Time Series Analysis:

5. Linear stochastic models (I)

The autoregressive process AR(p)

P. F. Góra

http://th-www.if.uj.edu.pl/zfs/gora/

2021

Let $\{y_n\}_{n=0}^{N}$ be a *stationary* time series. A stochastic model of this series is given by

$$y_n = \underbrace{\underbrace{\beta_1 y_{n-1} + \beta_2 y_{n-2} + \cdots + \beta_p y_{n-p}}_{\mathrm{AR}(p)} + \underbrace{\alpha_0 \eta_n + \alpha_1 \eta_{n-1} + \cdots + \alpha_q \eta_{n-q}}_{\mathrm{MA}(q)}}_{\mathrm{ARMA}(p,q)} \tag{1}$$

where $\{\eta_n\}$ is GWN. Eq. (1) is a causal IIR filter applied to the Gaussian white noise.

Let B be a time shift operator: $Bz_n = z_{n-1}$. Eq. (1) can be written as

$$\left(1 - \sum_{k=1}^{p} \beta_k B^k\right) y_n = \left(\sum_{k=0}^{q} \alpha_k B^k\right) \eta_n \tag{2}$$

.

Copyright © 2009-21 P. F. Góra

A “philosophy” of stochastic modelling

$\{y_n\}$ is our data, something that has been given to us and is an ordered sequence of measurement of some “real life” quantity. We do not know any mechanism, derived form first principles, that has generated this time series. All we know is the time series. We *assume* that the time series originates from some random events, but that there is also a mechanism that shapes the output. It may introduce regularities into the data, in particular correlations $\langle y_n y_m \rangle$. This mechanism is represented by the coefficients β_k, α_k. We do not want to find a “true” nature of the data. We merely want to find a model that reproduces the data sufficiently well.

Copyright © 2009-21 P. F. Góra

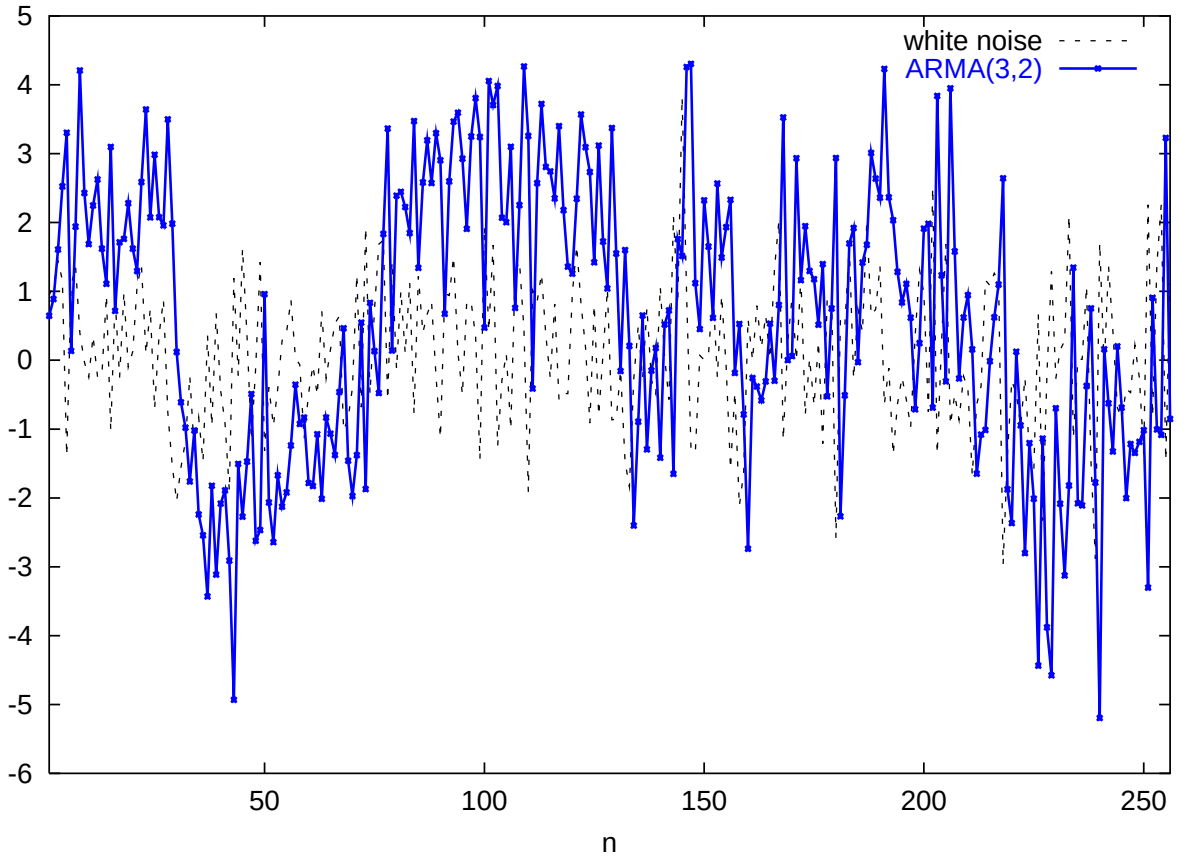
The “random events” in question may include

- thermal noise contaminating a signal
- decisions of individual investors to buy or sell stocks
- decisions of shoppers to purchase goods
- changes in population of an ant colony, reflecting environmental variability
- number of sunspots
- etc

$\{\eta_n\}$ can represent almost anything, as long as we can reasonably assume that this “anything” is modelled by equilibrium fluctuations. Even this assumption can be relaxed: $\{\eta_n\}$ can represent *any* stationary series of events, although in this case it no longer is a GWN, which, generally, shall not be covered in these lectures.

Copyright © 2009-21 P. F. Góra

Example of an ARMA(3,2) stochastic model



Copyright © 2009-21 P. F. Góra

Correlation function

Correlation function is the most important quantity that is used in analysis of linear models. For convenience, the correlation coefficient is used more frequently than the (not normalised) correlation function.

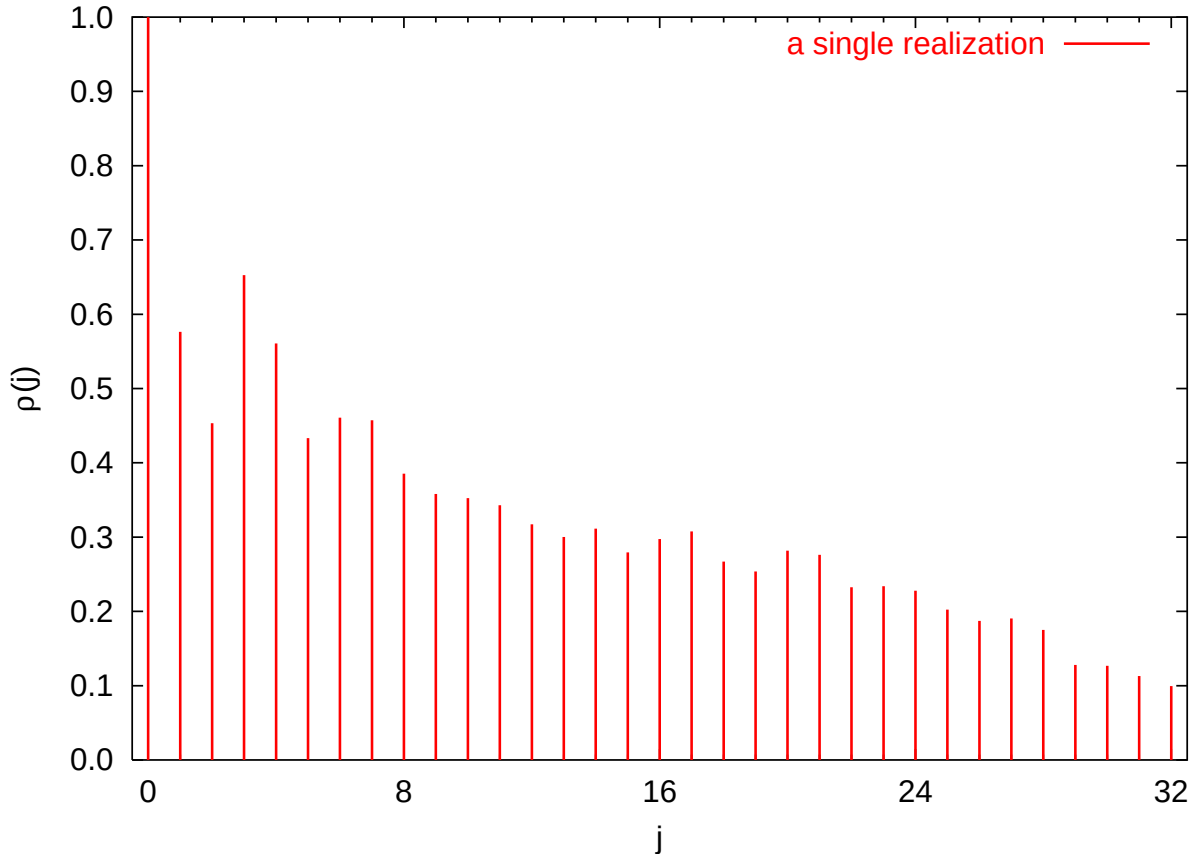
Theoretically the correlation function is calculated by averaging over realizations of the stochastic process that "generated" the series.

$$\rho(i) = \left\langle \left(\frac{1}{N-i} \sum_{j=i+1}^{N} y_{j-i} y_j \right) \Big/ \left(\frac{1}{N} \sum_{j=1}^{N} y_j^2 \right) \right\rangle \tag{3}$$

In reality we have only a single realization of the time series, and therefore, the correlation function estimated from the series *differs* from the theoretical correlation function.

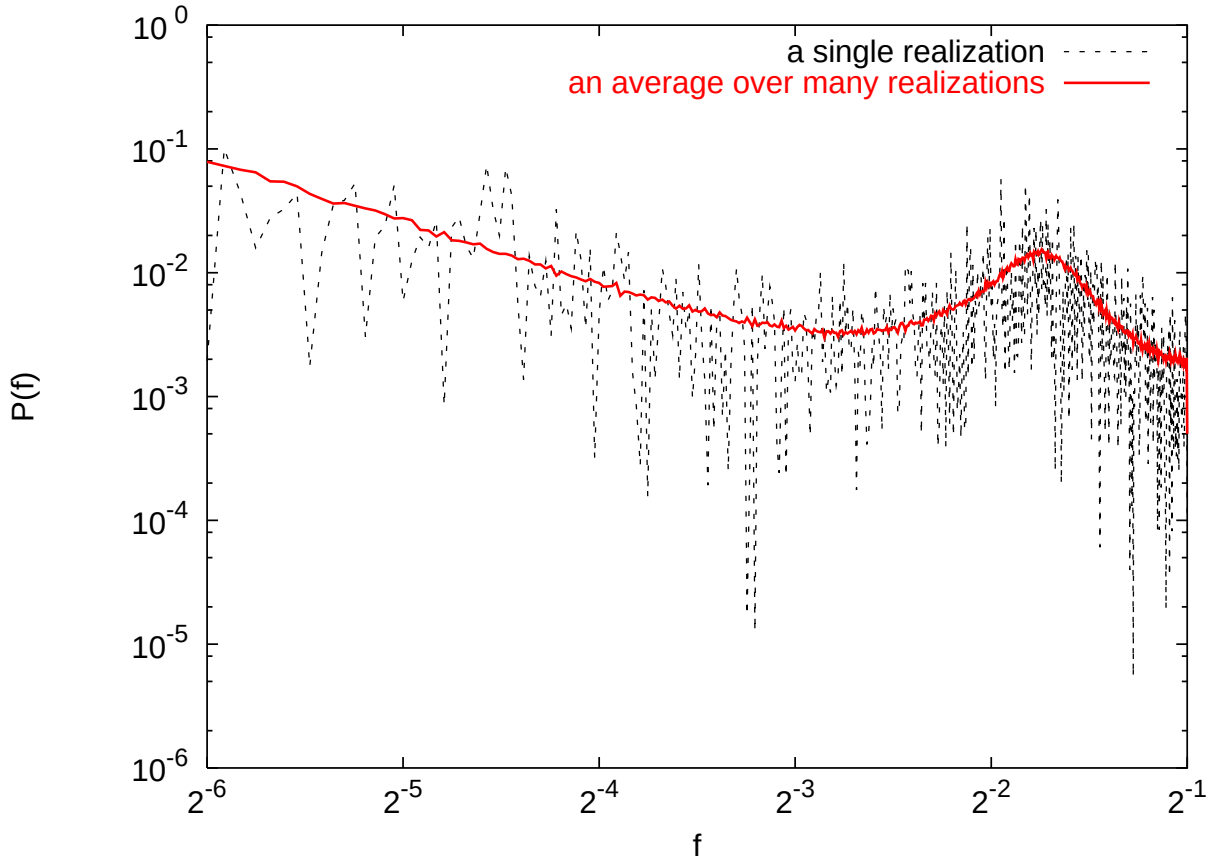
Copyright © 2009-21 P. F. Góra

Correlation function of the above process



Copyright © 2009-21 P. F. Góra

Power spectrum of the above process



Copyright © 2009-21 P. F. Góra

The goal of stochastic modelling

Given a stationary time series, its power spectrum and its correlation function, calculated from the only realization available, estimate the parameters p, q, β_k, α_k of the model (1).

By doing so, gain some insight on the mechanism that has generated the series. If you are lucky, use this mechanism to predict future behaviour (not necessarily the actual future values!) of the series.

☺ (or maybe ☹)

Copyright © 2009-21 P. F. Góra

Processes AR(p)

If the moving average part in (1) vanishes, $q = 0$, the model is called an Autoregressive Model of order p, AR(p):

$$y_n = \beta_1 y_{n-1} + \beta_2 y_{n-2} + \cdots + \beta_p y_{n-p} + \alpha_0 \eta_n \tag{4}$$

We demand that all roots of the polynomial

$$1 - \beta_1 u - \beta_2 u^2 - \cdots - \beta_p u^p = 0 \tag{5}$$

lie outside the unit circle, as otherwise the model (4) may produce nonstationary output.

Copyright © 2009-21 P. F. Góra

Autocorrelation function

Multiply both sides of (4) by y_{n-m} and take the average over realizations:

$$\langle y_{n-m} y_n \rangle = \beta_1 \langle y_{n-m} y_{n-1} \rangle + \cdots + \beta_p \langle y_{n-m} y_{n-p} \rangle + \alpha_0 \langle y_{n-m} \eta_n \rangle \ . \quad (6)$$

Because y_{n-m} is statistically independent from a *later* noise, the last term in (6) vanishes. Dividing by the variance, we get

$$\rho_m = \beta_1 \rho_{m-1} + \beta_2 \rho_{m-2} + \cdots + \beta_p \rho_{m-p} \quad (7)$$

Copyright © 2009-21 P. F. Góra

The general solution to a homogeneous difference equation (7) has a form

$$\rho_m = \sum_{j=1}^{p} A_j \lambda_j^m \,, \tag{8}$$

where λ_j are reciprocals of the roots of the polynomial (5)*. Because $\forall j: |\lambda_j|<1$, $\lambda_j = e^{-\tau_j} e^{2\pi i f_j}$, $\tau_j > 0$. Because the polynomial (5) has real coefficients, the constants A_j in (8) can be chosen such that ρ_m is real.

The autocorrelation function of an autoregressive process is a linear combination of vanishing exponentials and damped harmonic oscillations:

$$\rho_m = \sum A_j e^{-m\tau_j} + \sum{}' A_{j'} e^{-m\tau_{j'}} \sin(2\pi f_{j'} m + \phi_{j'}) \,. \tag{9}$$

*See Lecture 4, Eqns. (27)–(289). Additionally, we assume that all the roots are single.

Copyright © 2009-21 P. F. Góra

Power spectrum of an AR(p) process

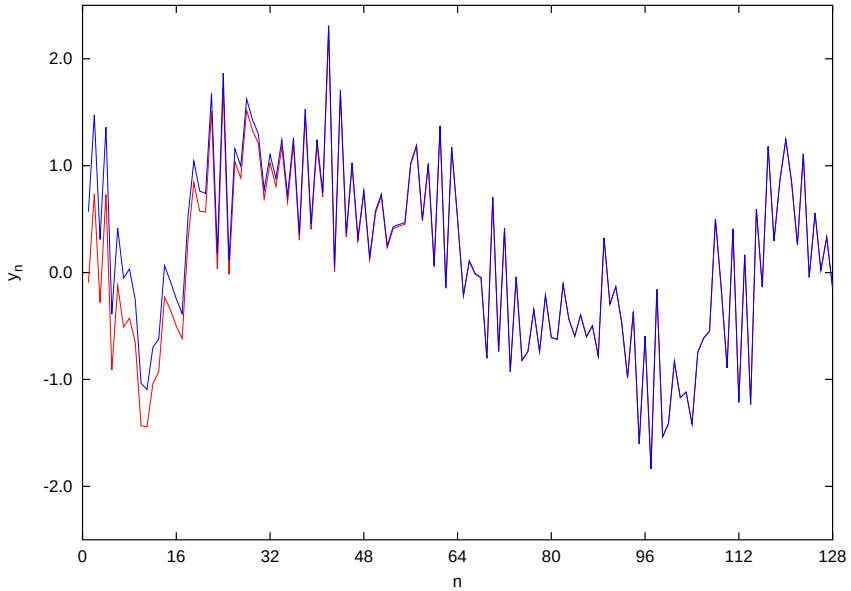
The power spectrum can be obtained immediately from the transfer function of the corresponding IIR filter:

$$P(f) = \frac{\alpha_0^2}{\left|1 - \sum_{n=1}^{p} \beta_n e^{2\pi i n f}\right|^2}, \quad 0 \leqslant f \leqslant \frac{1}{2},. \tag{10}$$

Copyright © 2009-21 P. F. Góra

Initialization

If we want to simulate an AR(p) process, we are supposed to know $y_{-1}, y_{-2}, \ldots, y_{-p}$. Actually, this is *not* a problem, as any realizations initialized with different conditions, but driven by the same noise, eventually converge to the same series. For convenience, we usually set $y_{-1} = 0$ etc.



Two realizations of the same AR(p) process, differing in initial conditions, but driven by the same noise.

Copyright © 2009-21 P. F. Góra

Process AR(1)

$$y_n = \beta_1 y_{n-1} + \alpha_0 \eta_n \tag{11}$$

$-1 < \beta_1 < 1$. The correlation function ($m > 0$):

$$\langle y_n y_{n-m} \rangle = \beta_1 \langle y_{n-1} y_{n-m} \rangle + \alpha_0 \langle \eta_n y_{n-m} \rangle \tag{12}$$

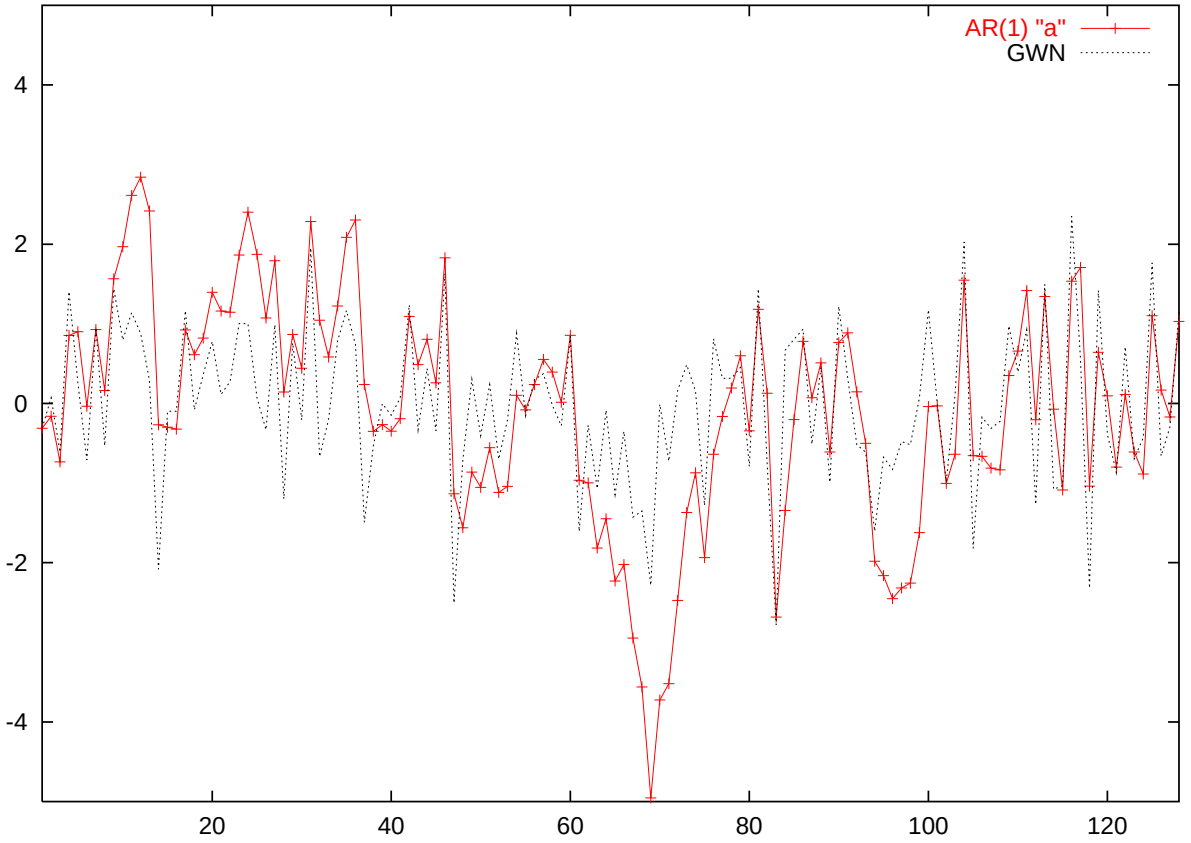
$$\rho_m = \beta_1 \rho_{m-1} \tag{13}$$

$$\rho_m = {\beta_1}^m = (\mathrm{sgn}(\beta_1))^m \, e^{m \ln |\beta_1|} \tag{14}$$

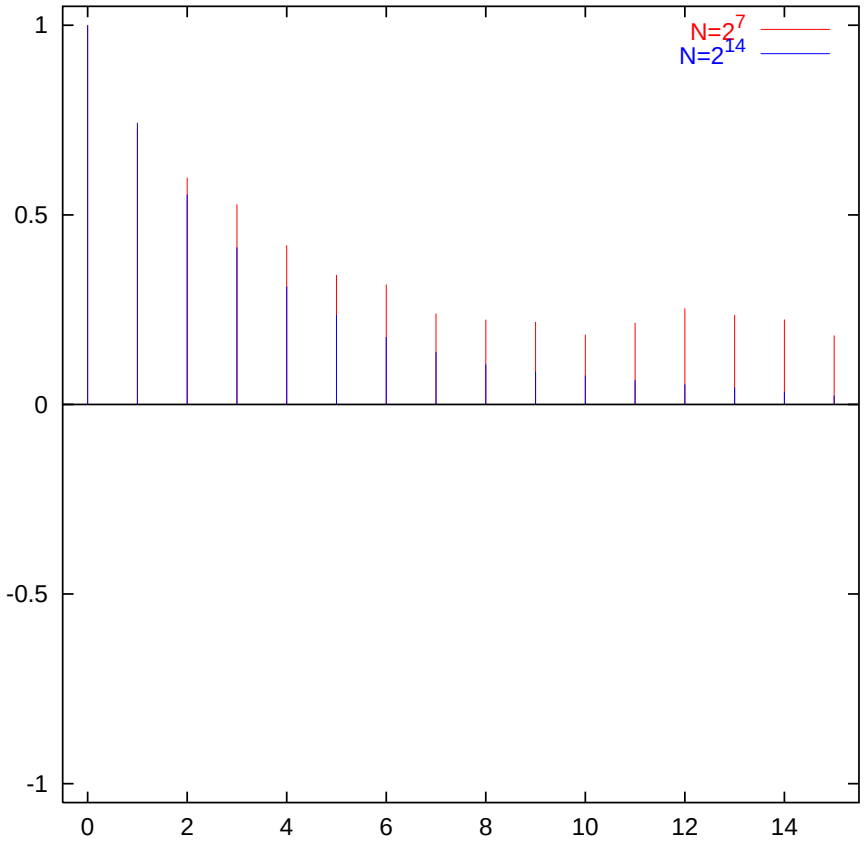
(Note that $\ln |\beta_1| < 0$.)

Copyright © 2009-21 P. F. Góra

$$y_n = 0.75 y_{n-1} + \eta_n$$

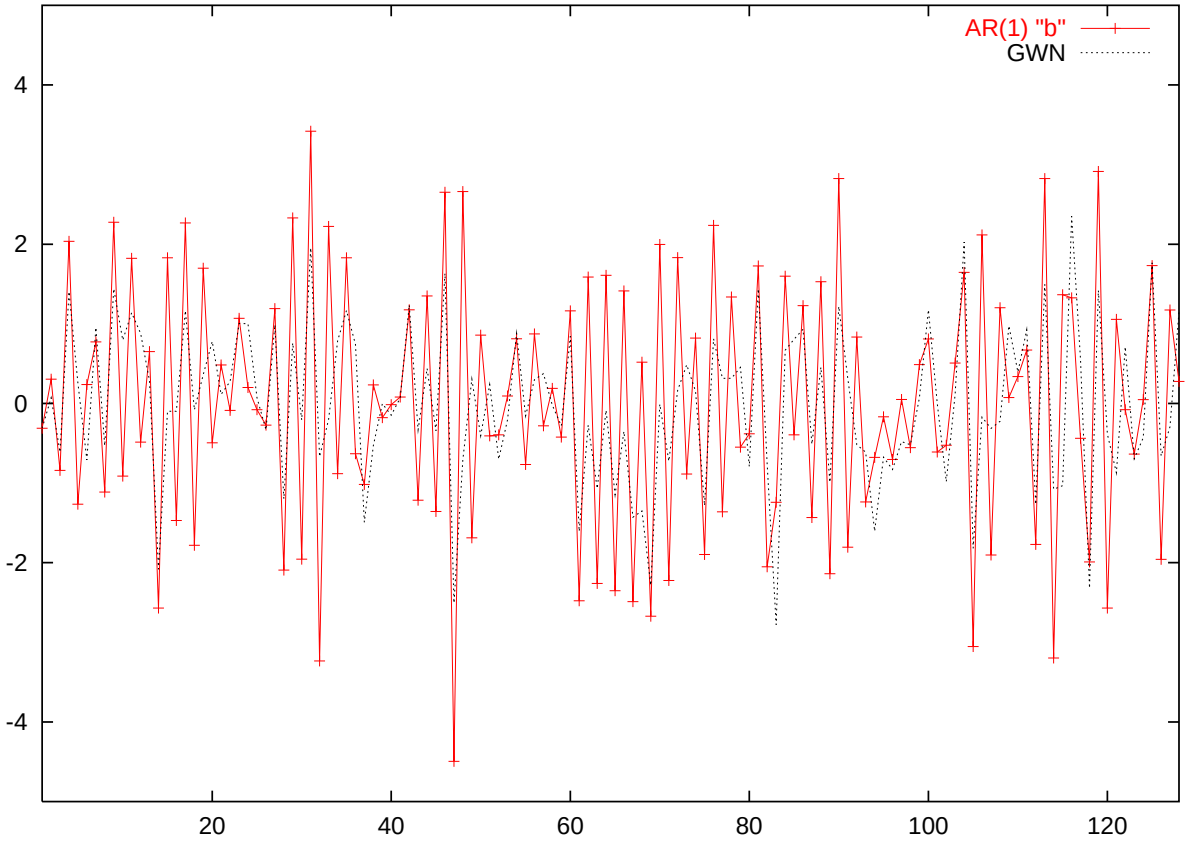


Copyright © 2009-21 P. F. Góra

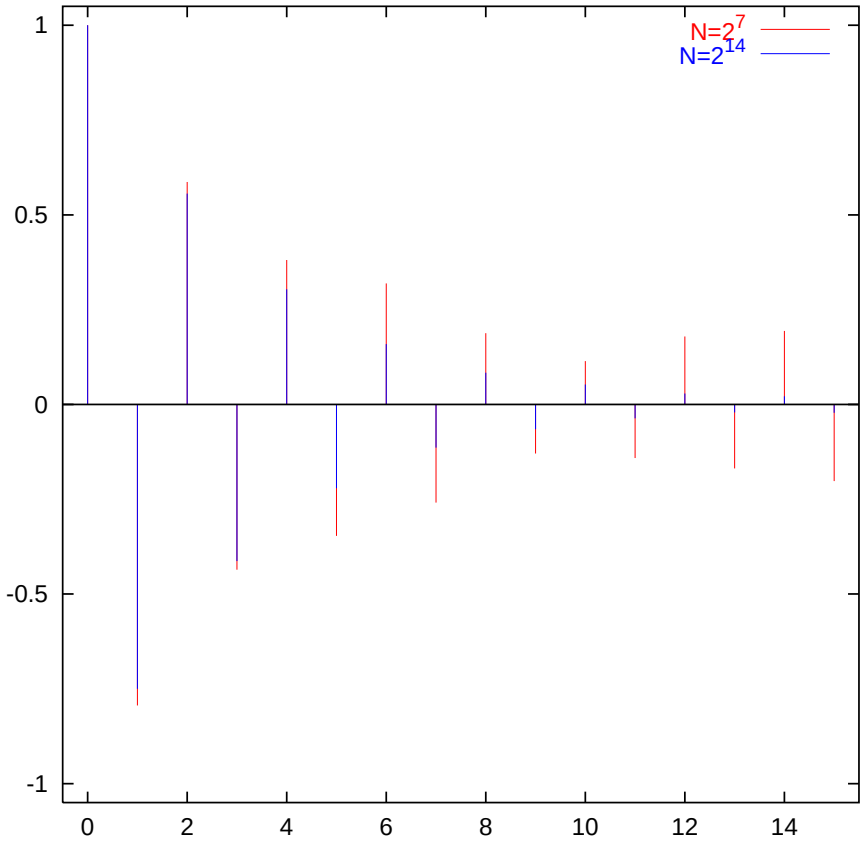


Copyright © 2009-21 P. F. Góra

$$y_n = -0.75 y_{n-1} + \eta_n$$



Copyright © 2009-21 P. F. Góra



Copyright © 2009-21 P. F. Góra

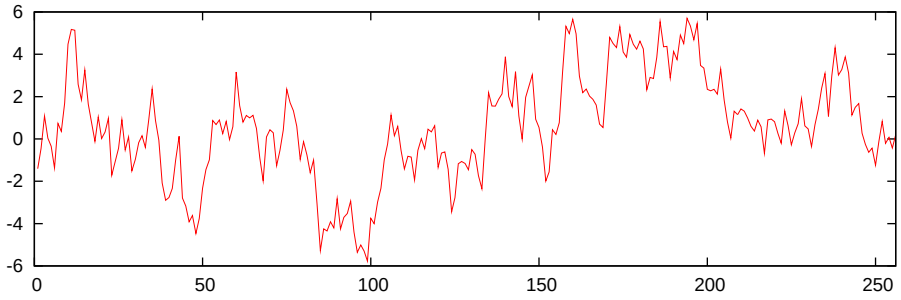
From (10) we can easily get that for AR(1)

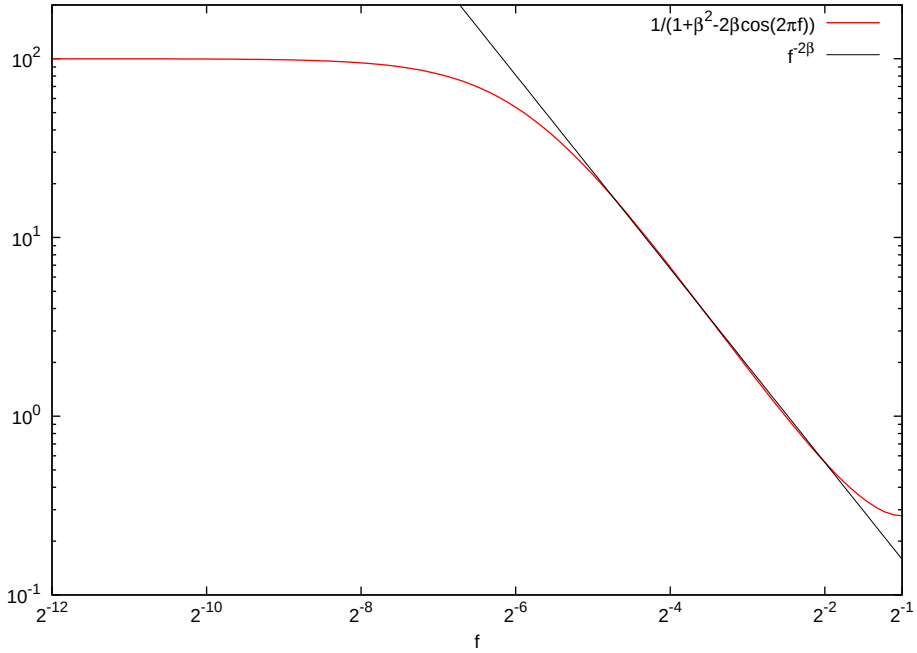
$$P(f) = \frac{\alpha_0^2}{1 + \beta_1^2 - 2\beta_1 \cos(2\pi f)} \tag{15}$$

the character of the process *strongly* depends on the sign of β_1.

Copyright © 2009-21 P. F. Góra

The Ornstein-Uhlenbeck process

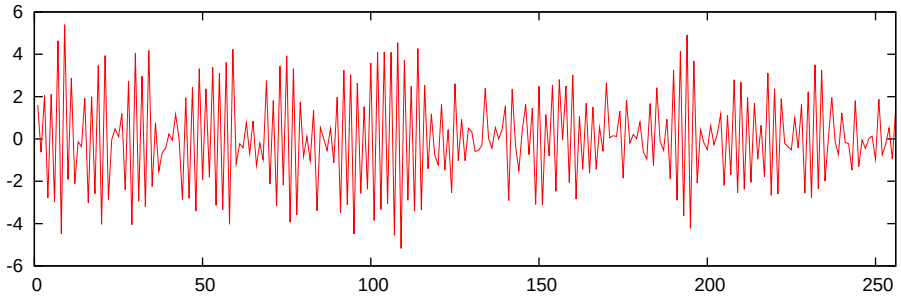


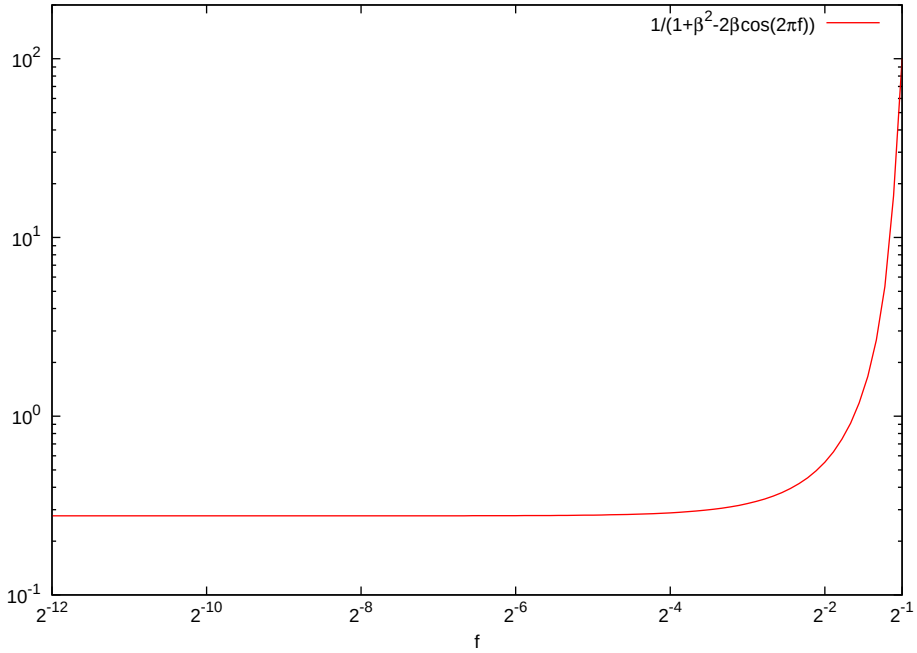


$$\beta_1 = 0.9 > 0$$

Copyright © 2009-21 P. F. Góra

A “blue” noise





$$\beta_1 = -0.9 < 0$$

Copyright © 2009-21 P. F. Góra

Note

A process AR(1) is Markovian: y_n depends on y_{n-1} only (plus random noise), and not on any previous values. If $0 < \beta_1 < 1$, the process (11) is also called the Ornstein-Uhlenbeck process.

A general AR(p) process is non-Markovian. However, an AR(p) process can be embedded in a p-dimensional space

$$\underbrace{\begin{bmatrix} y_n \\ y_{n-1} \\ y_{n-2} \\ \vdots \\ y_{n-p} \end{bmatrix}}_{\mathbf{y}_n} = \underbrace{\begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \dots & \beta_{p-1} & \beta_p \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} y_{n-1} \\ y_{n-2} \\ y_{n-3} \\ \vdots \\ y_{n-(p+1)} \end{bmatrix}}_{\mathbf{y}_{n-1}} + \alpha_0 \underbrace{\begin{bmatrix} \eta_n \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\eta_n} \tag{16}$$

Eq. (16) describes a p-variate AR(1) process $\mathbf{y}_n = \mathbf{B}\mathbf{y}_{n-1} + \alpha_0\eta_n$. This process is Markovian: a non-Markovian autoregressive process AR(p) of a *finite* order in one dimension becomes a Markovian process in p dimensions.

Copyright © 2009-21 P. F. Góra

Process AR(2)

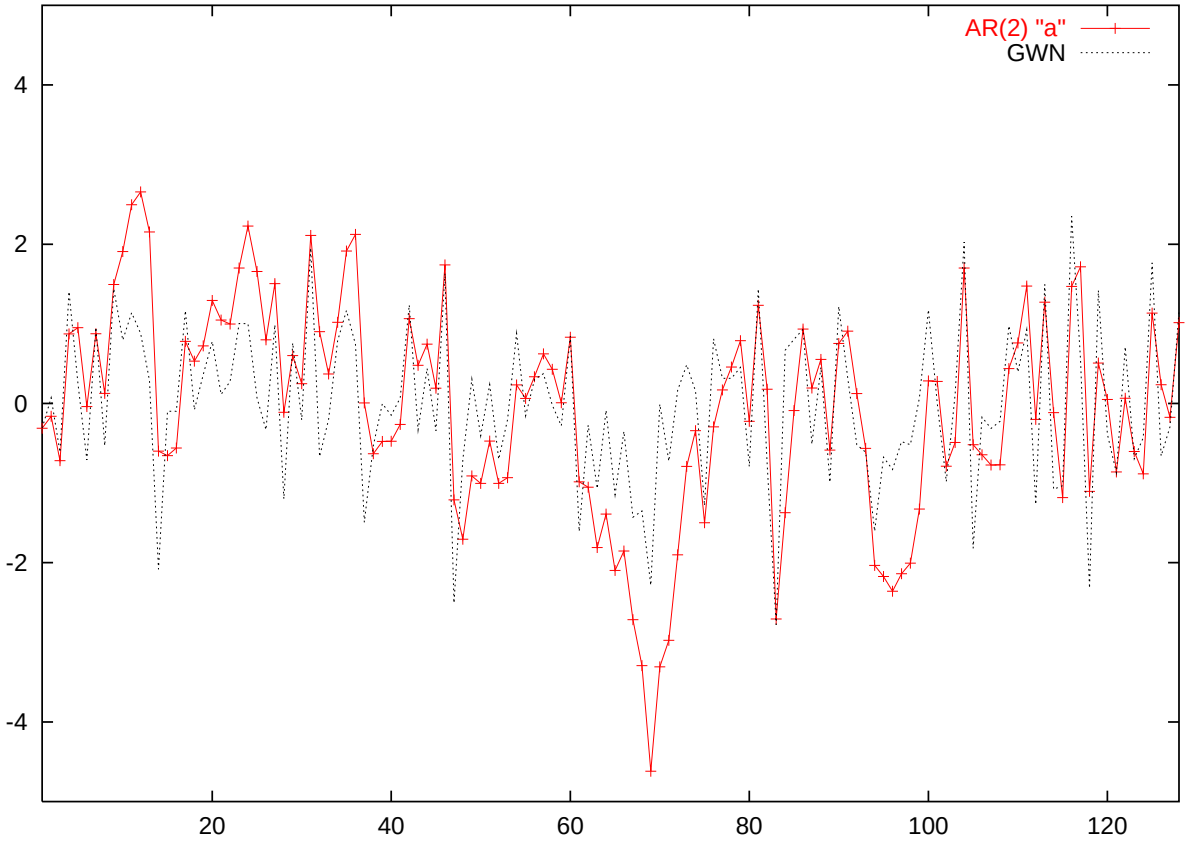
$$y_n = \beta_1 y_{n-1} + \beta_2 y_{n-2} + \alpha_0 \eta_n \tag{17}$$

$$\rho_i = \beta_1 \rho_{i-1} + \beta_2 \rho_{i-2} \tag{18}$$

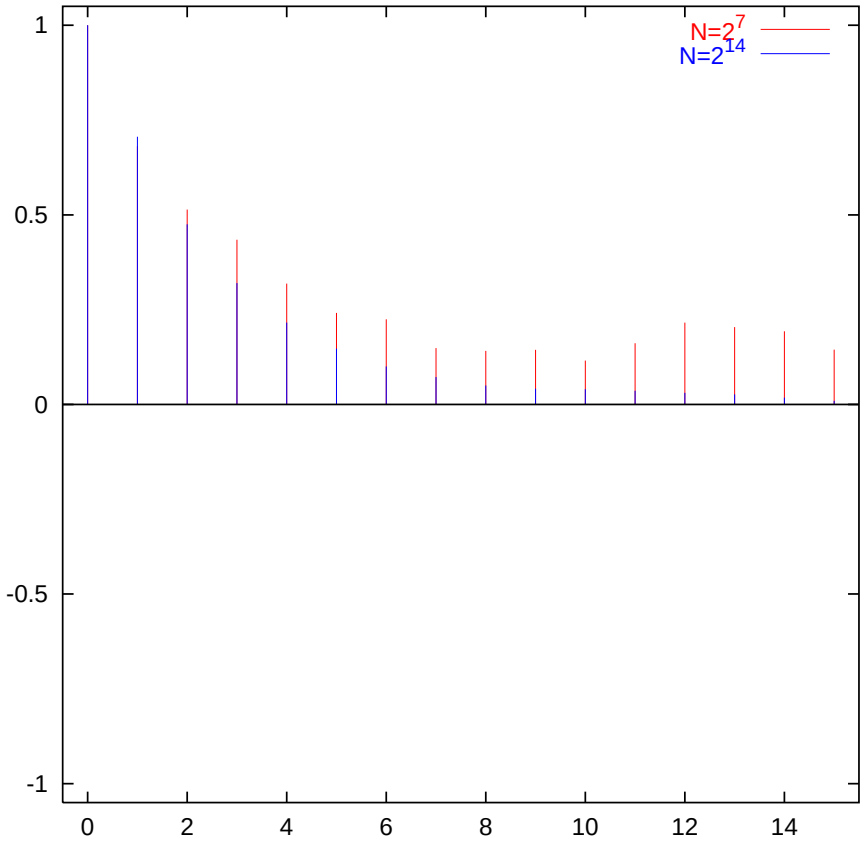
The solution to (18) is either a sum of two vanishing exponentials, or damped oscillations.

Copyright © 2009-21 P. F. Góra

$$y_n = 0.75 y_{n-1} - 0.05 y_{n-2} + \eta_n$$

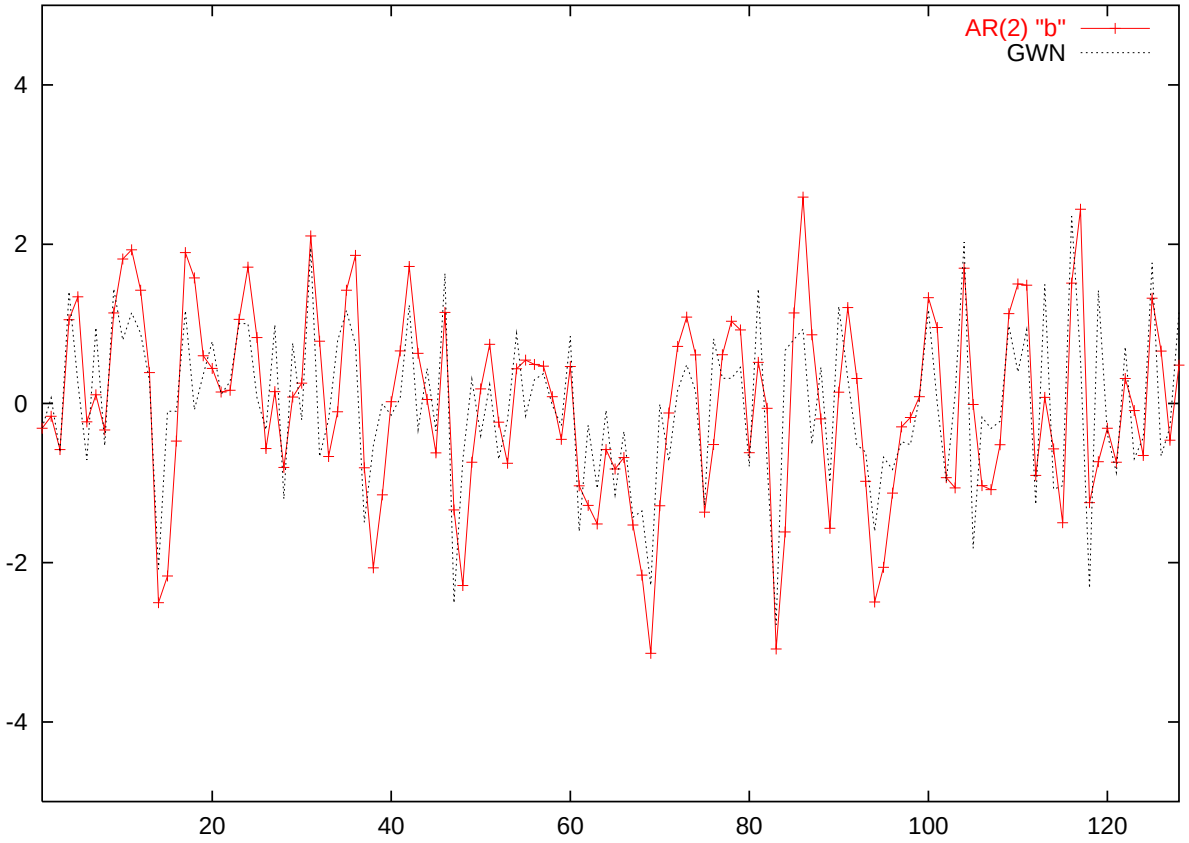


Copyright © 2009-21 P. F. Góra

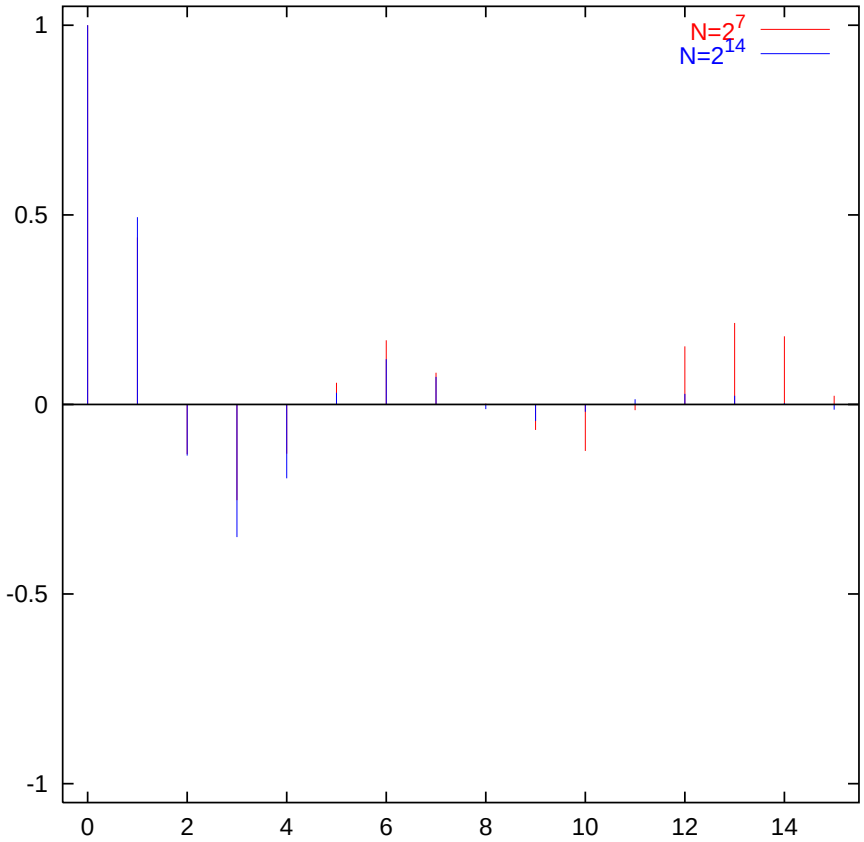


Copyright © 2009-21 P. F. Góra

$$y_n = 0.75 y_{n-1} - 0.5 y_{n-2} + \eta_n$$



Copyright © 2009-21 P. F. Góra



Copyright © 2009-21 P. F. Góra

Yule-Walker equations

Lets put $m = 1$ in Eq. (7):

$$\rho_1 = \beta_1 \rho_{1-1} + \beta_2 \rho_{1-2} + \ldots \beta_p \rho_{1-p} = \beta_1 + \beta_2 \rho_1 + \ldots \beta_p \rho_{p-1}\,, \qquad (19)$$

$\rho_j = \rho_{-j}$ by stationarity. If we do so for $m = 1, \ldots, p$, we get *Yule-Walker equations*:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}. \qquad (20)$$

Copyright © 2009-21 P. F. Góra

Theoretically, if we know the correlation function ρ_m, we can solve (20) for the parameters of the process, β_j. *In reality*, we do not know the autocorrelation. All we have is the "experimental" autocorrelation, calculated from the single realization available:

$$r_m = \left(\frac{1}{N-m} \sum_{i=m+1}^{N} y_{i-m} y_i \right) \Big/ \left(\frac{1}{N} \sum_{i=1}^{N} y_i^2 \right) . \tag{21}$$

If we substitute r_m for ρ_m in Yule-Walker equations, we can calculate *approximate* values of the coefficients β_i.

Note that the matrix in (20) is symmetrix an (is supposed to be) positive-definite. In terms of numerical linear algebra, this matrix is also *small* and there is no need for using algorithms tailored for large matrices.

Copyright © 2009-21 P. F. Góra

Partial correlation

There is one serious problem: How do we know the order of the process, or the dimension of (20)?

We can formally extend the Yule-Walker equation to the next order of correlations by adding a row and a column:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{p-1} & \rho_p \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{p-2} & \rho_{p-1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \rho_p & \rho_{p-1} & \rho_{p-2} & \cdots & \rho_1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p+1} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_{p+1} \end{bmatrix}. \tag{22}$$

Is it possible that $\beta_1, \ldots, \beta_p$ calculated either from (20) or from (22) are equal? Yes, if and only if $\beta_{p+1} = 0$.

Copyright © 2009-21 P. F. Góra

More formally, the autocorrelation of an AR(p) process has an infinite power series expansion but it depends on p linearly independent functions only. Lets assume that our process has an order k. Then Yule-Walker equations take the form

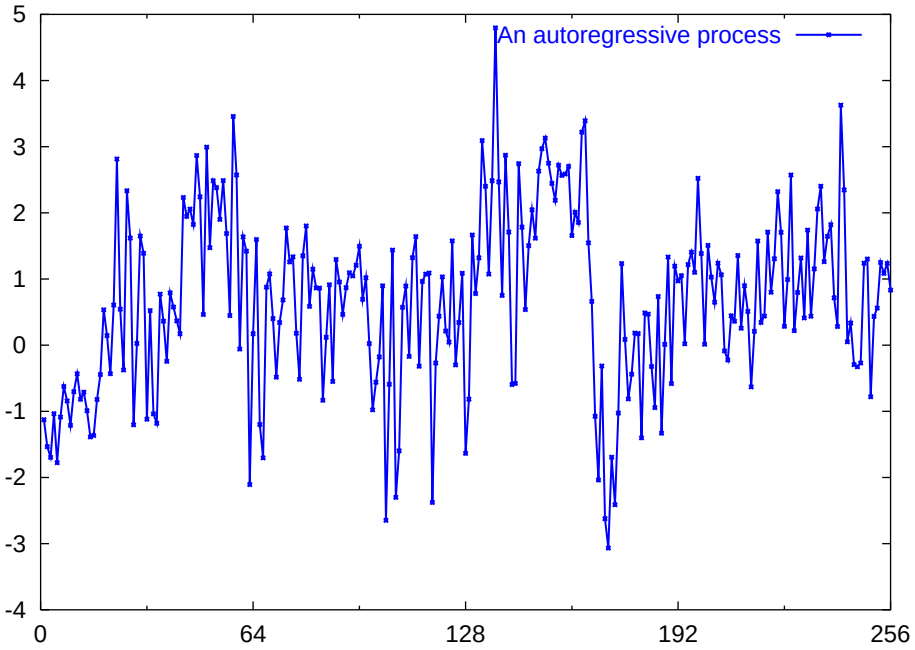
$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \varphi_{k1} \\ \varphi_{k2} \\ \vdots \\ \varphi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}. \tag{23}$$

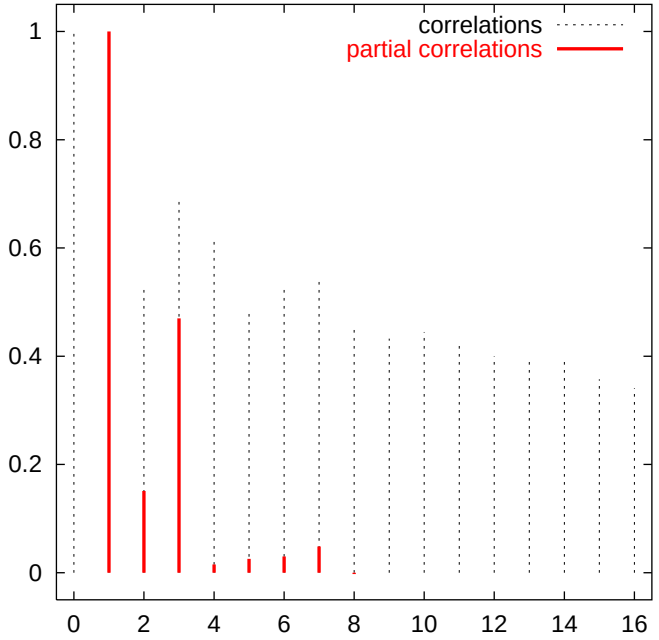
φ_{kk} is called a partial autocorrelation function.

For an AR(p) process, $\varphi_{kk} \neq 0$ for $k \leqslant p$ and $\varphi_{kk} \equiv 0$ for $k > p$. This suggests a procedure: increase the number of equations in (23) as long as $\varphi_{kk} \neq 0$. If $\varphi_{k'k'} = 0$, then $p = k' - 1$.

Copyright © 2009-21 P. F. Góra

Example





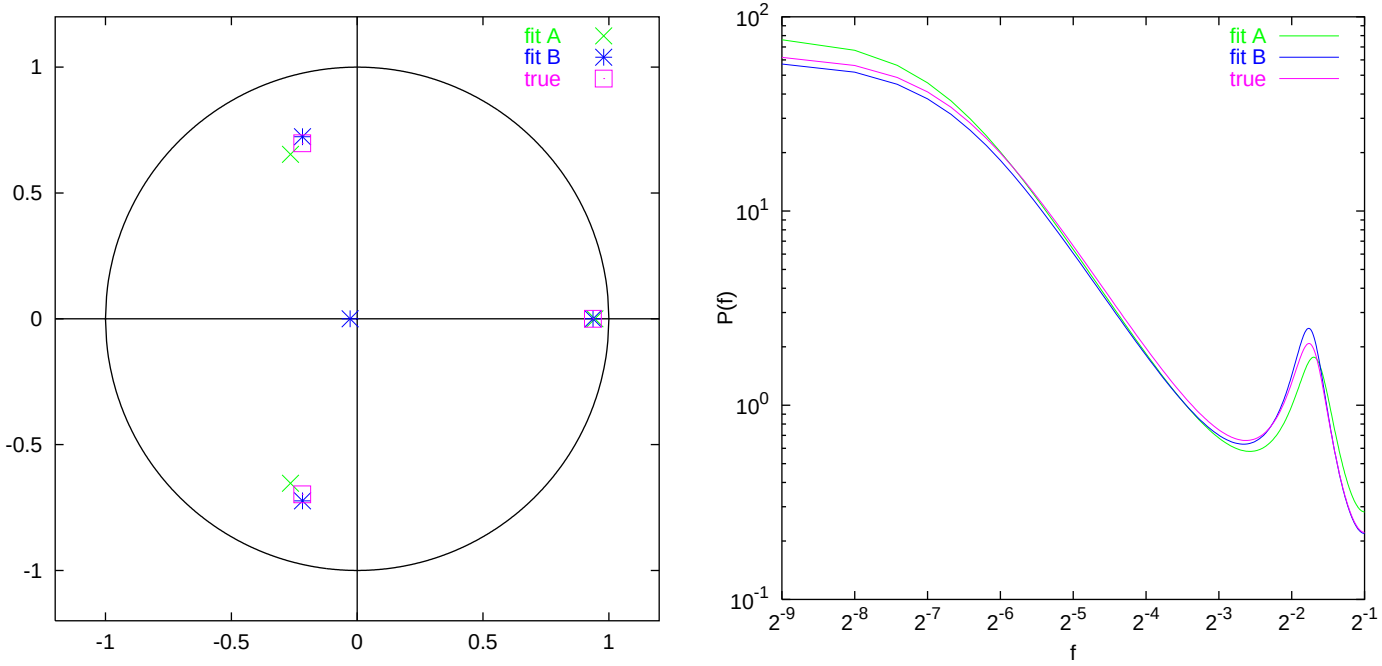
Estimated parameters: (A) $p = 3$: $\beta_1 = 0.415$, $\beta_2 = \quad 0.003$, $\beta_3 = 0.470$

(B) $p = 4$: $\beta_1 = 0.474$, $\beta_2 = -0.149$, $\beta_3 = 0.530$, $\beta_4 = 0.015$

"True" parameters: $p = 3$: $\beta_1 = 0.500$, $\beta_2 = -0.125$, $\beta_3 = 0.500$

Copyright © 2009-21 P. F. Góra

How much do these models *really* differ?



Reciprocals of roots of Eq. (5) (left) and the corresponding power spectra (right) for the models presented on the preceding page.

Copyright © 2009-21 P. F. Góra

Akaike Information Criterion

Sometimes it is not absolutely clear which partial correlation is *practically* zero and which is not. Several criteria have been proposed to decide in such cases.

It is obvious that the more parameters, or the higher the range of the AR(p) process, the better the fit to the actual data. However, models with too many parameters are "bad". Therefore, Hirotugu Akaike has proposed a criterion that pays for a better fit, but penalizes for too many parameters:

Copyright © 2009-21 P. F. Góra

$$AIC = \ln Q + \frac{2p}{N} \tag{24}$$

where p is the number of parameters, N is the sample size, and Q is the residual error

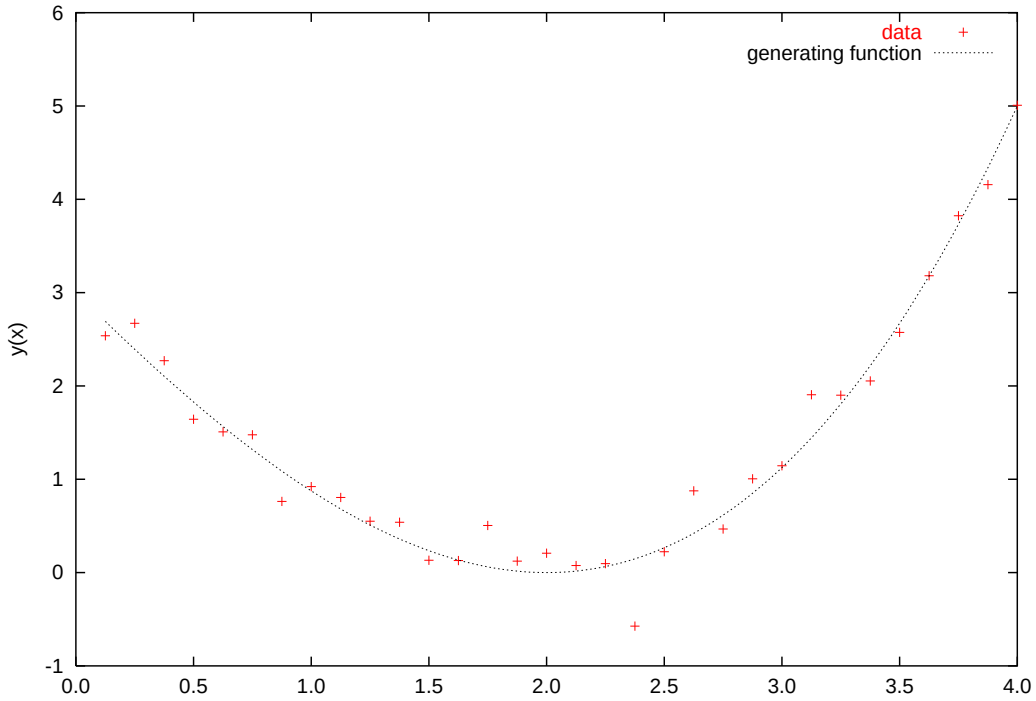
$$Q = \sum_{n=0}^{N-1} \left(y_n - \sum_{j=1}^{p} \beta_j^{(p)} y_{n-j} \right)^2 \tag{25}$$

where $\beta_j^{(p)}$'s are the parameters fitted *under the assumption* that the process has the order p.

Note: The residual error is calculated from the least squares because the noise is GWN.

Copyright © 2009-21 P. F. Góra

Akaike criterion in “ordinary” least squares

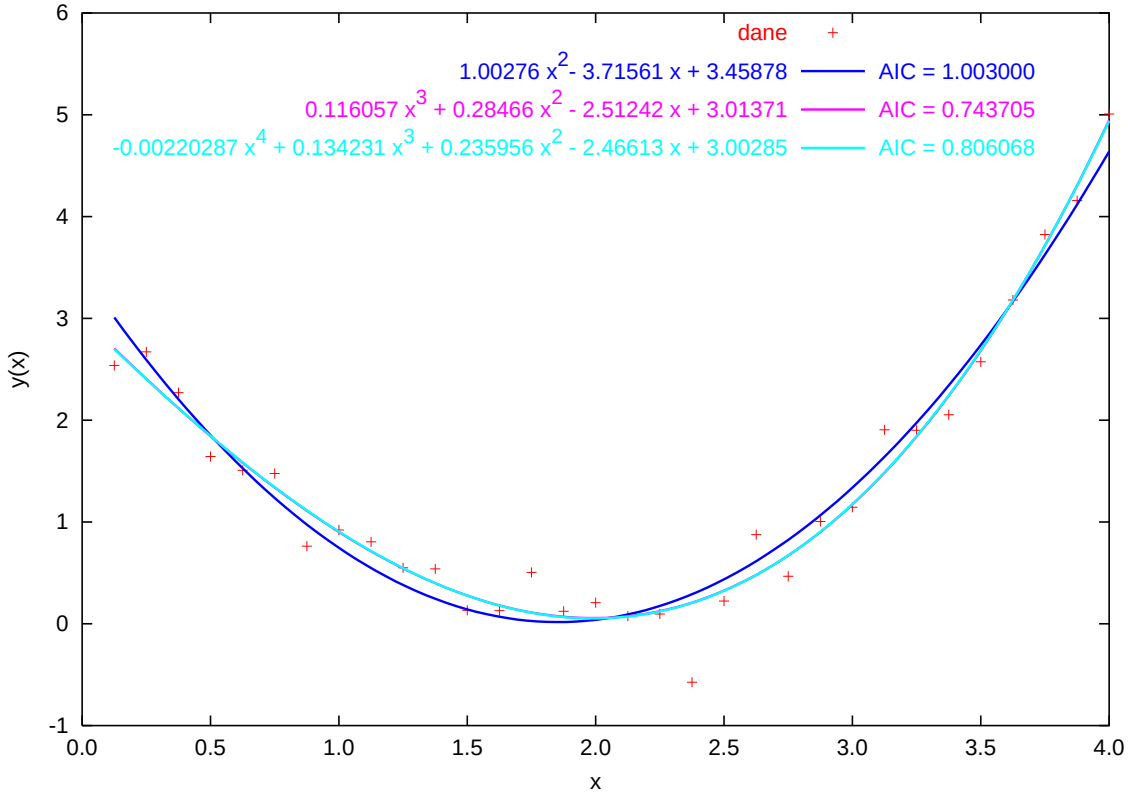


Data points generated according to

$$y_n = 0.125\,x_n^3 + 0.25\,x_n^2 - 2.5\,x_n + 3.0 + 0.5\,\eta_n$$

Copyright © 2009-21 P. F. Góra

Results of the fits



Fitted curves of the third and fourth orders are indistinguishable in the plotted range.

Copyright © 2009-21 P. F. Góra

How do we estimate the noise level?

We can do this from the power spectrum, in particular, from the power spectrum at zero frequency. From (10) we get

$$P(0) = \frac{\alpha_0^2}{\left|1 - \sum_{n=1}^{p} \beta_n\right|^2}. \tag{26}$$

Calculating α_0^2 is now straightforward. Note that the denominator of the above equation cannot vanish as all roots of Eq. (5) must lie *outside* the unit circle. However, a numerically calculated power spectrum carries some error with it. For low order processes, we can calculate the noise level without calculating the power spectrum first.

Copyright © 2009-21 P. F. Góra

Noise level in AR(1) process

For an AR(1) process

$$y_n = \beta_1 y_{n-1} + \alpha_0 \eta_n \tag{27}$$

we calculate

$$\langle \eta_n y_n \rangle = \beta_1 \langle \eta_n y_{n-1} \rangle + \alpha_0 \left\langle \eta_n^2 \right\rangle \tag{28}$$

$\langle \eta_n y_{n-1} \rangle = 0$ because y_{n-1} cannot depend on *future* noises. $\left\langle \eta_n^2 \right\rangle = 1$.

Further,

$$\left\langle y_n^2 \right\rangle = \beta_1 \langle y_n y_{n-1} \rangle + \alpha_0 \langle \eta_n y_n \rangle = \beta_1 \langle y_n y_{n-1} \rangle + \alpha_0^2 \tag{29}$$

Copyright © 2009-21 P. F. Góra

Finally

$$\langle y_n y_{n-1} \rangle = \beta_1 \left\langle y_{n-1}^2 \right\rangle + \alpha_0 \langle \eta_n y_{n-1} \rangle \tag{30}$$

The last term vanishes as before. By means of stationarity, $\left\langle y_{n-1}^2 \right\rangle = \left\langle y_n^2 \right\rangle$ and eventually

$$\alpha_0^2 = (1 - \beta_1^2) \left\langle y_n^2 \right\rangle \tag{31}$$

Copyright © 2009-21 P. F. Góra

Noise level in AR(2) process

For an AR(2) process we do as above, use the fact that $\langle y_{n-1}y_{n-2}\rangle = \langle y_n y_{n-1}\rangle$ and finally obtain

$$\alpha_0^2 = \left(1 - \frac{1+\beta_2}{1-\beta_2}\beta_1^2 - \beta_2^2\right)\left\langle y_n^2\right\rangle \tag{32}$$

The benefit of formulas like (31), (32) is that the noise level can be estimated directly from the variance of the process.

Copyright © 2009-21 P. F. Góra

What is it all good for?

We try to fit a stochastic model because

1. we want to get insight into the mechanism that has generated the process,
2. we want to do the forecasting.

Copyright © 2009-21 P. F. Góra

How do we forecast?

Having collected the time series $\{y_n\}_{n=0}^{N-1}$, and having fitted an AR(p) model (4)

$$y_n = \beta_1 y_{n-1} + \beta_2 y_{n-2} + \cdots + \beta_p y_{n-p} + \alpha_0 \eta_n$$

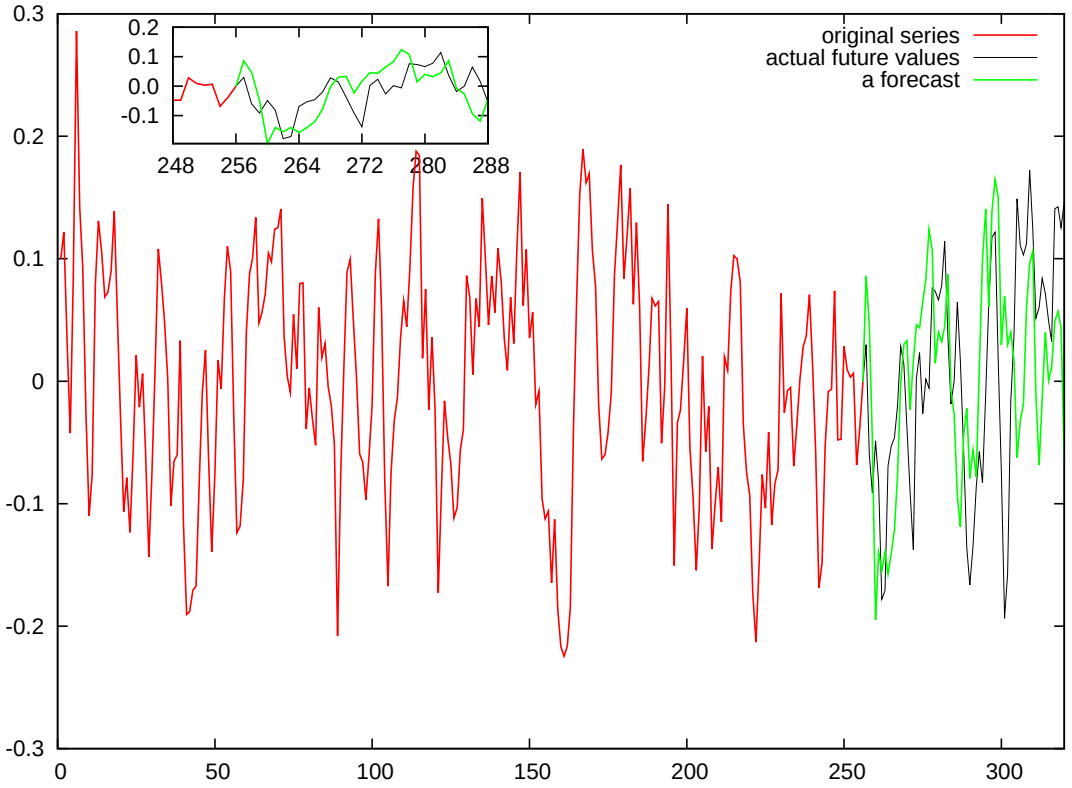
we generate the unknown, future noises, from a random number generator. We thus obtain a different realization of the process:

$$\begin{aligned}
\widehat{y}_N &= \beta_1 y_{N-1} + \beta_2 y_{N-2} + \cdots + \beta_p y_{N-p} + \alpha_0 \eta_N && \text{(33a)} \\
\widehat{y}_{N+1} &= \beta_1 \widehat{y}_N + \beta_2 y_{N-1} + \cdots + \beta_p y_{N-p+1} + \alpha_0 \eta_{N+1} && \text{(33b)} \\
\widehat{y}_{N+2} &= \beta_1 \widehat{y}_{N+1} + \beta_2 \widehat{y}_N + \cdots + \beta_p y_{N-p+2} + \alpha_0 \eta_{N+2} && \text{(33c)} \\
&\cdots
\end{aligned}$$

$\widehat{y}_{k \geqslant N}$ are the forecasts.

Copyright © 2009-21 P. F. Góra

Example of a forecast



Copyright © 2009-21 P. F. Góra

Note

If you train your model on a known series and want to compare the forecasts with actual values, you must *never* fit the parameters using values that will be forecasted.

Copyright © 2009-21 P. F. Góra