INDIAN INSTITUTE OF SCIENCE

STOCHASTIC HYDROLOGY

Lecture -10

Course Instructor : Prof. P. P. MUJUMDAR

Department of Civil Engg., IISc.

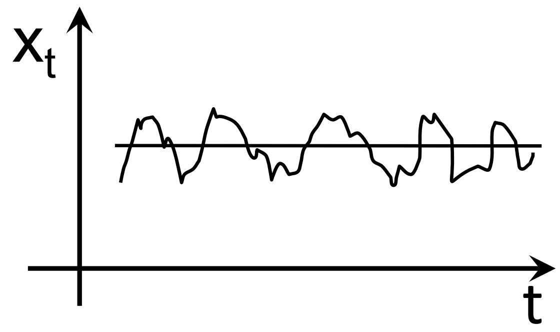
Summary of the previous lecture

- Data Generation
- Introduction to Time Series Analysis

Time Series Analysis

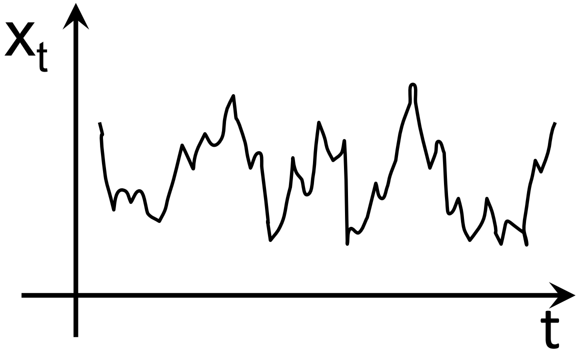
- Sequence of values of a random variable collected over time
- Discrete time series; Continuous time series
- Realization; Ensemble
- Hydrologic time series composed of deterministic and stochastic components

$$X_t = d_t + \varepsilon_t$$

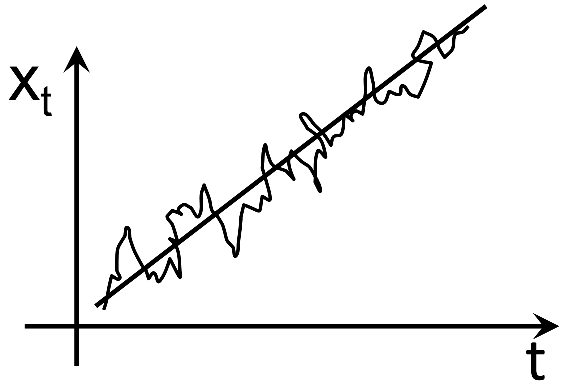


Long term mean

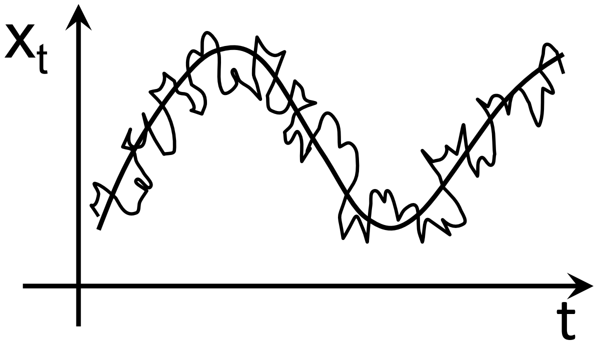
Time Series Analysis



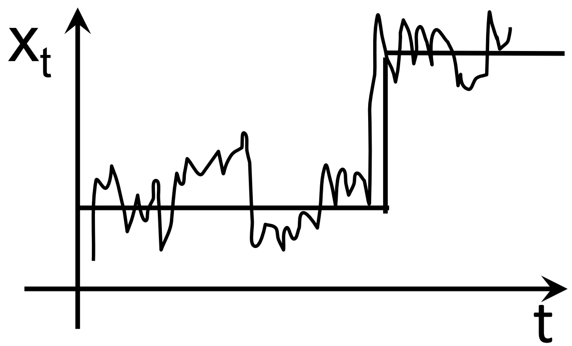
Stochastic



Stochastic + Trend



Stochastic + Periodic



Stochastic + Jump

$$X_t = d_t + \varepsilon_t$$

Time Series Analysis

- Deterministic component is a combination of a long term mean, trend, periodicity and jump.

- Time scale of time series – either discrete or continuous
- Discrete time scale: observations at specific times separated by Δt. (eg., average monthly stream flow, annual peak discharge, daily rainfall etc.)
- Continuous time scale: data recorded continuously with time (eg., turbulence studies, pressure measurements)

Time Series Analysis

- The pdf of a stochastic process $X(t)$ is $f(x; t)$
- $f(x; t)$ describes the probabilistic behavior of $X(t)$ at specified time ‘t’
- The time series is said to be stationary, if the properties do not change with time.
- $f(x; t) = f(x; t+\tau) \ \forall \ t$
- for stationary time series, pdf of X_t is same as that of $X_{t+\tau} \ \forall \ t$

Time Series Analysis

Time average for a realization

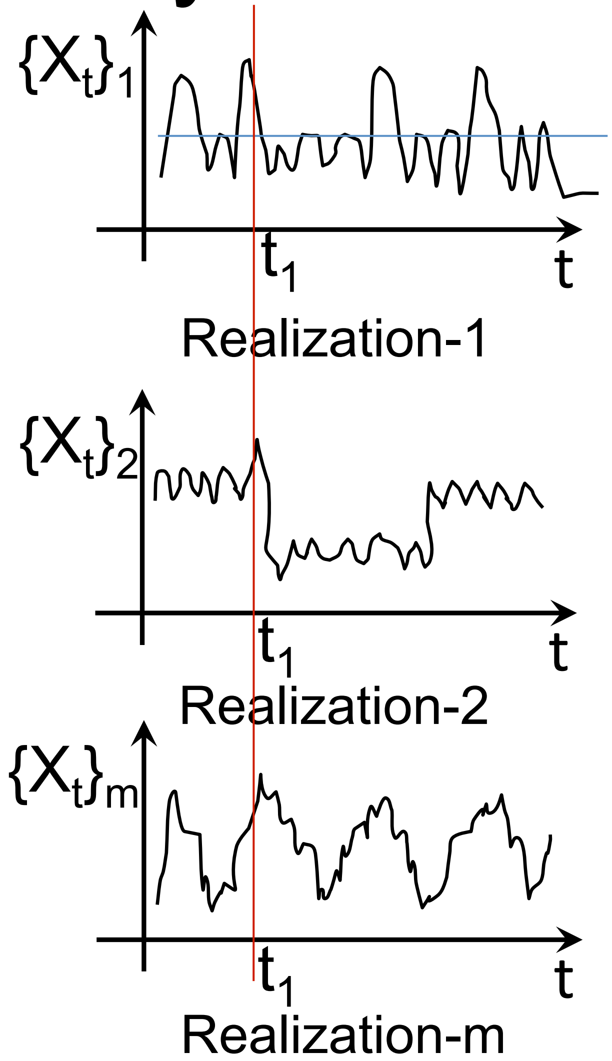
$$\bar{X}_1 = \frac{\sum_{j=1}^{n} \left\{ X_j(t) \right\}_1}{n}$$

n is no. of observations

Ensemble average at time t

$$\bar{X}_t = \frac{\sum_{i=1}^{m} X_i(t)}{m}$$

m is no. of realizations



Time Series Analysis

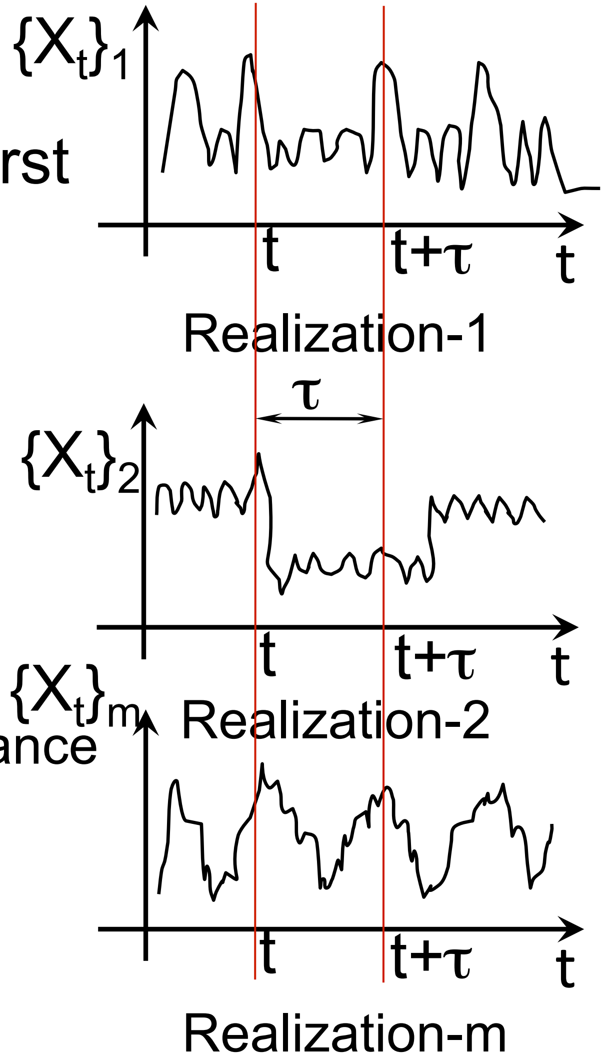
- If $\bar{X}_t = \bar{X}_{t+\tau}$ for all t, then the process is stationary in mean (first order stationary)
- If all the moments up to order 'f' are same for time t and t+τ, ∀ t then the time series is weakly stationary of order 'k'

  k = 1 Stationary in mean

  k = 2 Stationary in mean & covariance

- For a strictly stationary time series,

  $f(x_1) = f(x_2) = \ldots\ldots\ldots = f(x)$



Time Series Analysis

- Auto covariance

$$\gamma_k = \text{cov}\left(X_t, X_{t+k}\right)$$

$$= E\left[\left(X_t - \mu\right)\left(X_{t+k} - \mu\right)\right]$$

$$\gamma_0 = \sigma_X^2$$

- Auto correlation between X_t and $X_{t+\tau}$,

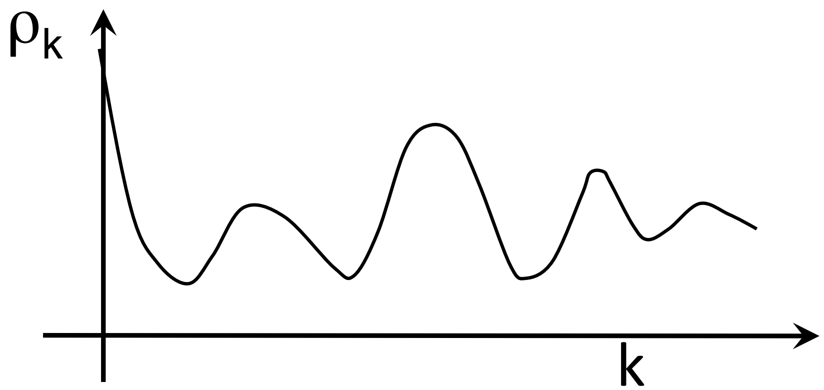
$$\rho_k = \frac{\text{cov}\left(X_t, X_{t+k}\right)}{\sigma_{X_t}\sigma_{X_{t+k}}}$$

$$= \frac{\text{cov}\left(X_t, X_{t+k}\right)}{\sigma_X^2} = \frac{\gamma_k}{\gamma_0}$$

$$\rho_0 = 1$$

If process is stationary
$\sigma_{X_t} = \sigma_{X_{t+k}}$

Time Series Analysis



Correlogram

- Auto correlation indicates the memory of a stochastic process

Time Series Analysis

- Auto covariance matrix

$$
\Gamma_n = \begin{array}{c} \\ X_1 \\ X_2 \\ X_3 \\ . \\ . \\ X_n \end{array}
\begin{array}{c} \begin{array}{cccccc} X_1 & X_2 & X_3 & . & . & X_n \end{array} \\
\begin{bmatrix}
\gamma_0 & \gamma_1 & \gamma_2 & \cdot & \cdot & \gamma_{n-1} \\
\gamma_1 & \gamma_0 & \gamma_1 & \cdot & \cdot & \gamma_{n-2} \\
\gamma_2 & & & & & \\
\cdot & & & & & \\
\cdot & & & & & \\
\gamma_{n-1} & \gamma_{n-1} & & & & \gamma_0
\end{bmatrix}_{n \times n} \end{array}
$$

Γ_n is symmetric and +ve definite matrix

Time Series Analysis

- Dividing the matrix Γ_n by γ_o, we get the auto correlation matrix P_n

$$P_n = \frac{\Gamma_n}{\gamma_0} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & . & . & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & . & . & \rho_{n-2} \\ \rho_2 & & & & & \\ . & & & & & \\ . & & & & & \\ \rho_{n-1} & \rho_{n-2} & . & . & . & 1 \end{bmatrix}_{n \times n}$$

P_n is symmetric and +ve definite matrix

Time Series Analysis

- Because P_n is +ve definite

$$\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} \geq 0$$

$$1 - \rho_1^2 \geq 0$$

$$-1 \leq \rho_1 \leq 1$$

Time Series Analysis

- Sample estimates:

$$\gamma_k = E\left[\left(X_t - \mu\right)\left(X_{t+k} - \mu\right)\right]$$

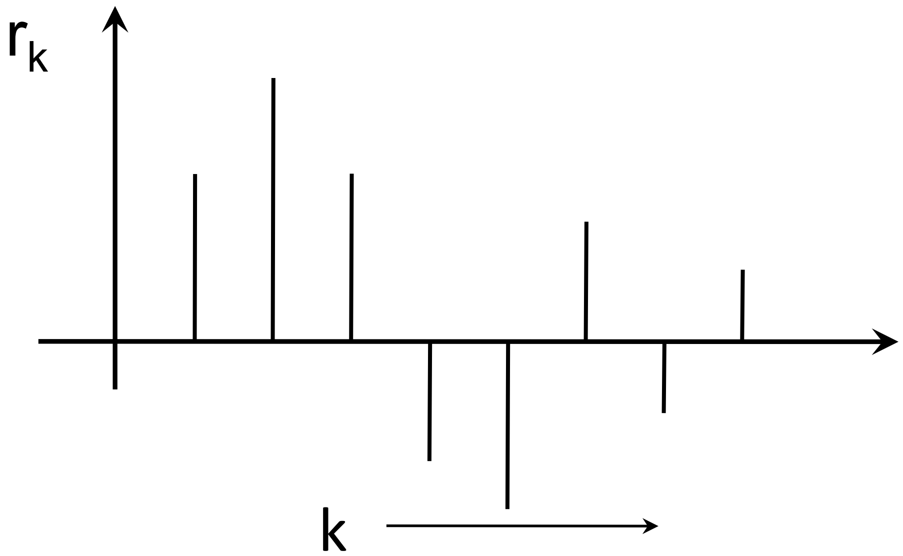
$$c_k = \frac{1}{N}\sum_{i=1}^{n-k}\left(X_t - \bar{X}\right)\left(X_{t+k} - \bar{X}\right)$$ ….. Sample estimate of auto covariance

$$r_k = \frac{c_k}{c_0}$$

$\rho_k = \frac{\gamma_k}{\gamma_0}$

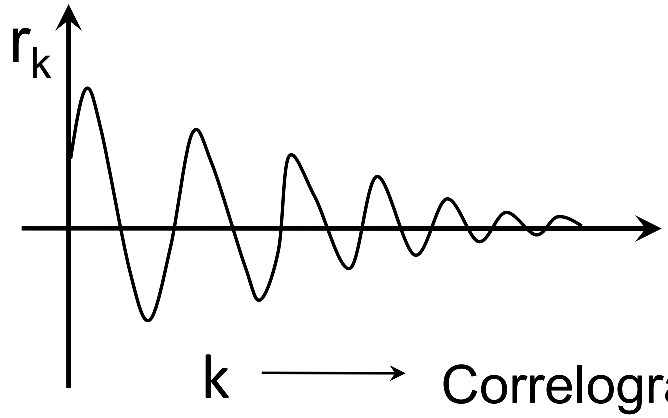
$$c_0 = S_X^2$$

(S_X^2 →) variance



Time Series Analysis

- Auto correlation function (r_k)



Correlogram

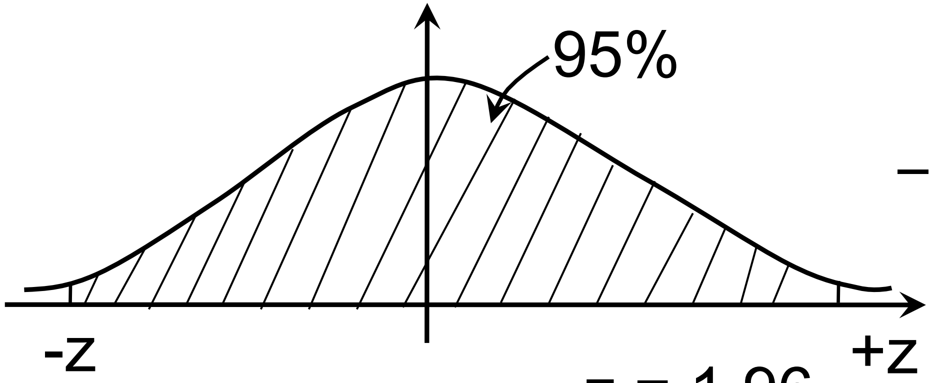
If it is purely stochastic (random) series,

$\rho_k = 0, \quad \forall \quad k = 1, 2, 3, \ldots\ldots..$

r_k = may not be zero (because r_k is a sample estimate)

$$r_k : \text{Normal Distribution}\left(0, \frac{1}{\sqrt{N}}\right)$$ For a random series

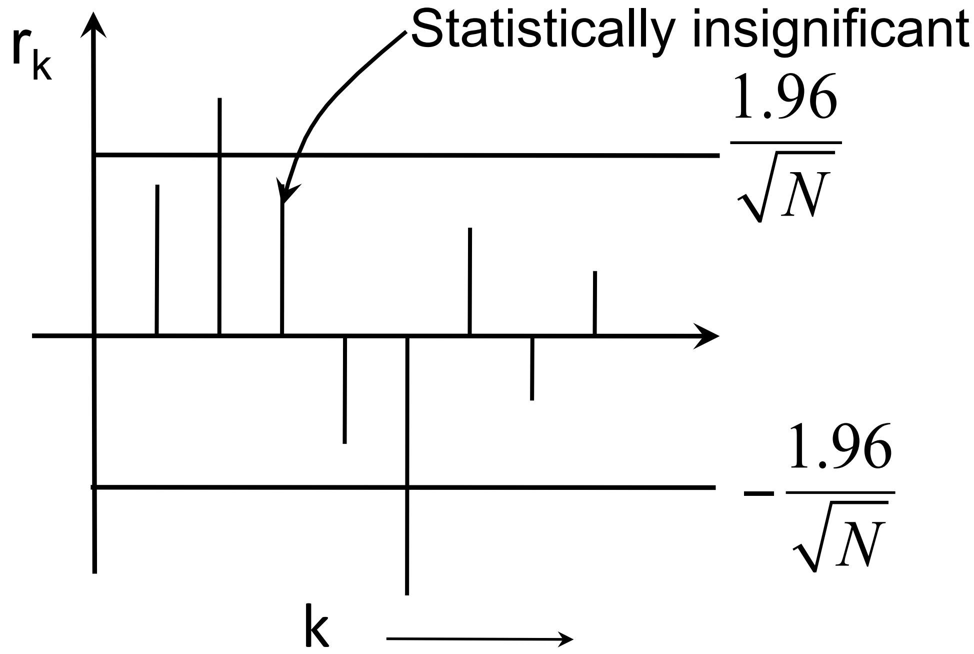
Time Series Analysis



z = 1.96

$$-\frac{1.96}{\sqrt{N}} \le r_k \le +\frac{1.96}{\sqrt{N}}$$

For a random series



Example-1

Obtain Auto correlation for k=1

S.No.	X_t	$(x_t - \bar{x})$	X_{t+1}	$(x_{t+1} - \bar{x})$	$(x_t - \bar{x}) \times (x_{t+1} - \bar{x})$
1	97	-10.50	110	2.5	-26.25
2	110	2.50	121	13.5	33.75
3	121	13.50	117	9.5	128.25
4	117	9.50	79	-28.5	-270.75
5	79	-28.50	140	32.5	-926.25
6	140	32.50	75	-32.5	-1056.25
7	75	-32.50	127	19.5	-633.75
8	127	19.50	90	-17.5	-341.25
9	90	-17.50	119	11.5	-201.25
10	119	11.50			
Σ	1075				-3293.75

Example-1 (contd.)

mean $\bar{x}$ = 1075/10

= 107.5

Variance, $c_0 = \dfrac{\sum_{t=1}^{n}(x_t - \bar{x})^2}{n-1} = \dfrac{4132.5}{10-1} = 459.2$

$c_1 = \dfrac{\sum_{t=1}^{n-1}(x_t - \bar{x})(x_{t+1} - \bar{x})}{n} = \dfrac{-3293.75}{10} = 329.375$

$r_1 = \dfrac{c_1}{c_0} = \dfrac{-329.375}{459.2} = -0.72$

$r_1 = -0.72$

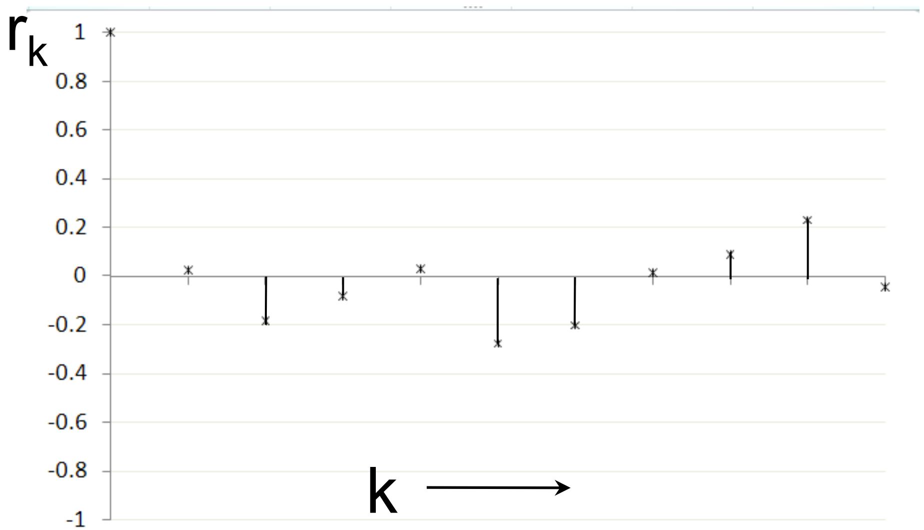
Example-2

Obtain correlogram for 40 uniformly distributed random numbers

S.No.	Data	S.No.	Data	S.No.	Data	S.No.	Data
1	98	11	73	21	25	31	89
2	69	12	36	22	49	32	70
3	30	13	11	23	73	33	36
4	50	14	54	24	38	34	42
5	93	15	31	25	14	35	84
6	1	16	74	26	4	36	82
7	66	17	23	27	87	37	55
8	99	18	88	28	99	38	93
9	76	19	82	29	69	39	2
10	65	20	92	30	57	40	43

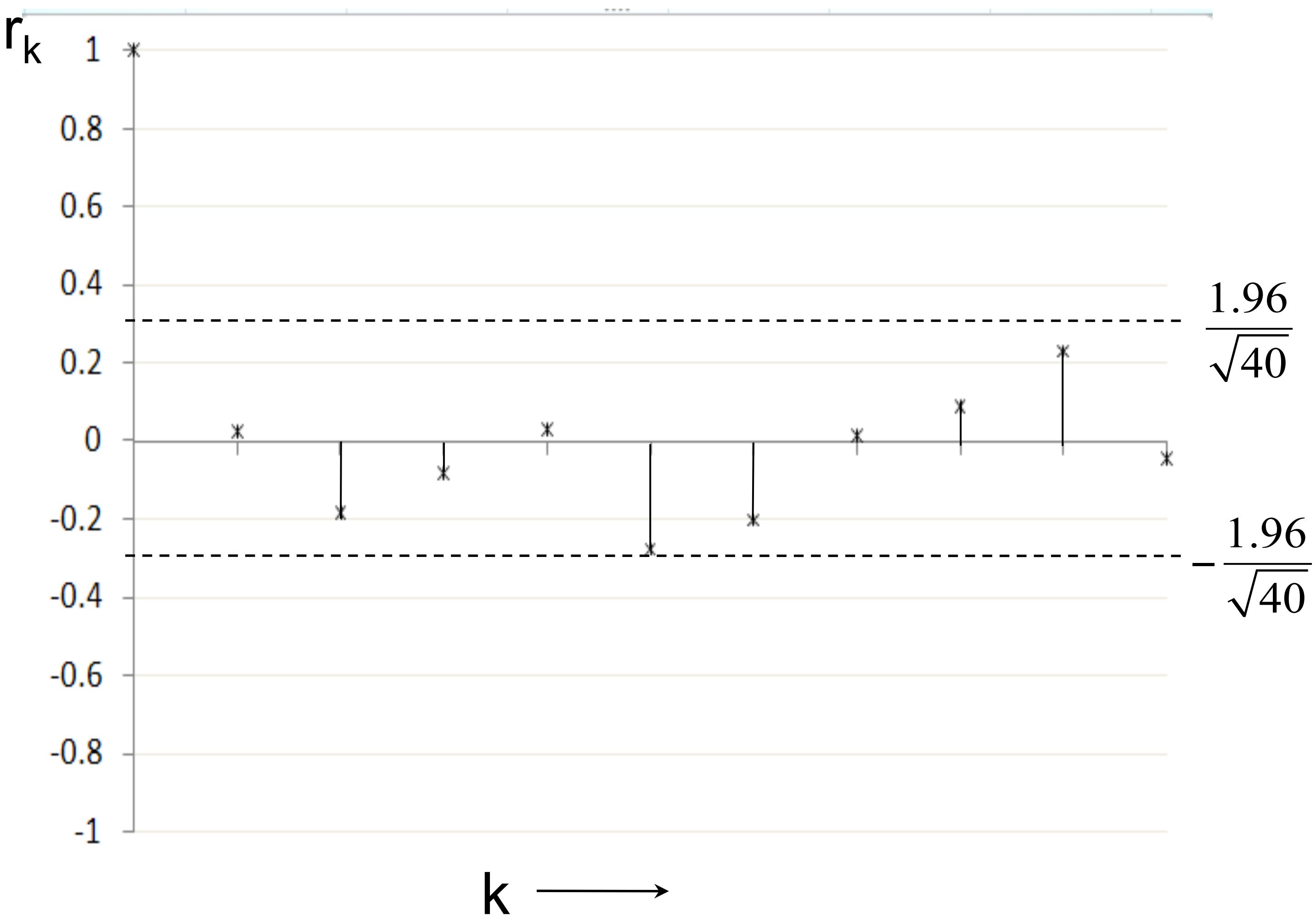
Example-2 (contd.)

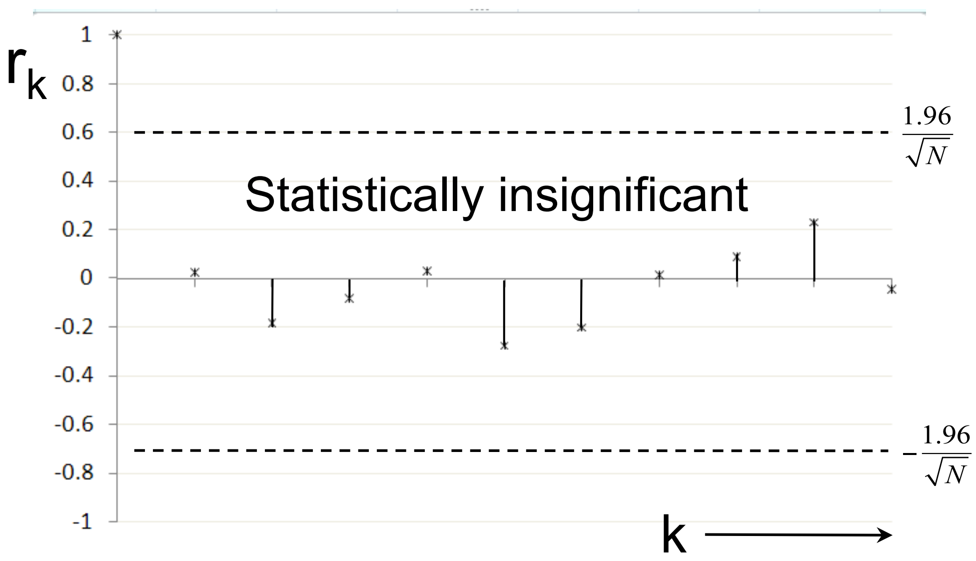
k	r_k
0	1
1	0.0235
2	-0.183
3	-0.0813
4	0.0315
5	-0.277
6	-0.202
7	0.0152
8	0.089
9	0.2304
10	-0.0435



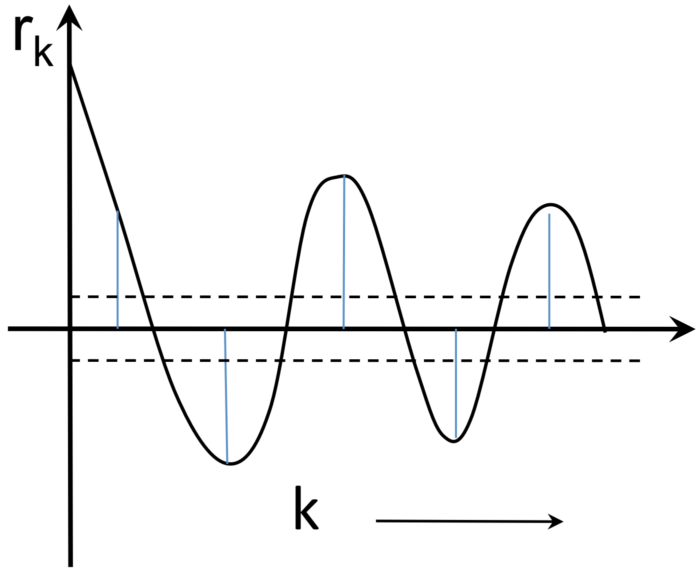
$$\frac{1.96}{\sqrt{N}} = \frac{1.96}{\sqrt{40}} = 0.31$$

Example-2 (contd.)





Purely stochastic process

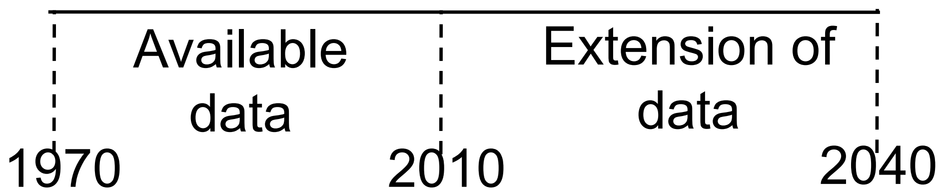


Periodic process

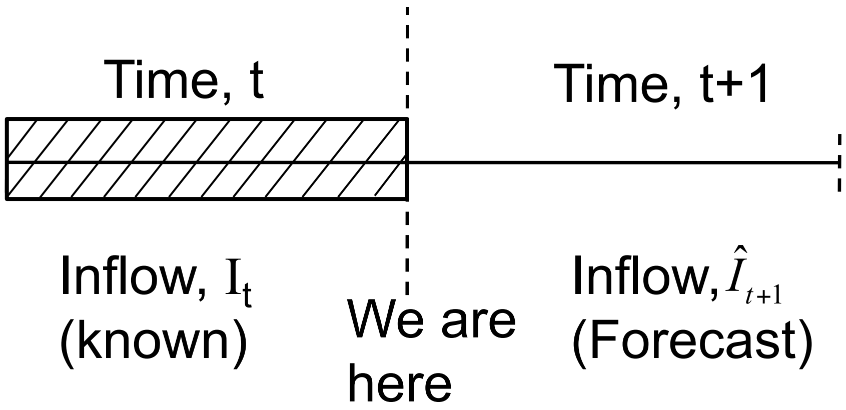
DATA EXTENSION & FORECASTING

Data Extension & Forecasting

e.g., Stream flow records for reservoir planning

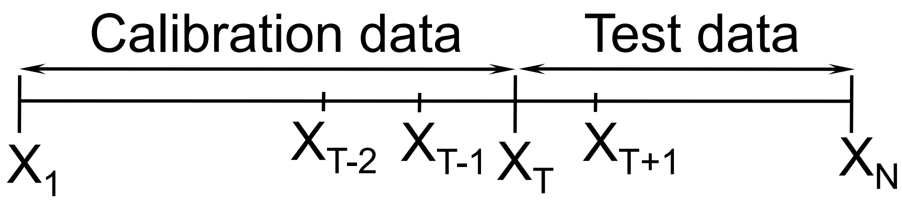


Data forecasting



$$\hat{I}_{t+1} = f\left[I_t, I_{t-1}, \ldots\ldots\ldots\right]$$

Data Extension & Forecasting



Use first 'T' values to build the model, use rest of data to validate it

F_{T+1}, F_{T+2} ……… F_N : forecasts obtained from the model

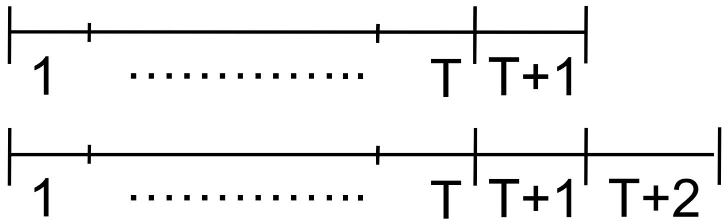
$(X_{T+1} – F_{T+1})$
$(X_{T+2} – F_{T+2})$
.
.
$(X_N – F_N)$

} Forecast errors

Data Extension & Forecasting

Method of simple averages: take the average of all the data up to period 'T' as the forecast for period (T+1)

$$\hat{X}_{t+1} = F_{T+1} = \frac{\sum_{t=1}^{T} X_t}{T}$$



$$\hat{X}_{t+2} = F_{T+2} = \frac{\sum_{t=1}^{T+1} X_t}{T+1}$$ and so on

For series with jumps & trends this is not a good procedure

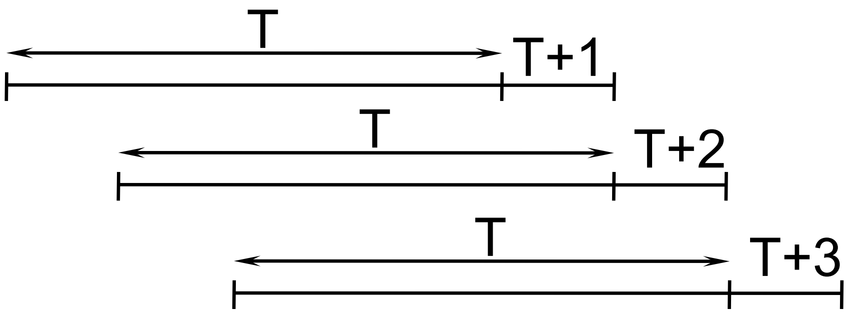
Example-3

Data	Forecast
105	-
115	110
103	107.67
108	107.75
120	110.2
97	108
110	108.28
121	109.87
117	110.67
79	107.5

Data Extension & Forecasting

Smoothening technique:

Moving Average (MA)



•As a new observation becomes available, new average is computed by dropping the oldest observation and including the newest one.

•No. of data points in each average remains constant

•Deals with the latest ‘T’ periods of known data

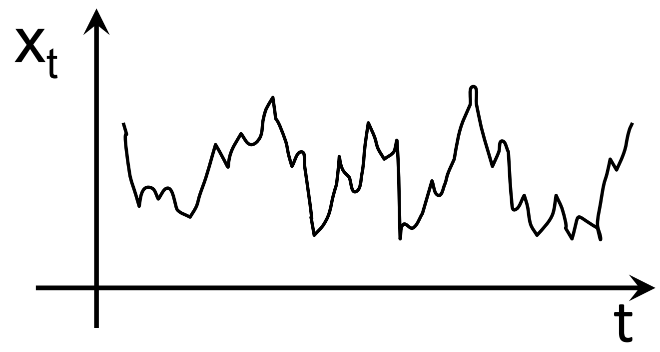
Example-4

Data	MA (3)	MA (3 x 3)
105	-	
115	-	
103	-	
108	107.67	
120	108.67	
97	110.33	108.89
110	108.33	109.11
121	109	109.22
117	109.33	108.89
79	116	111.44

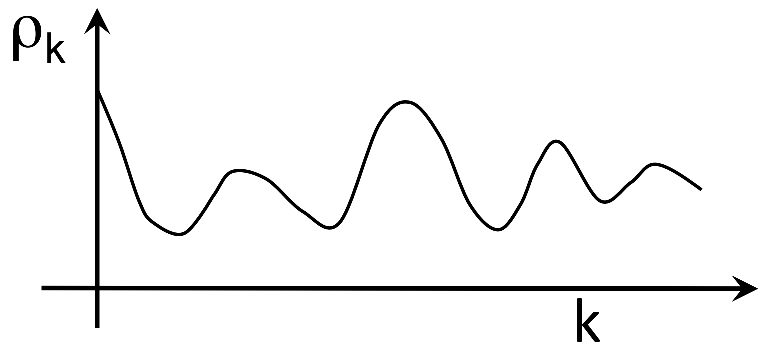
Data Generation – Serially Correlated Data

Purely random stochastic process:

Plot the time series

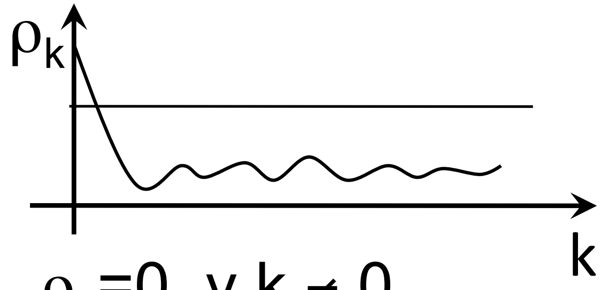


Plot the correlogram



Data Generation – Serially Correlated Data

If the correlogram indicate the time series is purely random



$\rho_k = 0, \forall k \neq 0$

- X_t , X_{t-k} are independent
- Distribution of X_t is known
- Generate X_t using data generation technique to follow given distribution with parameters estimated from sample

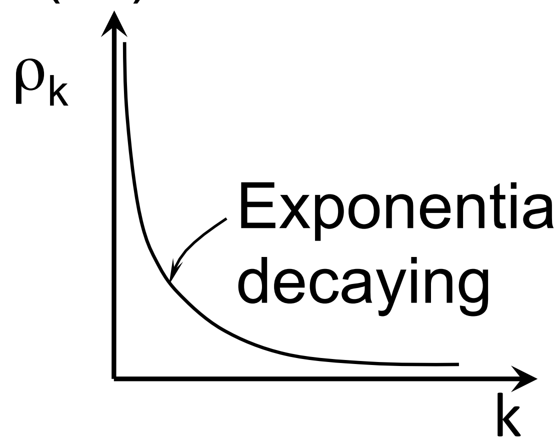
Data Generation – Serially Correlated Data

- Mainly used for flood peaks, storm intensities etc.
- Not useful for stream flows, seasonal rainfall.
- Most hydrologic time series exhibit serial dependence

  e.g., $X(t)$ correlated with $X(t-\tau)$

$\rho_k \cong (\rho_1)k$

$\rho_k \rightarrow 0,\ k \rightarrow \infty$



For First order Markov process

Data Generation – Serially Correlated Data

First order Markov process:

Random component

$$X_{t+1} = \underbrace{\mu_x + \rho_1 (X_t - \mu_x)}_{\text{Deterministic component}} + \varepsilon_{t+1}$$

Deterministic component

$\varepsilon \sim$ Mean 0 and variance σ_ε^2

This model is stationary w.r.t both mean and variance

Data Generation – Serially Correlated Data

$E[X_{t+1}] = E[\mu_x + \rho_1 (X_t - \mu_x) + \varepsilon_{t+1}]$

$= E[\mu_x] + \rho_1\{E[X_t] - E[\mu_x]\} + E[\varepsilon_{t+1}]$

$= \mu_x + \rho_1(\mu_x - \mu_x) + 0$

$= \mu_x$

$$\sigma_X^2 = E\left[X^2\right] - \left(E\left[X\right]\right)^2$$

$$= E\left[\left(\mu_x + \rho_1\left(X_t - \mu_x\right) + \varepsilon_{t+1}\right)^2\right] - \left(E\left[X_{t+1}\right]\right)^2$$

Data Generation – Serially Correlated Data

$$\sigma_X^2 = E\left[\mu_x^2 + \rho_1^2\left(X_t - \mu_x\right)^2 + \varepsilon_{t+1}^2 + 2\mu_x\rho_1\left(X_t - \mu_x\right) + \right.$$

$$\left. + 2\varepsilon_{t+1}\rho_1\left(X_t - \mu_x\right) + 2\mu_x\varepsilon_{t+1}\right] - \left(E\left[X_{t+1}\right]\right)^2$$

$$= E\left[\mu_x^2\right] + \rho_1^2 E\left[\left(X_t - \mu_x\right)^2\right] + E\left[\varepsilon_{t+1}^2\right] + 2\mu_x\rho_1 E\left[\left(X_t - \mu_x\right)\right]$$

$$+ 2\rho_1 E\left[\varepsilon_{t+1}\right] E\left[\left(X_t - \mu_x\right)\right] + 2\mu_x E\left[\varepsilon_{t+1}\right] - \left(E\left[X_{t+1}\right]\right)^2$$

$$= \mu_x^2 + \rho_1^2 E\left[\left(X_t - \mu_x\right)^2\right] + E\left[\varepsilon_{t+1}^2\right] + 0 + 0 + 0 - \mu_x^2$$

$$= \rho_1^2\sigma_X^2 + \sigma_\varepsilon^2$$

$$\sigma_\varepsilon^2 = \rho_1^2\left(1 - \sigma_X^2\right)$$

Data Generation – Serially Correlated Data

If $X \sim N(\mu_x, \sigma_x^2)$ then $\varepsilon \sim N(0, \sigma_\varepsilon^2)$

If $u \sim N(0, 1)$, $u\sigma_\varepsilon \left(\text{i.e., } u\sigma_x\sqrt{1-\rho_1^2}\right)$ is $N(0, \sigma_\varepsilon^2)$

$$X_{t+1} = \mu_x + \rho_1\left(X_t - \mu_x\right) + u\sigma_x\sqrt{1-\rho_1^2}$$

Standard normal deviate

First order stationary Markov model
Or
Thomas Fiering model (Stationary)

Data Generation – Serially Correlated Data

To generate data using First order Markov model,

$$X_{t+1} = \mu_x + \rho_1 \left(X_t - \mu_x \right) + u\sigma_x \sqrt{1-\rho_1^2}$$

- Known sample estimates of μ_x, σ_x, ρ_1
- Assume X_1 (normally assumed to be μ_x)
- Generate values from X_2………
- Generate large set of values and discard first 50-100 values to ensure that the effect of initial value dies down
- Negative value: retain it for generating next value, set it to zero.

Example-5

$\mu_x = 50$, $\sigma_x = 30$, $\rho_1 = 0.5$

Assume $X_1 = \mu_x = 50$

$$X_2 = \mu_x + \rho_1\left(X_t - \mu_x\right) + u\sigma_x\sqrt{1-\rho_1^2}$$
$$= 50 + 0.5\left(50 - 50\right) + \left(-0.464\right)30\sqrt{1-0.5^2}$$
$$= 37.165$$

$$X_3 = 50 + 0.5\left(37.165 - 50\right) + \left(0.335\right)30\sqrt{1-0.5^2}$$
$$= 52.3$$

Example-5 (contd.)

$$X_4 = 50 + 0.5\left(52.3 - 50\right) + \left(-0.051\right)30\sqrt{1 - 0.5^2}$$

$= 49.82$

$$X_5 = 50 + 0.5\left(49.82 - 50\right) + \left(1.226\right)30\sqrt{1 - 0.5^2}$$

$= 81.76$