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

yield some knowledge about a statistical population. In the case of the thermodynamic system discussed above the population from which the selection is made is the statistical ensemble and the individuals are the microscopic states. Usually each individual is chosen randomly such that each individual has the same probability of being chosen at any stage during the sampling process and each subset of k individuals has the same probability of being chosen for the sample as any other subset of k individuals. This technique is known as simple random sampling and it is only one of the probability sampling schemes.

First a *Monte Carlo experiment* defines a statistical ensemble (population in statistical language). Then a numerical method is used to generate several microscopic states (individual representatives) from the statistical ensemble (population). Finally the results of the individual computations are aggregated into the final result by a statistical analysis procedure. Usually the Monte Carlo statistical ensemble is defined not theoretically, but by an effective numerical generation algorithm of the microscopic states. By repeating the generation algorithm several times, we obtain a sample on which we do statistical analysis.

As an example, we evaluate the accuracy of a simple method to estimate the correlation parameter ϕ of a finite AR(1) noise $\{z_n\}$. From Eqs. (1.27) and (1.9) we obtain for the sample autocorrelation function the expression $\widehat{\rho}(h) = \phi^h$. It follows that for $h = 1$ the sample autocorrelation function is a statistical estimator $\widehat{\phi}_\rho = \widehat{\rho}(1)$ for the correlation parameter of a finite AR(1) noise. Using a Monte Carlo experiment we can determine the properties of this estimator without deriving them by a complicated theoretical analysis of its asymptotic properties.

We consider a statistical ensemble composed of S finite AR(1) time series $\{z_n^{(s)}\}$, $1 \leq s \leq S$, numerically generated by means of the algorithm described in Sect. 1.3. Figure 2.1a shows the histogram of the values obtained for the estimated values $\widehat{\phi}_\rho^{(s)}$ using $S = 2000$ time series with length $N = 10000$ and the correlation parameter $\phi = 0.7$. The estimated values are symmetrically distributed around the real value and for a given $\alpha < 1$ we characterize their spread by the interquantile interval $I_\alpha(\widehat{\phi}_\rho) = (Q_{\alpha/2}(\widehat{\phi}_\rho), Q_{1-\alpha/2}(\widehat{\phi}_\rho))$. The probability that $\widehat{\phi}_\rho^{(s)} \in I_\alpha(\widehat{\phi}_\rho)$ is equal with $1 - \alpha$.

In Fig. 2.1b we present for different values of ϕ and N the limits of the interquantile intervals with $\alpha = 0.05$ from which the real value of ϕ is subtracted. These quantities characterize the *sampling errors* of the estimation method. The errors decrease when ϕ increases, i.e., the higher the serial correlation, the more accurate its estimation is. It can be numerically checked that the standard deviation of the errors is inversely proportional to $\sqrt{N}$, i.e., the longer the series, the smaller the errors are. For short time series ($N = 100$) and large values of ϕ the interquantile interval becomes asymmetric, its upper limit being closer to the real value of ϕ. This behavior occurs because always $\widehat{\phi}_\rho < 1$ and when ϕ is close to 1, the maximum upper error is limited by the value $1 - \phi$. For long time series ($N \geq 1000$) the sampling errors are less than 0.1 for all values of ϕ and for short times series ($N = 100$) it is two times larger.

In applications we do not know the value of the parameter that we estimate and we cannot use the sampling error to characterize the accuracy of the estimation. In this

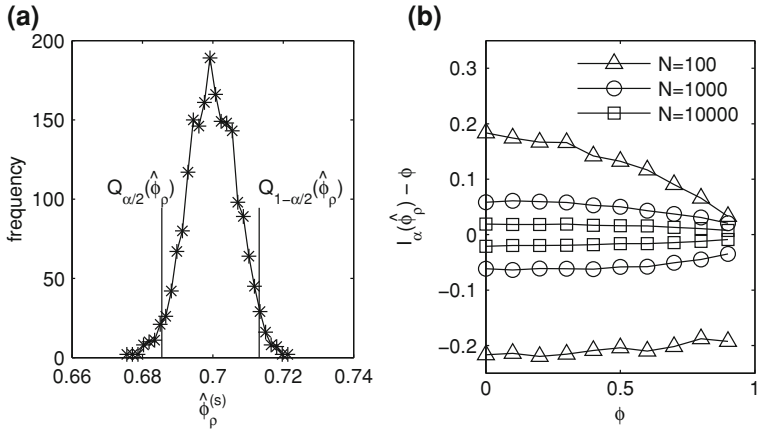

Fig. 2.1 **a** The histogram of the estimated values $\widehat{\phi}_\rho^{(s)}$ for finite AR(1) noises with $\phi = 0.7$ and $N = 10000$ and the interquantile interval for $\alpha = 0.05$; **b** the limits of the interquantile intervals with $\alpha = 0.05$ from which the real value of ϕ is subtracted

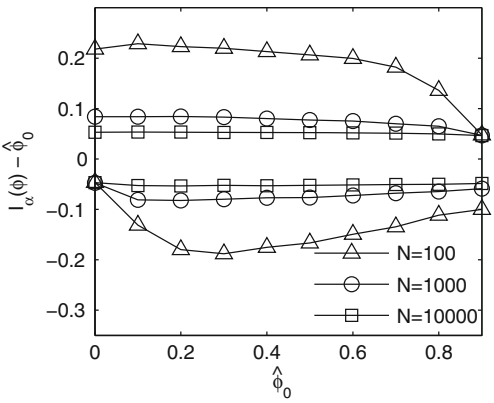

Fig. 2.2 The limits of the 95 % confidence intervals of the estimation $\widehat{\phi}_\rho$ from which the center $\widehat{\phi}_0$ of the variation range of the real values $\phi^{(s)}$ is subtracted

case we use the spreading of the possible values of ϕ for which we obtain a certain estimated value $\widehat{\phi}_\rho$. We generate a statistical ensemble of finite AR(1) time series with the values of real $\phi^{(s)}$, $1 \le s \le S$, uniformly distributed in a given interval $[\phi_{\min}, \phi_{\max}]$. For each time series we compute the estimated value $\widehat{\phi}_\rho^{(s)}$ and then we consider the time series for which $\widehat{\phi}_\rho^{(s)}$ belongs to an interval $[\widehat{\phi}_0 - \Delta\widehat{\phi}, \widehat{\phi}_0 + \Delta\widehat{\phi}]$ with given $\widehat{\phi}_0$ and $\Delta\widehat{\phi}$. The interquantile interval $I_\alpha(\phi)$ of the distribution of $\phi^{(s)}$ is called *confidence interval* and the number $1 - \alpha$ (or as a percentage $100\,\%(1 - \alpha)$) is the *confidence level*.

The 95 % confidence interval of the same estimator as that in Fig. 2.1 is presented in Fig. 2.2. The statistical ensemble contains 100000 finite AR(1) time series with $\phi^{(s)} \in [-0.05, 0.95]$. The estimated values $\widehat{\phi}_\rho^{(s)}$ are distributed in bins with the centers at $\widehat{\phi}_0 = 0, 0.1, \ldots, 0.9$ and with $\Delta\widehat{\phi} = 0.05$. In comparison with Fig. 2.1b

the independent variable is not the real value ϕ, but $\widehat{\phi}_0$, the center of the intervals containing the estimated values $\widehat{\phi}_\rho$. For long time series ($N = 10000$) the confidence intervals coincide with the bins $[\widehat{\phi}_0 - 0.05, \widehat{\phi}_0 + 0.05]$, hence if $\Delta\phi$ would be decreased, then the confidence intervals would become smaller. When $N = 1000$, the confidence intervals are slightly larger than the bins. For $N = 100$ the confidence intervals double and become slightly asymmetric because the values of $\phi^{(s)}$ are limited by $\phi_{\min} = -0.05$ and $\phi_{\max} = 0.95$.

2.2 Numerical Generation of Trends

Usually the trend estimation methods are evaluated by Monte Carlo experiments only for time series with monotonic trends (linear, power–law, exponential, and logarithmic) or periodic ones (sinusoidal). Such trends have a fixed functional form depending on several parameters. In the following chapters we test the trend estimation methods on numerically generated time series and we need artificial trends with shapes having a variability comparable with the trends of the real time series.

The numerical method to generate artificial trends should ensure effective control over their main features, as for example, the number and length of the monotonic segments of the artificial trend. Another important property of a trend is the amplitude of its variations over the monotonic segments, i.e., the difference between successive local extrema. The statistical ensemble of the generated trends should be homogeneous, the average distribution of the trends features should not vary with the location. If we need a nonhomogeneous ensemble of trends, then the trends can be easily rescaled to meet such a requirement.

The apparent choice of a complex nonmonotonic trend with significant variability is a high order polynomial trend with a large number of coefficients. If we choose the coefficients by means of a random algorithm, then the form of the generated trend is difficult to be controlled. Usually the generated trend has only a few parts with significant monotonic variation.

The number and length of monotonic segments of a polynomial trend can be manipulated by means of the position of its roots. Let us consider that we want to generate a polynomial trend with P monotonic parts. We select P distinct roots a_p, $p = 1, 2, \ldots, P$, with a uniform distribution within the definition interval $t \in (0, 1)$. The trend given by the polynomial

$$f(t) = \prod_{p=1}^{P} (t - a_p)$$

has $P - 1$ distinct local extrema. We have to choose the roots a_p not too close to each other. Suppose that we want to generate a time series with N values. Then we exclude the possibility that the interval between two successive roots has the length comparable to the time step $\Delta t = 1/N$, otherwise it would contain only a few

time steps. Therefore we impose the condition that each monotonic segment should contain more than a given number $\Delta N_{\min}$ of time steps.

In the following we describe a numerical algorithm that divides the interval (0, 1) into P subintervals satisfying the conditions previously imposed to the roots of the polynomial trend. The length of the subinterval p is chosen as a random number d_p uniformly distributed within the interval $[d_{\min}, 1]$, where $d_{\min}$ is a parameter that will be determined later. The union of all these subintervals is an interval with length $d = \sum_{p=1}^{P} d_p$ and each subinterval is divided by d such that the entire interval is rescaled to (0, 1). The interval (0, 1) is divided into $N - 1$ equal bins corresponding to the N values of the time series. Hence the superior limit of the subinterval p is equal to

$$N_p = 1 + \left[(N-1) d^{-1} \sum_{i=1}^{p} d_i \right], \tag{2.1}$$

where $[\cdot]$ is the integer part function. The subinterval p contains the time steps satisfying the condition $N_{p-1} < n \leq N_p$, where $N_0 = 0$. The number of time steps of any subinterval is equal with $\Delta N_p = N_p - N_{p-1}$.

We have to set the value of the parameter $d_{\min}$. If in Eq. (2.1) we do not take into account the integer part function, then the number of time steps of the segment p is approximately equal to $\Delta N_p \simeq d_p (N-1)/d$. The minimum of this quantity is obtained when $d_p = d_{\min}$ and $d_{p'} = 1$ for $p' \neq p$ and is equal to

$$\min(\Delta N_p) = \frac{d_{\min}(N-1)}{P - 1 + d_{\min}}.$$

Imposing the condition that $\min(\Delta N_p)$ equals the given minimum length of the monotonic segments $\Delta N_{\min}$, it follows

$$d_{\min} = \frac{(P-1)\Delta N_{\min}}{N - 1 - \Delta N_{\min}}.$$

Figure 2.3 contains some polynomial trends with the roots generated by the algorithm described above for $P = 10$ monotonic parts, $N = 1000$ values and $\Delta N_{\min} = 10$. The nonuniform distribution of the trend values is obvious. The variability of the trends near the boundaries is with almost two orders of magnitude greater than in the middle of the definition interval. The same nonuniform distribution is obtained if we choose the position of the local extrema of the polynomial trend, not of its roots. In general, all algorithms to generate polynomial trends suffer from this drawback which originates in the so-called "Runge phenomenon" [4].

It is possible to correct this problem if we introduce a mechanism to control the amplitude of monotonic variations. Such a solution is based on constructing monotonic semi-periods of sinusoid with random amplitude [11]. They are joined together such that the trend is continuous. The amplitude of the sinusoid associated

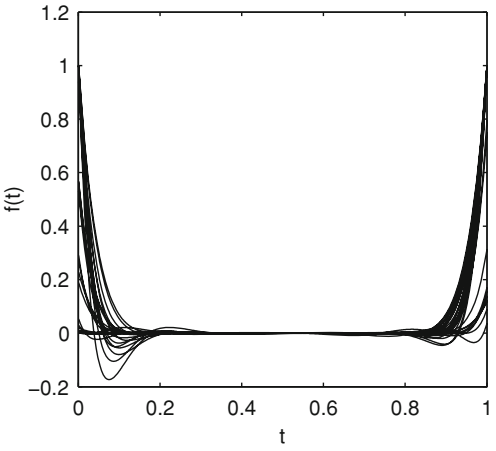

Fig. 2.3 Numerically generated polynomial trends with the roots uniformly distributed in the definition interval

with the subinterval p is a random number $A_p \in [0, 1]$ with a uniform probability distribution.

The sinusoid semi-periods are located in the definition interval [0, 1] by means of the numbers N_p generated by the algorithm previously used to set the polynomial roots. The value of the trend at a point n of the segment p, $N_p < n \leq N_{p+1}$, is given by the recurrence relation

$$f_n = f_{N_p} + c_p A_p \left[1 - \sin \frac{\pi}{2} \left(1 + 2\frac{n - N_p}{\Delta N_p}\right)\right] \tag{2.2}$$

where the term f_{N_p} is the last value of the previous segment $p - 1$ and for $p = 1$ we choose $f_0 = 0$. The coefficients c_p describe how the successive sinusoids are connected together. When $c_p = (-1)^p$, the first sinusoidal part is decreasing and the monotony of the other parts alternates such that the trend looks like a distorted sinusoid. Examples of such trends are given in Figs. 2.5, 3.2, and 4.5. If the values $c_p = \pm 1$ are randomly assigned, then the trend shape variability increases because it is possible to have several successive sinusoids with variation of the same sign (see Figs. 2.6, 6.2, 6.4, 7.5). Finally the mean of the series $\{f_n\}$ is removed from it.

The algorithm described above is characterized by three parameters: the length of the time series N, the number of monotonic segments P, and the minimum number of points in a monotonic segment $\Delta N_{\min}$. Figure 2.4 shows some numerically generated trends with $P = 10$, $N = 1000$, $\Delta N_{\min} = 10$ and the coefficient c_p alternately chosen. Unlike polynomial trends in Fig. 2.3, these trends have a large variability homogeneously distributed throughout the entire definition interval.

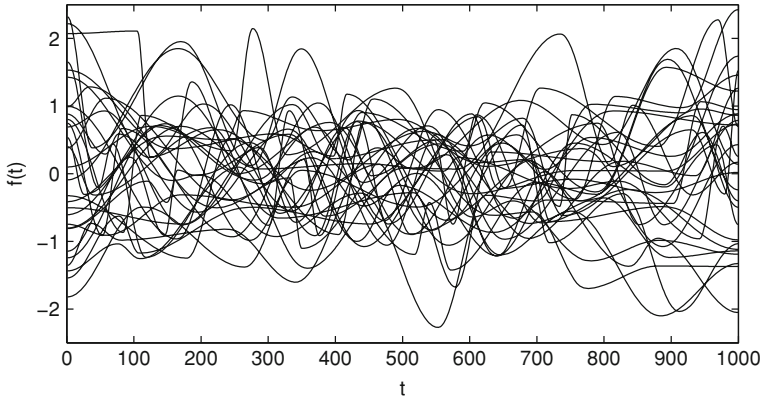

Fig. 2.4 Numerically generated trends as an alternate succession of random sinusoidal semi-periods

2.3 Numerical Generation of Noisy Time Series

In this section we present a numerical algorithm to generate a time series $\{x_n\}$ by superposing an AR(1) noise $\{z_n\}$ over a nonmonotonic trend $\{f_n\}$ generated by the algorithm described in the previous section. This algorithm is almost identical to that presented in [11] with the only difference that here we rescale the noise, not the trend. As discussed above, the trend $\{f_n\}$ is characterized by three parameters: N, $\Delta N_{\min}$, and P. The AR(1) noise is obtained by means of the numerical method described in Sect. 1.3 and has two parameters: ϕ for the serial correlation and σ_s^2 for its variance. We need an additional parameter r to characterize the ratio between the amplitudes of the trend variations and of the noise fluctuations defined as

$$r = \frac{\max\{f_n\} - \min\{f_n\}}{\max\{z_n\} - \min\{z_n\}}, \tag{2.3}$$

which is a more sensitive measurement of the classical signal to noise ratio. Because the noise variance depends on the new parameter, the noisy time series $\{x_n\}$ has five independent parameters: N, $\Delta N_{\min}$, P, ϕ, and r.

To obtain a time series characterized by these five parameters we generate a trend $\{f_n\}$ characterized by N, $\Delta N_{\min}$, and P and an AR(1) noise $\{z'_n\}$ with $\sigma_s = 1$ and a given ϕ. Then we rescale the noise according to the relation

$$z_n = \frac{z'_n}{r} \frac{\max\{f_n\} - \min\{f_n\}}{\max\{z'_n\} - \min\{z'_n\}}.$$

When $r > 1$ the time series $x_n = f_n + z_n$ is dominated by trend and when $r < 1$ by noise. Depending on the aim of the numerical experiment we generate time series

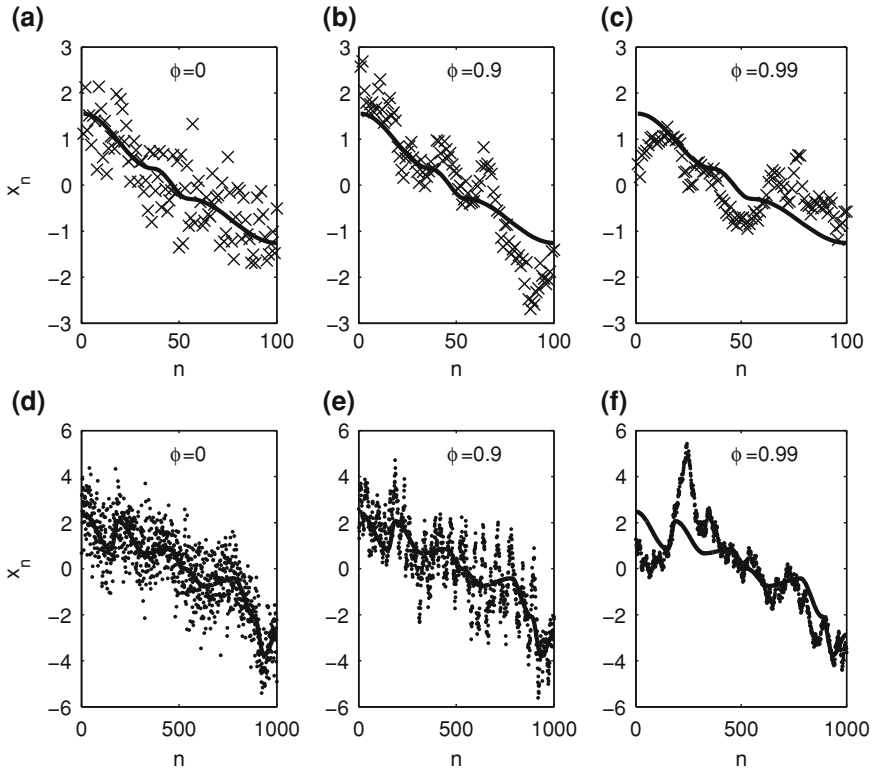

Fig. 2.5 Numerically generated time series for different values of N and ϕ with $r = 1$ and $\Delta N_{\min} = 10$

choosing different values for the five parameters. In the following we give some indication concerning the typical values of these parameters.

In order to obtain a description of each monotonic segment with an acceptable resolution we choose $\Delta N_{\min} > 10$, i.e., an order of magnitude greater than the time step. The value of $\Delta N_{\min}$ is related to the length of the time series and the number of monotonic parts through the relation $\Delta N_{\min} < N/P$. This condition also limits the values of N and P because for a given length N there is a maximum number of monotonic parts given by $P < N/\Delta N_{\min}$.

We impose the superior limit of ϕ to 0.9 because the AR(1) process with ϕ closer to unit has a special behavior similar to the Gaussian random walk (see Sects. 1.1 and 1.3), which must be analyzed with special methods [5]. In addition we consider for ϕ only positive values because few of the phenomena of interest are characterized by an anticorrelated noise. Hence the maximum range for the serial correlation parameter is $\phi \in [0, 0.9]$. When we choose the value of ϕ we have to take also into account the length of the time series. In Fig. 2.5 we present several noisy time series with $N = 100$ and $N = 1000$, different values of ϕ, and $r = 1$. For $\phi = 0.9$ and $N = 100$

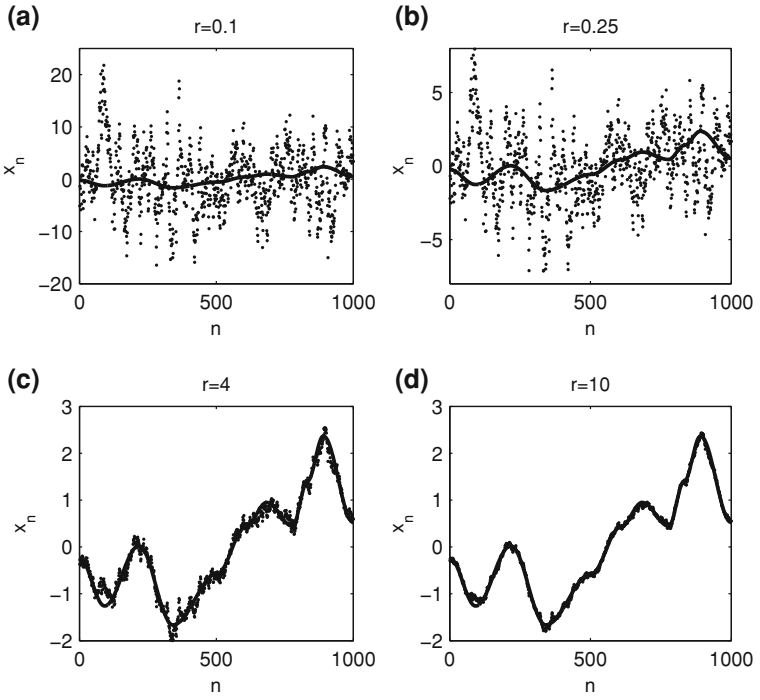

Fig. 2.6 Time series generated by superposing an AR(1) noise with $\phi = 0.9$ over the same trend but with different ratios r

a stochastic trend appears (Fig. 2.5b), i.e., the strongly correlated noise has variations that cannot be distinguished from those of the trend (see Sect. 1.1). For $N = 1000$ the time series with the same value $\phi = 0.9$ contains many noise fluctuations and the noise does not confound with the trend (Fig. 2.5e). But if we increase the serial correlation to $\phi = 0.99$, the noise variations are enhanced and the stochastic trend occurs even for $N = 1000$ (Fig. 2.5f).

Since for $r = 0$ the trend vanishes, we eliminate the small values of r. We choose the minimum value for r equal to 0.25 because, as shown in Fig. 2.6b, in this situation the shape of the total signal allows us to assume the presence of a trend even for strong serial correlation ($\phi = 0.9$). For smaller values of r the shape of trend is not observable (Fig. 2.6a). The maximum value of r that we use in our numerical simulations is 4 and corresponds to the time series in which the noise still has a large enough amplitude to allow the estimation of its parameters (Fig. 2.6c). For higher values of r the noise becomes negligible as in Fig. 2.6d. Hence the interval chosen for the variation of r is $r \in [0.25, 4]$.

Besides the previous criteria, we have chosen the variation ranges of the parameters of the artificial time series such that they also include the extreme situations for which the numerical algorithms fail. The most difficult problems occur for the short

time series and, in general, we have limited their length to $N = 100$ values. If such a short time series is characterized by the inferior limit of the variation range of the ratio $r = r_{\min} = 0.25$, i.e., it is strongly dominated by noise, then the trend estimated by any algorithm has large errors. Reversely, if the short time series is strongly dominated by trend ($r = r_{\max} = 4$), then the noise parameters are estimated with very large errors. If, in addition, the noise is strongly correlated ($\phi = \phi_{\min} = 0.9$), then the stochastic trend prevents any possibility to correctly separate the trend from the superposed noise.

For most of the numerical simulations in the next chapters we limit ourselves to AR(1) processes. Obviously, it is possible to choose other types of noises but it would mean an increase of the number of parameters describing the statistical ensemble. In order to preserve the presentation clarity we use noise models as simple as possible, retaining in the same time their essential characteristic, namely, the serial correlation. As shown in Fig. 2.5, the parameter ϕ controls the occurrence of the stochastic trend which is a major difficulty in an accurate trend estimation. The analysis of the noise models with more parameters is similar to that of the AR(1) noise, although more elaborate.

As a simple application of the algorithm presented above we analyze a method to estimate the parameter ϕ of a finite AR(1) noise. Unlike the method discussed in Sect. 2.1, this method can be applied to time series containing a trend. It allows a first guess of the serial correlation parameter without effectively detrending the time series. In some cases this estimation is not very accurate, but it can indicate to which category the time series belongs, for example, if it is strongly or weakly correlated. Such an approximate classification allows an adjustment of a more complex numerical algorithm to the characteristics of a particular time series.

If the approximation (1.18) is valid, then we can obtain information on the noise properties analyzing the differenced stochastic process $\{\nabla_d X_n\}$. From Eqs. (1.16) and (1.1), for a stationary noise with zero mean we have

$$\langle(\nabla_d Z_n)^2\rangle = \langle Z_{n+d}^2\rangle - 2\langle Z_{n+d} Z_n\rangle + \langle Z_n^2\rangle = 2\langle Z_n^2\rangle - 2\gamma(d).$$

Substituting Eq. (1.23) in the previous relation, we obtain for an AR(1) process the relation

$$\langle(\nabla_d Z_n)^2\rangle = 2\sigma_s^2(1 - \phi^d).$$

Because $\{Z_n\}$ is stationary, this relation holds for all indexes n. Applying it for $d = 1$ and $d = 2$ it results a system of equations from which the parameter ϕ can be computed

$$\phi = \frac{\langle(\nabla_2 Z_n)^2\rangle}{\langle(\nabla_1 Z_n)^2\rangle} - 1. \tag{2.4}$$

Since $\{Z_n\}$ is a stationary process, the ensemble average of one of its terms $\langle Z_n\rangle$ can be estimated by the temporal mean (1.7) of one of its realization $\{z_n\}$. For any n, the term $\langle(\nabla_d Z_n)^2\rangle$ is replaced by $(N - d)^{-1} \|\nabla_d z_n\|^2$ where

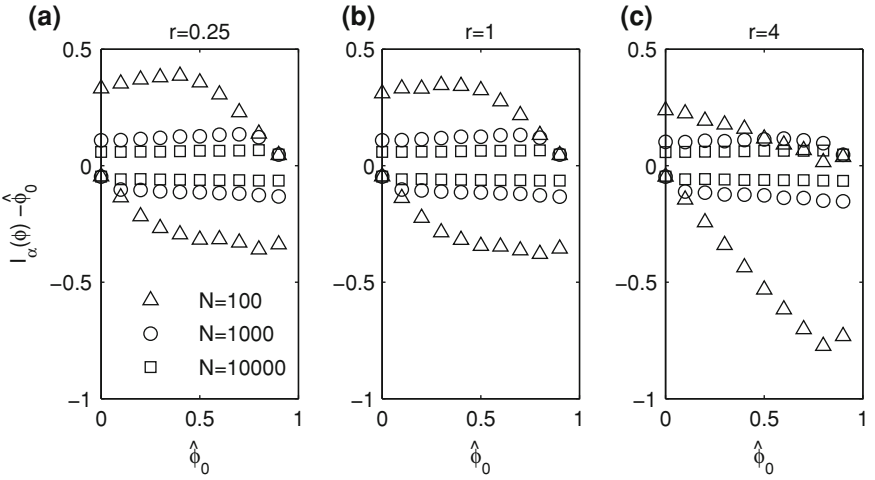

Fig. 2.7 The limits of the 95 % confidence intervals of the estimation $\widehat{\phi}_x$ from which the center $\widehat{\phi}_0$ of the variation range of the real values $\phi^{(s)}$ is subtracted

$$\|\nabla_d z_n\|^2 = \sum_{n=1}^{N-d} (\nabla_d z_n)^2 \tag{2.5}$$

is the usual quadratic norm of finite sequences. We use the approximation (1.18)

$$\nabla_d x_n \approx \nabla_d z_n,$$

written for a realization of the stochastic process $\{X_n\}$ which for small d is sufficiently accurate. Then the estimation (2.4) of ϕ using only the values of the time series $\{x_n\}$ is

$$\widehat{\phi}_x = \frac{N-1}{N-2} \frac{\|\nabla_2 x_n\|^2}{\|\nabla_1 x_n\|^2} - 1. \tag{2.6}$$

From this equation it is possible to result $\widehat{\phi}_x > 1$ corresponding to an acausal AR(1) noise (see Sect. 1.3). Such situations generally occur when ϕ is near 1 and then we correct the estimation of the serial correlation parameter by the maximum value of the parameter used to generate the time series $\widehat{\phi}_x = \phi_{\max} = 0.95$. Another correction is made when from the estimation we obtain $\widehat{\phi}_x < 0$, then we make the substitution $\widehat{\phi}_x = 0$.

In order to determine the confidence intervals of the estimation (2.6) we use Monte Carlo experiments. The results presented in Fig. 2.7 are obtained for the same AR(1) noises as those in Fig. 2.2, but superposed over numerically generated trends. Whatever the value of the ratio r, for long time series ($N \geq 1000$), the accuracy of the estimation $\widehat{\phi}_x$ is almost the same, showing that for such time series differencing allows the removal of the trend influence. The confidence intervals are symmetric with

respect to estimated values and approximately twice larger than those determined for the estimation $\widehat{\phi}_\rho$ (see Fig. 2.2). Hence the presence of the trend worsens the estimation quality, but the errors remain smaller than ± 0.1. The correction $\widehat{\phi}_x = 0.95$ when $\widehat{\phi}_x > 1$ causes the decrease of the superior limit of the confidence interval for $\widehat{\phi}_0 > 0.8$.

For the short time series ($N = 100$) the confidence intervals significantly increase and become asymmetric due to the influence of the corrections made at the boundaries of the variation range of $\phi^{(s)}$. For the time series dominated by trend ($r = 4$) the errors become unacceptably large. As discussed before in this section, this behavior indicates that for an efficient numerical processing, the nonstationary time series must have at least several hundreds of values.

2.4 Statistical Hypothesis Testing

Before processing a time series it is important to estimate the probability that it contains a significant deterministic trend on the basis of a statistical hypothesis test. First, we formulate the null hypothesis, in this case that there is no deterministic trend in the considered time series. Then we define a statistical quantity T, named statistic, which has the ability to detect the existence of a deterministic trend. We choose the value α of the *significance level* equal with the probability of incorrectly rejecting the null hypothesis. If we know the pdf of T, then we can compute the *critical value* t_α which delimits the values of T for which one can consider that the null hypothesis is true. The observed value t of T is compared with t_α and we can decide to reject or not the null hypothesis.

In the following we describe the Mann–Kendall test designed to detect a monotonic trend in a time series $\{x_n\}$. It is used especially in hydrology, other environmental sciences, and econometrics for short time series, even containing only several dozens of values [7]. It is based on the quantity $S_{MK} = P_{MK} - M_{MK}$, where P_{MK} is the number of the pairs $x_n > x_m$ for $n > m$ and M_{MK} is the number of the pairs $x_n < x_m$ for $n > m$. If $\{x_n\}$ are independent observations, then for $N > 10$ the random variable

$$T_{MK} = \begin{cases} (S_{MK} - 1)/\sigma_{MK} & \text{if} \quad S_{MK} > 0 \\ 0 & \text{if} \quad S_{MK} = 0 \\ (S_{MK} + 1)/\sigma_{MK} & \text{if} \quad S_{MK} < 0 \end{cases}$$

with $\sigma_{MK}^2 = N(N-1)(2N+5)/18$ follows a standard normal distribution. The null hypothesis that there is no monotonic trend is rejected when the computed value t_{MK} of T_{MK} is greater in absolute value than the critical value $t_{\alpha/2}$, where α is the chosen significance level. Hence, if $-t_{\alpha/2} < t_{MK} < t_{\alpha/2}$, then the probability that the time series $\{x_n\}$ does contain a monotonic trend is equal with α.

In general, the derivation of the probability distribution of a statistic T is difficult and usually it can be made only asymptotically, for infinite time series. Using Monte

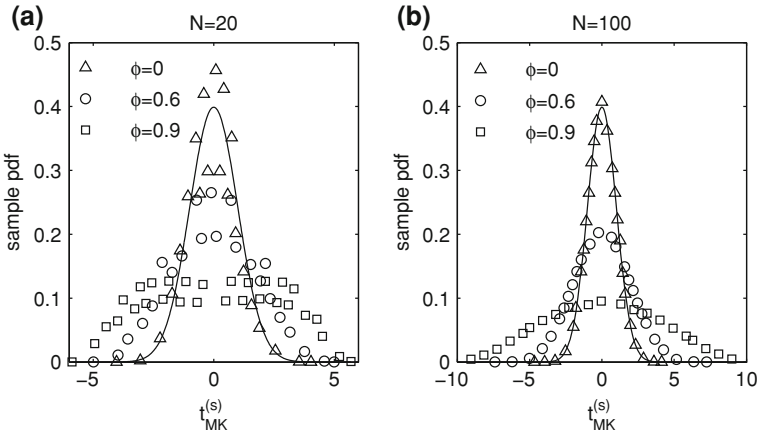

Fig. 2.8 The sample pdfs of the values $t_{MK}^{(s)}$ of the Mann–Kendall statistic for finite AR(1) noises without trend compared with the theoretical pdf for Gaussian white noise (*continuous line*)

Carlo experiments one can determine the sample pdf for any length of the time series and for any value of its parameters. For example in Fig. 2.8 we present the sample pdfs of $t_{MK}^{(s)}$ for $S = 100000$ finite AR(1) noises with $\sigma_s = 1$ and different values for N and ϕ. For $N = 100$ and $\phi = 0$ the sample pdf is indistinguishable from the theoretical pdf (Fig. 2.8b). As ϕ increases, the sample pdf is moving away from the theoretical pdf for uncorrelated noise. The pdf for correlated noise can be derived theoretically [6], but here we limit ourselves to the simplest case.

The asymptotic pdf becomes inadequate when the time series length decreases. For short uncorrelated time series ($N = 20$ and $\phi = 0$) the asymptotic pdf is correct only for $t_{MK} > 1$ (Fig. 2.8a). In the neighborhood of $t_{MK} = 0$ a significant fluctuation of the sample pdf occurs about the asymptotic pdf. The amplitude of these oscillations decreases when ϕ increases, but the oscillations extend at higher values of t_{MK}. Such a behavior is very difficult to be deduced theoretically, proving the usefulness of Monte Carlo experiments in hypothesis testing.

For a given significance level α we can determine the critical value $t_{\alpha/2}$ from the sample pdfs obtained by Monte Carlo experiments. Figure 2.9 shows $t_{\alpha/2}$ for $\alpha = 0.05$ and several values of N and ϕ. The dashed line represents $t_{\alpha/2}$ obtained from the asymptotic pdf of the uncorrelated noise and it coincides with that numerically determined for $\phi = 0$. When the serial correlation increases, the difference between the real and the asymptotic critical values becomes larger. For long time series we obtain a higher critical value, because for large N the probability distribution has a support that stretches to higher values of $t_{MK}^{(s)}$ (see Fig. 2.8).

Sometimes it is also important to analyze the possibility that the null hypothesis is not true and we erroneously decide that it is true. Therefore we have to determine the probability to accept the null hypothesis when it is false. The alternative hypothesis of the Mann–Kendall test is not clearly defined. If the null hypothesis (signal without

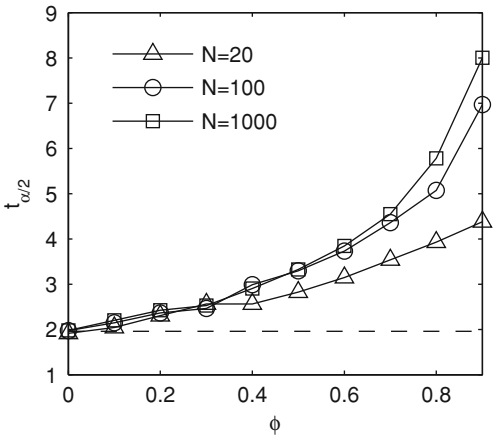

Fig. 2.9 The critical value $t_{\alpha/2}$ for a significance level $\alpha = 0.05$ of the Mann–Kendall test for AR(1) noises without trend in terms of the correlation parameter ϕ.The *dotted line* represents $t_{\alpha/2}$ obtained from the asymptotic pdf for Gaussian white noise

monotonic trend) is not true, then the signal contains a monotonic trend, but we have to mention what kind of trend from the infinity of possible functional forms is assumed. In this book we do not use this type of statistical test.

We are interested in statistical tests for the i.i.d. time series defined in Sect. 1.1. There are tests that check if the correlations between the terms of time series are negligible by means of the sample autocorrelation function (1.9). They are based on the fact that for i.i.d. time series the values of the sample autocorrelation function asymptotically form a normal i.i.d time series with mean zero and variance N^{-1} ([3], p. 223). This property is obtained as a special case of Bartlett's theorem that is valid for an arbitrary ARMA process ([3], p. 221). It is possible that, as in the case of financial time series (see Sect. 4.3), the time series is uncorrelated, but the amplitudes of its fluctuations are strongly correlated and then it is necessary to check if its absolute values are also i.i.d. It can be shown that in this case Bartlett's theorem remains valid and one can practically apply the same tests as for the initial time series.

In accordance with Bartlett's theorem, $\widehat{\rho}(h)$ has a normal distribution with mean zero and variance N^{-1}. A simple test that the time series is i.i.d. asks that less than 5 % of the values $\widehat{\rho}(h)$ should be outside the bounds $\pm 1.96 N^{-1/2}$. A more complex test is the Box–Pierce test [2] for ARMA residuals which is based on the statistic

$$Q = N \sum_{j=1}^{h} \widehat{\rho}^2(j)$$

that is approximately chi-squared distributed. These tests and their versions impose only the condition that the sample autocorrelation function should not have too large values, but in numerical processing it is possible that the sample autocorrelation function becomes too small. To avoid this possibility we compare the distribution of the sample autocorrelation function with the theoretical normal distribution.

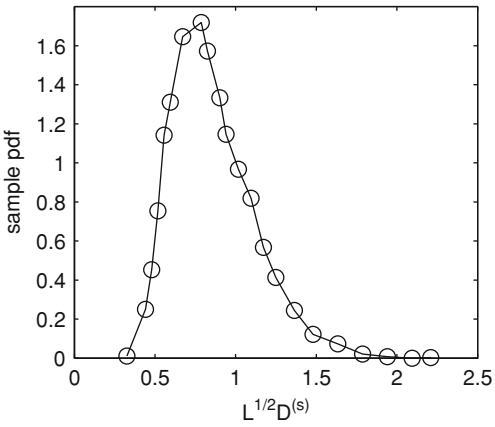

Fig. 2.10 The sample pdf of the Kolmogorov–Smirnov statistic $\sqrt{L}D^{(s)}$ for $S = 1000$ normal distributions

The Kolmogorov–Smirnov (KS) test establishes the resemblance between two distributions ([10], Sect. 14.3). If $\widehat{F}_L(x)$ is an empirical cdf with L values, then the KS statistic with respect to a theoretical cdf $F(x)$ is

$$D = \max_x \left| \widehat{F}_L(x) - F(x) \right| . \tag{2.7}$$

According to the Kolmogorov theorem the asymptotic pdf of this quantity is given by

$$\lim_{L\to\infty} P\{\sqrt{L}D < \alpha\} = \sum_{k=-\infty}^{\infty} (-1)^k e^{-k^2\alpha^2}$$

and it does not depend on the form of the distribution $F(x)$. The sum in this formula can be numerically calculated [8], but it can be also determined by a Monte Carlo experiment. In Fig. 2.10 we present the sample pdf of $\sqrt{L}D^{(s)}$ for $S = 10000$ normal distributions with zero mean, unit variance, and $L = 1000$ values. The mean of the quantity $\sqrt{L}D^{(s)}$ is 0.859 and its standard deviation is 0.259. Hence, even if $\widehat{F}_L(x)$ is obtained from a realization of a random variable with the cdf identical with $F(x)$, D is nonvanishing. Only asymptotically for $L \to \infty$ it becomes zero when $\lim_{L\to\infty} \widehat{F}_L = F$.

According to Bartlett's theorem, the sample autocorrelation function of a Gaussian i.i.d. time series forms a normal i.i.d. series with mean zero and variance N^{-1}. In order to test whether the time series is i.i.d. we calculate the KS statistic (2.7) for the first $N/4$ values of the sample autocorrelation function as recommended in [1]

$$\varepsilon = \max_{1\le h\le N/4} \left| \widehat{F}(\widehat{\rho}(h)) - G(\widehat{\rho}(h)) \right| , \tag{2.8}$$

where $G(x) = \Phi(x)/\sqrt{N}$ and $\Phi(x)$ is the normal cdf. Figure 2.11 shows the average $\langle\varepsilon\rangle$ over statistical ensembles with $S = 100$ for the autocorrelation function of the

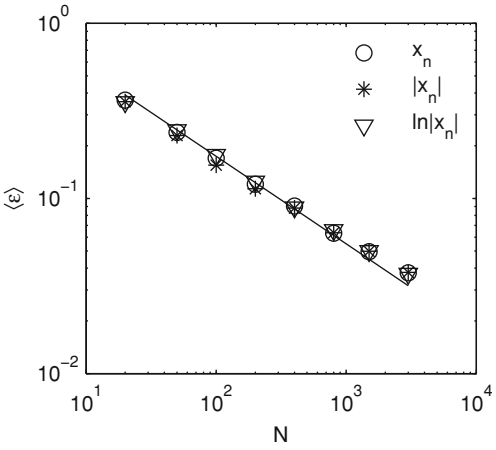

Fig. 2.11 The average of Kolmogorov–Smirnov statistic of the first $N/4$ values of the autocorrelation function of Gaussian i.i.d. white noise (o), of its absolute values (*), and of the logarithm of its absolute values (∇)

Gaussian white noises $\{x_n\}$, for their absolute values $\{|x_n|\}$, and for the logarithm of their absolute values $\{\ln|x_n|\}$. As mentioned before the three quantities are practically identical.

References

1. Box, G., Jenkins, G., Reinsel, G.: Time Series Analysis: Forecasting and Control, 3rd edn. Prentice-Hall, Upper Saddle River (1994)
2. Box, G.E.P., Pierce, D.A.: Distribution of the autocorrelations in autoregressive moving average time series models. J. Am. Stat. Assoc. **65**, 1509–1526 (1970)
3. Brockwell, P.J., Davies, R.A.: Time Series: Theory and Methods, 2nd edn. Springer, New York (1996)
4. Fornberg, B.: A Practical Guide to Pseudospectral Methods. Cambridge University Press, Cambridge (1998)
5. Hamilton, J.D.: Time Series Analysis. Princeton University Press, Princeton (1994)
6. Hirsch, R.M., Slack, J.R.: A nonparametric trend test for seasonal data with serial dependence. Water Resour. Res. **20**, 727 (1984)
7. Kendall, M.G.: Rank Correlation Methods. Griffin, London (1975)
8. Marsaglia, G., Tsang, W., Wang, J.: Evaluating kolmogorov's distribution. J. Stat. Softw. **8**(18), 1–4 (2003)
9. Metropolis, N., Ulam, S.: The Monte Carlo method. J. Am. Stat. Assoc. **44**, 335–341 (1949)
10. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in C. The Art of Scientific Computing, 2nd edn. Cambridge University Press, Cambridge (1992)
11. Vamoş, C., Crăciun, M.: Serial correlation of detrended time series. Phys. Rev. E **78**, 036707 (2008)

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