

ORDER 1 AUTOREGRESSIVE PROCESS OF FINITE LENGTH [‡]

CĂLIN VAMOȘ*, ȘTEFAN M. ȘOLTUZ*[†] and MARIA CRĂCIUN*

Abstract. The stochastic processes of finite length defined by recurrence relations request additional relations specifying the first terms of the process analogously to the initial conditions for the differential equations. As a general rule, in time series theory one analyzes only stochastic processes of infinite length which need no such initial conditions and their properties are less difficult to be determined. In this paper we compare the properties of the order 1 autoregressive processes of finite and infinite length and we prove that the time series length has an important influence mainly if the serial correlation is significant. These different properties can manifest themselves as transient effects produced when a time series is numerically generated. We show that for an order 1 autoregressive process the transient behavior can be avoided if the first term is a Gaussian random variable with standard deviation equal to that of the theoretical infinite process and not to that of the white noise innovation.

MSC 2000. 62M10, 62M15, 60G10.

Keywords. Autoregressive process, spectral analysis, time series.

1. INTRODUCTION

The purpose of the statistical processing of time series is to find a stochastic model which can generate time series having the same statistical properties as the observed series. One of the most common statistical model is the autoregressive moving average (ARMA) process having many applications in economy, finance, signal theory, geophysics, etc. [4]. Lately, especially the applications of the autoregressive (AR) models have become more frequent in physics as well. For example they are used in the analysis of the wind speed fluctuations [7], radar signals [5], climatic phenomena variability [13], electroencephalographic activity [12], heart interbeat time series [6], the daily temperature fluctuations [10], X-ray emission from the active galactic nuclei [17], sunspots variability [15], etc. As a result of their simple mathematical properties and their direct physical interpretation, the realizations of AR(1)

[‡] This work has been supported by MEdC under Grant 2-CEx06-11-96/19.09.2006.

*“T. Popoviciu” Institute of Numerical Analysis, Romanian Academy, P.O. Box 68, 400110 Cluj-Napoca, Romania, e-mail: {cvamos, ssoltuz, craciun@ictp.acad.ro}.

[†]Departamento Matemáticas, Universidad de los Andes, Carrera 1 No. 18A-10, Bogota, Columbia.

processes have been used as artificial series to analyze some numerical algorithms for monotonic trend removal [18], some surrogate data test for nonlinearity [11] or for renormalization group analysis [1].

A stochastic process $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ is called autoregressive process of order p , denoted $\text{AR}(p)$, if $\{X_t\}$ is stationary and for any t we have

$$(1) \quad X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t ,$$

where $\{Z_t\}$ is a Gaussian white noise with zero mean and variance σ^2 , and $\phi_1, \phi_2, \dots, \phi_p$ are real constants. The above relation may be rewritten in the compact form $\phi(B)X_t = Z_t$ where

$$(2) \quad \phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$$

and B is the shift operator, $BX_t = X_{t-1}$. Equation (1) has a unique solution if the polynomial (2) does not vanish for $|z| = 1$ [2]. If in addition $\phi(z) \neq 0$ for all $|z| < 1$, then the process is causal, i.e., there exists a sequence of constants $\{\psi_j\}$ such that $\sum_{j=1}^{\infty} |\psi_j| < \infty$ and

$$(3) \quad X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} , \quad t = 0, \pm 1, \dots$$

Hence, a causal process is characterized by the property that the random variable X_t can be expressed only in terms of noises at previous moments and at the same moment. The properties of $\text{AR}(p)$ processes have been studied in detail and they are the basis of the linear stochastic theory of time series ([2], [3], [8]).

The stochastic process defined above is an idealized mathematical object which has no direct correspondent in practice. Each time series obtained by measurements is a sequence of real numbers x_t and one makes the assumption that each of those numbers is a realization of a random variable X_t . To recover the statistical information regarding the stochastic process we need a large number of time series generated in the same conditions, a situation occurring rarely in practice. Usually we have only one single time series and the recognition of the stochastic process that generated it is very difficult. In addition the finite length (sometimes very short) of the data series render the problem even harder. Therefore it is important to analyze a finite stochastic process satisfying the same recurrence relation as the idealized infinite stochastic process (1).

In the following we analyze the properties of an $\text{AR}(1)$ stochastic process of finite length allowing us to make detailed analytical and numerical computations. We compare the results obtained with the ideal case of the infinite stochastic process and we discuss the conclusions which can be obtained from its power spectrum.

2. AR(1) PROCESS

A stochastic AR(1) process is defined by the simplified form of the relation (1)

$$(4) \quad X_t = \phi X_{t-1} + Z_t .$$

By definition an AR(1) process has an infinite length. In this case the polynomial (2) has the form $\phi(z) = 1 - \phi z$ and then the causality condition reduces to $|\phi| < 1$. Because the relation (4) is so simple, we can analyze more directly the significance of this condition. By successively applying (4) τ times we obtain

$$(5) \quad X_t = Z_t + \phi Z_{t-1} + \dots + \phi^{\tau-1} Z_{t-\tau+1} + \phi^\tau X_{t-\tau} .$$

Because the process $\{X_t\}$ is stationary, the random variables X_t and $X_{t-\tau}$ have the same norm and then the norm of the last term in (5) is ϕ^τ times the norm of the left term. So for any positive number $\varepsilon < 1$, there exists $\tau > \ln \varepsilon / \ln |\phi| > 0$ such that the last term from the right side can be neglected

$$(6) \quad X_t = Z_t + \phi Z_{t-1} + \dots + \phi^{\tau-1} Z_{t-\tau+1} + \mathcal{O}(\varepsilon)$$

for $\varepsilon \rightarrow 0$. This is the definition relation of a causal process (3) where $\psi_j = \phi^j$. From relation (6) one observes that the influence of the noise reduces as the time moves away from t .

In the nonstationary case $|\phi| = 1$, all the terms in (5) have unitary coefficients and for any delay τ none of them can be neglected. Every random variable X_t is an infinite sum of terms with the same norm, hence its norm is infinite.

2.1. Causal AR(1) process. In the following we analyze the basic properties of the causal stationary process AR(1) using the relation (4). If we take the mean of this relation we obtain $E\{X_t\} = 0$. Its square is

$$X_t^2 = \phi^2 X_{t-1}^2 + 2\phi X_{t-1} Z_t + Z_t^2 .$$

In accordance with relation (6) the random variables X_{t-1} and Z_t are independent, that is $E\{X_{t-1} Z_t\} = 0$. If we take the mean of the last relation we obtain a relation between the variance of successive random variables

$$\sigma_t^2 = \phi^2 \sigma_{t-1}^2 + \sigma^2 .$$

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

$$(7) \quad \sigma_s^2 = \frac{\sigma^2}{1-\phi^2} .$$

In order to compute the autocovariance function we multiply (4) with $X_{t-\tau}$ and we take the mean. Because the mean of X_t vanishes and $E\{X_{t-1} Z_t\} = 0$, we obtain

$$\gamma(\tau) \equiv E\{X_t X_{t-\tau}\} = \phi \gamma(\tau - 1) .$$

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

$$(8) \quad \gamma(\tau) = \sigma_s^2 \phi^\tau .$$

The spectral density is the Fourier transform of the autocovariance function

$$(9) \quad f(\nu) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} e^{-2\pi i \tau \nu} \gamma(\tau),$$

where the positive number ν is the frequency. By direct calculation with $\gamma(\tau)$ given by (8) we obtain

$$(10) \quad f(\nu) = \frac{\sigma^2}{2\pi} \frac{1}{1 - 2\phi \cos 2\pi\nu + \phi^2}.$$

Due to the periodicity of this function, we reduce its domain of definition to the interval $\nu \in [0, 0.5]$.

When $\phi \rightarrow 1$, the standard deviation σ_s given by (7) and the covariance function $\gamma(\tau)$ given by (8) become both of them infinite. The spectral density (10) becomes infinite only for $\nu = 0$.

2.2. Acausal AR(1) process. Let us analyze the basic properties for the stationary acausal AR(1) process, that is when $|\phi| > 1$ in (4). The relation (6) is no more true because now ϕ^τ increases when τ increases. If we divide (5) by ϕ^τ , then

$$\phi^{-\tau} X_t = \phi^{-\tau} Z_t + \phi^{-\tau+1} Z_{t-1} + \dots + \phi^{-1} Z_{t-\tau+1} + X_{t-\tau} .$$

For every positive $\varepsilon < 1$, there exists $\tau > -\ln \varepsilon / \ln |\phi| > 0$ such that the term from the left side can be neglected and then

$$(11) \quad X_{t-\tau} = -\phi^{-1} Z_{t-\tau+1} - \dots - \phi^{-\tau+1} Z_{t-1} - \phi^{-\tau} Z_t + \mathcal{O}(\varepsilon)$$

for $\varepsilon \rightarrow 0$. Hence, the random variable at moment $t - \tau$ can be expressed in terms of the noise values at the future moments. It means that the direction of this process is reversed to that of the causal process (3).

The same conclusion can be drawn from (4) if we write it in the form

$$X_{t-1} = \phi^{-1} X_t - \phi^{-1} Z_t .$$

In this way we obtain a stationary causal process AR(1) because $|\phi^{-1}| < 1$, but in reverse temporal direction and with the variance of the noise $\phi^{-1}\sigma$. It means that the formulas for the acausal AR(1) process can be obtained from those for the causal one by replacing ϕ with ϕ^{-1} and σ with $\phi^{-1}\sigma$. We shall derive these formulas directly as well in order to verify this assertion.

From (11) it follows that the random variables X_{t-1} and Z_t are not independent, that is $E\{X_{t-1}Z_t\} \neq 0$. In this case $E\{X_tZ_t\} = 0$, and then we write (4) in the form

$$\phi X_{t-1} = X_t - Z_t$$

and we square it

$$\phi^2 X_{t-1}^2 = X_t^2 - 2X_tZ_t + Z_t^2.$$

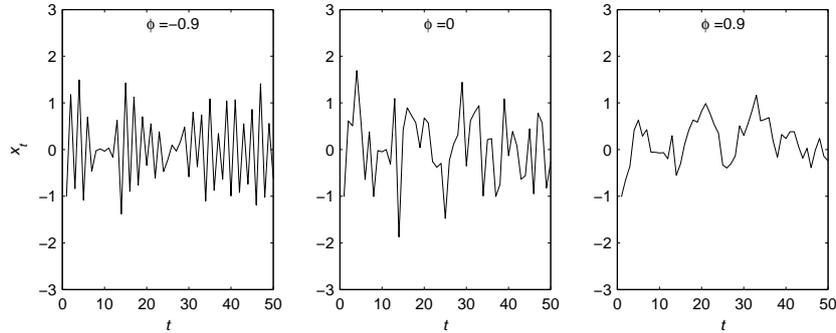


Fig. 1. Realizations of an anticorrelated AR(1) process ($\phi = -0.9$), of an uncorrelated one ($\phi = 0$), and of a correlated one ($\phi = 0.9$).

If we take the mean instead of (7) we obtain

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

In order to compute the autocovariance function we multiply (4) by $X_{t+\tau}$ and instead of (8) we obtain

$$\gamma(\tau) = \sigma_s^2 \phi^{-\tau}.$$

Finally, relation (10) is the same.

2.3. Anticorrelated AR(1) process. Let us remark that 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, if $\phi = 0$, the AR(1) process is reduced to a white noise with uncorrelated terms and a vanishing autocovariance function for $\tau > 0$. When $\phi > 0$, from (4) it follows that the fluctuations due to the white noise are superposed over the term $|\phi|X_{t-1}$ which memorizes a part of the previous value of the time series. Hence, the larger ϕ is, the closer from each other the successive values of the time series are, and the fluctuations due to the white noise are smaller. 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 (see Fig. 1). The autocovariance function (8) is positive and tends to zero for $\tau \rightarrow \infty$.

If $\phi < 0$, then the white noise is superposed over the term $-|\phi|X_{t-1}$ 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 with a larger amplitude than the white noise, as shown in Fig. 1. The successive values of the autocovariance function (8) are of opposite signs and the time series is called anticorrelated.

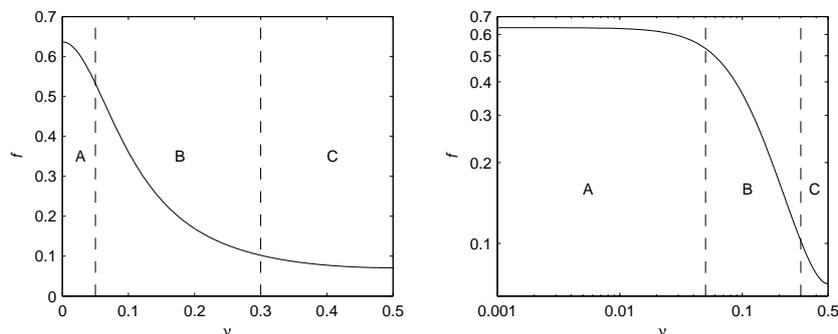


Fig. 2. The power spectrum of an AR(1) process for $\phi = 0.5$ and $\sigma = 1$ on a linear scale and on a log-log one.

3. AR(1) POWER SPECTRUM

In Fig. 2 we have plotted the power spectrum (10) for $\phi = 0.5$ and $\sigma = 1$ on a linear scale and on a log-log one. The logarithmic coordinates strongly distort the shape of the graphic because by taking the logarithm, the origin of the Ox axis is sent to $-\infty$ and any neighborhood of the origin is transformed into an infinite length interval. We have separated the graphic into three regions (A, B, and C) in order to evidence the deformations. For small frequencies (region A) the spectral density is strongly stretched such that a plateau appears with a value given by

$$(12) \quad f(0) = \frac{\sigma^2}{2\pi(1-\phi)^2} .$$

From relation (10) one observes that the plateau corresponds to the small values of ν , when the variable term at the denominator can be neglected in comparison with the constant term. Using the quadratic approximation for cosine function we obtain the condition that the graph of the AR(1) power spectrum has a plateau

$$(13) \quad \nu \ll \frac{1-\phi}{2\pi\sqrt{\phi}} .$$

One remarks that if ϕ tends to 1, then the plateau appears at smaller values of the frequency.

The region C of the spectral density for large frequencies is almost parallel to the Ox axis in linear coordinates. In logarithmic coordinates it is strongly squeezed and acquires a significant slope. The region B of median frequencies is not compressed so much, but its almost exponential shape in linear coordinates becomes a curve with a considerable part having a constant slope. The relative length of the three frequency regions depends on the minimum value ν_{\min} of the frequency scale in the plot. If ν_{\min} is not small enough, then the plateau may remain outside the graphic.

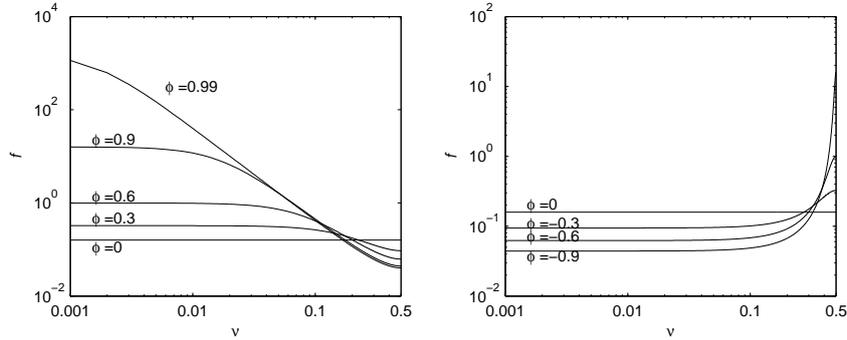


Fig. 3. Power spectra of an AR(1) process for $\sigma = 1$ and different values of ϕ .

As a global characteristic of the AR(1) power spectrum we introduce the difference of its extreme values, quantity which we call the *spectrum amplitude* and is denoted by A . For $\nu \in [0, 0.5]$ the spectral density (10) is monotonic and its extreme values occur at $\nu = 0$ and $\nu = 0.5$. Using (12) and

$$(14) \quad f(0.5) = \frac{\sigma^2}{2\pi(1+\phi)^2} ,$$

we have for the spectrum amplitude

$$(15) \quad A = |f(0) - f(0.5)| = \frac{4|\phi|\sigma^2}{(1-\phi^2)^2} .$$

The extreme values of the power spectrum (12) and (14) become equal for $\phi = 0$ when the power spectrum is a line parallel with Ox axis and $A = 0$. One can see that for $\phi \rightarrow 1$, both $f(0)$ and A tend to infinity.

In Fig. 3 we present the variation of the power spectrum with respect to ϕ . Although in linear coordinates the power spectrum corresponding to $-\phi$ is the reflection of that corresponding to ϕ with respect to a line parallel to Oy axis, in log-log scale they have very different shapes. This difference occurs because, as shown in Fig. 2, the small frequency region is stretched whereas that of large frequency is compressed. In accordance with (12), for $\phi > 0$ the plateau height $f(0)$ increases with ϕ and the extreme region for large frequencies of the power spectrum $f(0.5)$ given by (14) has smaller values for smaller ϕ .

From Fig. 3a and Fig. 2b it results that the AR(1) processes have some fractal features. For $\phi = 0.9$ and especially for $\phi = 0.99$, a large region of the power spectrum is linear with a slope near -2 . Also for small values of ϕ (for example $\phi = 0.5$ in Fig. 2b) a significant region of the power spectrum can be considered linear (fractal). In order to compute the slope of the power spectrum for arbitrary ϕ , we denote by $\xi = \log x$ and $\eta = \log y$ the double-logarithmic coordinates such that a function $y = f(x)$ is written in the new coordinates as

$$\eta = \log_{10} f(10^\xi).$$

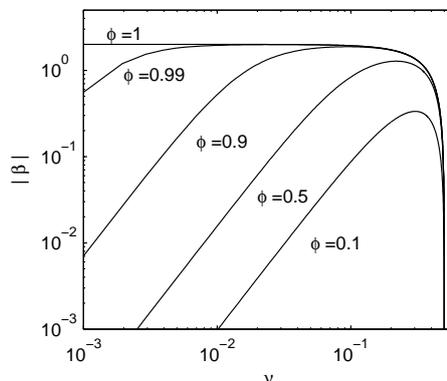


Fig. 4. The absolute value of the slope of the AR(1) power spectrum for $\sigma = 1$ and different values of ϕ .

The slope of the log-log plot is

$$\beta(x) = \frac{d\eta}{d\xi} = -x \frac{d}{dx} (\ln f(x)).$$

If we apply this relation to the function (10), we obtain the slope of the AR(1) power spectrum

$$(16) \quad \beta(\nu; \phi) = -\frac{4\pi\phi\nu \sin 2\pi\nu}{1+\phi^2-2\phi \cos 2\pi\nu}.$$

In Fig. 4 we have plotted the absolute value of this function in log-log scale and we observe that for $\phi \gtrsim 0.9$ there exist values of β near -2 . We can verify this behavior substituting $\phi = 1$ in (16)

$$\beta(\nu; 1) = -2\pi\nu \cot \pi\nu.$$

Obviously this case is artificial because for $\phi = 1$ we obtain the Brownian motion, not the AR(1) process. Then we have $\lim_{\nu \rightarrow 0} \beta(\nu; 1) = -2$ which corresponds to the plateau in Fig. 4. If $\phi < 1$, then $\beta(0; \phi) = 0$ and in Fig. 4 the curve is decreasing for small frequencies. For $\phi < 0.9$ there is only one maximum value for β that corresponds to the center of the "linear" (fractal) region of the power spectrum.

4. FINITE AR(1) PROCESS

The time series appearing in practice have a finite length and usually they are considered finite samples of an AR(1) process of infinite length. The first terms of the sample are correlated with the preceding terms of the realization which has not been recorded. But the first terms of a numerically generated time series can not be related with realizations of other preceding random variables. Therefore, a numerically generated time series is never strictly a realization of a finite sample of an ideal stationary stochastic process of infinite length. For instance, the index t in relations (1) and (4) cannot be an arbitrary integer. Because the relations defining the process are recursive, the first terms

must be defined by additional relations. As we shall show in the following, the manner in which these additional relations are chosen can essentially modify the properties of the stochastic process. We shall call *finite AR(1) process* a stochastic process of finite length satisfying a recursive relation (1).

Let T denote the length of a finite AR(1) process, that is $t = 1, 2, \dots, T$. The first term X_1 must be chosen independently and then we obtain another process instead of that studied in the previous section, denoted by $\{\hat{X}_t\}$. Because \hat{X}_t satisfies the relation (4) for $t > 1$, if we apply this relation successively we can express the terms of the stochastic process as a finite sum

$$(17) \quad \hat{X}_t = Z_t + \phi Z_{t-1} + \dots + \phi^{t-2} Z_2 + \phi^{t-1} \hat{X}_1.$$

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

If \hat{X}_1 is a Gaussian random variable with variance $\hat{\sigma}_1$ and zero mean, then from (17) it follows that \hat{X}_t is the sum of t Gaussian random variables, hence it has also a Gaussian distribution with variance

$$\hat{\sigma}_t^2 = \sigma^2 (1 + \phi^2 + \phi^4 + \dots + \phi^{2(t-2)}) + \hat{\sigma}_1^2 \phi^{2(t-1)}.$$

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

$$(18) \quad \hat{\sigma}_t^2 = \sigma_s^2 + (\hat{\sigma}_1^2 - \sigma_s^2) \phi^{2(t-1)},$$

where we have used (7). The variance of the finite AR(1) process has a constant term equal with the variance of the infinite AR(1) process (7) and a variable term which tends asymptotically to zero because $|\phi| < 1$. In this case the finite AR(1) process is nonstationary presenting transient effects, i.e., its variance approximates the theoretical one $\hat{\sigma}_t \simeq \sigma_s$ only after a time interval t_0 for which ϕ^{2t_0} can be neglected.

4.1. Quasistationary finite AR(1) process. For $\hat{\sigma}_1 = \sigma_s$ the variable term in (18) vanishes and $\hat{\sigma}_t = \sigma_s$ for all $t \leq T$. Hence, if for a finite AR(1) process we choose $\hat{X}_1 = (\sigma_s/\sigma)Z_1$, 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. Because for $\hat{\sigma}_1 \neq \sigma_s$ the calculations become more complicated, we deal only with the case $\hat{\sigma}_1 = \sigma_s$.

Let us show that for $\hat{\sigma}_1 = \sigma_s$ the properties of the finite AR(1) process are very identical to those of a finite sample of a stationary infinite AR(1) process. The autocovariance function $\hat{\gamma}(\tau; t) = E\{\hat{X}_t \hat{X}_{t-\tau}\}$ can be calculated only if $|\tau| < T$ and $\tau < t \leq T + \tau$. Unlike the autocovariance function (8) the quantity $\hat{\gamma}(\tau; t)$ does not depend only on τ since it exists only for certain values of t . Therefore $\{\hat{X}_t\}$ is not a stationary stochastic process in a strict mathematical meaning. However, when it exists, we can show that

$E\{\widehat{X}_t\widehat{X}_{t-\tau}\} = \phi E\{\widehat{X}_{t-1}\widehat{X}_{t-\tau}\}$ proceeding in the same way as for (8). Then instead of (8) we obtain

$$(19) \quad \widehat{\gamma}(\tau; t) = \phi^\tau \widehat{\sigma}_{t-\tau}^2 .$$

If we choose $\widehat{\sigma}_1 = \sigma_s$, then $\widehat{\sigma}_{t-\tau} = \sigma_s$ is constant and when $\widehat{\gamma}(\tau; t)$ exists it is identical to the covariance function $\gamma(\tau)$ in (8). Hence, if we want to numerically model a stationary infinite AR(1) process, then we have to use a finite AR(1) process $\{\widehat{X}_t\}$ with $\widehat{\sigma}_1 = \sigma_s$.

4.2. Brownian motion. Let us analyze now the finite AR(1) process satisfying (4) for $\phi = 1$. This is the well known Brownian motion. In this case $\widehat{X}_1 = Z_1$ and then (17) becomes

$$(20) \quad \widehat{X}_t = Z_t + Z_{t-1} + \dots + Z_2 + Z_1 .$$

Because \widehat{X}_t is the sum of t gaussian random variables, it results that its variance is the sum of the variances of the terms of the sum

$$(21) \quad \widehat{\sigma}_t^2 = t\sigma^2 .$$

If we write the relation (20) in the form

$$\widehat{X}_{t+\tau} = Z_{t+\tau} + \dots + Z_{t+1} + \widehat{X}_t ,$$

multiply it with \widehat{X}_t and take the mean, then we obtain

$$(22) \quad E\{\widehat{X}_{t+\tau}\widehat{X}_t\} = \widehat{\sigma}_t^2 = t\sigma^2 .$$

The relation between the Brownian motion and the quasistationary finite AR(1) process can be clarified if in (18) we take $\widehat{\sigma}_1 = \sigma$ corresponding to the choice of the first term for the Brownian motion $\widehat{X}_1 = Z_1$

$$(23) \quad \widehat{\sigma}_t^2 = \sigma_s^2 (1 - \phi^{2t}) .$$

Figure 4 shows the variation of $\widehat{\sigma}_t^2$ for 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, i.e., the Brownian motion. So the Brownian motion corresponds to the transient region 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 very close to 1 the only possibility is to choose $\widehat{\sigma}_1 = \sigma_s$ completely eliminating the transient region.

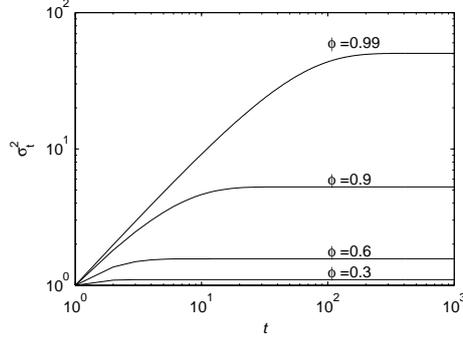


Fig. 5. The standard deviation of a finite AR(1) process for different values of ϕ when the first term coincides with the white noise $\hat{\sigma}_1 = \sigma$. The transient region extends when ϕ takes greater values.

5. PERIODOGRAM

For a finite AR(1) process the definition (9) can not be applied. Therefore we use the periodogram which is defined as the square of the absolute value of the discrete Fourier transform of the stochastic process of finite length [16]. Here we consider only a finite equidistant set of frequencies for which we define the amplitude

$$(24) \quad A_n = \frac{1}{\sqrt{T}} \sum_{t=1}^T \hat{X}_t \omega_T^{(t-1)(n-1)}, \quad n = 1, 2, \dots, T,$$

where $\omega_T = \exp(-2\pi i/T)$. The corresponding terms of the periodogram are

$$(25) \quad I_n = |A_n|^2.$$

Replacing (24) in (25) we have

$$I_n = \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \hat{X}_t \omega_T^{(t-1)(n-1)} \hat{X}_s \omega_T^{-(s-1)(n-1)}.$$

Because $\omega_T^{t-1} \omega_T^{-s+1} = \exp\left\{-\frac{2\pi i}{T}(t-s)\right\}$, we introduce a new variable $\tau = t-s$ and then

$$(26) \quad I_n = \frac{1}{T} \sum_{t=1}^T \hat{X}_t^2 + \frac{1}{T} \sum_{\tau=1}^{T-1} \left[\omega_T^{\tau(n-1)} + \omega_T^{-\tau(n-1)} \right] \sum_{t=1}^{T-\tau} \hat{X}_t \hat{X}_{t+\tau}.$$

The terms I_n of the periodogram are random variables and their mean $E\{I_n\}$ tends to the spectral density $f(\nu)$ given by (10) when $T \rightarrow \infty$ [2]. We shall show this property only for the two particular cases considered in the previous section for which the calculations are shorter.

5.1. The periodogram of the quasistationary finite AR(1) process.

According to the analysis in the previous section, if $\hat{\sigma}_1 = \sigma_s$, then $E\{\hat{X}_t^2\} = \sigma_s^2$ and $E\{\hat{X}_t \hat{X}_{t+\tau}\} = \sigma_s^2 \phi^\tau$ and the average of relation (26) reads

$$E\{I_n\} = \sigma_s^2 + \frac{1}{T} \sigma_s^2 \sum_{\tau=1}^{T-1} (T - \tau) \phi^\tau \left[\omega_T^{\tau(n-1)} + \omega_T^{-\tau(n-1)} \right].$$

The sums in this formula are equal to

$$(27) \quad \sum_{\tau=1}^{T-1} (T - \tau) x^\tau = \frac{xT}{1-x} - \frac{x(1-x^T)}{(1-x)^2}$$

where $x = \phi \omega_T^{\pm(n-1)}$. Since the first term is proportional to T , it is larger than the second term, therefore we write the periodogram mean as a sum of a dominant term and a correction

$$(28) \quad E\{I_n\} = \frac{\sigma^2}{1 - 2\phi \cos \frac{2\pi}{T}(n-1) + \phi^2} (1 + \rho_n).$$

When the correction ρ_n can be neglected, this formula is an approximation of the spectral density of the stationary infinite AR(1) process (10)

$$E\{I_n\} \simeq 2\pi f(\nu_n),$$

where $\nu_n = (n-1)/T$. Since (28) is a periodic function, we shall consider for the index n only the values $n = 1, 2, \dots, [(T+1)/2]$, where $[\cdot]$ is the integer part function, so that $\nu_n \in [0, 0.5]$. Relation (28) can be written as well as

$$(29) \quad E\{I_n\} = \frac{\sigma^2}{(1-\phi)^2 + 4\phi \sin^2 \pi \nu_n} (1 + \rho_n).$$

Now we shall analyze the dependence of the correction ρ_n on the parameters n , T and ϕ . Taking into account that in (27) $x^T = \phi^T$, from a direct computation it follows that ρ_n can be written as a product of two factors depending each of them only on two of the three parameters

$$(30) \quad \rho_n = r(\phi, T) \eta(\phi, \nu_n),$$

where

$$(31) \quad r(\phi, T) = \frac{1}{T} \frac{2\phi(1-\phi^T)}{1-\phi^2}$$

and

$$(32) \quad \eta(\phi, \nu_n) = \frac{2\phi - (1+\phi^2) \cos 2\pi \nu_n}{1+\phi^2 - 2\phi \cos 2\pi \nu_n}.$$

The derivative of the function η with respect to ν_n indicates that it is monotonic increasing for any fixed value $\phi \in (-1, 1)$. For the extreme values of ν_n , the function η is independent on the values of ϕ , i.e., $\eta(\phi, 0) = -1$ and $\eta(\phi, 0.5) = 1$. In Fig. 6 we have plotted η with respect to ν_n for different values of ϕ . For $\phi = 0$ the function reduces to a cosine function $\eta(0, \nu_n) = -\cos 2\pi \nu_n$ which, as $|\phi|$ increases, is distorted into a constant function, i.e.,

$$(33) \quad \lim_{\phi \rightarrow \pm 1} \eta(\phi, \nu_n) = \pm 1.$$

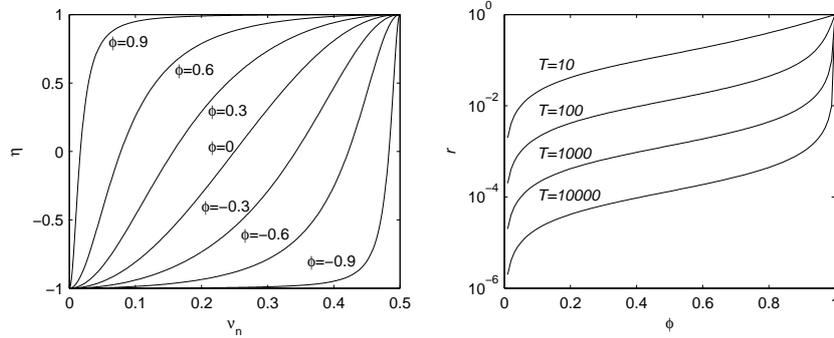


Fig. 6. Correction of the periodogram of the quasistationary finite AR(1) process in comparison with the power spectrum of the stationary infinite AR(1) process.

The absolute value of this factor is always smaller or equal to 1.

The second factor in the correction (30) given by (31) is plotted in Fig. 6 for $\phi > 0$ and a few values of T . It is a monotonic increasing function with respect to ϕ and

$$(34) \quad \lim_{\phi \rightarrow 1} r(\phi, T) = 1 .$$

From (33) and (34) it follows that when $\phi \rightarrow 1$ the correction (30) is also approximately equal to 1, so that the approximation (28) of the AR(1) power spectrum is wrong in the very dominant order. Reversely, for a given value of ϕ , the relation (31) allows us to find the value of T so that the correction r should have the desired value.

If T is even, then these conclusions hold for $\phi < 0$ too, since r changes only its sign $r(-\phi, T) = -r(\phi, T)$. But if T is odd, then ϕ^T changes its sign and the function $r(\phi, T)$ acquires a different form. The most important modification is that, for $\phi \rightarrow -1$ the factor r becomes infinite. Thus, in this case, the periodogram of the quasistationary finite AR(1) process is completely different from the power spectrum of the stationary infinite AR(1) process.

If the length T is large enough such that the correction ρ_n in (29) could be neglected, then the graphical representation of the finite AR(1) power spectrum is identical to that discussed in Section 3. For example, when $T = 1000$ and $\phi = 0.9$ the correction (30) is $\rho_n = 0.0095 \eta(\phi, \nu_n) < 0.0095$ and it can be neglected. But for $T = 1000$ and $\phi = 0.99$ we have a much greater correction $\rho_n = 0.0995 \eta(\phi, \nu_n)$. In Fig. 6 one can see that for ϕ near 1, the factor η is approximately 1 excepting for very small frequencies. Therefore the correction r is constant for almost all frequencies and the periodogram (29) differs from the spectral density (10) by the constant factor $1 + \rho_n$. In log-log scale this constant factor shifts the periodogram parallel to the AR(1) power spectrum (see Fig. 7).

Because $\nu_1 = 0$, in a log-log scale the minimum frequency for a periodogram of a finite AR(1) process is $\nu_2 = 1/T$. Then the condition that the periodogram

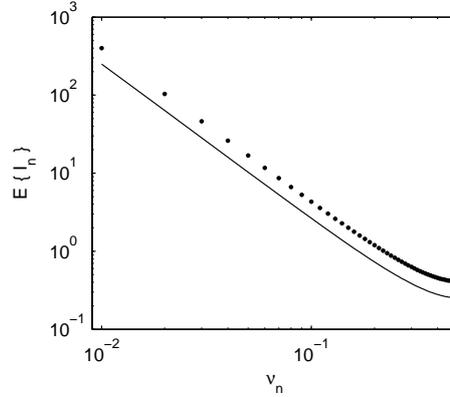


Fig. 7. The periodogram of an finite AR(1) process with $T = 100$ and $\phi = 0.99$ (point markers) in comparison with the power spectrum of an infinite AR(1) process with the same ϕ (continuous line).

has a plateau can be obtained from (13)

$$(35) \quad T \gg \frac{2\pi\sqrt{\phi}}{1-\phi} .$$

For example, when $\phi = 0.99$ we have $T \gg 600$ and in Fig. 7 one can see that for $T = 100$ the periodogram has no plateau.

5.2. The periodogram of the Brownian motion. The periodogram of the Brownian motion is obtained by replacing (21) and (22) in the formula resulting by averaging (26)

$$(36) \quad E \{I_n\} = \frac{\sigma^2}{2} (T + 1) + \frac{\sigma^2}{2T} \sum_{\tau=1}^{T-1} (T - \tau) (T - \tau + 1) \cdot (\omega_T^{\tau(n-1)} + \omega_T^{-\tau(n-1)}) .$$

After some calculations we obtain

$$(37) \quad E \{I_n\} = \frac{\sigma^2 T}{8 \sin^2\left(\frac{\pi}{T}(n-1)\right)} + \frac{\sigma^2 \cos\left(\frac{2\pi}{T}(n-1)\right)}{2 \sin^2\left(\frac{\pi}{T}(n-1)\right)} .$$

The dominant term in (37) is different from that in (29) for $\phi = 1$ only by the factor $T/2$. The relation (37) does not hold for $n = 1$ because in the calculations we used relations incompatible with that special value. If in (36) we take $n = 1$, then

$$(38) \quad E \{I_1\} = \frac{\sigma^2}{3} (T + 1) (T^2 - T + 3) .$$

In Fig. 8 we have plotted the dominant term of the periodogram (37) for $T = 1000$. The most part of the graphic is a straight line since, if in (37) we consider $n \ll T$, then $E \{I_n\} \propto n^{-2}$. So in a log-log scale we obtain a straight line with the slope -2 and since the small frequencies region is strongly delated

the most part of the graphic has this property. Figure 8 shows that outside of the plateau, the periodogram of the Brownian motion is parallel to the periodograms of quasistationary finite AR(1) processes with ϕ close to 1 and the distance between them equals $\log(T/2)$ as shown above. This behavior is due to the fact that for $1 - \phi \ll 1$ the value of the sine function at the denominator in (29) is dominant. Only for ν_n small enough the quantities (29) and (37) are significantly different.

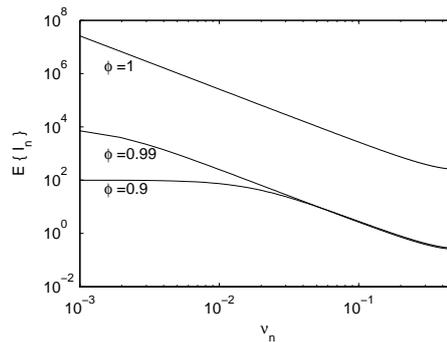


Fig. 8. The periodogram of the Brownian motion in comparison with the periodograms of the quasistationary finite AR(1) process with $\phi = 0.9$ and 0.99 .

6. CONCLUSIONS

Although the finite and infinite AR(1) processes are defined by the same recurrence relation, their properties can be very different when the serial correlation is large and the length of the finite AR(1) process is small. This difference is minimized if the first term of the finite AR(1) process is chosen such that its standard deviation should be equal to the standard deviation of the infinite AR(1) process. However, even in this case the periodogram of the finite AR(1) process for ϕ close to ± 1 can significantly differ from the power spectrum of the infinite AR(1) process. That is why to numerically generate a time series as a realization of an AR(1) process we must take into account that the series length must be larger than the threshold value (35).

Although the AR(1) process is the simplest stochastic process describing the serial correlation only by means of a single parameter, it has remarkable properties which make it very useful as a first step in time series modeling. However, one has to take care that it is only one in the infinity of existing stochastic models and it is possible that an autoregressive structure to be incorrectly assigned to a time series. For example, although the autocovariance function (8) has an exponential decay, sometimes one attempted to model a $1/f$ noise characterized by a power law decay with autoregressive processes [9], [14]. Therefore, it is necessary a thorough analysis of the AR(1) process properties not only in its infinite idealized form, but in the finite one as well. If

the time series is not long enough, it is very likely that essential characteristics of the AR(1) process should be lost, as for example the existence of the plateau at small frequencies, and misinterpretations can occur.

REFERENCES

- [1] BLENDER, R., *Renormalization group analysis of autoregressive processes and fractional noise*, Phys. Rev. E, **64**, 067101 (2001).
- [2] BROCKWELL, P.J. and DAVIS, R., *Time Series: Theory and Methods*, Springer-Verlag, New York, 1991.
- [3] BROCKWELL, P.J. and DAVIS, R., *Introduction to Time Series and Forecasting*, Springer-Verlag, New York, 1996.
- [4] BOX, G. E. P. and JENKINS, G. M. *Time Series Analysis: Forecasting and Control*, 2nd ed., Holden-Day, San Francisco, 1976.
- [5] GAO, J., HU, J., TUNG, W., CAO, Y., SARSHAR, N. and 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. and 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. and KANTZ, H., *Precursors of extreme increments*, Phys. Rev. E, **75**, 016706 (2007).
- [8] HAMILTON, J.D., *Time Series Analysis*, Princeton University Press, 1994.
- [9] KAULAKYS, B., *Autoregressive model of 1/f noise*, Physics Letters A, **257**, 37 (1999).
- [10] KIRÁLY, A. and JÁNOSI, I. M., *Stochastic modeling of daily temperature fluctuations*, Phys. Rev. E, **65**, 051102 (2002).
- [11] KUGIUMTZIS, D., *Statically transformed autoregressive process and surrogate data test for nonlinearity*, Phys. Rev. E, **66**, 025201 (2002).
- [12] LILEY, D. T., CADUSCH, P. J., GRAY, M. and NATHAN, P. J., *Drug-induced modification of the system properties associated with spontaneous human electroencephalographic activity*, Phys. Rev. E, **68**, 051906 (2003).
- [13] MARAUN, D., RUST, H.W. and TIMMER, J., *Tempting long-memory \hat{U} on the interpretation of DFA results*, Nonlinear Processes in Geophysics, **11**, 495-503 (2004).
- [14] MORARIU, V.V. and COZA, A., *Nonlinear properties of the atomic vibrations in protein backbones*, Physica A, **320**, 449 (2003).
- [15] PALUS, M. and NOVOTNA, D., *Sunspot Cycle: A Driven Nonlinear Oscillator?* Phys. Rev. Lett., **83**, 3406 (1999).
- [16] STOICA, P. and MOSES, R. L., *Introduction to Spectral Analysis*, Prentice-Hall, New Jersey, 1997.
- [17] TIMMER, J., SCHWARZ, U., VOSS, H.U., WARDINSKI, I., BELLONI, T., HASINGER, G., VAN DER KLIS, M and KURTHS, J., *Linear and nonlinear time series analysis of the black hole candidate Cygnus X-1*, Phys. Rev. E, **61**, 1342 (2000).
- [18] VAMOŞ, C., *Automatic algorithm for monotone trend removal*, Phys. Rev. E, **75**, 036705 (2007).

Received by the editors: February 14, 2007.