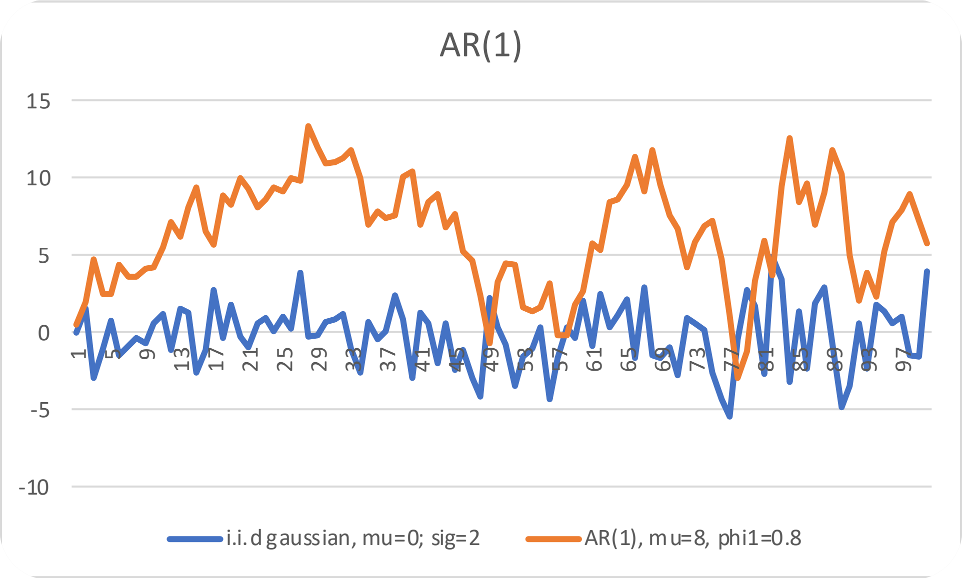
PLATEFORME DE CONSTRUCTIONS HYDRAULIQUES

Water Resources Engineering and Management

(CIVIL-466, A.Y. 2024-2025)

5 ETCS, Master course



Lecture 7-1: Univariate linear Autoregressive Model

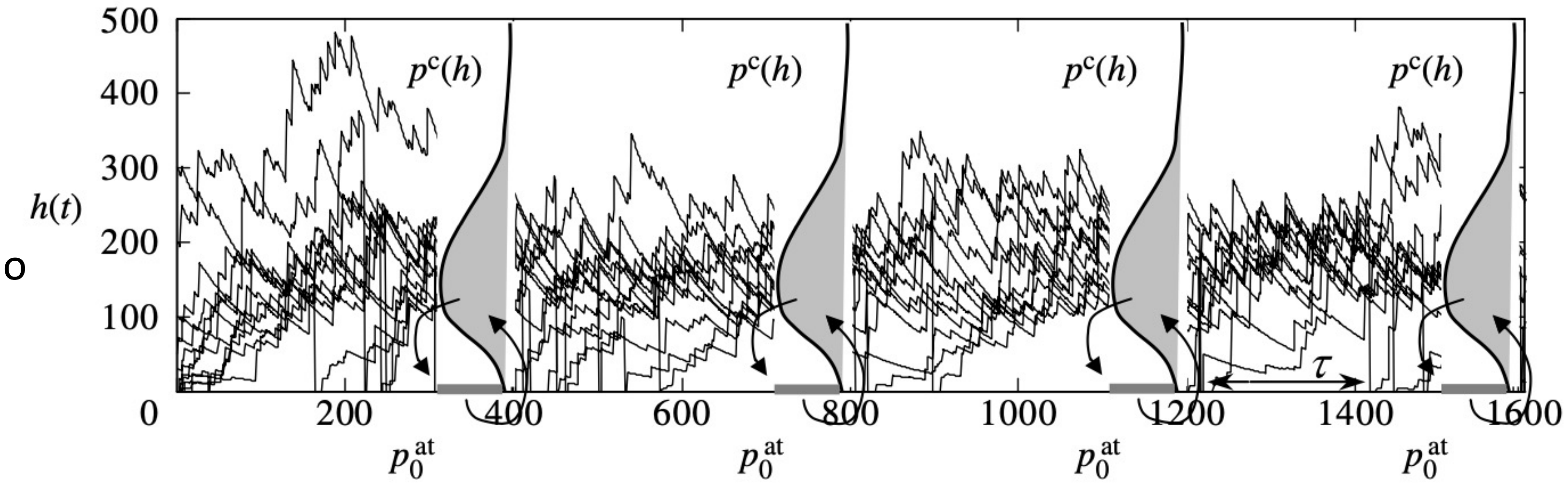
Prof. P. Perona
Platform of hydraulic constructions

EPFL

Stochastic modelling of time series

A FEW IMPORTANT CONCEPTs: we assume the time series to be modeled is stationary and ergodig, hence the available data is representative of the full variability of the population universe

STATIONARITY: Implies that the statistical properties of the series (and of the model representing it) do not change with time



ERGODICITY: Implies that one single realisation is long enough to be representative of the whole ensemble

NOTA BENE: ergodicity cannot be stricktly proved with measured time series. This implies that it is assumed as a working hypothesis for which

Sample mean = ensemble mean
Sample variance = ensemble average …and so on

Autoregressive (Linear) models

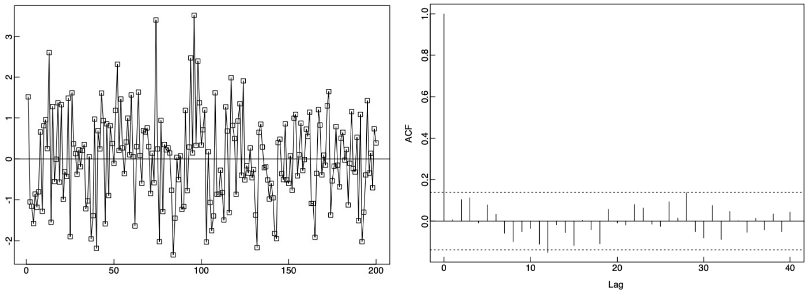
Linear Autoregressive models

Consider the following mathematical model

$$y_t = \mu + \sum_{j=1}^{p} \phi_j (y_{t-j} - \mu) + \varepsilon_t$$

This is called AR(p) model, i.e. AutoRegressive model of order p with $0 \leq \phi_j < 1$ (when $|\phi_j|$ ≥1 the model is non stationary and diverges)

ε_t is the **noise** term (or innovation) , purely random, i.i.d and gaussian distributed with zero mean and variance σ_ε^2



MODEL STATISTICS: are all analytically exact

y_t is also normally distributed

$E(y) = \mu$ Expected mean of the model

$\text{Var}(y) = \sigma^2 = \frac{\sigma_\varepsilon^2}{\left(1 - \sum_{j=1}^{p} \phi_j \rho_j\right)}$ Variance of the model

$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \cdots + \phi_p \rho_{k-p}$ Autocorrelation function (Yule-Walker equation)

This vector is called "partial autocorrelation function"

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & & \rho_{p-2} \\ \vdots & \vdots & & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}$$

AR(1) model

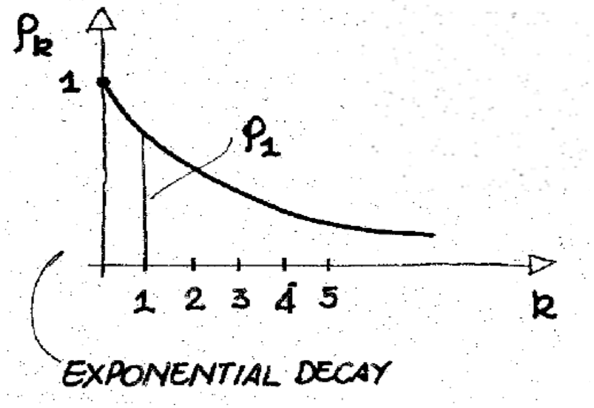
The Autoregressive model order 1 is (p=1)

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \varepsilon_t$$

$$E(y) = \mu$$

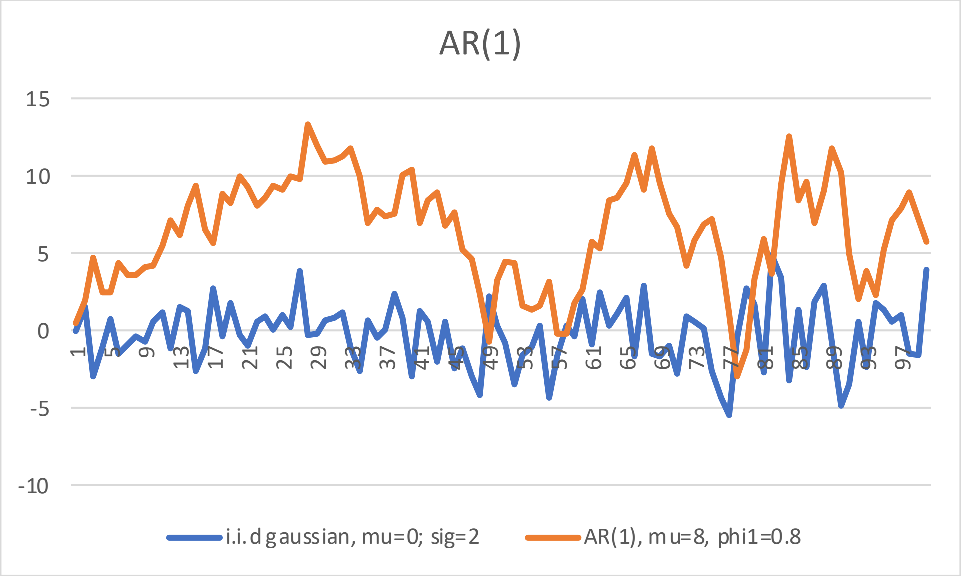
$$\text{Var}(y) = \sigma^2 = \frac{\sigma_\varepsilon^2}{(1-\phi_1\rho_1)}$$

$$\rho_k = \phi_1\rho_{k-1} = \phi_1^k$$



Hence

$$\text{Var}(y) = \sigma^2 = \frac{\sigma_\varepsilon^2}{(1-\rho_1^2)} = \frac{\sigma_\varepsilon^2}{(1-\phi_1^2)}$$



Hence, from these statistics it is also easy to identify the model from statistical data properties

lag

	0	1	2	3	4
ρ_k	1	0.8	0.64	0.512	0.409
r_k	1	0.81577	0.6287	0.50043	0.4246

AR(1) model (example)

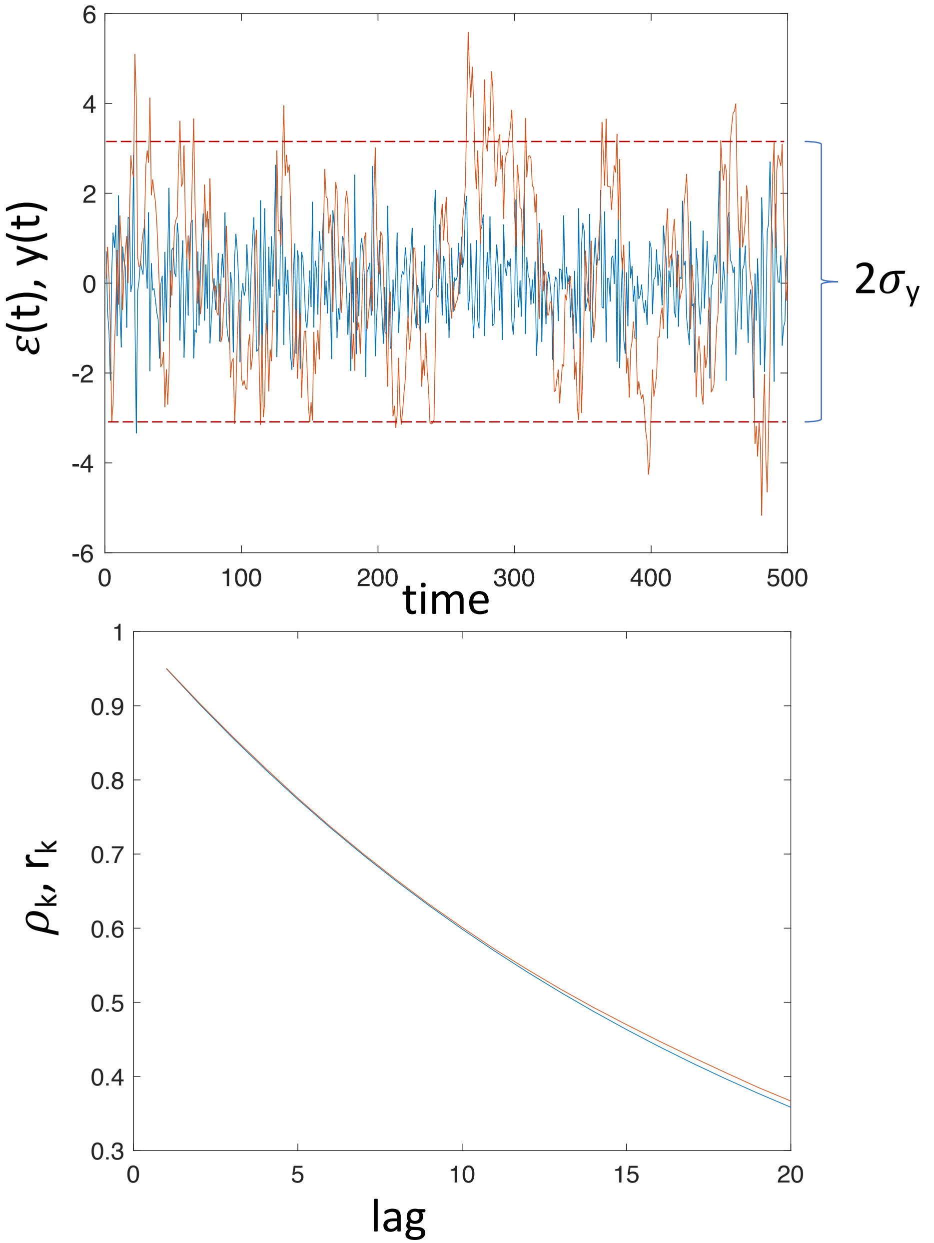
$\phi_1 = 0.95$

$E[\varepsilon] = 0$, $\sigma_\varepsilon{}^2$=1

E[y]=0 , $\sigma_y{}^2 = \frac{\sigma_\varepsilon^2}{(1-\phi_1^2)}$ = 10.25

y(t) is also normally distributed

Here shows something about what the variance of the generated series becomes when starting from a gaussian distributed noise with mean zero and std=1



Example: synthetic data generation with AR(1)

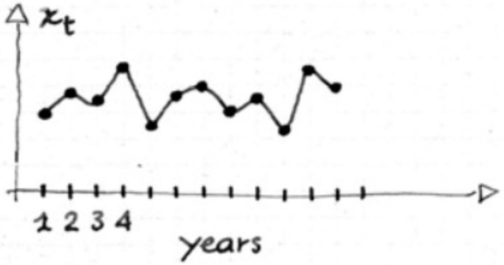
PLATEFORME DE CONSTRUCTIONS HYDRAULIQUES

VARIABLE: annual river flow, x_t

MEAN: 470 mm, $\bar{x}$

STANDARD DEVIATION: 95.8 mm, S_x

AUTOCORRELATION COEFF., LAG 1: 0.324, r_1



- GENERATION OF x_t FOR n years → t, t+1, t+2, t+3
- MODEL AR(1)
- $\tilde{x}_t = \dfrac{x_t - \bar{x}}{S_x}$ IS NORMAL DISTRIBUTED
- $x_{t-1} = \bar{x} = 470$ mm

$\tilde{x}_t$ is standardized and has therefore unitary variance, hence $\sigma_\varepsilon^2 = (1 - \rho_1^2)$

AR(1) → $\tilde{x}_t = \phi_1 \tilde{x}_{t-1} + \varepsilon_t$

GAUSSIAN, $E[\varepsilon_t] = 0$

$\sigma^2[\varepsilon_t] = \sigma_\varepsilon^2 = 1 - r_1^2$

$\phi_1 = r_1$

generating equation: $\tilde{x}_t = \phi_1 \tilde{x}_{t-1} + \sqrt{1-r_1^2}\,\tilde{\varepsilon}_t$

Standard gaussian variate

substituting the above assumptions

$$x_t = \bar{x}(1-r_1) + r_1 x_{t-1} + S_x \sqrt{1-r_1^2}\,\tilde{\varepsilon}_t$$

$$x_t = 470(1-0.324) + 0.324 \cdot x_{t-1} + 95.8\sqrt{1-0.324^2} \cdot \tilde{\varepsilon}_t$$

$\tilde{\varepsilon}_t$ RANDOMLY GENERATED

$\tilde{\varepsilon}_t = 0.87$

$\tilde{\varepsilon}_{t+1} = -0.65$

$\tilde{\varepsilon}_{t+2} = 1.15$

$\tilde{\varepsilon}_{t+3} = 0.05$

t → $x_t = 549$ mm

t+1 → $x_{t+1} = 437$ mm

t+2 → $x_{t+2} = 563$ mm

t+3 → $x_{t+3} = 505$ mm

Example: AR(2) model

The Autoregressive model order 2 is (p=2)

Stability conditions
1. $(\phi_1 + \phi_2) < 1$
2. $(\phi_2 - \phi_1) < 1$
3. $|\phi_2| < 1$

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \phi_2 (y_{t-2} - \mu) + \varepsilon_t$$

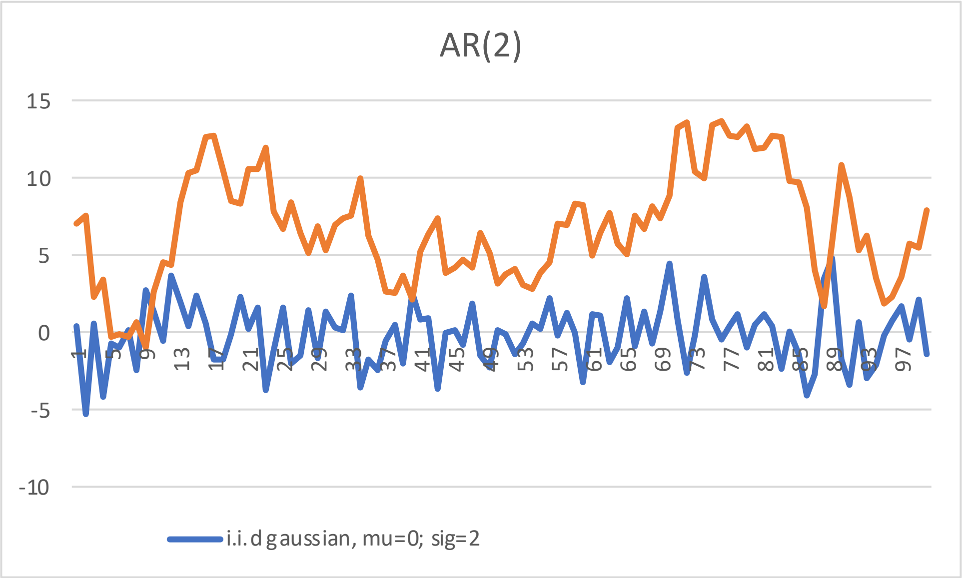
$$E(y) = \mu$$

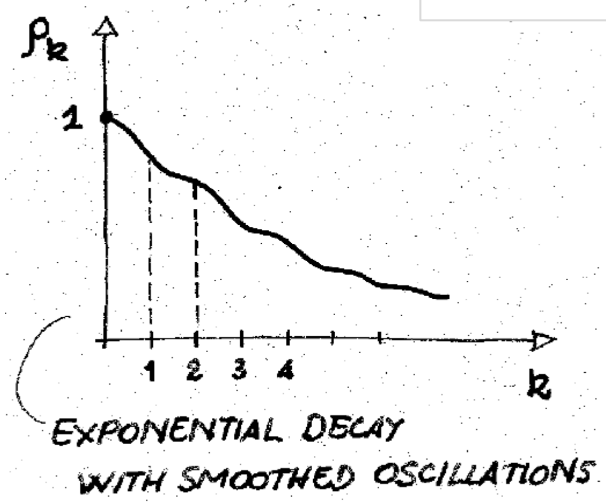
$$\text{Var}(y) = \sigma^2 = \frac{\sigma_\varepsilon^2}{(1-\phi_1\rho_1-\phi_2\rho_2)}$$

$$\rho_1 = \phi_1\rho_0 + \phi_2\rho_{-1} \qquad (\rho_{-k} = \rho_k)$$

$$\rho_2 = \phi_1\rho_1 + \phi_2\rho_0$$

$$\rho_1 = \phi_1(1 - \phi_2)$$
$$\rho_2 = \phi_2 + \frac{\phi_1^2}{1-\phi_2}$$





Parameter estimation can be done by means several approaches. The most obvious is the method of moments, where one replaces model statistics with sample statistics and solves for the parameters

Moving Average (Linear) models

Moving Average model – MA(q)

PLATEFORME DE CONSTRUCTIONS HYDRAULIQUES

Consider the following mathematical model

$$y_t = \mu + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

Despite the name "Moving Average", this model should not be confused with the homonymous smoothing technique!

This is called MA(q) model, i.e. Moving Average model of order q with $0 \le \theta_j \le 1$

ε_t is the noise term (or innovation) , purely random, i.i.d and gaussian distributed with zero mean and variance σ_ε^2

MODEL STATISTICS: are all analytical

$E(y) = \mu$

Expected mean of the model

$\text{Var}(y) = \sigma^2 = \sigma_\varepsilon^2\left(1 + \sum_{j=1}^{q} \theta_j^2\right)$

Expected variance does not depend on autocorrelation

$\rho_k = \dfrac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{\left(1 + \sum_{j=1}^{q} \theta_j^2\right)}$

Autocorrelation function depends on model parameters only

The MA model is always stationary and ρ_k=0 , k>q!!

Example: MA(1) model

PLATEFORME DE CONSTRUCTIONS HYDRAULIQUES

The Moving Average model order 1 is (q=1)

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

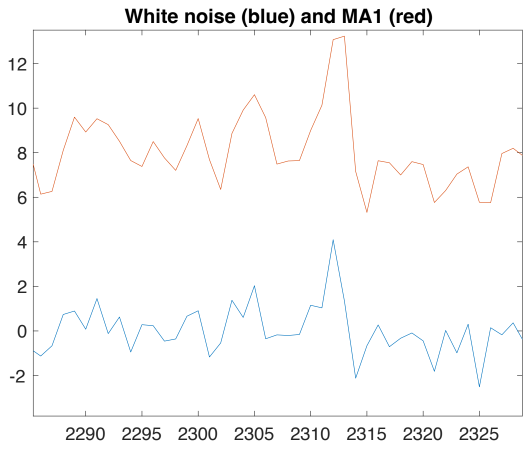
$$E(y) = \mu$$

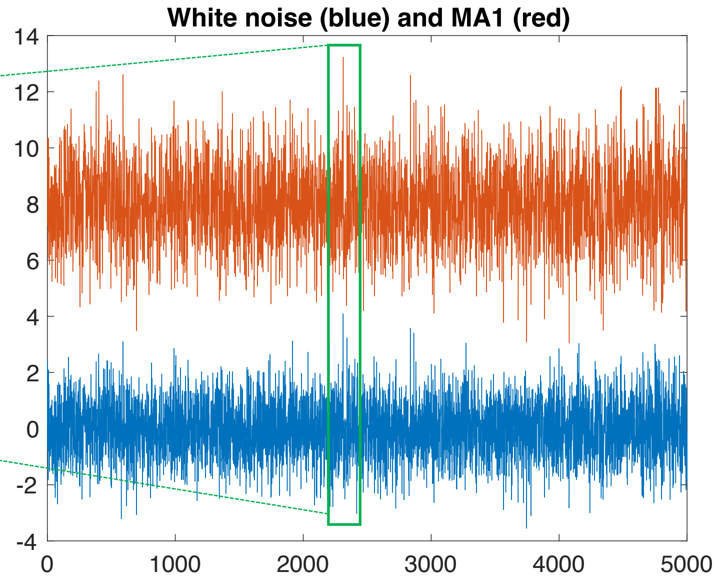
$$\text{Var}(y) = \sigma^2 = \sigma_\varepsilon^2 (1 + \theta_1^2)$$

$$\rho_1 = \frac{\theta_1}{(1 + \theta_1^2)}$$

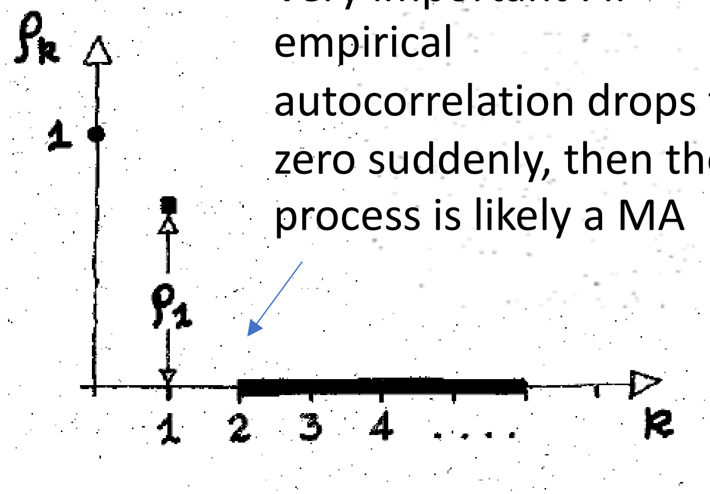
VERY IMPORTANT: ρ_k=0 , k>1

Max $(\rho_1) = 0.5$

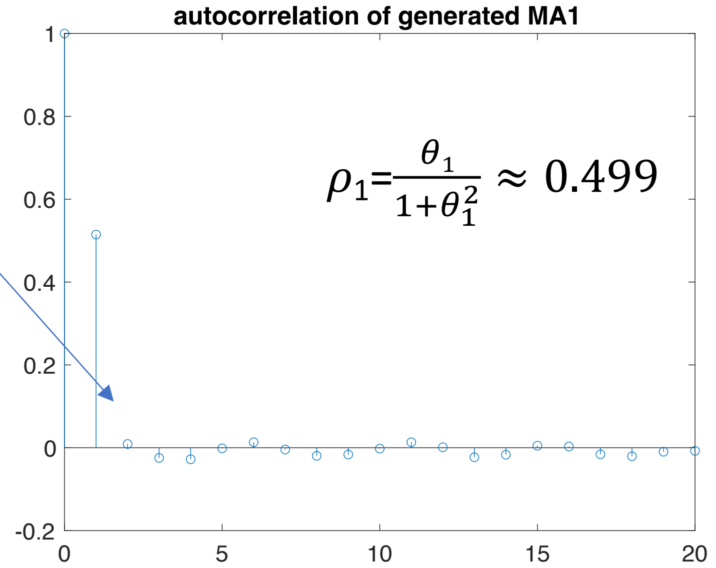




μ=8
θ_1=0.95



Very important : if empirical autocorrelation drops to zero suddenly, then the process is likely a MA



Example: MA(2) model

The Moving Average model order 2 is (q=2)

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

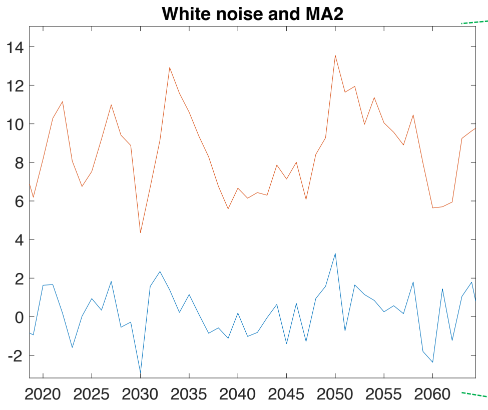
$$E(y) = \mu$$

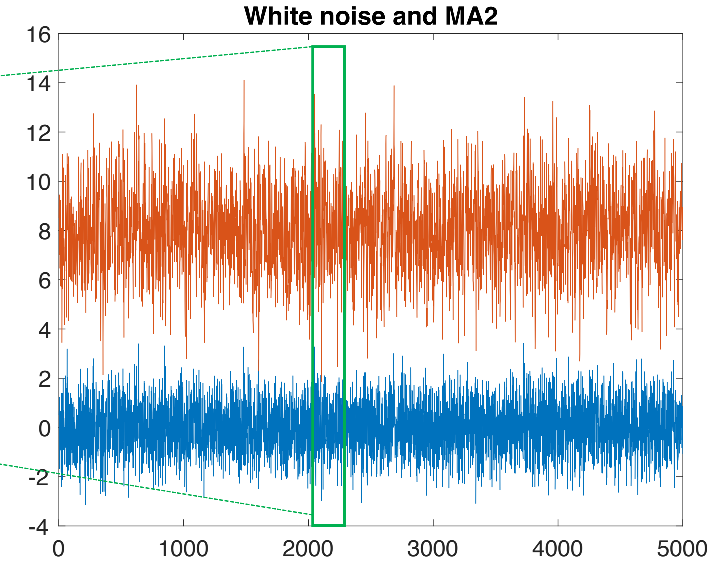
$$\text{Var}(y) = \sigma^2 = \sigma_\varepsilon^2 (1 + \theta_1^2 + \theta_2^2)$$

$$\rho_1 = \frac{\theta_1 (1 + \theta_2)}{(1 + \theta_1^2 + \theta_2^2)}$$

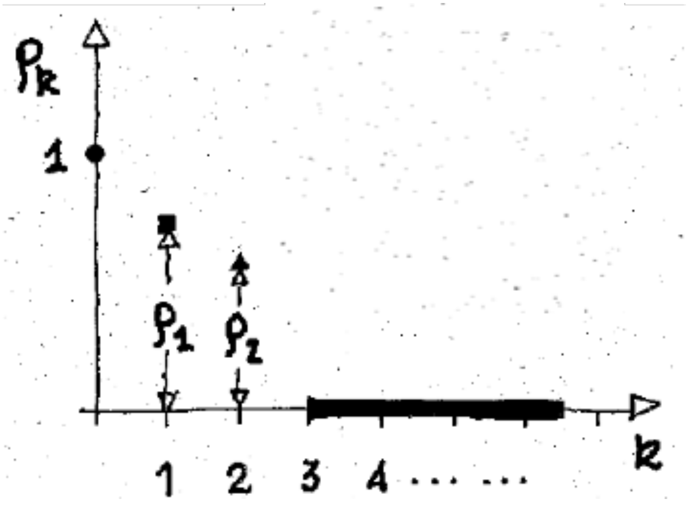
$$\rho_2 = \frac{\theta_2}{(1 + \theta_1^2 + \theta_2^2)}$$

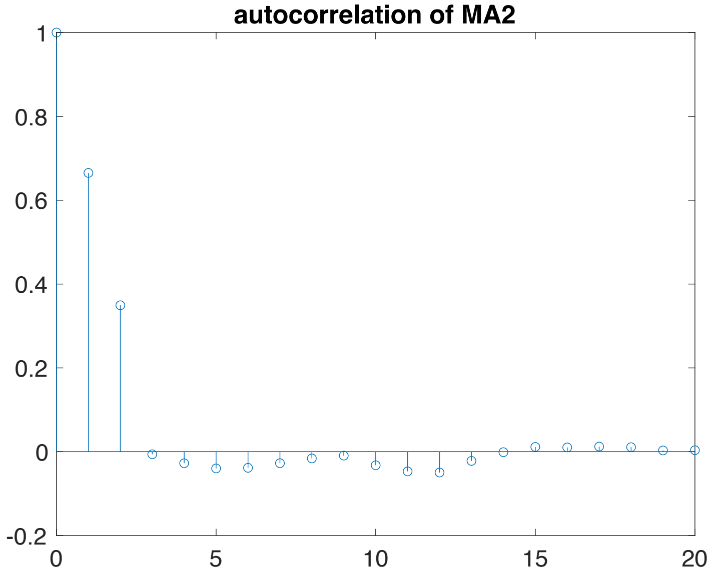
VERY IMPORTANT: ρ_k=0 , k>2





μ=8
θ_1=0.9
θ_2=0.9





AutoRegressive Moving Average (Linear) models

AutoRegressive Moving Average – ARMA(p,q) models

Consider the following mathematical model

$$y_t = \mu + \sum_{j=1}^{p} \phi_j (y_{t-j} - \mu) + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

This is called ARMA(p,q) model, i.e. AutoRegressive Moving Average model of order p,q

ε_t is the noise term (or innovation) , purely random, i.i.d and gaussian distributed with zero mean and variance σ_ε^2

$$\left.\begin{array}{l} u^p - \phi_1 u^{p-1} + \cdots - \phi_p = 0 \\ u^q - \theta_1 u^{q-1} + \cdots - \theta_q = 0 \end{array}\right\}$$

ARMA parameters must fulfill some mathematical constraints in order to ensure model stability. Such constraints are given by the characteristic equations

This means that the model parameters must all lie within the unit radius circle

Example: ARMA (1,1)

The ARMA (1,1) model is

$$y_t = \mu + \phi_1(y_{t-1} - \mu) + \varepsilon_t + \theta_1\varepsilon_{t-1}$$

Stability is guarranteed by the stability of the AR(1) process, i.e. $|\phi| < 1$

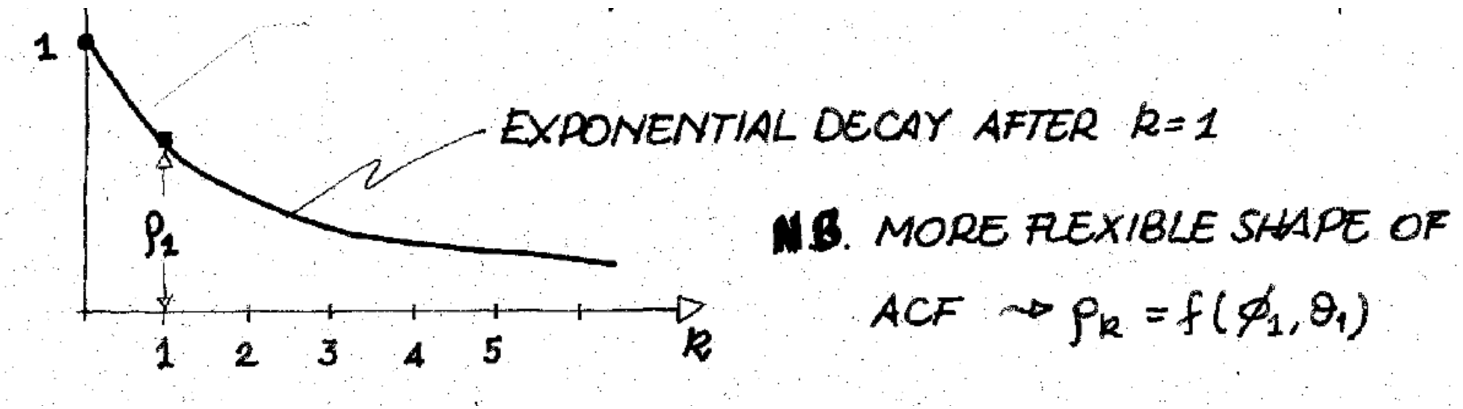
$$E(y) = \mu$$

$$\text{Var}(y) = \sigma^2 = \sigma_\varepsilon^2 \frac{(1+2\phi_1\theta_1+\theta_1^2)}{(1+\theta_1^2)}$$

$$\rho_1 = \frac{(\phi_1 + \theta_1)(1 + \theta_1\phi_1)}{(1 + 2\phi_1\theta_1 + \theta_1^2)}$$

$$\rho_k = \phi_1\rho_{k-1}$$

Again exponential decay caused by the AR(1) influence



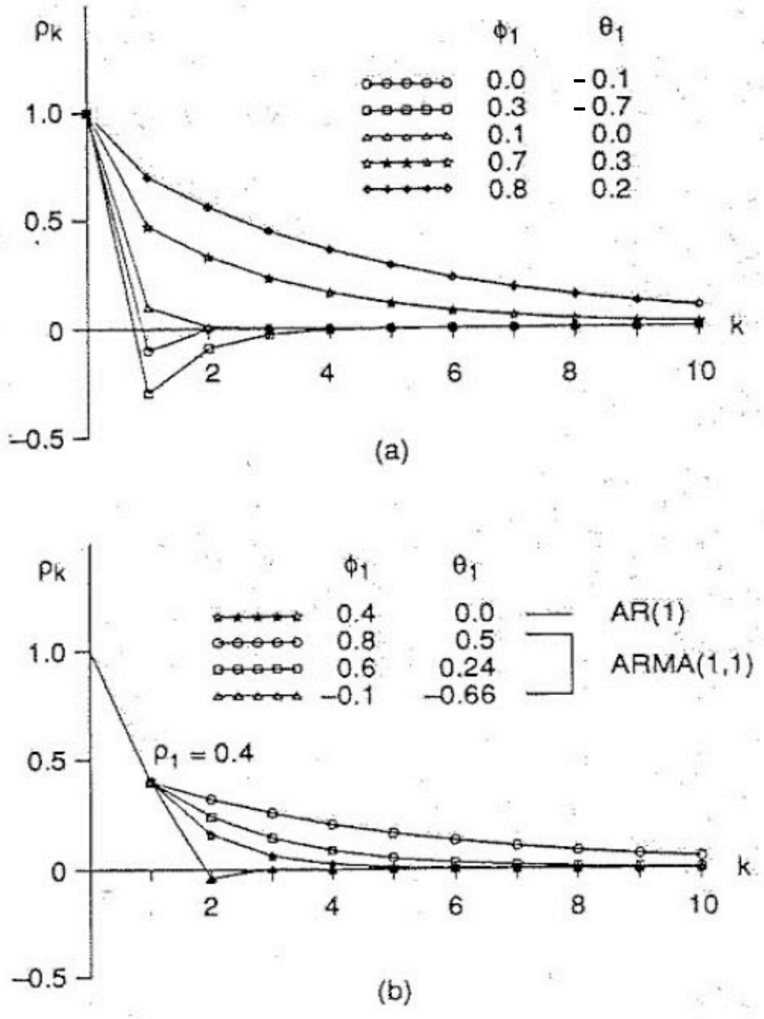


FIGURE 19.3.1 Correlograms ρ_k for (*a*) the ARMA(1, 1) process for various sets of parameters ϕ_1 and θ_1 and (*b*) the AR(1) and ARMA(1, 1) processes for which $\rho_1 = 0.4$.

ARMA processes are more flexible than **AR**

e.g. ←

- ρ_k of AR(1) $f(\phi_1)$
- ρ_k of ARMA(1,1) $f(\phi_1, \theta_1)$

AR models represent better short memory processes

ARMA models represent better long memory processes

Courtesy of P. Burlando