Aalto University

T61.3040

Introduction to ARMA processes

Amaury Lendasse and
Yoan Miche

Welcome

The new Environmental and Industrial Machine Learning (EIML) group is a sub-group of the Adaptive Informatics Applications (AIA) group. It is part of both the Department of Information and Computer Science and the Adaptive Informatics Research Centre, **Centre of Excellence of the Academy of Finland**.

The EIML group is based on the former Time Series Prediction and Chemoinformatics group. The EIML group is developing **new** Machine Learning techniques:

- to model environment (using e.g. time series prediction, variable selection and ensemble modeling)
- to solve industrial problems (for example in the fields of chemometrics, electricity production and distribution, bankruptcy prediction and information security.

The EIML group has been created and is lead by Dr. Amaury Lendasse, Docent. The Environmental modeling is directed by Dr. Federico Pouzols. The Industrial Machine Learning is under the responsibility of Dr. Francesco Corona, Docent.

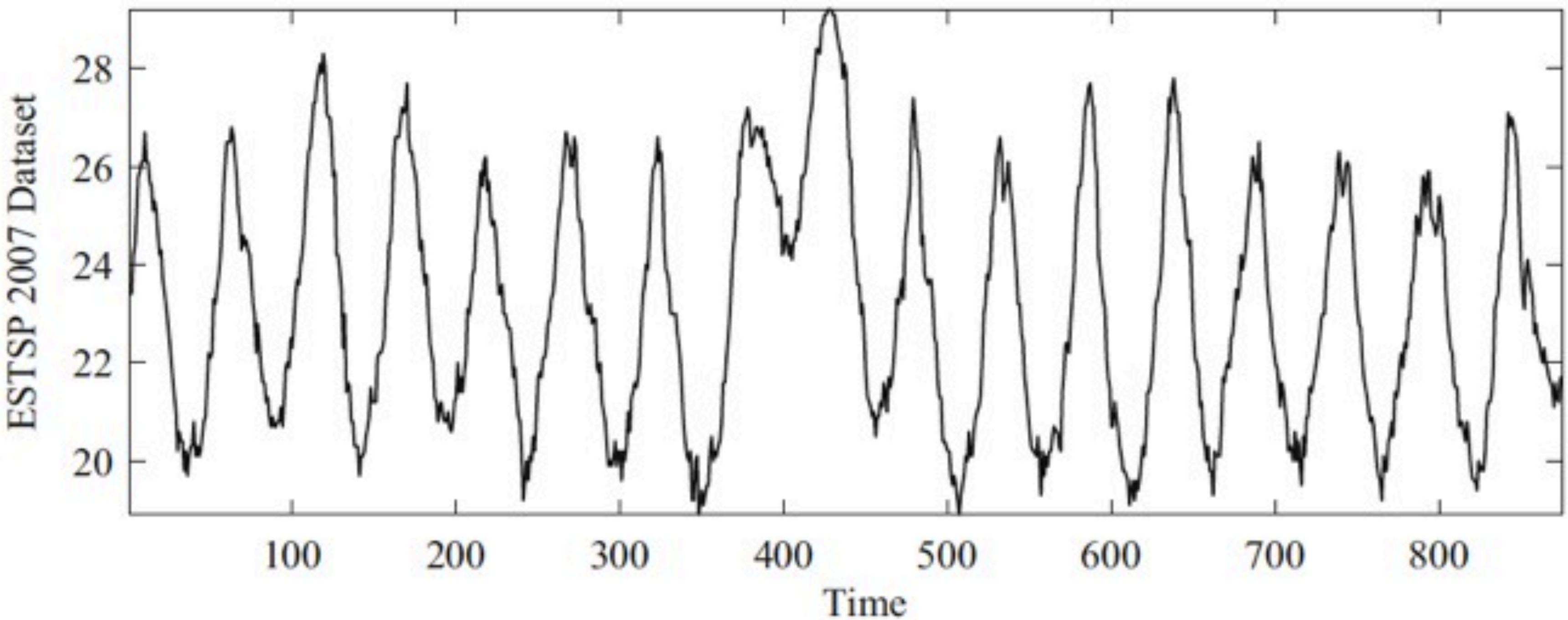
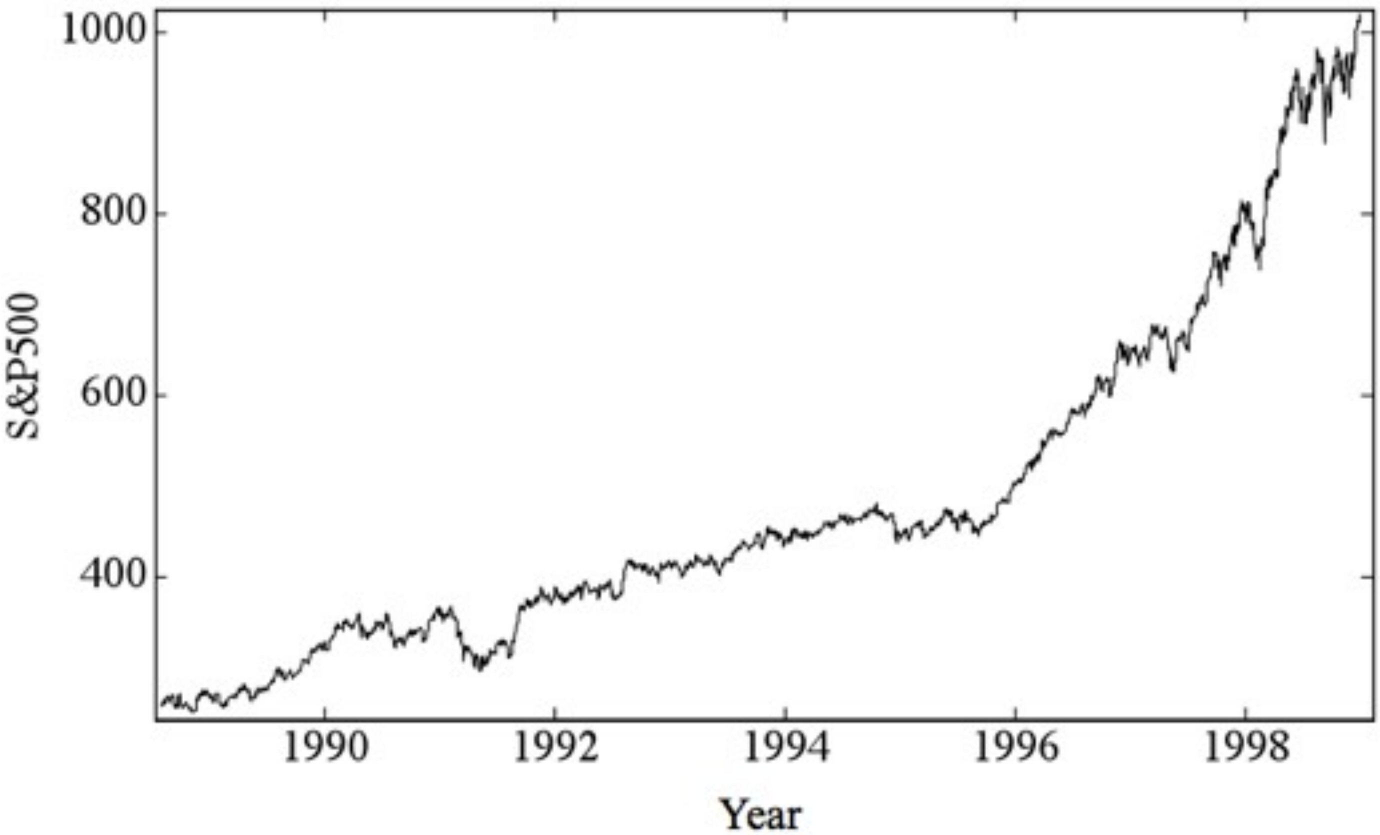
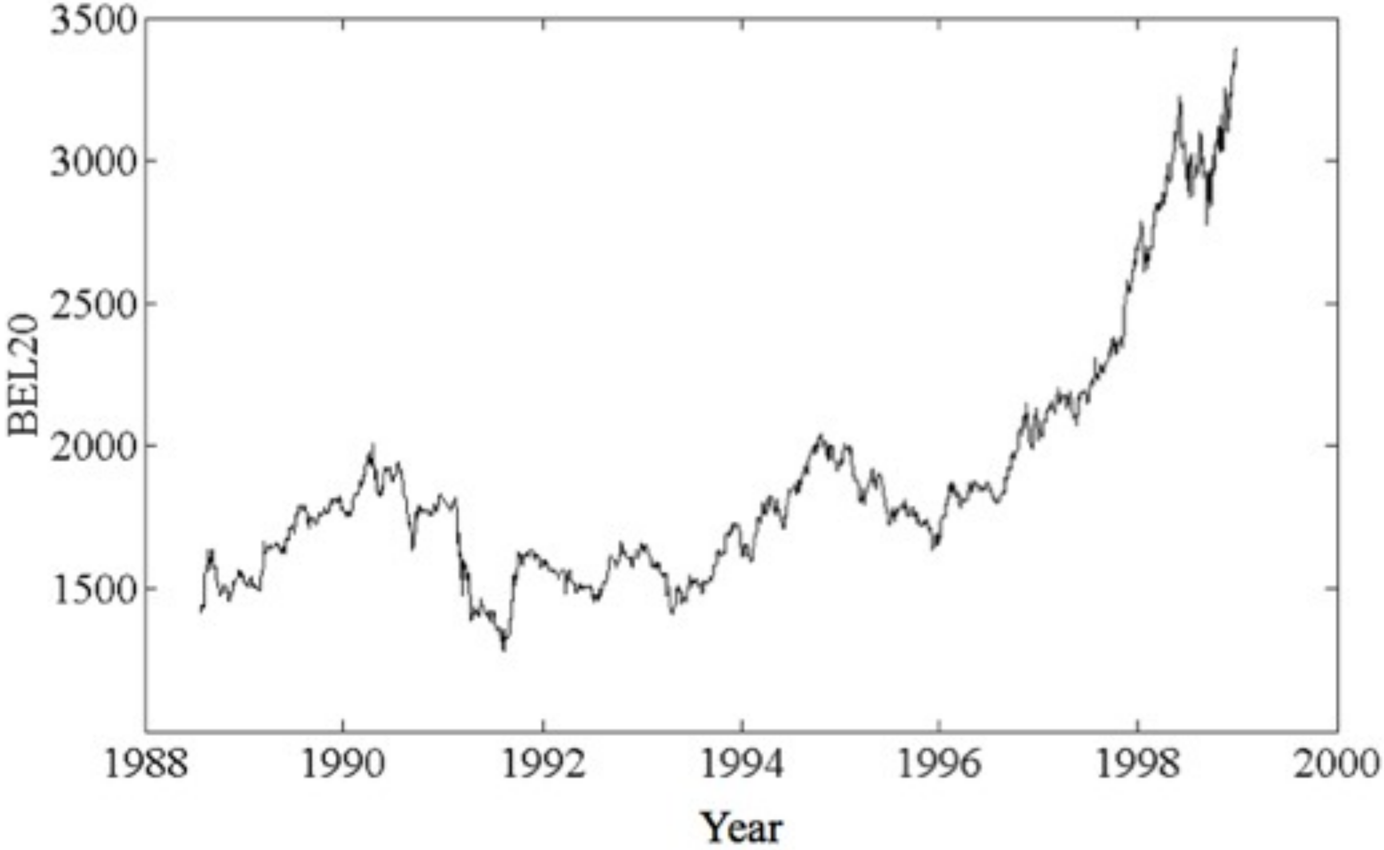


Fig. 3. ESTSP 2007 competition data.


S&P500
1000
800
600
400
1990
1992
1994
1996
1998
Year


3500
3000
2500
2000
1500
BEL20
1988
1990
1992
1994
1996
1998
2000
Year

Introduction to ARMA processes

- We examine ARMA process, some example
- We learn the required calculation techniques at the same time
- Later we will justify the use of ARMA processes in detail

Noise

- The random process *v*(*n*) which is "difficult" to predict
- We assume that noise is Gaussian and uncorrelated:

$$v(n) \sim N(0, \sigma^2), \quad [v(n_1), v(n_2), \ldots, v(n_N)] \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

- $n \neq m \Rightarrow v(n), v(m)$ uncorrelated

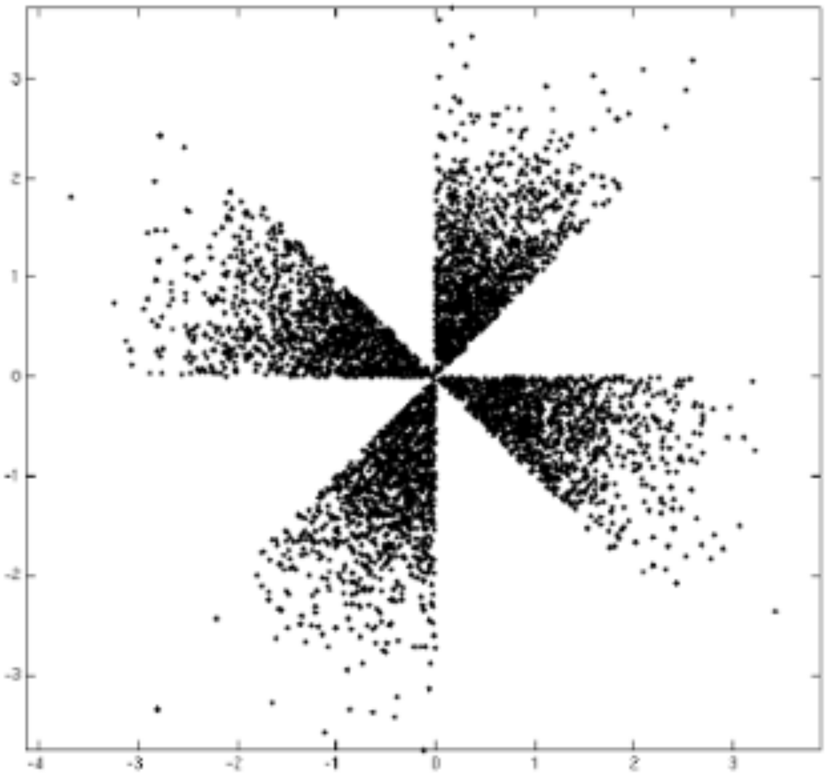
- A special case of white noise
- Noise does not need to be Gaussian

Noise

- $v(n)$ is a sequence of uncorrelated jointly normally distributed random variables
- an important property: the observed values $v(n-1)$, $v(n-2)$, ... do not help to predict the value of $v(n)$
- Note that for jointly normally distributed random variables: Uncorrelated implies independence
- Just uncorrelated implies that noise is linearly unpredictable (see exercises)

Noise

- conditional distribution: $p(v(n) \mid v(n\text{-}1), v(n\text{-}2), \ldots)$
- Interpretation: How $v(n)$ is distributed, when values $v(n\text{-}1)$, $v(\text{n-}2)$, ... are known
- Jointly Gaussian, white noise:
  $p(v(n) \mid v(n\text{-}1), v(n\text{-}2), \ldots) = p(v(n))$

- not enough that the individual variables are normally distributed, cf. image.
- 

  variables are individually normal. distrib. and uncorrelated but not independent!
- However, in this course "normally distributed noise" refers to jointly Gaussian noise.

- from the conditional distributions, we can calculate parameters such as mean and variance

- with jointly normally distributed white noise *v*(*n*), the conditional distribution can be presented by using the expectation and the variance:

$$\mathrm{E}(v(n) \mid v(n-1), v(n-2), \ldots) = \mathrm{E}(v(n)) = 0$$

$$\mathrm{Var}(v(n) \mid v(n-1), v(n-2), \ldots) = \mathrm{Var}(v(n)) = \sigma^2$$

- the values of the noise can be added together to form the "sum process"

$$M(n) = \sum_{i=0}^{n} v(i)$$

- it satisfies the condition:

$$\mathrm{E}(M(n)|M(n-1), M(n-2), \ldots) = M(n-1)$$

- processes under this condition can be used in applications where the best guess on the future value is the latest observation (the financial time series)

AR-process

- typically want to predict the future values using observations (noise is therefore useless)

- define the random process as difference equation

$$x(n) = -ax(n-1) + v(n)\,, \quad v(n)$$ normal noise

- this defines a random process $x(n)$, $n = 0, 1, 2, \ldots$ Also, $x(n)$, $n < 0$ are random variables which satisfy the same difference equation, i.e. the process is thought to have been under way long before the time of the simulation, $n = 0$, can be set for example using $x(-1) = 0$.

The process can be simulated as follows:

1. chosen the initial value of $x(-1) = 0$ and set $n = 0$
2. generate the value of the noise v(n)
3. calculated as $x(n) = -ax(n-1) + v(n)$
($x(n-1)$ has been calculated previously)
4. set $n := n + 1$, and continue to step 2

- Previous slide is an example of an AR-process
- selected a few parameter values and simulate the corresponding AR(1)-Processes
- a special case of the value of $a = 0$ yielding the noise process
- demo: ar_1.R

- general AR (p) process can be written

$$x(n) = -\sum_{k=1}^{p} a(k)x(n-k) + v(n),$$

$v(n)$ normal noise

- p is the degree of the process (in practice $p<\infty$, but in principle can have $p=\infty$)

- AR(p), can be simulated as AR(1), but we need more initial values ($x(n) = 0$ when $n<0$)

- The term AR comes from the word autoregressive

- Value of $x(n)$ of process is formed by linear regression from the p previous values

- AR-process model is equivalent to the linear relationship of the previous values

MA-process

- a linear model can also be formed from previous noise values. Example:

$$x(n) = b(0)v(n) + b(1)v(n-1)$$

- Is it useful to sum the noise values?
- i.e. is the $x(n)$ also noise?

- Let's check process $x(n) = v(n) - v(n\text{-}1)$
- It is clear that $x(n)$ is a normal distribution with expectation value zero
- Calculated correlation of successive values (check yourself the steps):

$$\mathrm{E}(x(n)x^*(n-1)) = -\sigma^2 < 0$$

- So process $x(n)$ is not noise

- MA term stands for moving average: let's calculate moving average from the noise

- A general MA(q) process is

$$x(n) = \sum_{k=0}^{q} b(k)v(n-k)$$

where $v(n)$ is white noise

- Simulation is easy, simulate the noise and then get $x(n)$
- demo: ma_1.R

- We can take linear combination of both previous values and the noise

- Let's have a ARMA(p,q) process

$$x(n) = -\sum_{k=1}^{p} a(k)x(n-k) + \sum_{l=0}^{q} b(l)v(n-l),$$

$v(n)$ normal noise

- A linear model with respect to parameters *a* and *b*
- demo: arma_11.R

Presentation in the z-plane (space)

- In signal processing (course?), people studied z-transform
- The value $x(n)$ multiplied by z^{-k} is comparable to the delay in time domain:

$$z^{-1}x(n) = x(n-1)$$

$$\vdots$$

$$z^{-k}x(n) = x(n-k)$$

- We interpret z^{-k} as symbols that carry out the process time shifting

- Using z^{-k}, the MA and AR processes can be written simply:

$$
\begin{aligned}
x(n) &= B(z)v(n) \\
&= (b(0) + b(1)z^{-1} + \cdots + b(q)z^{-q})v(n)
\end{aligned}
$$

$$A(z)x(n) = v(n)$$

$$(1 + a(1)z^{-1} + \cdots + a(p)z^{-p})x(n) = v(n)$$

- ARMA (p,q)-process can be written

$$x(n) = \frac{B(z)}{A(z)} v(n)$$

- This is somewhat imprecise, because multiplying by the z^{-k} is interpreted as a delay in the process

- The form $B(z)/A(z)$ can be justified taking z-transformation of ARMA difference equation

- polynomials $A(z)$ and $B(z)$ are determined by the parameters: $a(1). . . , a(p), b(0). . . , b(q)$
- When the parameters are known, then the corresponding ARMA process is fully defined (if the noise is Gaussian)

- Are different parameters for the ARMA process, defining always a different random processes?

- No: the processes can be converted so that the values of the parameters are changing, but the process does not change

- example: let's convert the AR(1)-process

$$x(n) = -a(1)\, x(n-1) + v(n) \qquad \text{(Eq. 1)}$$

- into (Eq. 1) we can substitute

$$x(n-1) = -a(1)\, x(n-2) + v(n-1)$$

- We obtain

$$x(n) = a^2(1)\, x(n-2) - a(1)\, v(n-1) + v(n)$$

- by repeating the same k times, we get ARMA (k,k-1) process

$$x(n) = (-a(1))^k x(n-k) + (-a(1))^{k-1} v(n-k+1) + (-a(1))^{k-2} v(n-k+2) + \ldots + v(n)$$

- if $|a(1)|$ is smaller than 1 then

$$x(n) = \sum_{k=0}^{\infty} (-a(1))^k v(n-k)$$

which may be interpreted as a MA(∞)-process

- Transforming AR(1) to MA(∞) may also be made more directly

- Write the AR(1)-process using the polynomial $A(z)$:

$$x(n) = \frac{1}{A(z)} v(n) = \frac{1}{1+a(1)z^{-1}} v(n)$$

- The term $\frac{1}{1+a(1)z^{-1}}$ can be developed into a series of z^{-1} ratio

- Can be obtained easily

$$\frac{1}{1+a(1)z^{-1}} = 1 + \left(-a(1)\right)z^{-1} + \left(-a(1)\right)^2 z^{-2} + \ldots$$

- Then we can write the AR(1)-process in the form of

$$x(n) = \sum_{k=0}^{\infty} \left(-a(1)\right)^k v(n-k)$$

as before

- More complex transformations require more complicated calculations, see example in exercise session
- The important point here is that the AR-process can be transformed into the MA-process, and vice versa
- Although the number of parameters generally change from finite to infinite

- Can the ARMA-process parameters have any values?
- demo: $x(n) = 1.1x(n\text{-}1) + v(n)$ (ar_infty.R)
- Process "explodes" towards infinity, because the process not “weak stationary” (defined more accurately in the future)
- AR-process, the system $1/A(z)$ defined the AR-process, is not stable because it has a pole outside the unit circle: $z = 1.1$

- Stability “is not a result” from the fact that the absolute values of parameters are less than unity

- demo: $x(n) = -0.7x(n-1) + 0.6x(n-2) + v(n)$ (ar_infty2.R)

- Later, we see that a large number of random processes can be always be defined as stable ARMA-process

- different models have certain advantages:
  - AR: easy to solve the parameters
  - MA: easy to calculate statistics of the process
  - ARMA: often the smallest number of parameters

- if in an application can be assumed that any of the above models are suitable for both conversion examples showed that, in principle, the model can be freely chosen between these ones

Sine Signal

- In addition to ARMA processes, will also discuss sine signals
- These are random processes if they are correctly defined
- Simplest sine signal is a process

$$x(n) = A\sin(n\omega + \phi)$$

- What makes this process random?

- If the amplitude A, frequency ω and phase ϕ are constants so it is not a random process

- Let's assume the amplitude and frequency are constants

- Instead, the phase ϕ is assume to be uniformly distributed on the interval $[0, 2\pi[$

- Note! despite of the randomness, the phase is the same for all n

- One realization can be obtained by simulating once the value of ϕ and calculating

$$x(n) = A \sin(n\omega + \phi)$$

- Simulated sine signal seems to be exactly like a deterministic signal

- Randomness of phase only translate the sine along the time axis

- A complex sine signal

$$x(n) = A\exp(jn\omega), \;\; A = |A|\exp(j\phi)$$

- A is a complex number that contains the phase f and the "real" amplitude |*A*|

- We always use the sine complex signals, because the real sine can be written

$$\sin(\omega) = \frac{1}{2j}[\exp(j\omega) - \exp(-j\omega)]$$

- So a real sine-signal with frequency ω corresponds to two complex sines with frequencies $\pm\omega$

- Later, we will see how to estimate the frequencies of complex sine signal

- When the frequencies are known, the amplitudes and phases are still to be resolved

- Solution for the complex sine signal is enough

- Let's assumes that the we observed

$$x(n) = A\cos(n\omega + \phi) + v(n)$$

and the frequency ω is known

- We solved amplitude and phase by minimizing the

$$\sum_n [x(n) - A\cos(n\omega + \phi)]^2$$

- The phase ϕ is in the cosine function, so we do not get a direct linear solution (except if you develop the cosine)

- It can be written

$$x(n) = \frac{1}{2}A\exp(jn\omega) + \frac{1}{2}A^*\exp(-jn\omega) + v(n)$$

- Now we have to minimize

$$\sum_n \left[x(n) - \frac{1}{2}A\exp(jn\omega) - \frac{1}{2}A^*\exp(-jn\omega)\right]^2$$

- Unknown complex variable *A* can be solve now linearly

- Usually we consider the sum of sine signals with added white noise:

$$x(n) = \sum_{k=1}^{L} A_k \exp(jn\omega_k) + v(n)$$

- Every phase ϕ_k, included in the amplitude of the $A_k = |A_k| \exp(j\phi_k)$, is equally distributed and independent from other phases

- What makes the model interesting for many applications where the frequency estimation does not work sufficiently well using Fourier transform based methods