Time Series Analysis

Week 6 Material | 21 August 2017 | Due: 28 August 2017

Contents

[Introduction 2](#_Toc491940980)

[Plotting Time Series Data 2](#_Toc491940981)

[Additive and Multiplicative Models 3](#_Toc491940982)

[Additive Model 3](#_Toc491940983)

[Multiplicative Model 3](#_Toc491940984)

[Parts of time series 4](#_Toc491940985)

[Deterministic 4](#_Toc491940986)

[Stochastic 4](#_Toc491940987)

[Functions in Time Series Analysis 5](#_Toc491940988)

[General Parameters 5](#_Toc491940989)

[Mean and Variance function 5](#_Toc491940990)

[Covariance 5](#_Toc491940991)

[Auto-covariance function (AC**V**F) 5](#_Toc491940992)

[Autocorrelation function (ACF) 6](#_Toc491940993)

[Purely random Process (White noise) 7](#_Toc491940994)

[Simulating in R 7](#_Toc491940995)

[Examples 7](#_Toc491940996)

[Random Walk Process 8](#_Toc491940997)

[Mean Value 10](#_Toc491940998)

[Variance 10](#_Toc491940999)

[Autocovariance 10](#_Toc491941000)

[Autocorrelation 10](#_Toc491941001)

[Drift 10](#_Toc491941002)

[Simulating in R 10](#_Toc491941003)

[Moving Average Process 11](#_Toc491941004)

[Variance of 11](#_Toc491941005)

[Autocovariance 11](#_Toc491941006)

[Autocorrelation 11](#_Toc491941007)

[Example Problem from Lecture Notes 12](#_Toc491941008)

[Solving the Mean value function 12](#_Toc491941009)

[Solving the Mean Value function in R 12](#_Toc491941010)

[Stationary Time Series 13](#_Toc491941011)

[Example 14](#_Toc491941012)

[Appendix 15](#_Toc491941013)

[Plot of Various Time Series Functions 15](#_Toc491941014)

[Comparison of White Noise and Random Walk Data 16](#_Toc491941015)

# Introduction

Classical Statistical methods require the following assumptions:

1. Observations are independent
2. Observations are from the same population (i.e. probability distribution)
3. If the Statistical Method is *parametric*:
   * The data follows a particular probability distribution (e.g. Normal or Chi)

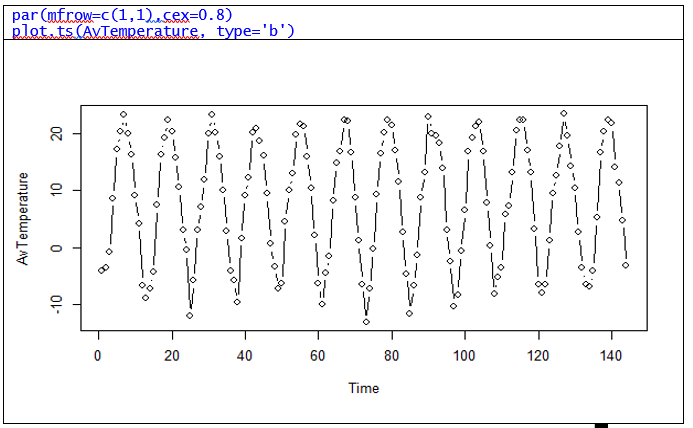
Where observations are non-independent due to the fact that they were collected sequentially over time classical statistical methods may not be appropriate and Time Series Analysis comes into play.

Imagine, rolling a dice, if at 10am a dice was rolled and the result was a 6, the probability of rolling a 6 at 11 am would not be affected. However if at 10am the thermometer read 35 OC it would be quite unlikely for the thermometer to read 5 OC at 11am, due merely to the fact of the past observation. This is why Time Series Analysis is useful.

Where observations are non-independent due to the fact that they were collected in close proximity (space) techniques known as spatial statistics are used.

## Plotting Time Series Data

Time series data is an observation ( plotted against the date , e.g.:



## Additive and Multiplicative Models

Time series can be broken up into four types of components:

* Trend (T)
* Cycle (C)
* Seasonal (S)
* Irregular (I)

These are best illustrated by way of a diagram:



It is assumed that these components interact in either an additive or multiplicative fashion

|  |  |
| --- | --- |
| Additive Model This is used where data is of similar magnitude (trend-free or over a short run) with constant absolute growth or decline | Multiplicative Model This is used where data is of increasing or decreasing magnitude (long run or trended data) with constant percent growth or decline. |

The multiplicative model becomes additive as logarithms are taken (of non-negative data).

## Parts of time series

Time series have two major parts, a **deterministic** part and a **stochastic** part.

### Deterministic

The **deterministic** part may consist of various effects, such as,

* Long-term Trend-The general movement over all years;
* Cyclical effect – repetitive up and down movements about a trend that covers several years; Cyclical trends will go up and down randomly.
* Seasonal effect – repetitive cyclical pattern within a year (or a week or other smaller time period), Seasonal effects will have a constant pattern.

### Stochastic

The **stochastic** component is the random variation.

* Essentially this is just random error and fluctuation, irregular difficult to explain movement of data. (Also donated by ).

If data is *independent and identically distributed* (i.i.d.) then the additive time series model becomes a multiple linear regression (from wk. 5 material; how to use a multiple linear regression to model this is discussed in Wk. 7).

Generally the data is a sequence of successive dependent observations and special time series models are used (discussed in Wk. 8).

# Functions in Time Series Analysis

## General Parameters

### Mean and Variance function

The ***mean*** and ***variance*** function of a time series is defined by:

|  |  |
| --- | --- |
|  |  |

refers to the expected value of random variable , these are both functions of .

### Covariance

In the previous notes Covariance was defined as:

Covariance can also be defined more broadly as the expected product of their deviations from their individual expected values:[[1]](#footnote-2)

The reference book provides:[[2]](#footnote-3)

### Auto-covariance function (AC**V**F)

The ***auto covariance*** function of is defined by:

This is a function of and ;

### Autocorrelation function (ACF)

The autocorrelation is a measure of how a random variable (say), varies over time relative to past observations (of say). E.g. If the temperature was higher than average at 12 pm, we would expect the temperature to be above average at 1pm.

The ***autocorrelation function*** (ACF) is defined by:

The lag-1 auto correlation for some time series is:

The lag- auto correlation is defined:

Where:

* denotes the lag-1 autocovariance,
* is the lag-0 autocovariance and is equal to the variance.
* The lag-0 auto correlation is 1 (because obviously each value is correlated with itself).

## Purely random Process (White noise)

A white noise process has a mean value of zero and NO correlation between its values at different times, hence any random distribution (discrete or continuous) can be white noise (e.g. Normal, Uniform, Poisson, Binomial, Log-Normal, Chi, Gamma, Beta, Weibull, Exponential etc.).

So the definitions are:

Purely Random Process (AKA White Noise):

* Autocorrelation of 0 for all values
  + This is for all the data lag-1, lag-2, lag-3 …
* Auto covariance of 0
  + This is for all the data lag-1, lag-2, lag-3 …
* Constant Mean value (e.g. 0)

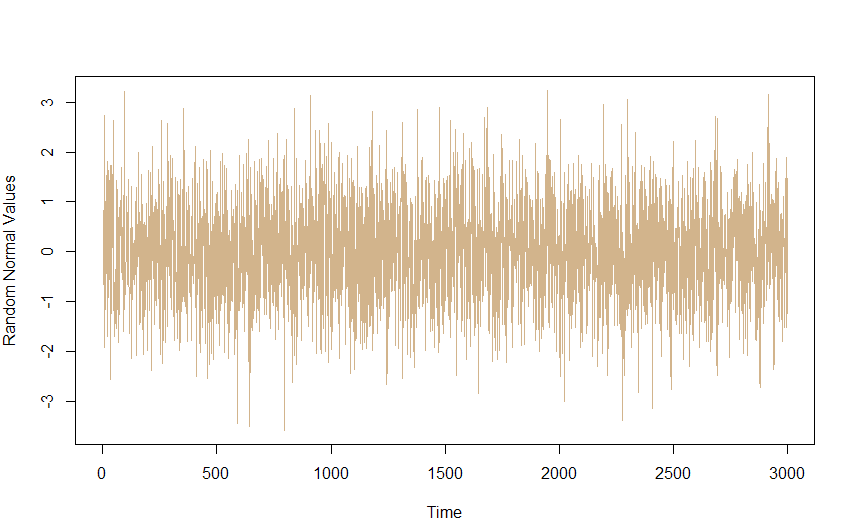
A white noise process may represent the error observed in a model, in this case the white noise process will have a mean value of 0.

Expressed more formally a white noise process is:

## Simulating in R

WN\_1 <- arima.sim(model=list(order=c(0,0,0)), n=50, mean=0, sd=1)

as.xts(WN\_

A white noise process would look like:

WN\_1 <- arima.sim(model=list(order=c(0,0,0)), n=300)

plot(as.xts(WN\_1), lwd=0.01, col="tan", grid.ticks.on = FALSE)

plot.ts(rnorm(3000), ylab="Random Normal Values", col="tan

## Examples

* if contained 300 values it would be a white noise process.
  + Because the values have
* However a vector containing those 300 values and another 200 values from a normal distribution with would NOT be a white noise process because the variation is not constant over time.

## Random Walk Process

A random walk process, is such that:

And is such that:

* There is no specified mean
* Strong dependence over time
* It’s changes are white noise

A Random Walk process () requires a starting point, usually.

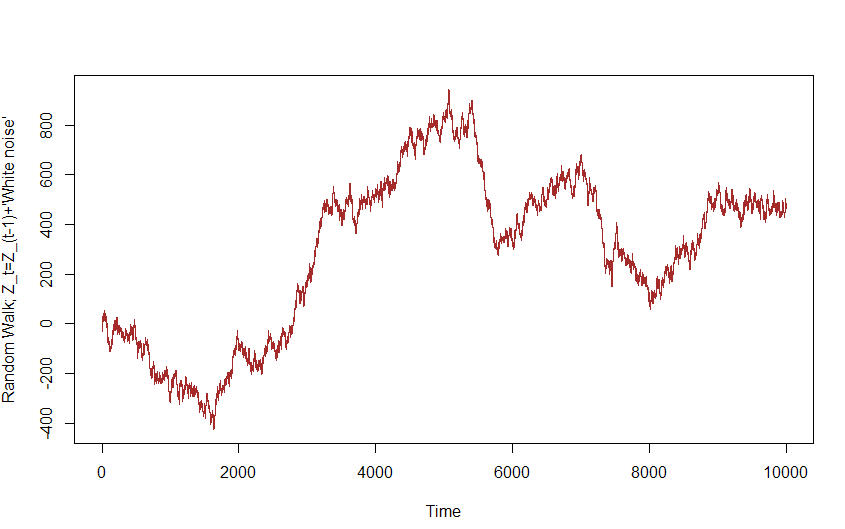
A Random Walk Process has only one parameter, the variance of the white noise.

The difference between each value on a random walk, is white noise, i.e. diff(Z)=WN.

If is a white noise process with and a constant variance, then is a random walk process:

For

…

A random walk process could look like this:

### Mean Value

### Variance

### Autocovariance

I need to prove this rather than just writing it out.

### Autocorrelation

## Drift

A random Walk with drift is:

And merely adds a constant value to the form.

A white noise process with a drift has two parameters:

* White Noise Variance
* Constant
  + The mean of the white noise is

## Simulating in R

RW\_1 <- arima.sim(model=list(order=c(0,1,0)), n=50)

as.xts(RW\_1

# Moving Average Process

A moving average process of order 1 ( ***MA (1)*** ) is:

Where and are constants, mean is 0 and variance is constant (.

## Variance of

Not totally clear

## Autocovariance

It can be shown that a *Moving Average Process* of order 1 has autocovariance:[[3]](#footnote-4)

A *Moving Average Process* of order has autocovariance:[[4]](#footnote-5)

Where:

## Autocorrelation

It can be shown that a *Moving Average Process* of order 1 has autocorrelation:[[5]](#footnote-6)

A *Moving Average Process* of order has auto correlation:[[6]](#footnote-7)

Where

## Example Problem from Lecture Notes

In the lecture notes the following time series is proposed:

Where:

* constant values
* is a variable representing the time
  + We will assume that
* is a stationery series, with a mean value of 0.

### Solving the Mean value function

The mean value in this case is the expected value:

|  |  |  |
| --- | --- | --- |
| is just a constant value, e.g. if I had a constant value, say 5, I would expect that value to be 5 because, well, it’s 5. |  |  |

By substituting this into the equation:

# Stationary Time Series

A process is strictly stationary if all probabilistic behaviour is unchanged by shifts in time.[[7]](#footnote-8)

This is a very strong assumption, it will often suffice to assume that a process is weakly stationary.

A process is weakly stationary if its mean, variance and covariance are unchanged by shifts in time.

A time series is stationary if the properties of the underlying model do not change through time,[[8]](#footnote-9) i.e. the properties do not depend on the time at which the series is observed. [[9]](#footnote-10)

A time series is said to be stationary if:[[10]](#footnote-11)

* The mean function is a constant value, it does not change through time and
  + ; i.e. constant
* The auto covariance function does not change through time.

Observe that:

* Purely random processes are stationary
* MA(1) is Stationary
* Random Walk Processes are not stationary
  + Because is dependent on time.

A time series that is stationary in this sense is known as a weak stationary, covariance stationary, second-order stationary or wide sense stochastic process. [[11]](#footnote-12)

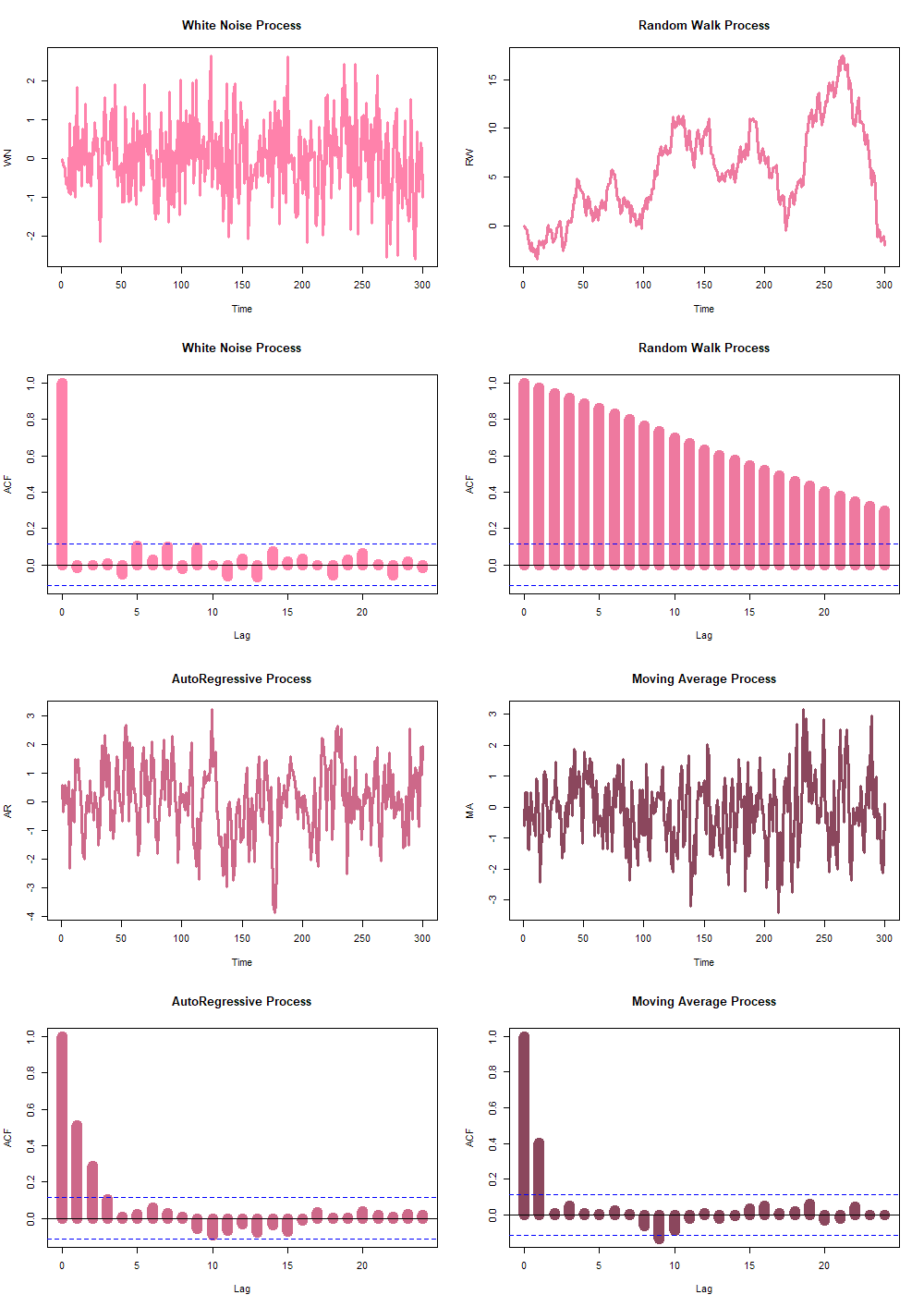
A stationary process can be modelled with fewer parameters, that’s why it is useful.

## Example

For a stationary process at time, the following should be true:

# Appendix

## Plot of Various Time Series Functions



## Comparison of White Noise and Random Walk Data

# Use arima.sim() to generate WN data

white\_noise <- arima.sim(model=list(order=c(0,0,0)), n=100)

# Use cumsum() to convert your WN data to RW

random\_walk <- cumsum(white\_noise)

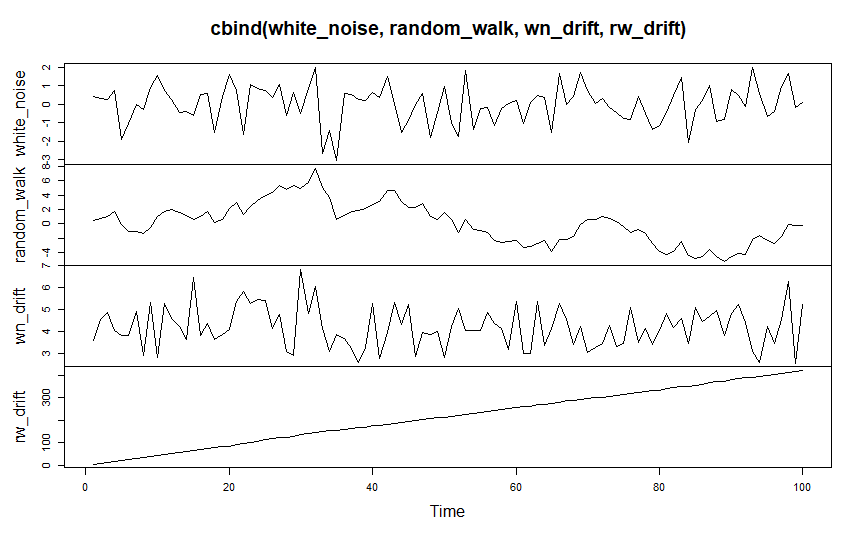
# Use arima.sim() to generate WN drift data

wn\_drift <- arima.sim(model=list(order=c(0,0,0)), n=100, mean=4)

# Use cumsum() to convert your WN drift data to RW

rw\_drift <- cumsum(wn\_drift)

# Plot all four data objects

plot.ts(cbind(white\_noise, random\_walk, wn\_drift, rw\_drift))

1. Oxford Dictionary of Statistics, Oxford University Press, 2002, p. 104. [↑](#footnote-ref-2)
2. Chatfield, C. (2000). *Time-series forecasting*. Boca Raton: Chapman & Hall/CRC.

   Equation (2.6.1), (p. 40 of 265) [↑](#footnote-ref-3)
3. Chatfield, C. (2000). *Time-series forecasting*. Boca Raton: Chapman & Hall/CRC.

   Equation (2.5.6), (p. 39 of 265); Compare this with the lecture notes and the formula to conclude that the numerator of the autocorrelation corresponds to the autocovariance. [↑](#footnote-ref-4)
4. Chatfield, C. (2000). *Time-series forecasting*. Boca Raton: Chapman & Hall/CRC.

   Equation (3.1.2), (p. 46 of 265) [↑](#footnote-ref-5)
5. Chatfield, C. (2000). *Time-series forecasting*. Boca Raton: Chapman & Hall/CRC.

   Equation (2.5.6), (p. 39 of 265) [↑](#footnote-ref-6)
6. Chatfield, C. (2000). *Time-series forecasting*. Boca Raton: Chapman & Hall/CRC.

   Equation (3.1.2), (p. 46 of 265) [↑](#footnote-ref-7)
7. https://s3.amazonaws.com/assets.datacamp.com/production/course\_1143/slides/ch2\_4\_supplementary.pdf [↑](#footnote-ref-8)
8. Chatfield, C. (2000). *Time-series forecasting*. Boca Raton: Chapman & Hall/CRC.

   Equation (2.4.1), (p. 34 of 265) [↑](#footnote-ref-9)
9. Hyndman Rob J. and Athana­sopou­los George (2013) F*orecasting: Principles and Practice*, OTexts,

   <https://www.otexts.org/fpp> [↑](#footnote-ref-10)
10. Chatfield, C. (2000). *Time-series forecasting*. Boca Raton: Chapman & Hall/CRC.

    Equation (2.4.1), (p. 34 of 265) [↑](#footnote-ref-11)
11. Imdadullah, M. (2016). *Stationary Stochastic Process*. [online] Basic Statistics and Data Analysis. Available at: http://itfeature.com/time-series-analysis-and-forecasting/stationary-stochastic-process [Accessed 18 Aug. 2017]. [↑](#footnote-ref-12)