Apprentissage de modèles dynamiques

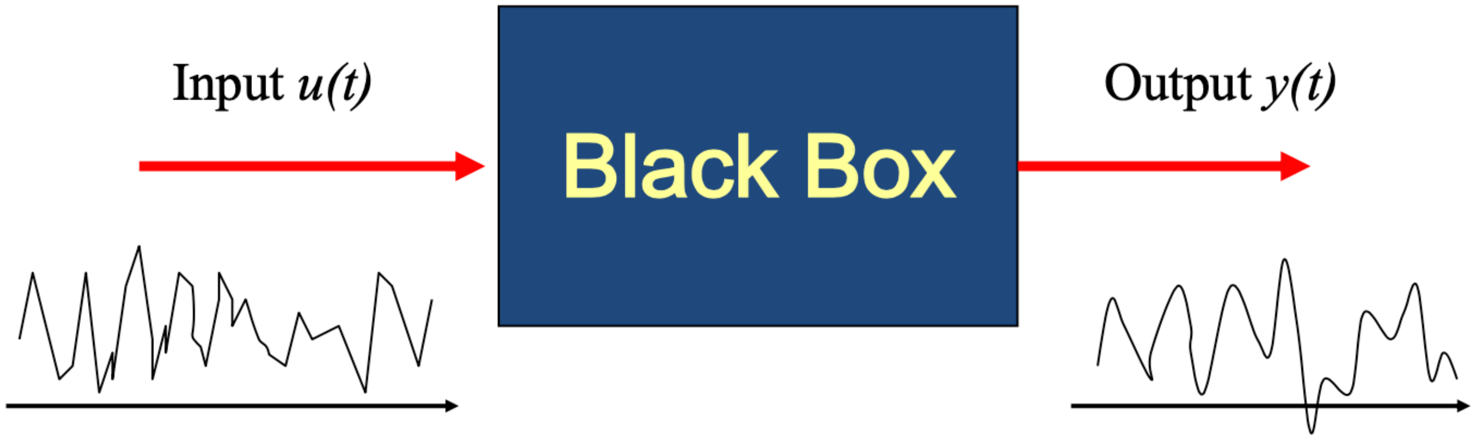
Models for disturbances: stochastic models

Hugues GARNIER

hugues.garnier@univ-lorraine.fr

Data-driven model learning = System identification

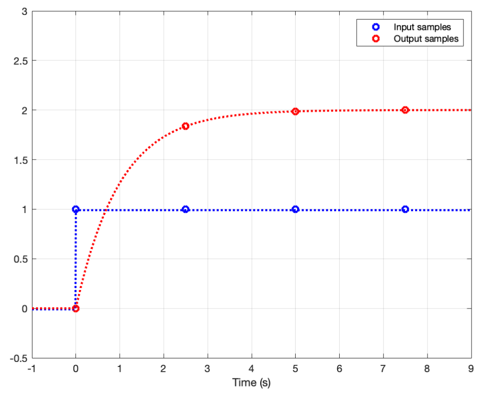
System Identification;
"Let the data speak about the system".



$$G(s)=\frac{Y(s)}{U(s)}=\frac{b_0s^m+b_1s^{m-1}+\cdots+b_m}{s^n+a_1s^{n-1}+\cdots+a_n}$$

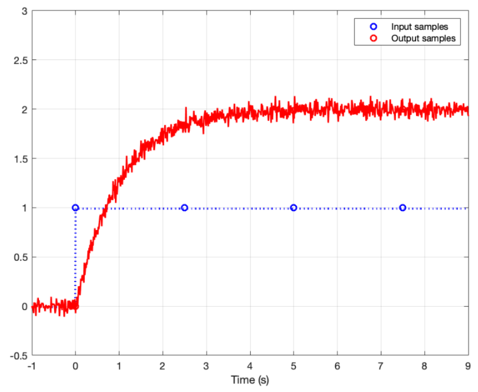
Find a model structure and determine parameter values that fit the data.

Use of simple linear regression for transfer function model learning – A brief recap



When the model structure is known

- In the ideal noise-free or deterministic case, use of simple least squares works fine !
  - the transfer function model parameter can be estimated by linear regression



- In practice, the simple least squares method breaks down
  - The output is not perfectly known. It is contaminated by measurement noise

$\Longrightarrow$ Incorrect least squares estimates (whatever the continuous or discrete-time model form)

How can we model the noisy measurement output ?

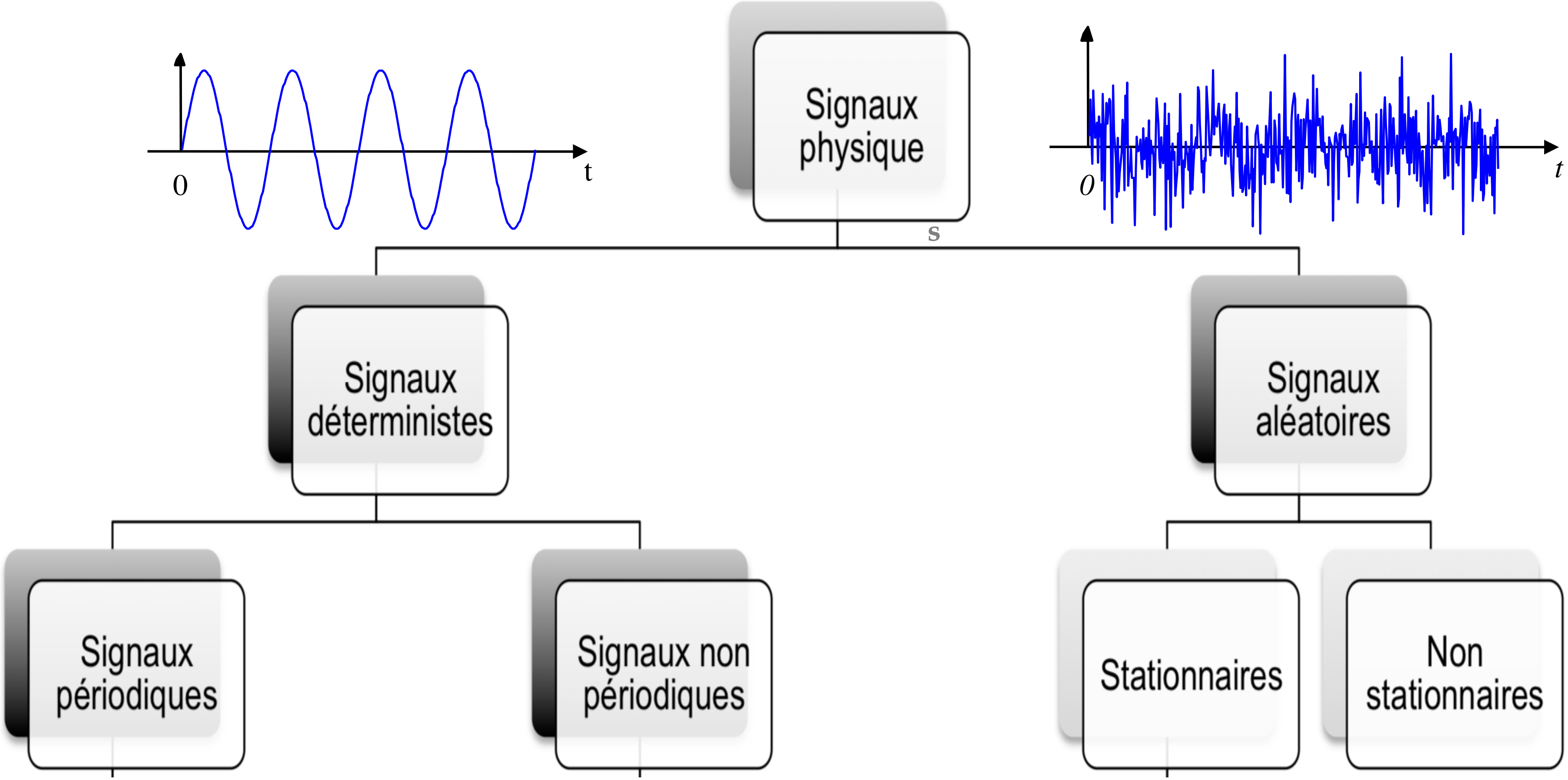
H. Garnier

Identifying measurement noise models

- In control design and estimation/prediction, it is often important to identify not only the dynamics from input to output, but also the measurement noise dynamics
  - how noise and disturbance perturb the system
  - where noise comes in
  - whether the noise is colored and correlated

- In system identification, various techniques are available for identifying both input-output dynamics and noise dynamics

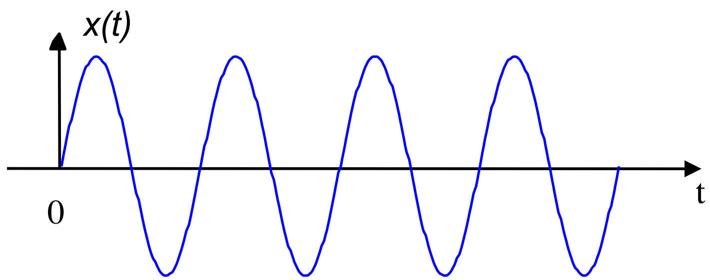
Classification of signals



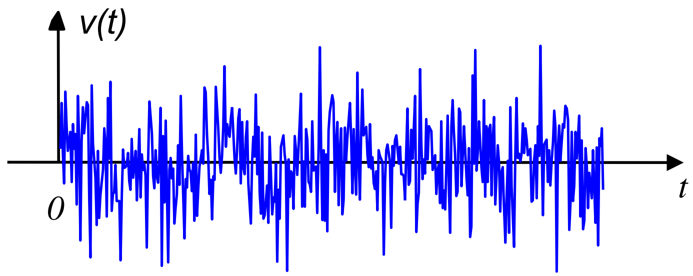
Model for noisy measurement output signals

- often modelled as

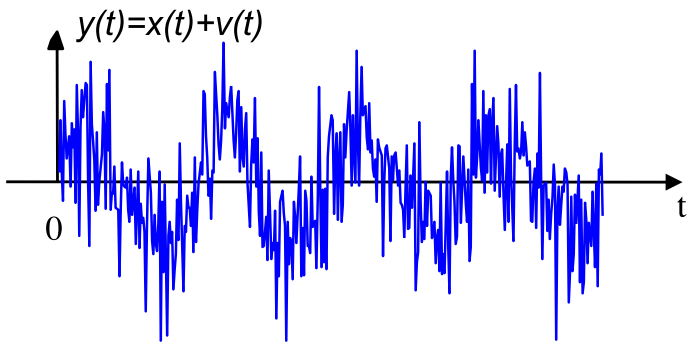
a deterministic signal



+ a stochastic signal

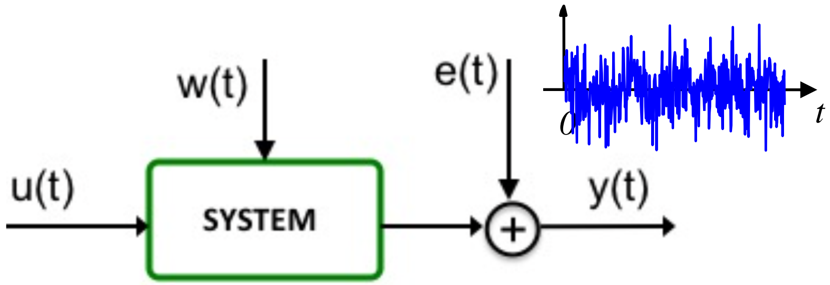


= noisy measurement signal



Models for disturbance and measurement noise

So far: only deterministic models



$u(t) =$ input

$w(t) =$ disturbance

$e(t) =$ measurement noise

$y(t) =$ output

- disturbance and measurement noise are stochastic signals
- Stochastic models:
  - means, covariances
  - spectra (energy or power)

Moments of a probability distribution
A brief review

- Moments of a random variable X with density $f_X(x)$:
  - *l-th moment*

$$m'_l = \mathrm{E}[X^l] = \int_{-\infty}^{\infty} x^l f_X(x)\,dx$$

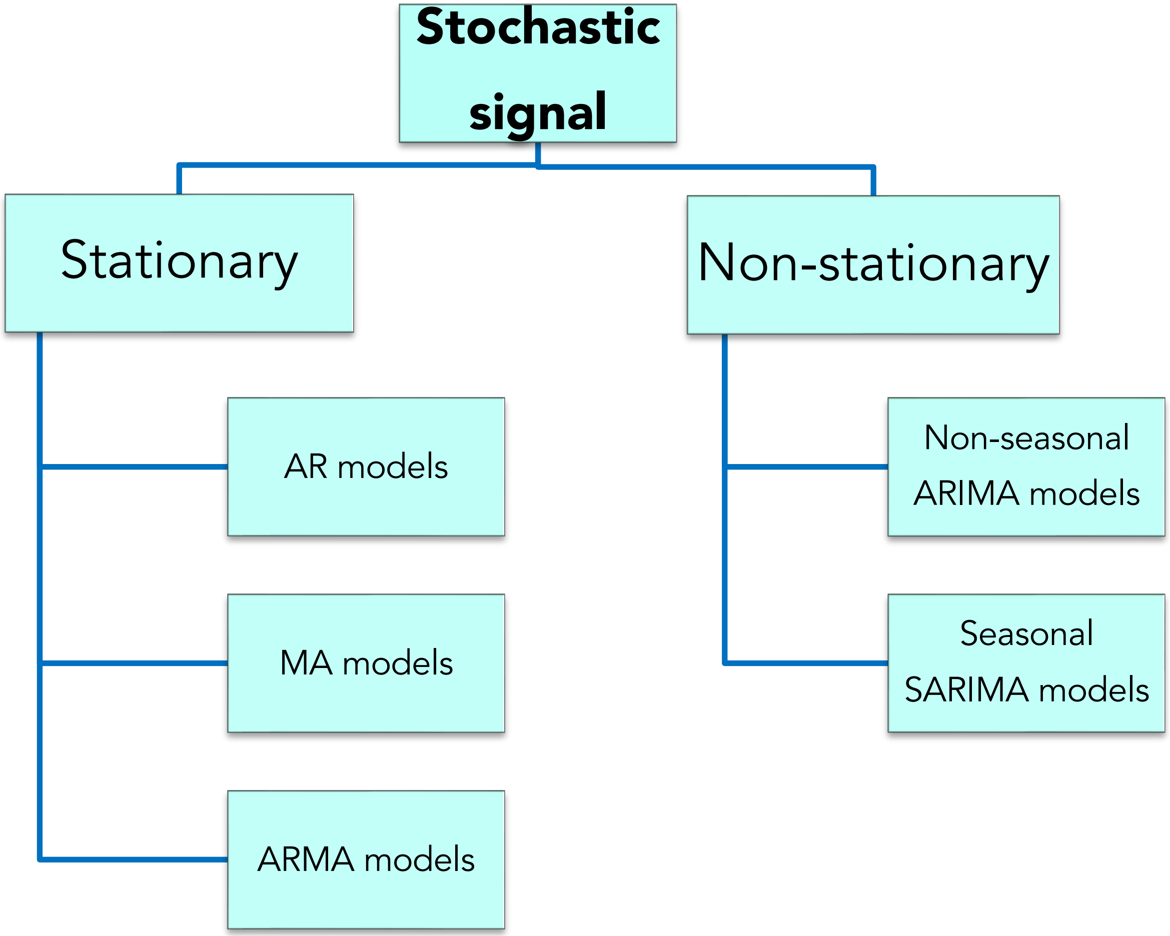
  - *l-th central moment*

$$m_l = \mathrm{E}[(X-\mu)^l] = \int_{-\infty}^{\infty} (x-\mu)^l f_X(x)\,dx$$

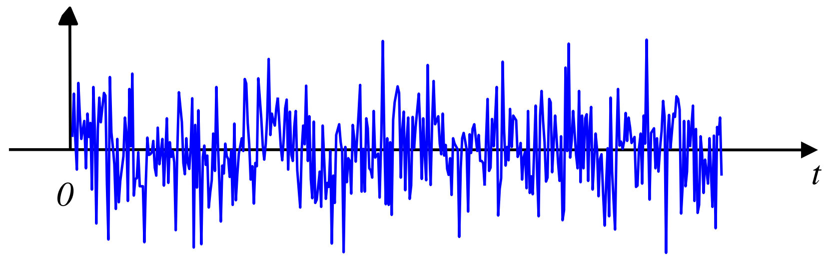
- Low-order moments: example
  - Expectation (mean): $m_1 = \mu = \mathrm{E}[X]$
  - Variance: $m_2 = \sigma^2 = \mathrm{E}[(X-\mu)^2]$

Family of ARMA/ARIMA models

- ARMA/ARIMA models are a class of ***black-box*** models that is capable of representing <u>stationary</u> ***as well as*** <u>non-stationary</u> stochastic signals



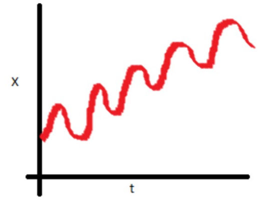
Modelling the disturbances: the stationary assumption



- The stochastic measurement process is assumed to be stationary
  - Its probability distribution does not change when shifted in time
  - Realizations of a stationary stochastic process, vary over time in a stable manner about a fixed mean
  - It is *(weakly) stationary* if it can be described by its first two moments only
    - Mean, variance
    - AutoCorrelation Function (ACF)

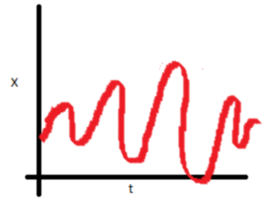
Major assumption: **stationarity** of the signals

- The properties of one section of a data are much like the properties of the other sections. The future is "similar" to the past *(in a probabilistic sense)*
- A stationary stochastic signals has
  - no trend / no seasonality



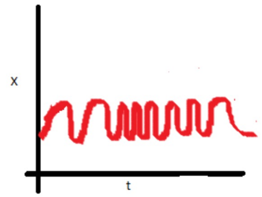
Non-Stationary series

  - no systematic change in variation



Non-Stationary series

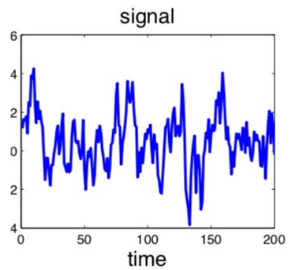
  - no periodic fluctuations

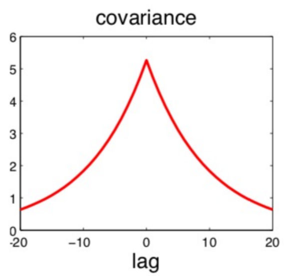


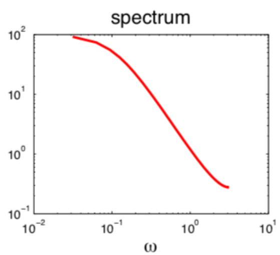
Non-Stationary series

Models for stationary stochastic signals

- Signaux aléatoires stationnaires
  - Modèles non paramétriques
    - Evolution temporelle
    - Fonction d'autocorrélation
    - Densité spectrale de puissance
  - Modèles paramétriques
    - Modèles AR, MA, ARMA







$$y_t = \frac{C(q^{-1})}{D(q^{-1})} e_t$$

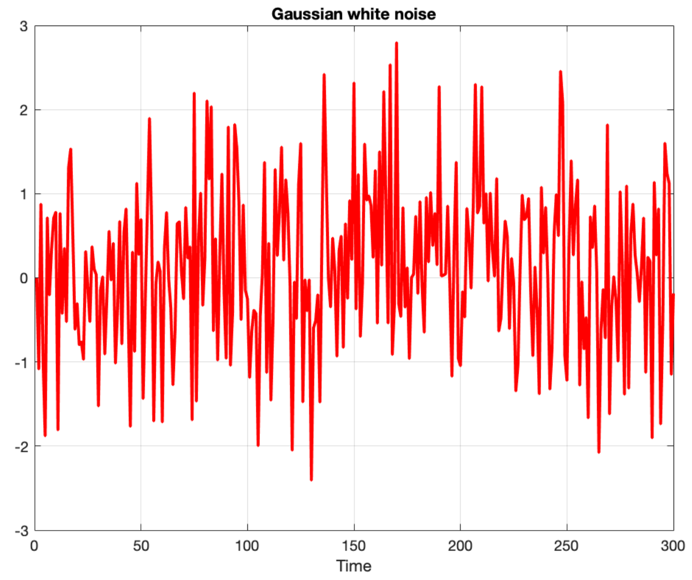
Autocorrelation function (ACF)

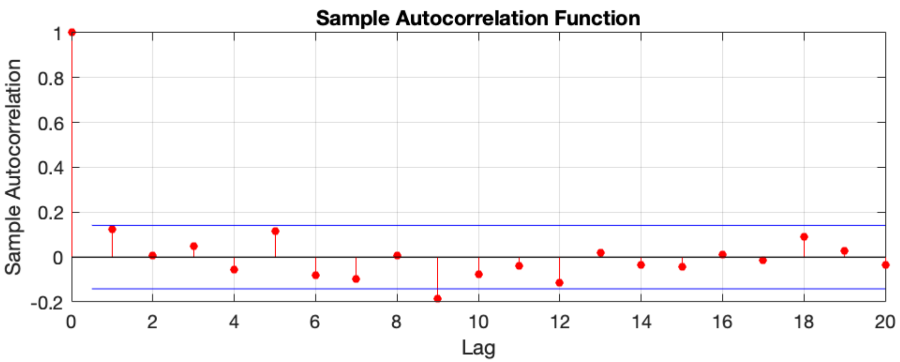
- Statistical correlation summarizes the strength of the relationship between two different variables

- We can calculate the correlation for time series observations with observations with previous time instants, called lags. This is called an autocorrelation

- A plot of the autocorrelation of a time series in terms of lags is called the AutoCorrelation Function, or its acronym ACF

- Sample ACF at lag h, *denoted as* $\gamma_y(h)$, measures the linear correlation between y_t and y_{t+h}



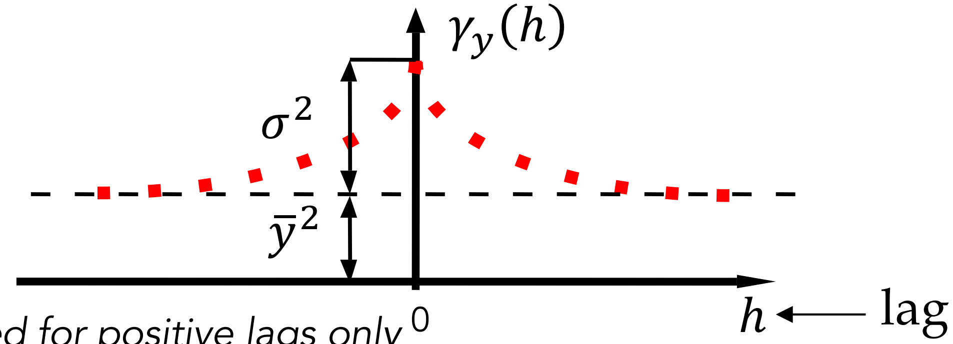


ACF: stationary case

- Autocovariance function of a stationary time series $\{y_t\}$

$$\gamma_y(h) = \text{Cov}(y_{t+h}, y_t) = \text{E}[(y_{t+h} - \mu)(y_t - \mu)] \quad |h| < N$$

with the following 3 properties



1. $\gamma_y(0) \geq 0,$
2. $|\gamma_y(h)| \leq \gamma_y(0)$
3. $\gamma_y(h) = \gamma_y(-h)$

⇒ even function. ACF is usually plotted for positive lags only

- Autocorrelation function of a stationary signal $\{y_t\}$

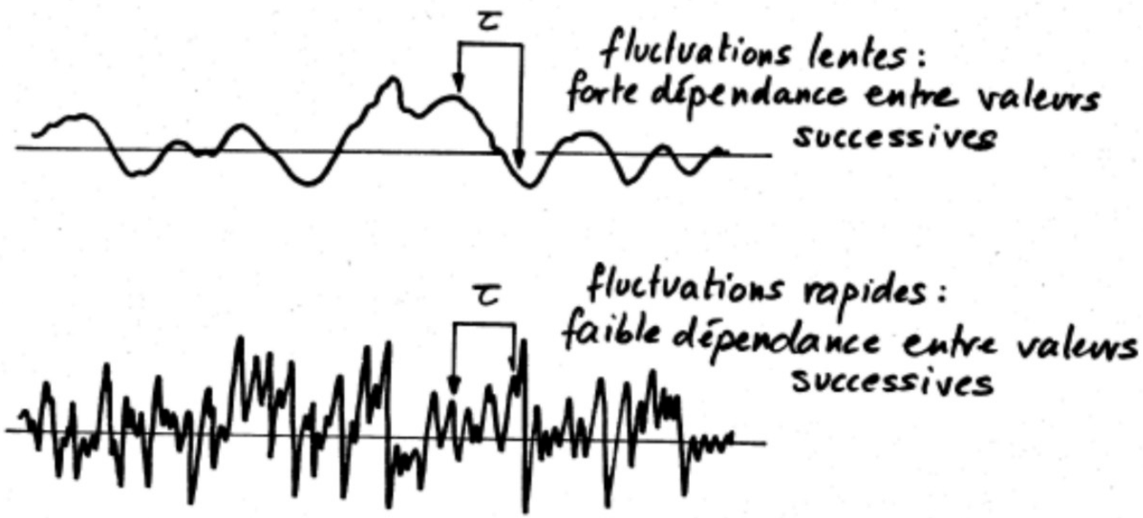
$$\rho_y(h) = \frac{\gamma_y(h)}{\gamma_y(0)} \qquad 0 \leq h < N$$

with all the properties of the autocovariance function, except $\rho_y(0) = 1$

- It measures the linear correlation between y_t and y_{t+h}

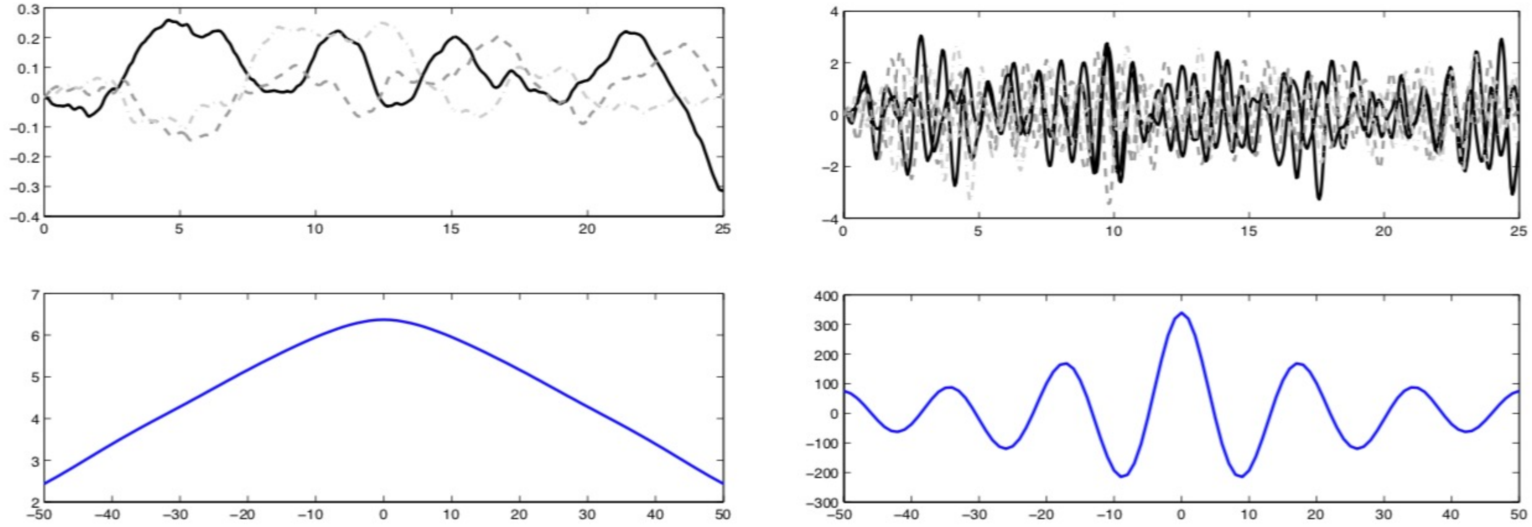
Autocorrelation function (ACF)

- ACF: *measures the speed of variation of temporal evolutions*
  - we compare the time series with itself but shifted by τ (*or* h)
  - it allows us to see how the time series at a given time is influenced (*linear autocorrelation*) by what happened at a previous time



Autocorrelation function (ACF)

Slowly varying autocorrelation function – slowly varying process
Quickly varying autocorrelation function – quickly varying process



Finite sample statistics

- Given $\{y_1, \ldots, y_N\}$ observations of a stationary signal $\{y_t\}$, estimate the finite sample mean, variance, autocovariance and ACF
  - Sample mean

$$\hat{\mu} = \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

  - Sample variance

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \hat{\mu})^2$$

  - Sample autocovariance function

$$\hat{\gamma}_y(h) = \frac{1}{N}\sum_{j=1}^{N-h}(y_{j+h} - \bar{y})(y_j - \bar{y}), \qquad 0 \leq h < N,$$

with $\hat{\gamma}_y(h) = \hat{\gamma}_y(-h), \qquad -N < h \leq 0$

  - Sample autocorrelation function (ACF)

$$\hat{\rho}_y(h) = \frac{\hat{\gamma}_y(h)}{\hat{\gamma}_y(0)}, \qquad |h| < N$$

Finite sample ACF - Example

$y = [0\ 1\ 1\ 1\ 0]$ **$N=5$**

$$\bar{y} = \frac{1}{5}\sum_{i=1}^{5} y_i = 0.6$$

$$\hat{\gamma}_y(h) = \frac{1}{5}\sum_{j=1}^{5-h}(y_{j+h}-\bar{y})(y_j-\bar{y}), \qquad h = 0,1,2,3,4$$

$$\hat{\rho}_y(h) = \frac{\hat{\gamma}_y(h)}{\hat{\gamma}_y(0)}, \qquad h = 0,1,2,3,4$$

$$\hat{\rho}_y = [1\ \ -0.13\ \ -0.26\ \ -0.4\ \ 0.3]$$

In Matlab :

```
y=[0 1 1 1 0];
[rho_hat_y,Lag]=xcov(y,'norm');
stem(Lag,rho_hat_y)
Or
autocorr(y)
```

(Matlab econometrics toolbox)

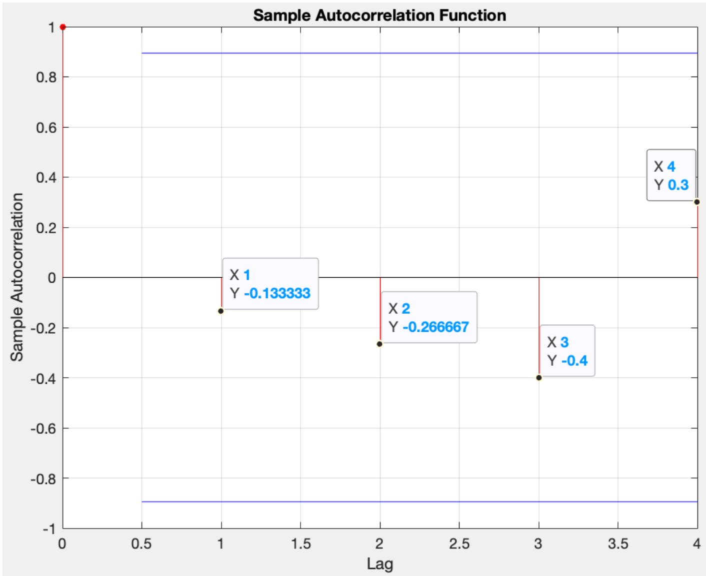
$$\hat{\gamma}_y(0) = \frac{1}{5}\sum_{j=1}^{5}(y_j-\bar{y})(y_j-\bar{y}) = 0.24$$

$$\hat{\gamma}_y(1) = \frac{1}{5}\sum_{j=1}^{4}(y_{j+1}-\bar{y})(y_j-\bar{y}) = -0.0320$$

$$\hat{\gamma}_y(2) = \frac{1}{5}\sum_{j=1}^{3}(y_{j+2}-\bar{y})(y_j-\bar{y}) = -0.0620$$

$$\hat{\gamma}_y(3) = \frac{1}{5}\sum_{j=1}^{2}(y_{j+3}-\bar{y})(y_j-\bar{y}) = -0.0960$$

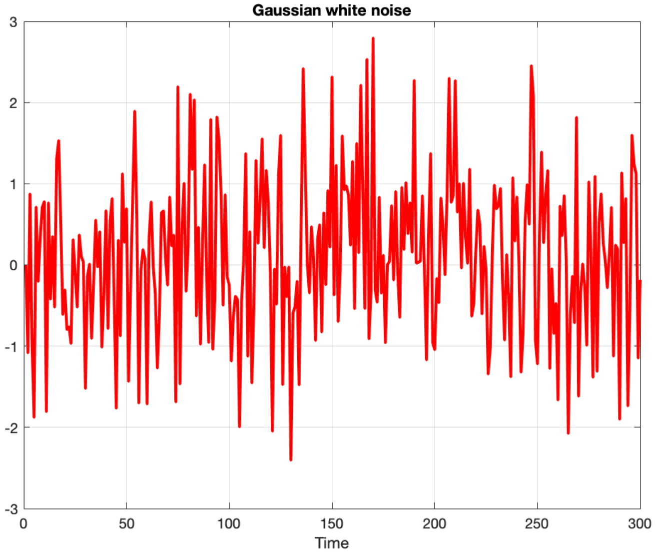
$$\hat{\gamma}_y(4) = \frac{1}{5}\sum_{j=1}^{1}(y_{j+4}-\bar{y})(y_j-\bar{y}) = 0.0720$$

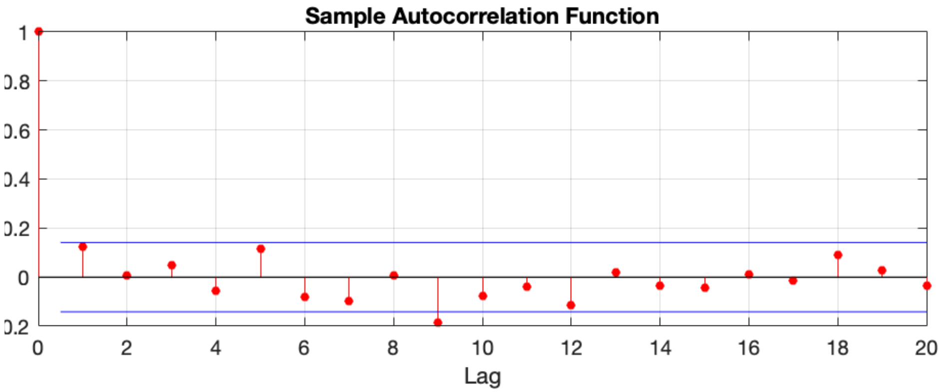


The white noise process

The most fundamental example of stationary signal

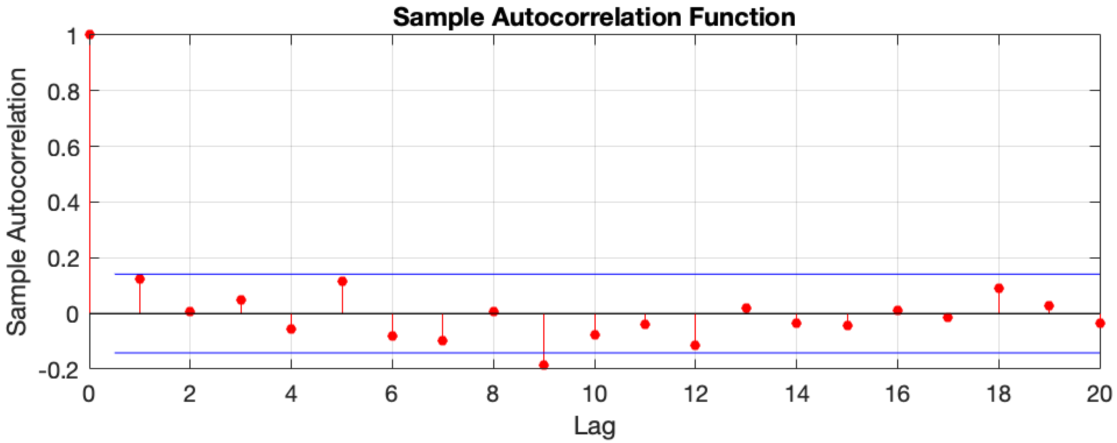
- A white noise is a sequence of independent and identically distributed *(i.i.d)* random variables
  - The sequences are *uncorrelated*, have zero mean, and constant variance
  - A Gaussian white noise are *i.i.d* observations from $\mathcal{N}(0, \sigma^2)$
  - Because independence implies that its variables are uncorrelated at different times, its ACF looks like a Kronecker impulse





Finite sample distribution of sample ACF

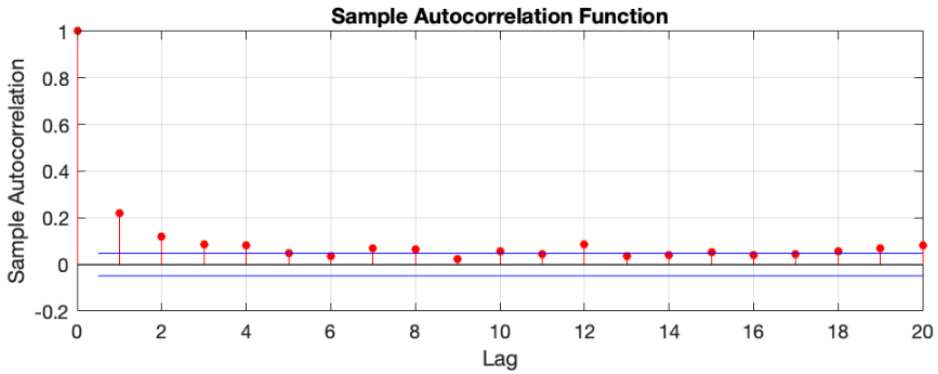
- Finite sample distribution of ACF for a white noise is asymptotically Gaussian $\mathcal{N}\left(0, \frac{1}{N}\right)$
  - 95% of all ACF coefficients for a white noise must lie within $\pm\frac{1.96}{\sqrt{N}}$
  - It is common to plot horizontal limit lines at $\pm\frac{1.96}{\sqrt{N}}$ when plotting the ACF
- When $N = 125$, limit line values at $\pm\frac{1.96}{\sqrt{125}} = \pm 0.175$
  - All ACF coefficients lie within these limits, confirming that the data are white noise (*more precisely, the data cannot be distinguished from white noise*)



Properties of white noise

- Best prediction of a white noise
  - If a signal is white noise, it is unpredictable and so there is nothing to forecast. Or more precisely, *the best prediction is its mean value which is zero*

- Whitening test of the residuals
  - At the validation stage of the system identification methodology, we will check whether the *prediction errors=residuals* are a white noise by plotting its sample ACF



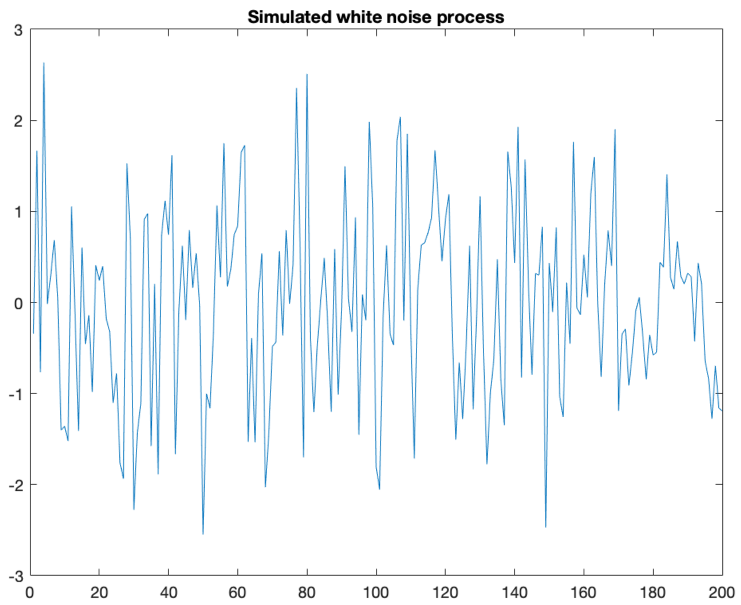
Sample ACF shows some significant autocorrelations at lags 1, 2, 3 and 4. This shows the residuals are not white here

  - If the residual ACF does not resemble to the ACF of a white noise, it suggests that improvements could be made to the predictive model
  - If the residual ACF resembles to the ACF of a white noise, the modelling procedure is finished. There is nothing else to capture in the residuals

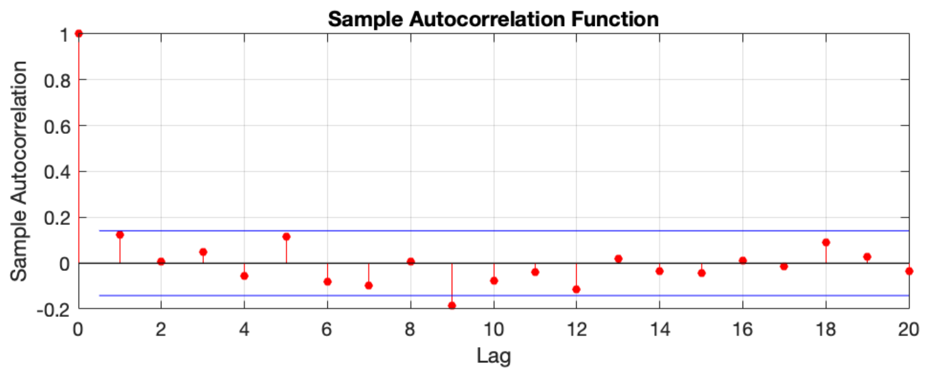
Gaussian white noise

- It is a white noise whose probability function is Gaussian
- In Matlab

```
>> e=randn(200,1);
>> plot(e)
```




```
>> autocorr(e)
```



Gaussian probability function
A brief review

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

m : mean

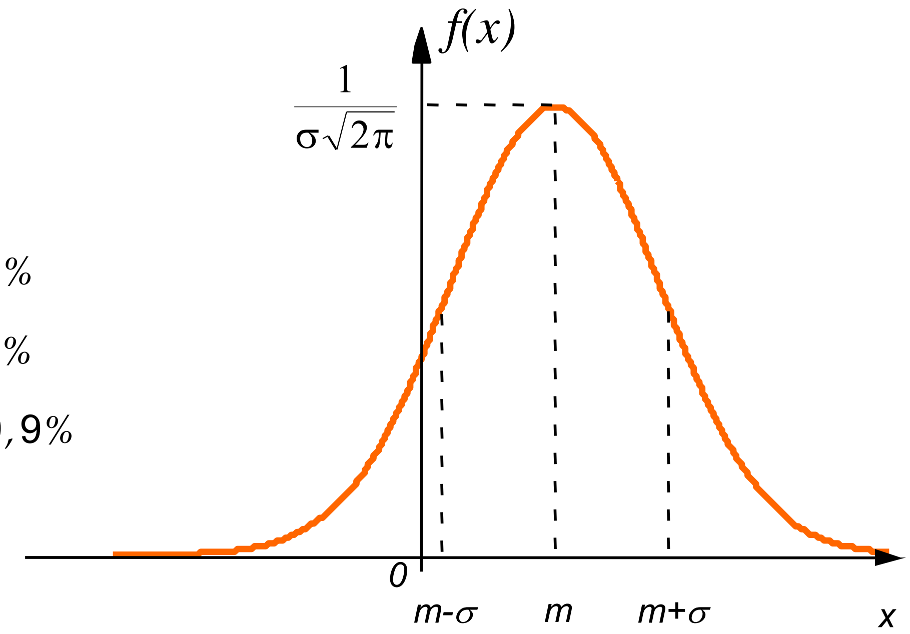
σ : standard-deviation

$P(m-\sigma \leq x \leq m+\sigma) = 67\%$

$P(m-2\sigma \leq x \leq m+2\sigma) = 95\%$

$P(m-3\sigma \leq x \leq m+3\sigma) = 99\%$

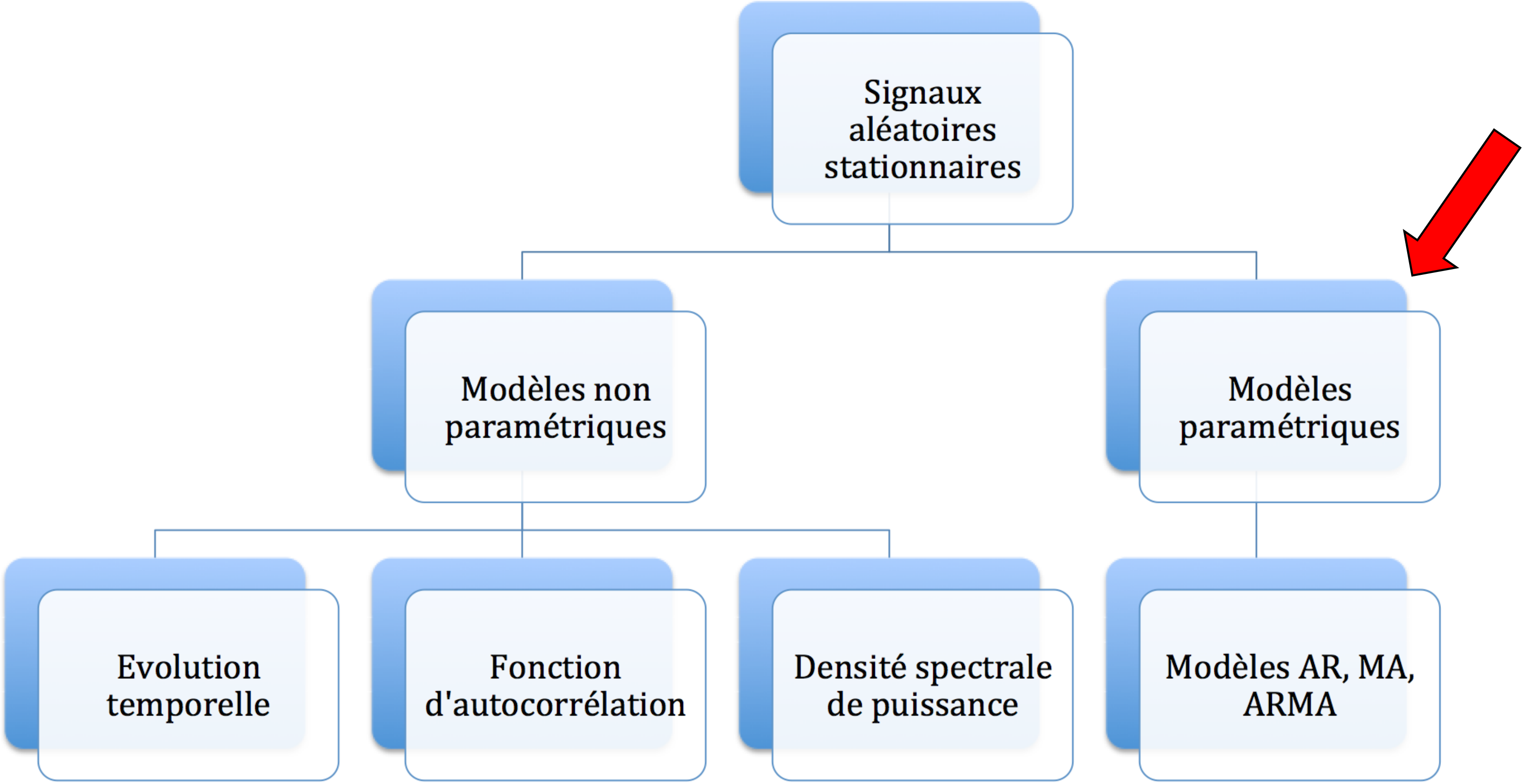
$P(m-4\sigma \leq x \leq m+4\sigma) = 99,9\%$

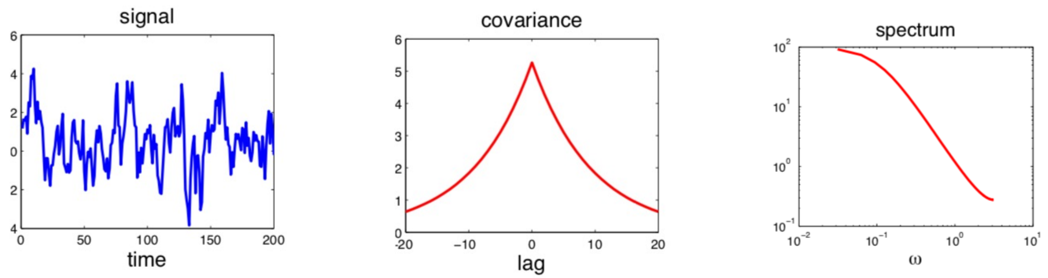


Gaussian probability function
Review

- Important properties
  - Two Gaussian random signals x_k and x_l for $k \neq l$ are uncorrelated (property of white noise) and therefore independent (property of Gaussian probability density)
  - The Gaussian probability density is the only law for which there is equivalence between non-correlation and independence
  - Gaussian laws preserve their Gaussian character in any linear operation: derivation, integration, convolution, filtering

Models for stationary random signals





$$y_t = \frac{C(q^{-1})}{D(q^{-1})}\varepsilon_t$$

General linear parametric model of stationary signals

- Box and Jenkins in 1970 *(following Yule and Slutsky 1927)*
  - Many time series *(or their derivatives)* can be considered as a special class of stochastic processes: *(weakly)* stationary stochastic processes
    - First two moments are finite and constant over time
    - Defined completely by the mean, variance and autocorrelation function

- General parametric model of stationary stochastic processes *(Wold 1938)*
  - All *(weakly)* stationary stochastic processes can be written as

$$y_k = c + \sum_{i=1}^{+\infty} \psi_i e_{k-i} + e_k$$

where c is a constant and e_k is a white Gaussian noise

  - e_k is often called the *innovation process* because it captures all new information in the series at time *k*

Backward shift q^{-1} operator

- The ***backward shift*** operator, q^{-1} , is defined as

$$q^{-1}\,\varepsilon_k = \varepsilon_{t-1}$$

$$q^{-i}\,\varepsilon_k = \varepsilon_{k-i}$$

- The general linear model of a stationary stochastic process can be written as

$$y_k = c + \sum_{i=1}^{+\infty} \psi_i \varepsilon_{k-i} + \varepsilon_k$$

$$y_k \;= c + \psi(q^{-1})\,\varepsilon_k$$

$$\Psi(q^{-1}) = 1 + \sum_{i=1}^{+\infty} \psi_i L^i$$

- This model has an infinite-degree polynomial $\Psi(q^{-1})$ with infinite coefficients which cannot be estimated from a finite amount of data in the time series 😩

Towards AR, MA and ARMA models for stationary signals

- If $H(q)$ is a rational polynomial, we can write it (at least approximately) as the quotient of two finite-degree polynomials

$$H(q) = \frac{C(q^{-1})}{D(q^{-1})}$$

$$C(q^{-1}) = 1 + c_1 q^{-1} + \cdots + c_{n_c} q^{-n_c}$$

$$D(q^{-1}) = 1 + d_1 q^{-1} + \cdots + d_{n_d} q^{-n_d}$$

(Matlab System Identification toolbox notations)

- ***Wold's theorem***: every zero-mean stationary stochastic process can be written as

$$y_k = \frac{C(q^{-1})}{D(q^{-1})} e_k$$

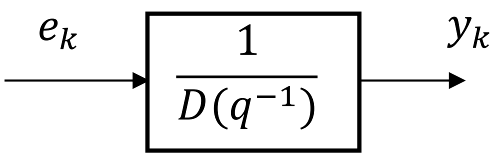
  - which has a finite number $(n_c + n_d)$ of coefficients
  - e_k is a white Gaussian noise

- This leads to the use of <u>parsimonious</u> models : **AR, MA** and **ARMA** models
  - They are most useful for practical applications since these models can be quite easily estimated from a finite number of signal samples

Family of ARMA models
for stationary stochastic signals

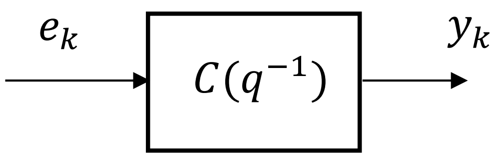
- ARMA models: a way to "see" stationary stochastic signals as *filtered white noise*
  - The filter takes different forms according to the signal properties

AR models



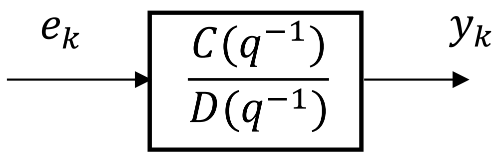
$$y_k = \frac{1}{D(q^{-1})} e_k$$

MA models



$$y_k = C(q^{-1}) e_k$$

ARMA models



$$y_k = \frac{C(q^{-1})}{D(q^{-1})} e_k$$

$$e_k \sim \mathcal{N}(0, \sigma^2)$$