INDIAN INSTITUTE OF SCIENCE

STOCHASTIC HYDROLOGY

Lecture -16

Course Instructor : Prof. P. P. MUJUMDAR

Department of Civil Engg., IISc.

Summary of the previous lecture

- Box-Jenkins time series models
- Differencing the time series
- ‘B’ Operator

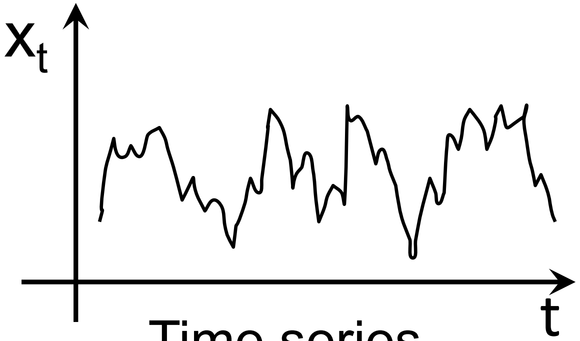
ARIMA Models

1. Identification of the model structure:

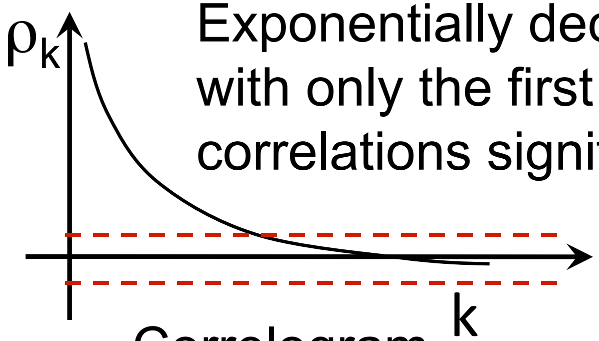
- Identify if the series is stationarity.
    - Plot correlogram (correlogram shows a rapid decay for a stationary series)
- Remove non-stationarity if any by differencing/ standardization.
- Obtain the order of AR and MA components of the model.
- PAC determines the order of the AR process

ARIMA Models

For example, AR(1) process:

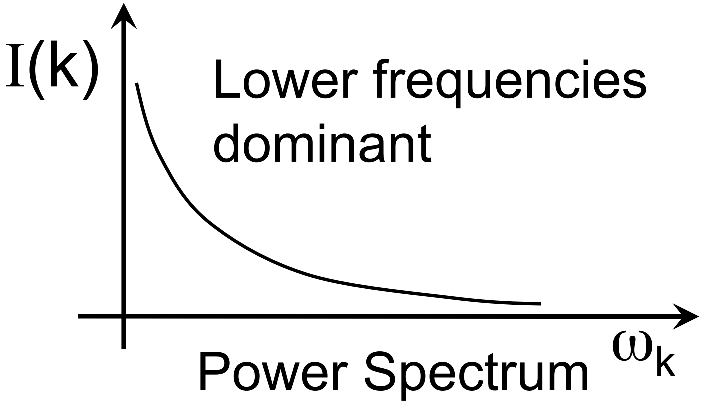


Time series



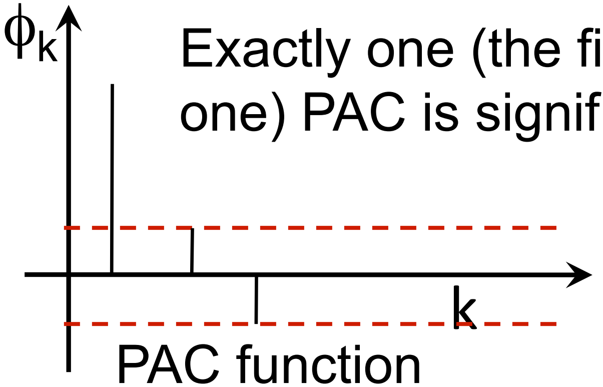
Exponentially decaying with only the first few correlations significant

Correlogram



Lower frequencies dominant

Power Spectrum

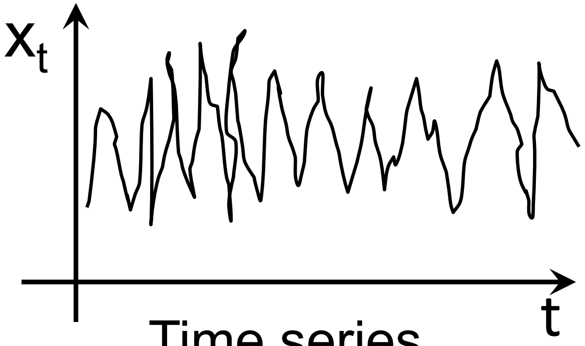


Exactly one (the first one) PAC is significant

PAC function

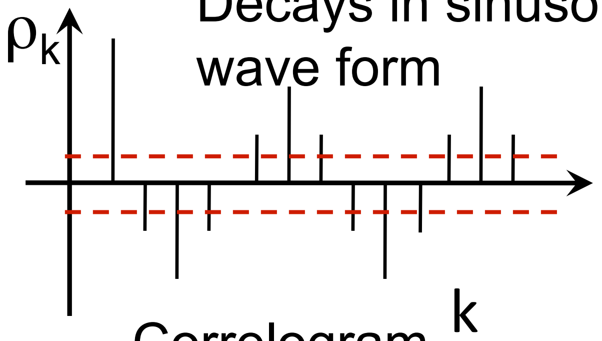
ARIMA Models

AR(2) process:

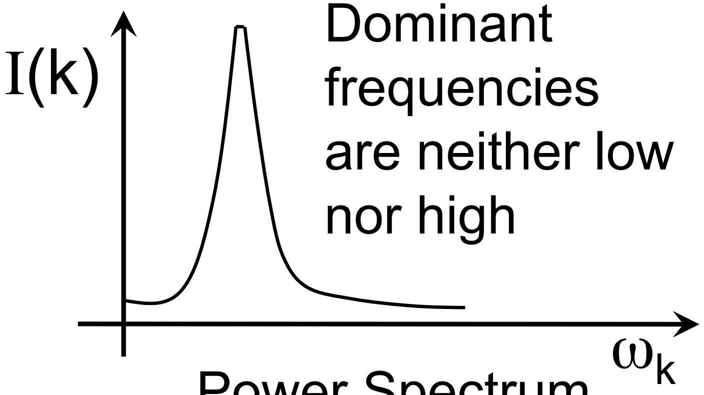


Time series

Decays in sinusoidal wave form



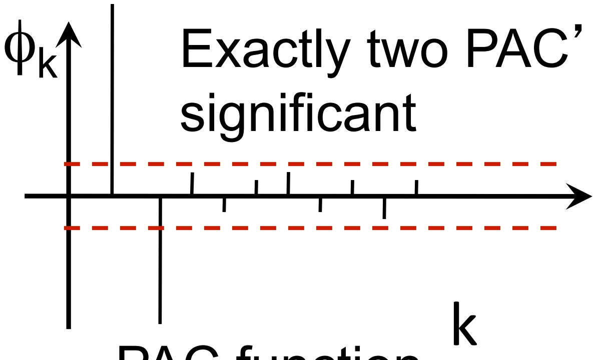
Correlogram



Dominant frequencies are neither low nor high

Power Spectrum

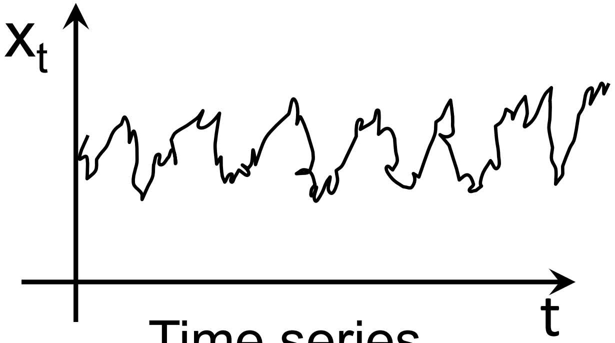
Exactly two PAC's significant



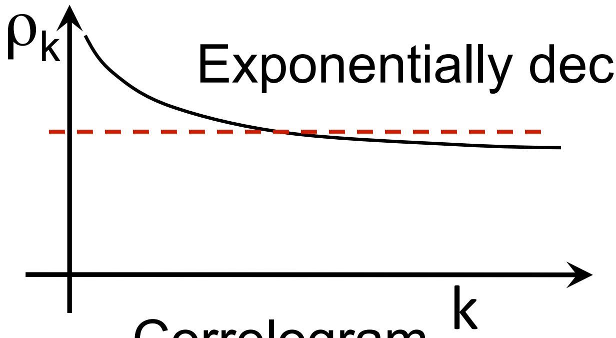
PAC function

ARIMA Models

Another AR(2) process:

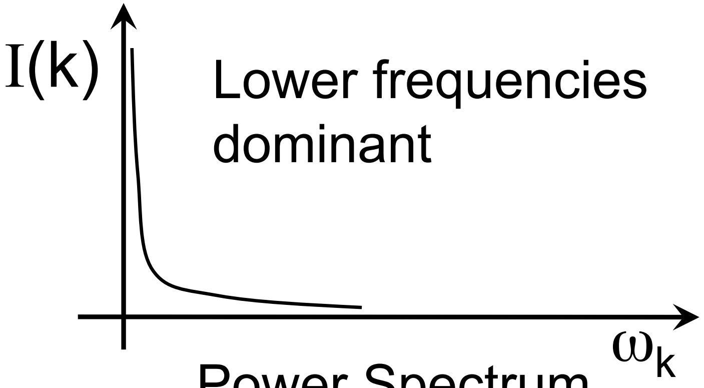


Time series



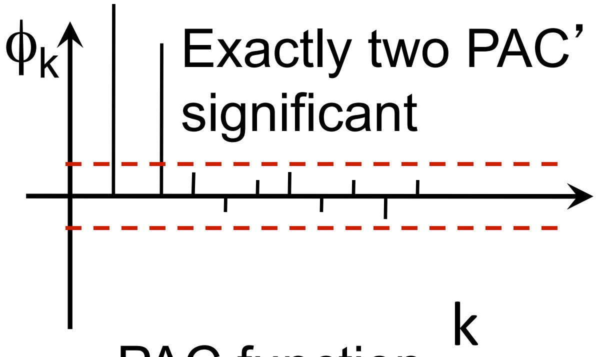
Exponentially decaying

Correlogram



Lower frequencies dominant

Power Spectrum



Exactly two PAC's significant

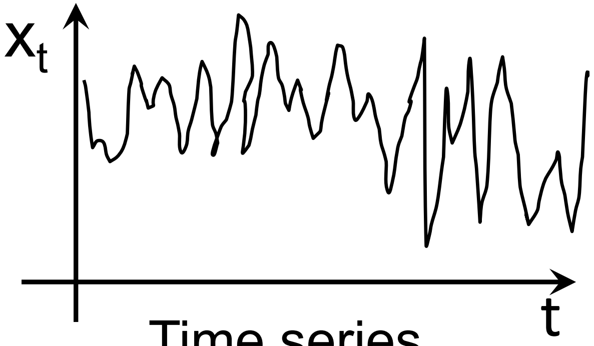
PAC function

ARIMA Models

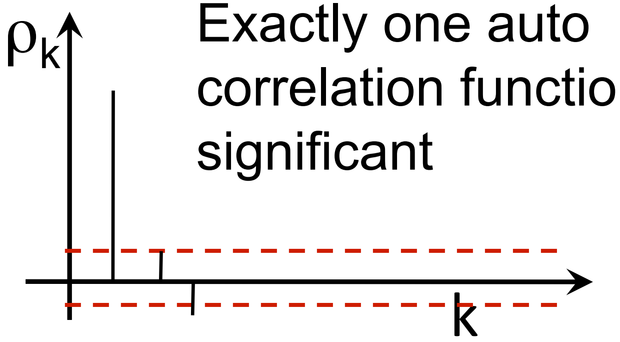
- Behavior of AR process:
  - Decaying auto correlation function (either exponentially or in a dampened sine wave)
  - Order of AR determined by the significant PAC' s

ARIMA Models

MA(1) process:

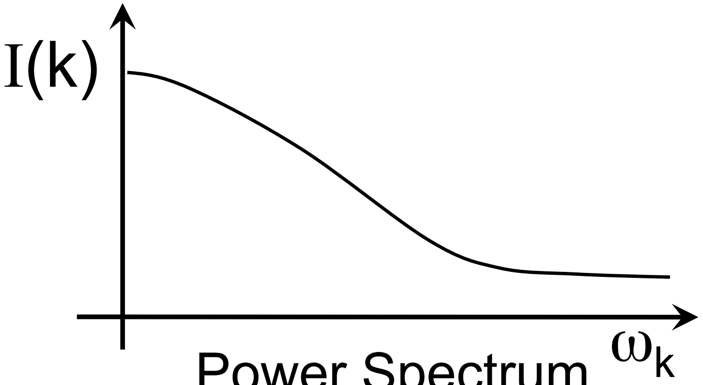


Time series

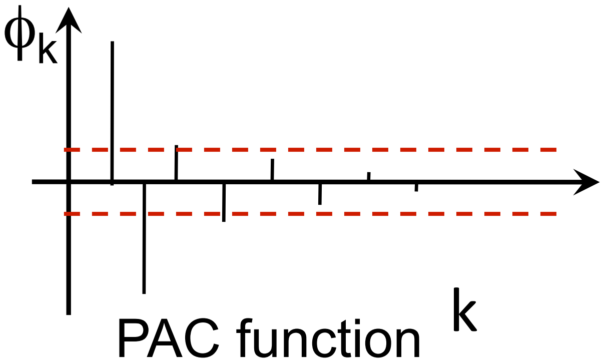


Exactly one auto correlation function is significant

Correlogram



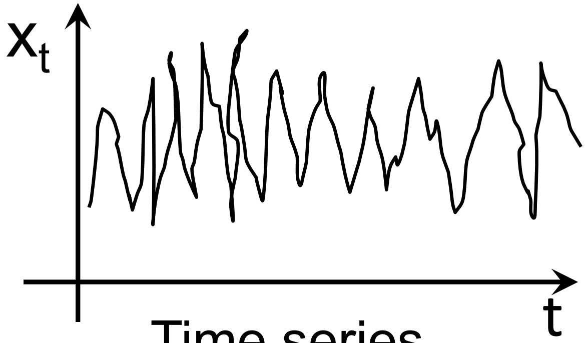
Power Spectrum



PAC function

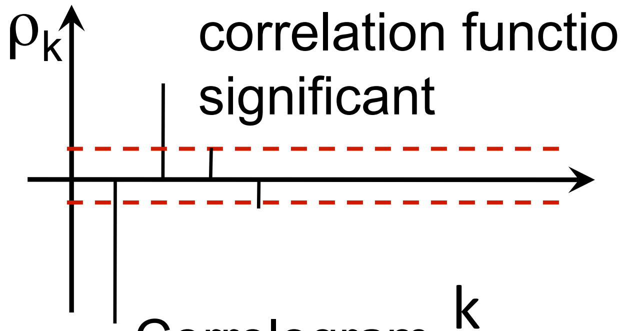
ARIMA Models

MA(2) process:

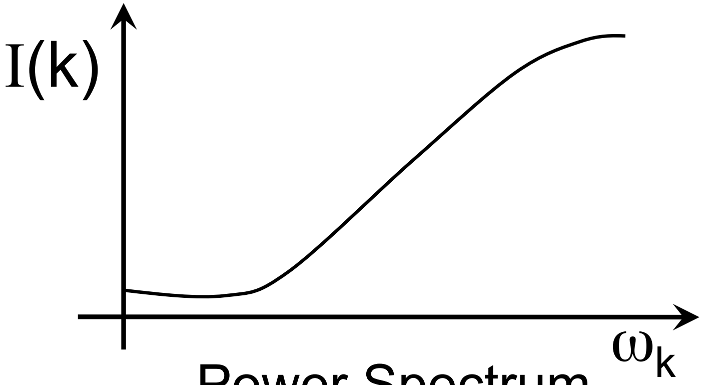


Time series

Exactly two auto correlation functions significant

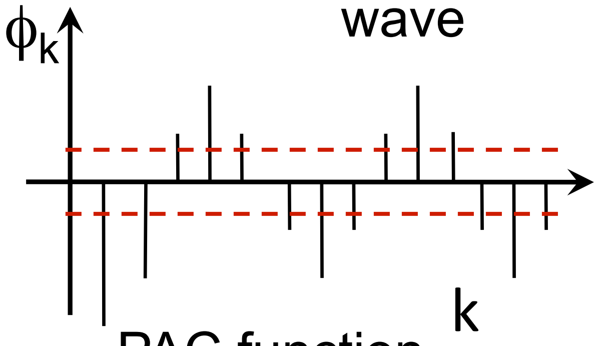


Correlogram



Power Spectrum

Decays in sinusoidal wave



PAC function

ARIMA Models

- Behavior of MA process:
    - The order of MA is determined by the number of significant auto correlations
    - Decaying PAC function (either exponentially or in a dampened sine wave)

ARIMA Models

2. Parameter estimation and calibration:

- Algorithms are available for parameter estimation
  - e.g., Marquadt's algorithm, available in most statistical tool boxes, "armax" toolbox in Matlab.
- For some algorithms, initial values of the parameters need to be supplied based on the Yule-Walker equations
- Solve the Yule-Walker equations of order 'p' and give the resulting ϕ_1, ϕ_2,….. ϕ_p as initial values of the AR parameters.

ARIMA Models

Estimation of initial values of MA parameters:

$$X_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \;\ldots\ldots\ldots - \theta_q e_{t-q}$$

$$\rho_k = \frac{-\theta_k + \theta_1\theta_{k-1} + \theta_1\theta_{k-2} + \ldots.. + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \ldots.. + \theta_q^2}$$

k = 1, 2, ….. q

$= 0$ k > q

Ref: Forecasting methods and applications by Markridakis, Wheelwright, McGee, John Willey 1978

Example – 1

Obtain MA parameters for $r_1 = 0.37$

$$\rho_k = \frac{-\theta_k + \theta_1\theta_{k-1} + \theta_1\theta_{k-2} + \ldots.. + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \ldots.. + \theta_q^2}$$

For k = 1,

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

$$\rho_1 + \rho_1\theta_1^2 + \theta_1 = 0$$

$$0.37\theta_1^2 + \theta_1 + 0.37 = 0$$

$$\theta_1 = -0.443$$

Example – 2

Matlab function “armax” syntax:

m = armax(data, orders)

‘data’ : array of timeseries data

orders = [na, nc]

na = order of AR parameters

nc = order of MA parameters

Example – 2

For example, for ARMA(1, 2) model,

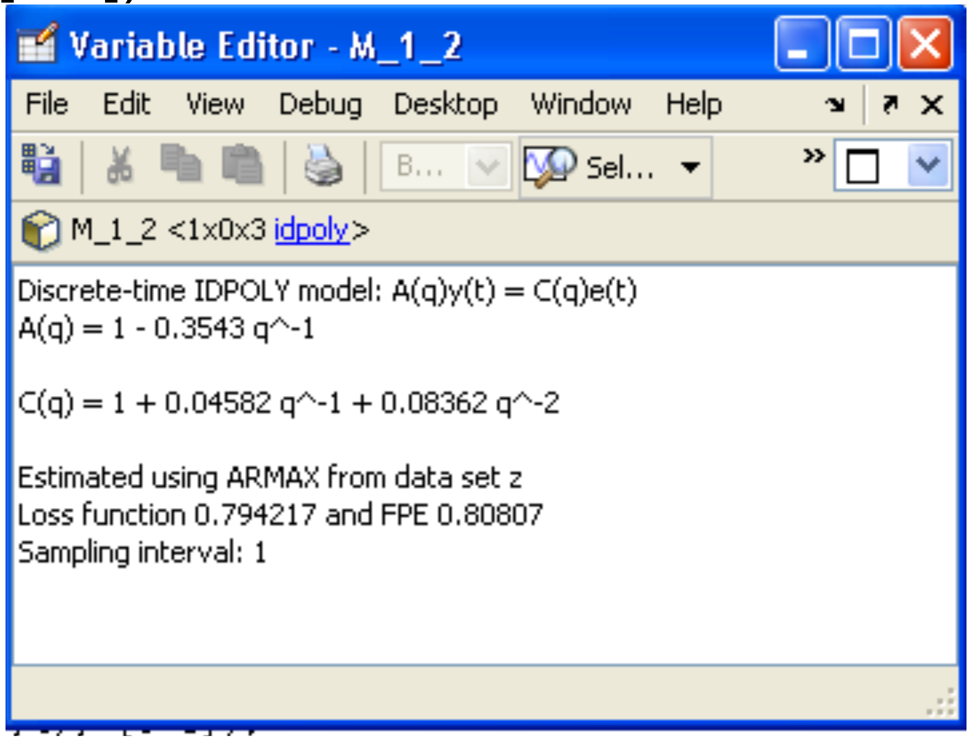
m_1_2 = armax(datax, [1 2]);

The output is as shown

ϕ_1 = -0.3543

θ_1 = 0.04582

θ_2 = 0.0836



```
Discrete-time IDPOLY model: A(q)y(t) = C(q)e(t)
A(q) = 1 - 0.3543 q^-1

C(q) = 1 + 0.04582 q^-1 + 0.08362 q^-2

Estimated using ARMAX from data set z
Loss function 0.794217 and FPE 0.80807
Sampling interval: 1
```

ARIMA Models

Model selection:

- Model selection is important in time series analysis as there are infinitely many possible models
- In general, AR parameters of order up to 6 and MA parameters of order up to 2 serve the purpose in most hydrologic applications. The models may be contiguous or non-contiguous.
- A model may be selected by using the following two criteria from among several candidate models
  - Maximum likelihood rule (ML)
  - Mean square error (MSE)

ARIMA Models

Maximum likelihood rule:

- A likelihood value for each of the candidate models is evaluated.
- The model with highest likelihood value is chosen.
- The general form of log-likelihood function for the ith model for a Gaussian process is

$$L_i = \ln\left(p\left[z, \hat{\phi}_i\right]\right) - n_i$$

A specific likelihood function in this general class I may be approximated as,

$$L_i = -\frac{N}{2}\ln\left(\sigma_i\right) - n_i$$

Ref: Kashyap R.L. and Ramachandra Rao.A, “Dynamic stochastic models from empirical data”, Academic press, New York , 1976

ARIMA Models

Where L_i is the likelihood value,

z is the vector of historical series

$\hat{\phi}_i$ is the vector of parameters and residual variance

$(\theta_1,\theta_2,\ldots\ldots..; \phi_1,\phi_2, \ldots\ldots; \sigma_i)$

σ_i is the residual variance and

n_i is the number of parameters

- As the number of parameters increase, the likelihood value decreases.
- The ML rule selects the models with a small number of parameters (principle of parsimony)

ARIMA Models

Mean square error (Prediction approach):

- Using a portion of available data (N/2) estimate the parameters of different models
- Forecast the series one step ahead by using the candidate models
- Estimate the MSE corresponding to each model
- The model with least value of MSE is selected for prediction

ARIMA Models

The one step ahead forecast for ARMA(p, q) is

$$\hat{X}_{t+1} = \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{j=1}^{q} \phi_j e_{t-j}$$

The error for one step ahead forecast is

$$e_{t+1} = X_{t+1} - \hat{X}_{t+1}$$

If the series consists on N observations, the first N/2 observations are used for parameter estimation and N/2+1 to N are used for error series calculation.

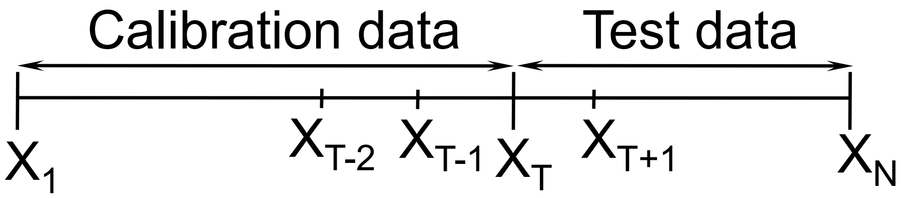
ARIMA Models

The MSE for model is

$$MSE = \frac{\sum_{i=\frac{N}{2}+1}^{N} e_i^2}{N/2}$$

ARIMA Models

3. Model testing / Validation:



First ‘T’ values are used to build the model (say 50% of the available data) and the rest of data is used to validate the model.

All the tests are carried out on the residual series only.

ARIMA Models

The tests are performed to examine whether the following assumptions used in building the model are valid for the model selection

- The residual series has zero mean
- No significant periodicities are present in the residual series
- The residual series is uncorrelated

$$\underset{\text{Residual}}{e_t} = \underset{\text{Data}}{X_t} - \underbrace{\left(\sum_{j=1}^{m_1} \phi_j X_{t-j} + \sum_{j=1}^{m_2} \theta_j e_{t-j} \right)}_{\text{Simulated from the model}}$$

ARIMA Models

Validation tests are listed here

- Significance of residual mean
- Significance of periodicities
- Cumulative periodogram test or Bartlett’s test
- White noise test
  - Whittle’s test
  - Portmanteau test

ARIMA Models

Significance of residual mean:

- This test examine the validity of the assumption that the error series e(t) has zero mean
- A statistic η(e) is defined as

$$\eta\left(e\right)=\frac{N^{1/2}\overline{e}}{\hat{\rho}^{1/2}}$$

Where

$\overline{e}$ is the estimate of the residual mean

$\hat{\rho}$ is the estimate of the residual variance

Ref: Kashyap R.L. and Ramachandra Rao.A, “Dynamic stochastic models from empirical data”, Academic press, New York , 1976

ARIMA Models

- The statistic $\eta(e)$ is approximately distributed as t (α, N–1), where α is the significance level at which the test is being carried out.
- If the value of $\eta(e) \leq t(\alpha, N-1)$, then the mean of the residual series is not significantly different from zero – series passes the test.

ARIMA Models

Significance of periodicities:

- This test ensures that no significant periodicities are present in the residual series
- The test is conducted for different periodicities and the significance of each of the periodicities is tested.
- A statistic η(e) is defined as

$$\eta\left(e\right)=\frac{\gamma^{2}\left(N-2\right)}{4\hat{\rho}_{1}}$$

ARIMA Models

Where $\gamma^2 = \alpha^2 + \beta^2$

$$\hat{\rho}_1 = \frac{1}{N}\left[\sum_{t=1}^{N}\left\{e_t - \hat{\alpha}\cos\left(\omega_k t\right) - \hat{\beta}\sin\left(\omega_k t\right)\right\}^2\right]$$

$$\alpha_k = \frac{2}{N}\sum_{t=1}^{n} e_t \cos\left(\omega_k t\right)$$

$$\beta_k = \frac{2}{N}\sum_{t=1}^{n} e_t \sin\left(\omega_k t\right)$$

$2\pi/\omega_k$ is the periodicity for which test is being carried out.

ARIMA Models

- The statistic $\eta(e)$ is approximately distributed as $F_{\alpha}(2, N–2\ )$, where α is the significance level at which the test is being carried out.
- If the value of $\eta(e) \leq F_{\alpha}(2, N–2\ )$, then the periodicity is not significant.

ARIMA Models

Cumulative periodogram test or Bartlett's test :

- This test is also carried out to ensure that no significant periodicities are present in the residual series
- This test is conducted to detect the first significant periodicity in the series.
- If significant periodicity is observed, the first periodicity is removed and new series is obtained for which the test is repeated and checked for periodicity and so on.

ARIMA Models

$$\gamma_k^2 = \left\{\frac{2}{N}\sum_{t=1}^{N} e_t \cos\left(\omega_k t\right)\right\}^2 + \left\{\frac{2}{N}\sum_{t=1}^{N} e_t \sin\left(\omega_k t\right)\right\}^2$$

k = 1,2,……N/2

$$g_k = \frac{\sum_{j=1}^{k} \gamma_j^2}{\sum_{k=1}^{N/2} \gamma_k^2}$$

$0 \leq g_k \leq 1$

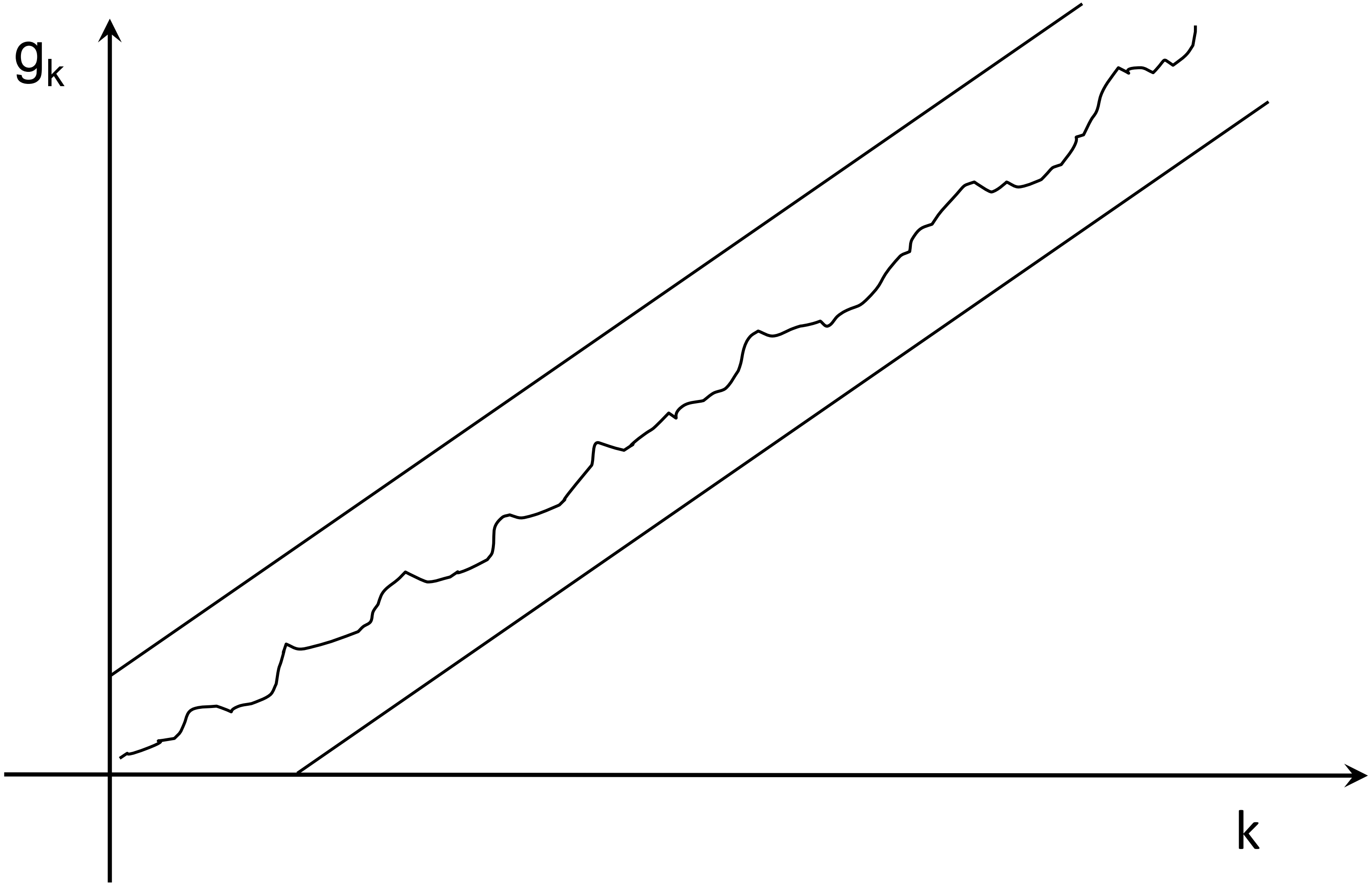
The plot of g_k vs k is called as cumulative periodogram

Ref: Kashyap R.L. and Ramachandra Rao.A, “Dynamic stochastic models from empirical data”, Academic press, New York , 1976

ARIMA Models

- On the cumulative periodogram two confidence limits ($\pm\lambda/(N/2)^{1/2}$) are drawn
- The value of λ prescribed for 95% confidence limits is 1.35 and for 99% confidence limits is 1.65
- If all the values of g_k lie within the significance band, there is no significant periodicities in the series.
- If a value of g_k lies outside the significance band, the periodicity corresponding to that value of g_k is significant.

ARIMA Models



ARIMA Models

White noise test (Whittle's test):

- This test is carried out to test the absence of correlation in the series.
- The covariance r_k at lag k of the error series e(t)

$$r_k = \frac{1}{N-k} \sum_{j=k+1}^{N} e_j e_{j-k} \qquad k = 0, 1, 2, \ldots\ldots k_{max}$$

- The value of k_{max} is normally chosen as 0.15N

Ref: Kashyap R.L. and Ramachandra Rao.A, "Dynamic stochastic models from empirical data", Academic press, New York

ARIMA Models

- The covariance matrix is

$$\Gamma_{n1} = \begin{bmatrix} r_0 & r_1 & r_2 & . & . & r_{k_{\max}} \\ r_1 & r_0 & r_1 & . & . & r_{k_{\max}-1} \\ r_2 & & & & & \\ . & & & & & \\ . & & & & & \\ r_{k_{\max}} & r_{k_{\max}-1} & & & & r_0 \end{bmatrix}_{k_{\max} \times k_{\max}}$$

ARIMA Models

- A statistic η(e) is defined as

$$\eta(e) = \frac{N}{n1-1}\left(\frac{\hat{\rho}_0}{\hat{\rho}_1} - 1\right)$$

Where $\hat{\rho}_0$ is the lag zero correlation and

$$\hat{\rho}_1 = \frac{\det \Gamma_{n1}}{\det \Gamma_{n1-1}}$$

The matrix $\Gamma_{\text{n1-1}}$ is constructed by eliminating the last row and the last column from the Γ_{n1} matrix.

ARIMA Models

- The statistic $\eta(e)$ is approximately distributed as $F_{\alpha}(n1, N–n1\ )$, where α is the significance level at which the test is being carried out.
- If the value of $\eta(e) \leq F_{\alpha}(n1, N–n1\ )$, then the residual series is uncorrelated.

ARIMA Models

White noise test (Portmanteau test):

- This test is also carried out to test the absence of correlation in the series.
- This test also uses the covariance r_k defined earlier.
- A statistic η(e) is defined as

$$\eta(e) = (N - n1) \sum_{k=1}^{n1} \left(\frac{r_k}{r_0} \right)^2$$

Ref: Kashyap R.L. and Ramachandra Rao.A, "Dynamic stochastic models from empirical data", Academic press, New York

ARIMA Models

- The statistic η(e) is approximately distributed as χ^2_α(n1), where α is the significance level at which the test is being carried out.
- The value of n1 is normally chosen as 0.15N
- If the value of η(e) $\leq$ χ^2_α(n1), then the residual series is uncorrelated.

ARIMA Models

Data Generation:

Consider AR(1) model,

$$X_t = \phi_1 X_{t-1} + e_t$$

ϕ_1 = 0.5 therefore AR(1) model is

$X_t = 0.5X_{t-1} + e_t \longrightarrow$ Choose e_t terms with zero mean and uncorrelated

Let us choose standard normal deviates

ARIMA Models

Say $X_1 = 3.0$

$X_2 = 0.5*3.0 + 0.335$
$= 1.835$

$X_3 = 0.5*1.835 + 1.226$
$= 2.14$

And so on…

ARIMA Models

Consider ARMA(1, 1) model,

$$X_t = \phi_1 X_{t-1} + \theta_1 e_{t-1} + e_t$$

$\phi_1 = 0.5$, $\theta_1 = 0.4$ therefore the model is

Standard normal deviates

$$X_t = 0.5X_{t-1} + 0.4e_{t-1} + e_t$$

Choose e_{t-1} terms as previous e_t and set initial value as zero

ARIMA Models

Say $X_1 = 3.0$

$$X_2 = 0.5*3.0 + 0.4*0 + 0.667$$
$$= 2.167$$

$$X_3 = 0.5*2.167 + 0.4*0.667 + 1.04$$
$$= 2.39$$

$$X_4 = 0.5*2.39 + 0.4*1.04 + 2.156$$
$$= 3.767$$

and so on...

ARIMA Models

Data Forecasting:

Consider AR(1) model,

$X_t = \phi_1 X_{t-1} + e_t$

Expected value is considered.

$$E\left[X_t\right] = \phi_1 E\left[X_{t-1}\right] + E\left[e_t\right]$$

$$\hat{X}_t = \phi_1 X_{t-1}$$

Expected value of e_t is zero

ARIMA Models

Consider ARMA(1, 1) model,

$$X_t = \phi_1 X_{t-1} + \theta_1 e_{t-1} + e_t$$

$$E[X_t] = \phi_1 X_{t-1} + \theta_1 e_{t-1} + 0$$

Error in forecast in the previous period

$\phi_1 = 0.5$, $\theta_1 = 0.4$ therefore the model is

$$X_t = 0.5X_{t-1} + 0.4e_{t-1}$$

ARIMA Models

Say $X_1 = 3.0$

Initial error assumed to be zero

$$\hat{X}_2 = 0.5 \times 3.0 + 0.4 \times 0$$
$$= 1.5$$

$X_2 = 2.8$

Error $e_2 = 2.8 – 1.5 = 1.3$

$$\hat{X}_3 = 0.5 \times 2.8 + 0.4 \times 1.3$$
$$= 1.92$$

Actual value to be used

ARIMA Models

$X_3 = 1.8$

Error $e_3 = 1.8 - 1.92 = -0.12$

$$\begin{aligned}\hat{X}_4 &= 0.5 \times 1.8 + 0.4 \times (-0.12) \\ &= 0.852\end{aligned}$$

and so on...

Markov Chains

Markov Chains:

- Markov chain is a stochastic process with the property that value of process X_t at time t depends on its value at time t-1 and not on the sequence of other values (X_{t-2} , X_{t-3},……. X_0) that the process passed through in arriving at X_{t-1}.

$$P\left[X_t / X_{t-1}, X_{t-2}, ...X_0\right] = P\left[X_t / X_{t-1}\right]$$

Single step Markov chain

Markov Chains

$$P\left[X_t = a_j / X_{t-1} = a_i\right]$$

- The conditional probability gives the probability at time t will be in state ‘j’, given that the process was in state ‘i’ at time t-1.
- The conditional probability is independent of the states occupied prior to t-1.
- For example, if X_{t-1} is a dry day, what is the probability that X_t is a dry day or a wet day.
- This probability is commonly called as transitional probability

Markov Chains

$$P\left[X_t = a_j / X_{t-1} = a_i\right] = P_{ij}^t$$

- Usually written as P_{ij}^t indicating the probability of a step from a_i to a_j at time ‘t’.
- If P_{ij} is independent of time, then the Markov chain is said to be homogeneous.

i.e., $P_{ij}^t = P_{ij}^{t+\tau}$ $\forall$ t and τ

the transitional probabilities remain same across time

Markov Chains

Transition Probability Matrix(TPM):

$$P = \begin{array}{cc} \begin{matrix} t+1 \rightarrow \\ t \\ \downarrow \end{matrix} & \begin{array}{cccccc} 1 & 2 & 3 & . & . & m \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ . \\ . \\ m \end{array} & \begin{bmatrix} P_{11} & P_{12} & P_{13} & . & . & P_{1m} \\ P_{21} & P_{22} & P_{23} & . & . & P_{2m} \\ P_{31} & & & & & \\ . & & & & & \\ . & & & & & \\ P_{m1} & P_{m2} & & & & P_{mm} \end{bmatrix}_{m \times m} \end{array}$$

Markov Chains

$$\sum_{j=1}^{m} P_{ij} = 1 \quad \forall \text{ j}$$

- Elements in any row of TPM sum to unity (stochastic matrix)
- TPM can be estimated from observed data by tabulating the number of times the observed data went from state ‘i’ to ‘j’
- $P_j^{(n)}$ is the probability of being in state ‘j’ in the time step ‘n’.

Markov Chains

- $p_j^{(0)}$ is the probability of being in state ‘j’ in period t = 0.

$$p^{(0)} = \begin{bmatrix} p_1^{(0)} & p_2^{(0)} & . & . & p_m^{(0)} \end{bmatrix}_{1\times m}$$

…. Probability vector at time 0

$$p^{(n)} = \begin{bmatrix} p_1^{(n)} & p_2^{(n)} & . & . & p_m^{(n)} \end{bmatrix}_{1\times m}$$

…. Probability vector at time ‘n’

- Let $p^{(0)}$ is given and TPM is given

$$p^{(1)} = p^{(0)} \times P$$

Markov Chains

$$p^{(1)} = \begin{bmatrix} p_1^{(0)} & p_2^{(0)} & . & . & p_m^{(0)} \end{bmatrix} \begin{bmatrix} P_{11} & P_{12} & P_{13} & . & . & P_{1m} \\ P_{21} & P_{22} & P_{23} & . & . & P_{2m} \\ P_{31} & & & & & \\ . & & & & & \\ P_{m1} & P_{m2} & & & & P_{mm} \end{bmatrix}$$

$$= p_1^{(0)} P_{11} + p_2^{(0)} P_{21} + ... + p_m^{(0)} P_{m1}$$ …. Probability of going to state 1

$$= p_1^{(0)} P_{12} + p_2^{(0)} P_{21} + ... + p_m^{(0)} P_{m2}$$ …. Probability of going to state 2

And so on…

Markov Chains

Therefore

$$p^{(1)} = \begin{bmatrix} p_1^{(1)} & p_2^{(1)} & . & . & p_m^{(1)} \end{bmatrix}_{1\times m}$$

$$p^{(2)} = p^{(1)} \times P$$

$$= p^{(0)} \times P \times P$$

$$= p^{(0)} \times P^2$$

In general,

$$p^{(n)} = p^{(0)} \times P^n$$

Markov Chains

- As the process advances in time, $p_j^{(n)}$ becomes less dependent on $p^{(0)}$
- The probability of being in state 'j' after a large number of time steps becomes independent of the initial state of the process.
- The process reaches a steady state ay very large n

$$p = p \times P^n$$

- As the process reach steady state, TPM remains constant

Example – 3

Consider the TPM for a 2-state (state 1 is non-rainfall day and state 2 is rainfall day) first order homogeneous Markov chain as

$$TPM = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$

Obtain the

1. probability of day 1 is non-rainfall day / day 0 is rainfall day
2. probability of day 2 is rainfall day / day 0 is non-rainfall day
3. probability of day 100 is rainfall day / day 0 is non-rainfall day

Example – 3 (contd.)

1. probability of day 1 is non-rainfall day / day 0 is rainfall day

$$TPM = \begin{array}{r} \\ \text{No rain} \\ \text{rain} \end{array} \begin{array}{cc} \text{No rain} & \text{rain} \\ \left[\begin{matrix} 0.7 \\ 0.4 \end{matrix}\right. & \left.\begin{matrix} 0.3 \\ 0.6 \end{matrix}\right] \end{array}$$

The probability is 0.4

2. probability of day 2 is rainfall day / day 0 is non-rainfall day

$$p^{(2)} = p^{(0)} \times P^2$$

Example – 3 (contd.)

$$p^{(2)} = \begin{bmatrix} 0.7 & 0.3 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}$$

$$= \begin{bmatrix} 0.61 & 0.39 \end{bmatrix}$$

The probability is 0.39

3. probability of day 100 is rainfall day / day 0 is non-rainfall day

$$p^{(n)} = p^{(0)} \times P^n$$

Example – 3 (contd.)

$$P^2 = P \times P$$

$$= \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{bmatrix}$$

$$P^4 = P^2 \times P^2 = \begin{bmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{bmatrix}$$

$$P^8 = P^4 \times P^4 = \begin{bmatrix} 0.5715 & 0.4285 \\ 0.5714 & 0.4286 \end{bmatrix}$$

$$P^{16} = P^8 \times P^8 = \begin{bmatrix} 0.5714 & 0.4286 \\ 0.5714 & 0.4286 \end{bmatrix}$$

Example – 3 (contd.)

Steady state probability

$$p = \begin{bmatrix} 0.5714 & 0.4286 \end{bmatrix}$$

For steady state,

$$p = p \times P^n$$

$$= \begin{bmatrix} 0.5714 & 0.4286 \end{bmatrix} \begin{bmatrix} 0.5714 & 0.4286 \\ 0.5714 & 0.4286 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5714 & 0.4286 \end{bmatrix}$$