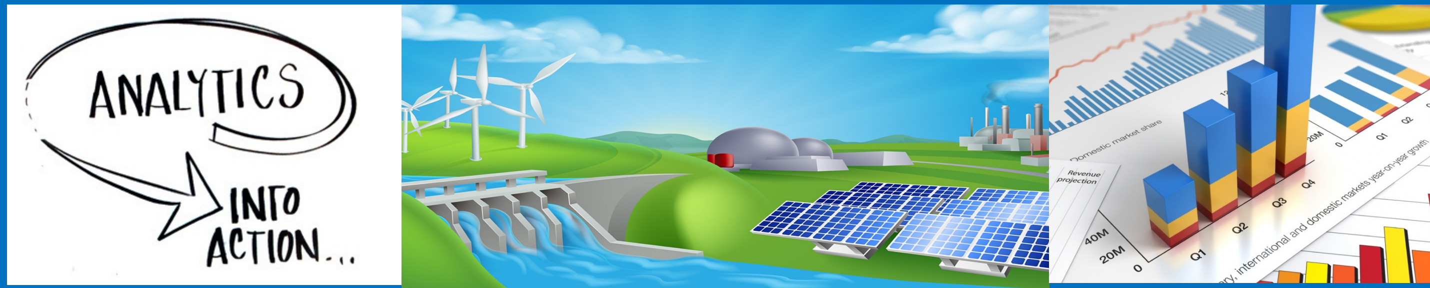


ENV797 - TIME SERIES ANALYSIS FOR ENERGY AND ENVIRONMENT APPLICATIONS

M2- Autocovariance and Autocorrelation Function

Prof. Luana Medeiros Marangon Lima, Ph.D.

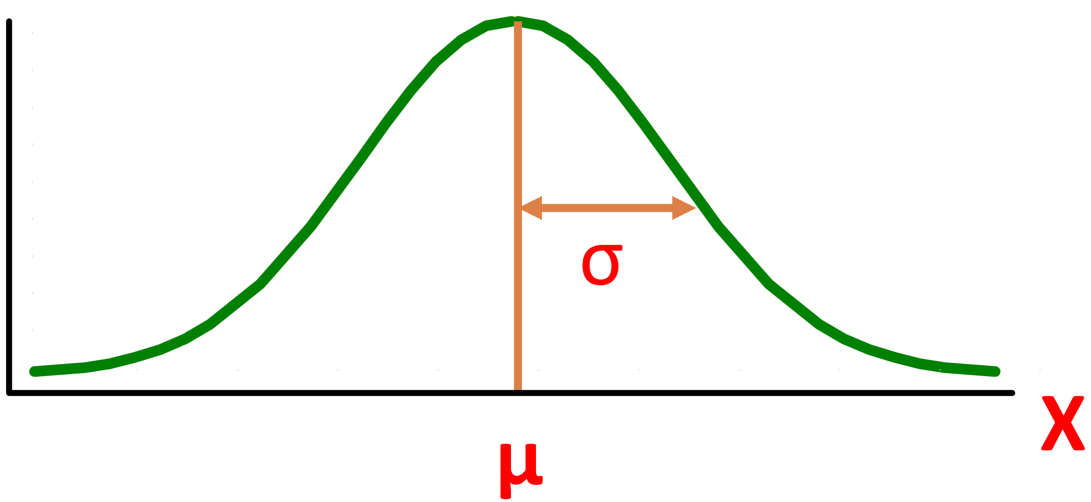
Nicholas School of the Environment - Duke University

Learning Goals

- Mean, Variance & Std. Deviation
- Stochastic Processes
- Autocovariance Function
- Autocorrelation Function (ACF)
- Stationary Process
- Partial Autocorrelation Function (PACF)

Mean, Variance and Stand. Deviation

- Mean is average of a group of numbers
- Variance is the average of squared differences from mean
- Standard Deviation measure how spread out are the numbers



Simple Sequence

- Suppose we have a sequence of numbers $y_1, y_2, \dots, y_T$

Mean

$$\mu = \frac{\sum_{i=1}^{T} y_i}{T}$$

Standard Deviation

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

- But what happens when we have a stochastic process ?

Stochastic Processes *(Ch. 2 of Cryer and Shan)*

- In this case instead of a simple sequence of variables, we have a random variable
- The sequence of random variable is called stochastic process and is a model for an observed time series
- When dealing with time series we talk about
  - Mean function
  - Variance fucntion
  - Autocovariance function
  - Autocorrelation function

Because they are a function of time

Mean and Variance

- The mean function is defined by

$$\mu_t = \mathrm{E}(\mathrm{Y}_t)$$

is the expected value of the process at time t

- The variance is defined by

$$\sigma_t^2 = E(Y_t - \mu_t)^2 = E(Y_t^2) - \mu_t^2$$

**adapted from Ch. 2 of Cryer and Shan*

Variance function explained

$$\sigma_t^2 = E(Y_t - \mu_t)^2 =$$
$$E(Y_t^2 - 2Y_t\mu_t + \mu_t^2) =$$
$$E(Y_t^2) - E(2Y_t\mu_t) + E(\mu_t^2)$$

μ_t^2

But μ_t is a constant, therefore $E(\mu_t) = \mu_t$ and $E(\mu_t^2) = \mu_t^2$

$$\sigma_t^2 = E(Y_t^2) - 2\mu_t E(Y_t) + \mu_t^2$$

Recall $E(Y_t) = \mu_t$, then

$$\sigma_t^2 = E(Y_t^2) - 2\mu_t\mu_t + \mu_t^2 = E(Y_t^2) - 2\mu_t^2 + \mu_t^2$$
$$\boldsymbol{\sigma_t^2 = E\left(Y_t^2\right) - \mu_t^2}$$

Meaning of Autocorrelation Function

- Recap: What is correlation?

*From stats: covariance and correlation measure **joint variability** of two variables.*

Meaning of Autocorrelation Function

- Recap: What is correlation?

Is a measure of linear dependence between two variables

- In TSA: What is autocorrelation?

Is a measure of dependence between two adjacent values of the same variables

- The prefix *auto* is to convey the notion of self-correlation, that is, correlation between variables from the same time series

Autocovariance & Autocorrelation Function

The **autocovariance** function is defined as

$$\boxed{\gamma_{t,s}} = Cov(Y_t, Y_s)$$
$$= E[(Y_t - \mu_t)(Y_s - \mu_s)]$$
$$= E[Y_t Y_s] - \mu_t \mu_s$$

The **autocorrelation** function is defined as

$$\boxed{\rho_{t,s}} = Corr(Y_t, Y_s)$$
$$= \frac{Cov(Y_t, Y_s)}{\sqrt{Var(Y_t)Var(Y_s)}}$$
$$= \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\,\gamma_{s,s}}}$$

**adapted from Ch. 2 of Cryer and Shan*

Autocovariance function explained

$$\begin{aligned}
\gamma_{t,s} &= E[(Y_t - \mu_t)(Y_s - \mu_s)] \\
&= E(Y_t Y_s - Y_t \mu_s - Y_s \mu_t + \mu_t \mu_s) \\
&= E(Y_t Y_s) - \mu_s E(Y_t) - \mu_t E(Y_s) + \mu_t \mu_s \\
&= E(Y_t Y_s) - \mu_s \mu_t - \mu_t \mu_s + \mu_t \mu_s
\end{aligned}$$

$$\boldsymbol{\gamma_{t,s} = E(Y_t Y_s) - \mu_s \mu_t}$$

Autocorrelation function explained

From stats, correlation between two variables X and Y is given by

$$\rho_{XY} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

Also from stats $Var(Y) = cov(\mathrm{Y},\mathrm{Y}) = \gamma_{YY}$

$$\rho_{t,s} = \frac{cov(Y_t, Y_s)}{\sqrt{Var(Y_t)Var(Y_s)}} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}}$$

How to compute autocorrelation?

- In the context of a single variable, Y_t is the original series and Y_s is a lagged version of the series

Y_t	Y_s
Y_1	Y_2
Y_2	Y_3
Y_3	Y_4
Y_4	Y_5
$\vdots$	$\vdots$
Y_{N-3}	Y_{N-2}
Y_{N-2}	Y_{N-1}
Y_{N-1}	Y_N
Y_N	

Compute lag 1 autocorrelation

$$\rho_{t,s} = Corr(Y_t, Y_s)$$

How to compute autocorrelation?

- In the context of a single variable, Y_t is the original series and Y_s is a lagged version of the series

Y_t	Y_s
Y_1	Y_3
Y_2	Y_4
Y_3	Y_5
Y_4	Y_6
$\vdots$	$\vdots$
Y_{N-3}	Y_{N-1}
Y_{N-2}	Y_N
Y_{N-1}	
Y_N	

Compute lag 2 autocorrelation

$$\rho_{t,s} = Corr(Y_t, Y_s)$$

Main Conclusion

Autocovariance and autocorrelation function give information about the dependence structure of a time series

Properties

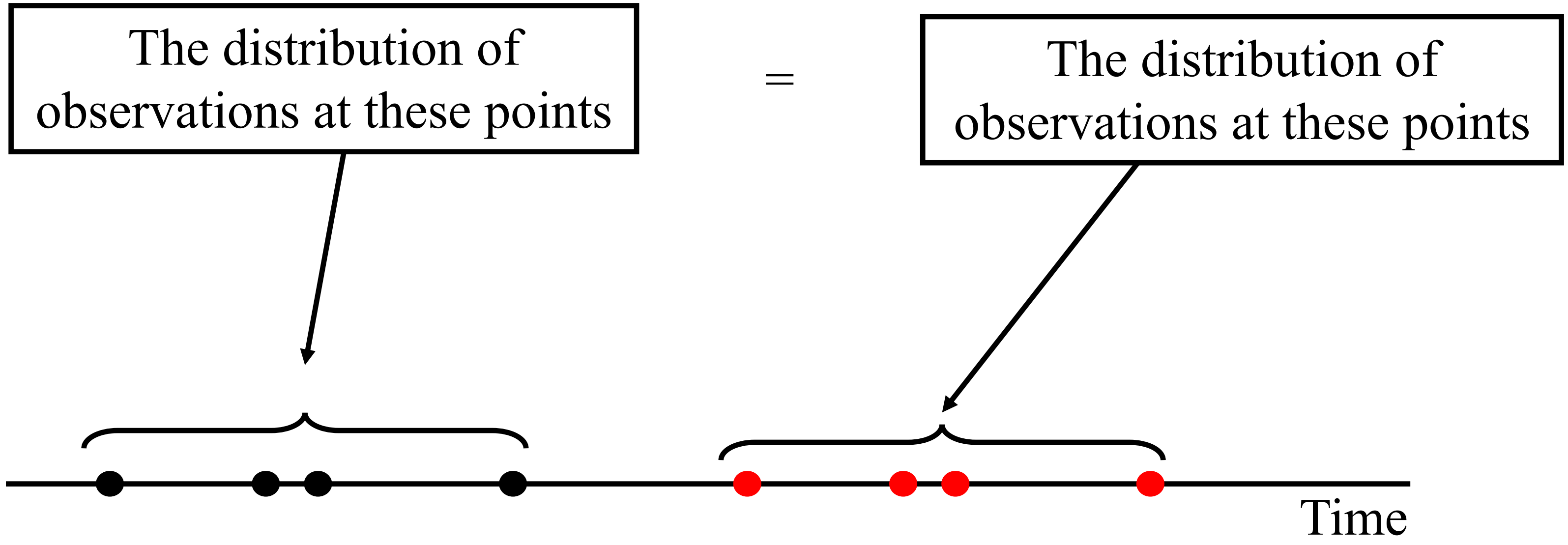
$\gamma_{t,t} = Var(Y_t)$	$\rho_{t,t} = 1$
$\gamma_{t,s} = \gamma_{s,t}$	$\rho_{t,s} = \rho_{s,t}$
$\lvert\gamma_{t,s}\rvert \leq \sqrt{\gamma_{t,t}\,\gamma_{s,s}}$	$\lvert\rho_{t,s}\rvert \leq 1$

Homework: try to understand why the six expressions in the table are true!!

- Values of $\rho_{t,s}$ close to ± 1 indicate strong linear dependence
- Values of $\rho_{t,s}$ close to 0 indicate weak linear dependence
- If $\rho_{t,s} = 0$, then Y_t and Y_s are uncorrelated

Stationary Process

- The basic idea of stationarity is that the probability laws that govern the behavior of the process do not change over time



Consequences of Stationarity

- Distribution of Y_t is the same of Y_{t-k} for all t and k
- Then,
  - $E(Y_t) = E(Y_{t-k})$ for all t and k so the **mean function is constant** for all time
  - $Var(Y_t) = Var(Y_{t-k})$ for all t and k so the **variance is also constant** over time

- And what happens with the autocovariance function?

**adapted from Ch. 2 of Cryer and Shan*

Consequences of Stationarity (cont'd)

- If the process is stationary, then

$$\gamma_{t,s} = Cov(Y_t, Y_s) = Cov(Y_{t-k}, Y_{s-k})$$

For $k = s \quad \rightarrow \quad Cov(Y_t, Y_s) = Cov(Y_{t-s}, Y_0)$

For $k = t \quad \rightarrow \quad Cov(Y_t, Y_s) = Cov(Y_0, Y_{s-t})$

Thus, $\quad \gamma_{t,s} = Cov\left(Y_0, Y_{|t-s|}\right) = \gamma_{0,|t-s|}$

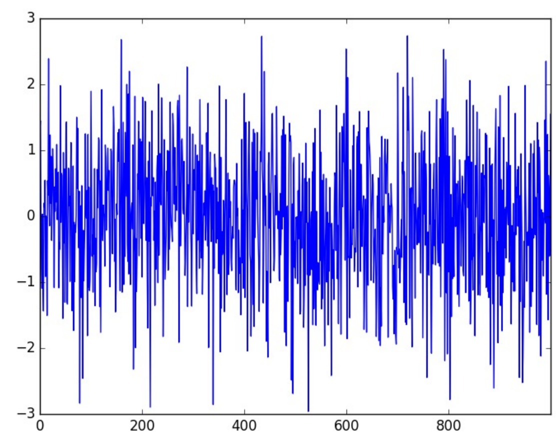
- In other words, the covariance between Y_t and Y_s depends only on the time difference $|t - s|$ and not on the actual times *t* and *s*

**adapted from Ch. 2 of Cryer and Shan*

White Noise Series

- Example of a stationary process: **white noise** series
- The white noise series is a sequence of independent, identically distributed (i.i.d.) random variables $\{e_t\}$
- $\{e_t\}$ is a stationary process, then

$$\boxed{\begin{aligned}
\mu_t &= \mathrm{E}(\mathrm{e}_t) \\
\gamma_k &= \begin{cases} Var(e_t) & for\ k = 0 \\ 0 & for\ k \neq 0 \end{cases} \\
\rho_k &= \begin{cases} 1 & for\ k = 0 \\ 0 & for\ k \neq 0 \end{cases}
\end{aligned}}$$



- In time series modeling we usually assume that the white noise process has mean zero and $Var(e_t) = \sigma_e^2$

**adapted from Ch. 2 of Cryer and Shan*

Partial Autocorrelation Function

Recap: The ACF of a stationary process Y_t at lag h

$$\rho_{t,t-h} = Corr(Y_t, Y_{t-h})$$

measures the linear dependency among the process variables Y_t and Y_{t-h}.

But the dependency structure among the **intermediate variables**

$$Y_t, \boxed{Y_{t-1}, Y_{t-2}, \cdots Y_{t-h+2}, Y_{t-h+1},} Y_{t-h}$$

also plays an important role on the value of the ACF.

Partial Autocorrelation Function (cont'd)

Imagine if you could **remove** the influence of all these intermediate variables…

You would have only the directly correlation between Y_t and Y_{t-h}

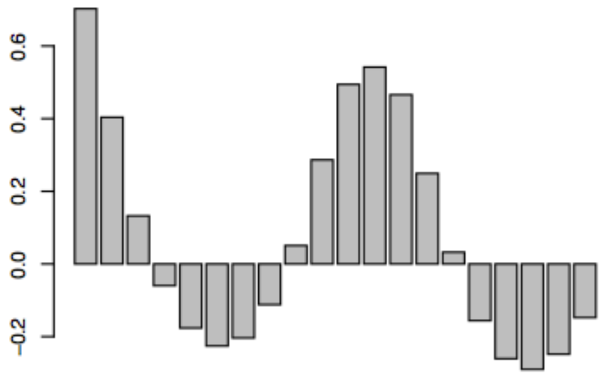
That's the so called **partial autocorrelation function (PACF)**

Partial Autocorrelation Function (cont'd)

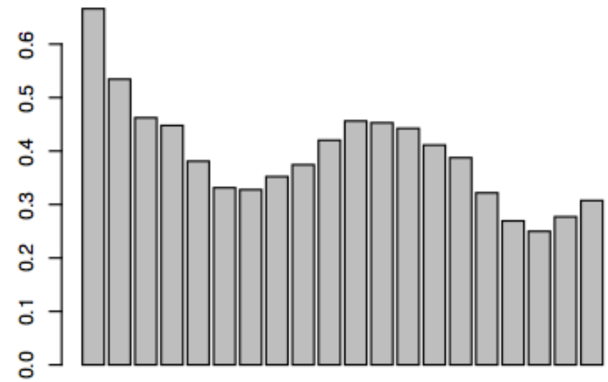
- The PACF is a little more difficult to compute
- We will talk about that later when we discuss the Yule Walker equations
- In summary:
  - The ACF and PACF measure the temporal dependency of a stochastic process
  - You will always build the ACF and PACF before fitting a model to a stochastic process
  - The ACF and PACF give us information about the **auto-regressive component** of the series

Examples of ACF and PACF plots

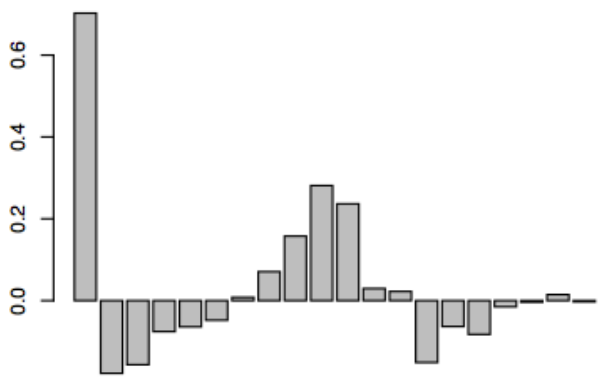
ACF plot



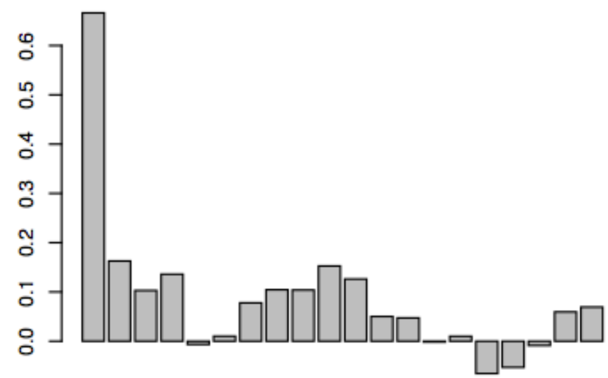
ACF plot

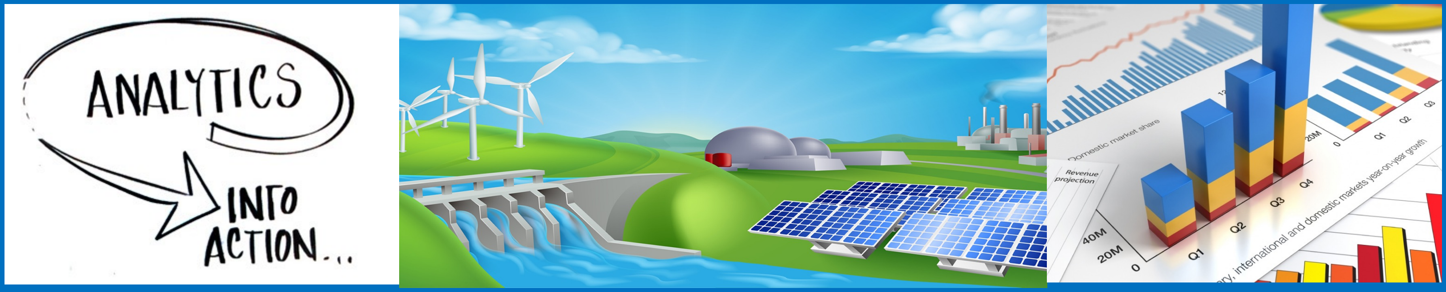


PACF plot



PACF plot





THANK YOU !

luana.marangon.lima@duke.edu