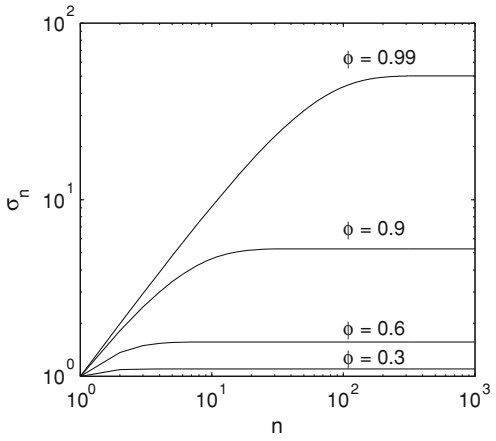

Fig. 1.2 The standard deviation of a finite AR(1) process for different values of ϕ when the first term coincides with the white noise $\widehat{Z}_1 = W_1$

where we have used Eq. (1.22). The variance of the finite AR(1) process has a constant term σ_s^2 equal with the variance of the infinite AR(1) process and a variable term which tends asymptotically to zero because $|\phi| < 1$. Hence the finite AR(1) process is nonstationary presenting transient effects, i.e., its variance approximates the theoretical one $\widehat{\sigma}_n \simeq \sigma_s$ only after a time interval t_0 for which ϕ^{2t_0} can be neglected.

For $\widehat{\sigma}_1 = \sigma_s$ the variable term in Eq. (1.26) vanishes and $\widehat{\sigma}_n = \sigma_s$ for all $n \leq N$. Hence, if for a finite AR(1) process we choose $\widehat{Z}_1 = (\sigma_s/\sigma)W_1$, then all the terms have the same variance. This choice is natural because it is more reasonable to take the first term of the finite AR(1) process similar to the stationary infinite AR(1) process and not to the white noise. In the following we show that for $\widehat{\sigma}_1 = \sigma_s$ the properties of the finite AR(1) process are identical to those of a finite sample of a stationary infinite AR(1) process.

The autocovariance function $\widehat{\gamma}(h, n) = \langle \widehat{Z}_n \widehat{Z}_{n-h} \rangle$ can be calculated only if $|h| < N$ and $h < n \leq N + h$. Unlike the autocovariance function (1.23) the quantity $\widehat{\gamma}(h, n)$ depends also on n since it exists only for certain values of n. Therefore $\{\widehat{Z}_n\}$ is not a stationary stochastic process in a strict mathematical meaning. However, when it exists, we can show with the same method as for Eq. (1.23) that $\langle \widehat{Z}_n \widehat{Z}_{n-h} \rangle = \phi \langle \widehat{Z}_{n-1} \widehat{Z}_{n-h} \rangle$. Then instead of Eq. (1.23) we obtain

$$\widehat{\gamma}(h, n) = \phi^h \widehat{\sigma}_{n-h}^2. \tag{1.27}$$

If we choose $\widehat{\sigma}_1 = \sigma_s$, then $\widehat{\sigma}_{n-h} = \sigma_s$ is constant and when $\widehat{\gamma}(h, n)$ exists it is identical to the covariance function $\gamma(h)$ in Eq. (1.23). Hence, if we want to numerically model a stationary infinite AR(1) process, then we have to use a finite AR(1) process $\{\widehat{Z}_n\}$ with $\widehat{\sigma}_1 = \sigma_s$. A more detailed analysis of the properties of the finite AR(1) process is available in [16].

The finite AR(1) process satisfying Eq. (1.19) for $\phi = 1$ and $\widehat{Z}_1 = W_1$ becomes the Gaussian random walk (see Sect. 1.1). From Eq. (1.4) the sample variance is equal to

$$\widehat{\sigma}_n^2 = \gamma(n, n) = n\sigma^2.$$

The relation between the Gaussian random walk and the quasistationary finite AR(1) process can be clarified if in Eq. (1.26) we take $\widehat{\sigma}_1 = \sigma$ corresponding to the choice of the first term for the Gaussian random walk $\widehat{Z}_1 = W_1$

$$\widehat{\sigma}_n^2 = \sigma_s^2 \left(1 - \phi^{2n}\right).$$

Figure 1.2 shows the variation of $\widehat{\sigma}_n^2$ for $\sigma = 1$ and different values of ϕ. For a given ϕ, at the beginning there is a nonstationary transient period before the stationary state of the AR(1) process is reached. As ϕ tends to 1, the transient region is expanded and at the limit it becomes infinite, such that for $\phi = 1$ an entirely nonstationary process is obtained. Hence the Gaussian random walk corresponds to the transient region of the finite AR(1) process extended to the infinity, whereas the stationary infinite AR(1) process corresponds to the stationary part of the graph. Therefore, to obtain a quasistationary finite AR(1) process for ϕ close to 1, the only possibility is to choose $\widehat{\sigma}_1 = \sigma_s$ completely eliminating in this way the transient region.

References

1. Alexandrov, T., Bianconcini, S., Dagum, E.B., Maass, P., McElroy, T. S.: A review of some modern approaches to the problem of trend extraction. Research Report Series US Census Bureau, Statistics 3. http://www.census.gov/srd/papers/pdf/rrs2008-03.pdf (2008)
2. Blender, R.: Renormalisation group analysis of autoregressive processes and fractional noise. Phys. Rev. E **64**, 067101 (2001)
3. Brockwell, P.J., Davies, R.A.: Introduction to Time Series and Forecasting. Springer, New York (1996)
4. Brockwell, P.J., Davies, R.A.: Time Series:Theory and Methods. Springer, New York (1996)
5. Gao, J., Hu, J., Tung, W., Cao, Y., Sarshar, N., Roychowdhury, V.P.: Assessment of long range correlation in time series: How to avoid pitfalls. Phys. Rev. E **73**, 016117 (2006)
6. Guzman-Vargas, L., Angulo-Brown, F.: Simple model of the aging effect in heart interbeat time series. Phys. Rev. E **67**, 052901 (2003)
7. Hallerberg, S., Altmann, E.G., Holstein, D., Kantz, H.: Precursors of extreme increments. Phys. Rev. E **75**, 016706 (2007)
8. Király, A., Jánosi, I.M.: Stochastic modeling of daily temperature fluctuations. Phys. Rev. E **65**, 051102 (2002)
9. Kugiumtzis, D.: Statically transformed autoregressive process and surrogate data test for nonlinearity. Phys. Rev. E **66**, 025201 (2002)
10. Liley, D.T., Cadusch, P.J., Gray, M., Nathan, P.J.: Drug-induced modification of the system properties associated with spontaneous human electroencephalographic activity. Phys. Rev. E **68**, 051906 (2003)
11. Maraun, D., Rust, H.W., Timmer, J.: Tempting long-memory—on the interpretation of DFA results. Nonlinear Proc. Geoph. **11**, 495–503 (2004)

12. Palus, M., Novotna, D.: Sunspot cycle: a driven nonlinear oscillator? Phys. Rev. Lett. **83**, 3406–3409 (1999)
13. Stephen, D., Pollock, G.: Statistical signal extraction and filtering: a partial survey. In: Belsley, D.A., Kontoghiorghes, E. (eds.) Handbook of Computational Econometrics, pp. 321–376. Wiley, New York (2009).
14. Timmer, J., Schwarz, U., Voss, H.U., Wardinski, I., Belloni, T., Hasinger, G., van der Klis, M., Kurths, J.: Linear and nonlinear time series analysis of the black hole candidate Cygnus X-1. Phys. Rev. E **61**, 1342–1352 (2000)
15. Vamoş, C.: Automatic algorithm for monotone trend removal. Phys. Rev. E **75**, 036705 (2007)
16. Vamoş, C., Şoltuz, Ş.M., Crăciun, M.: Order 1 autoregressive process of finite length. Rev. Anal. Numer. Theor. **36**, 199–214 (2007)
17. Wu, Z., Huang, N.E., Long, S.R., Peng, C.K.: On the trend, detrending, and variability of nonlinear and nonstationary time series. PNAS **18**, 14889–14894 (2007)
18. Wentzell, A.D.: A Course in the Theory of Stochastic Processes. McGraw-Hill, New York (1981)

Chapter 2
Monte Carlo Experiments

In this chapter we design a numerical algorithm to generate nonmonotonic trends with a diversity of shapes comparable to those encountered in practice. This original algorithm is essential for all the rest of the book because it provides the numerical trends on which the estimation methods are tested. Over these trends finite AR(1) noises (see Sect. 1.3) are superposed so that the resulting artificial time series depend on five independent parameters. In the case of the trend estimation algorithms the complexity of the problem is reduced because the accuracy of the estimated trend significantly depends only on three parameters: the time series length, the noise serial correlation, and the ratio between the amplitudes of the trend variations and noise fluctuations. Using Monte Carlo experiments we derive the accuracy of a simple method to estimate the serial correlation of an AR(1) noise.

2.1 Monte Carlo Statistical Ensembles

The Monte Carlo method does not have a rigorous and exhaustive definition. Even in the original paper where it was firstly presented as a general numerical method, instead of a definition, Metropolis and Ulam gave only a few examples [9]. In terms of statistical physics, a Monte Carlo experiment provides an approximation of the statistical ensemble associated to the phenomenon under investigation. For example, the *statistical ensemble* associated to a macroscopic state of a thermodynamic system is the set of all microscopic states which are compatible with the considered macroscopic state and form a multidimensional volume in phase space. By Monte Carlo simulations one intends to obtain a finite number of microscopic states randomly distributed in phase space which describe the volume associated to the statistical ensemble as accurately as possible.

From the point of view of the theoretical statistics, the Monte Carlo methods are computational algorithms that perform repeated random sampling. Generally by sampling we mean the selection of individual observations of a sample intended to

C. Vamoş and M. Crăciun, *Automatic Trend Estimation*, SpringerBriefs in Physics,
DOI: 10.1007/978-94-007-4825-5_2, © The Author(s) 2012