**My note for Time Series Analysis and Its**

**Applications using R**

Author of the book: Robert H. Shumway & David S. Stoffer

# Chapter 1: Characteristics of time series

* Definition of time series analysis: systematic approach by which one goes about answering mathematical and statistical questions posed by these time correlations
* Unique problem we are trying to solve: Obvious correlation introduced by sampling of adjacent points int time can severely restrict the applicability of the many conventional statistical models traditionally dependent on the assumption that these adjacent observations are independent and identically distributed
* First step is always about plotting the data for careful examination of the recorded data plotted over time
* Two approach of time series analysis:
  + The time domain approach 🡪 investigation of lagged relationship as most important
  + The frequency domain approach 🡪 investigation of cycles as most important

## The Nature of time Series Data

* Example 1.1: Johnson & Johnson Quarterly earnings showing gradually increasing trend and the rather regular variation superimposed on the trend that seems to repeat over quarter
* Example 1.2: Global warning with an apparent upward trend which is a more interesting question than any periodicities
* Example 1.3: Speech data with repetitive nature of the signal rather regular periodicities. Speech recognition and spectral analysis are in used in this context.
* Example 1.4: Down Jones Industrial Average showing typical of return data with stable mean with an average approximately zero, however, highly volatile (variable) periods tend to be clustered together. Issue of volatility forecasting of future return using ARCH and GARCH models
* Example 1.5 El Niño and Fish Population 🡪 Analysis of several series at once. Periodic behaviour can be discovered. Possible model is some versions of regression analysis as a procedure for relating the series (transfer function modelling)
* Example 1.6 fMRI Imaging 🡪 Analysis of variance techniques accomplish identification of different responses in classical statistics
* Example 1.7 Earthquakes and Explosions 🡪 time series discriminant analysis

## Time series statistical models

* The primary job of a time series analysis is to develop mathematical models that provide plausible descriptions for sample data.
* Time series is defined as a collection of random variables indexed according to the order they are obtained in time
* In general, a collection of random variables, , indexed is referred to as a stochastic process
  + Time index could be ***continuous*** or ***discrete***
* Observed value of the stochastic process is called a *realization*
* Topic to dig if needed: ***aliasing***
* A fundamental visual characteristic distinguishing the different time series is the ***different degree of smoothness***
* One possible explanation for this smoothness is that it is being induced by the supposition that adjacent points in time are ***correlated,*** so that the value of the series at time , say , depends in some ways on the past values , , …
* First fundamental time series is ***White Noise (WN)***
  + Collection of uncorrelated random variables, , with mean 0 and finite variance
  + Notation
* Sometimes, we will require the noise to be independent and identically distributed *iid so* ***iid noise***
  + Notation
* A particularly useful white noise is the ***Gaussian white Noise*** wherein the are independent normal random variable with mean 0 and variance
  + Notation
* **Definition:** A linear combination of values in a time series is referred to, generically, as a ***filter series***
* Autoregression model is a model for which the current value is a function of past values
* Other notable series:
  + White noise with or without drift
  + Signal in noise

## Measures of dependence

* A complete description of a time series, observed as a collection of n random variables at arbitrary time point for any positive integer , is provided by the joint ***distribution function***, evaluated as the probability that the values of the series are jointly less than the constants : i.e.,
* Unfortunately, these multidimensional distribution functions cannot usually be written easily ***unless the random variable are jointly normal*** in which case the joint density has the well-known form displayed as follow:

With

* + for , where | · | denotes the determinant.
  + The mean vector
  + The covariance matrix as , which is assumed to be positive definite,
* The distribution function is virtually impossible to plot and analysing time series. The marginal distribution functions:

Or the corresponding marginal density functions:

When they exist, are often informative for examining the marginal behaviour of a series.

* Another informative marginal descriptive measure is the ***mean function*** if It exists defined as:
  + We are averaging over the realization of x at t fix
* The lack of independence between two adjacent values and can be assessed numerically, as in statistics, using the notions of ***covariance*** and ***correlation.***
  + The ***autocovariance function*** is defined as the second moment product:
  + For all and – also we could drop the subscript when no ambiguity
  + The measures linear dependence between two points on the same series observed at different time 🡪 smooth series exhibit high autocovariance that stays large even for t and s far apart
* ***Covariance of linear combination***
  + if the random variables
  + are linear combinations of (finite variance) random variables and , then
* As in classical statistics, it is more convenient to deal with a measure of association between −1 and 1, and this leads to the following definition.
  + The ***autocorrelation function (ACF)***
  + The ACF measure the linear predictability of the series at time t, say , using only
  + Cauchy-Schwarz inequality shows easily that
* Often, we would like to measure the predictability of another series from the series . Assuming both series as finite variances, we have
  + The ***cross-covariance function*** between two series and
  + The ***cross-correlation function (CCF)***
* We could easily extend the above definition to the case of more than two series, say a multivariate time series with r components.

## Stationarity Time series

* Definition of ***strictly stationarity***
  + The probabilistic behaviour of every collection of values is identical to that of the time shifted set
  + Mathematically
  + If a time series is strictly stationary, then all the multivariate distribution functions for subsets of variables must agree with their counterparts in the shifted set for all values of the shift parameter
* The strict stationarity is too strong for many applications thus we will introduce ***a milder version only posing conditions on the first two moments***
* Definition of ***a weakly stationarity***
  + Weakly stationarity time series, , is a ***finite variance*** process such that:
    - (1) the mean value function, , is constant and does not depend on time
    - (2) the autocovariance function, , depends only on and through their difference
  + Henceforth, we will use the term ***stationarity*** to mean ***weakly stationarity;*** if the process is stationarity in the strict sense, we will use the term ***strictly stationary***
* Strictly stationarity finite variance leads to weakly stationarity, but the converse isn’t necessarily true unless further conditions.
* **Property: One important** case where stationarity implies strict stationarity is if the time series is Gaussian [meaning all finite distributions of the series are Gaussian]
* We may simplify the notation for stationary time series
  + Mean function
  + Autocovariance
  + Autocorrelation Function (ACF)
* is non-negative definite ensuring that the variances of linear combination of the variates will never be negatives
* The autocovariance function of a stationary timeseries is ***symmetric*** around the origin
* We can also extend the definition to
  + Jointly stationary series
  + The cross-correlation function (CCF):
* With Cauchy-Schwarz,
* Linear process
  + A linear process, , is defined as a linear combination of white noise variance , and is given by:
  + We can show that the autocovariance function is:
* A casual linear process is a linear process which depends on the future has for 🡪 not a focus
* Definition: A process, is said to be a ***Gaussian process*** if the n-dimensional vectors ’, for every collection of distinct time series and every positive integer n, have a multivariate normal distribution (distribution defined up in the file)
  + Easy to prove If a Gaussian time series, , is weakly stationary thus this series is strictly stationary
  + A result called the ***world decomposition*** states that a stationary non-deterministic time series is a casual linear process (but w )
* A linear process need not be Gaussian, but if a time series is Gaussian, then it is a causal linear process with
  + It is not enough for the marginal distributions to be Gaussian for the process to be Gaussian
  + It is easy to construct a situation where *X* and *Y* are normal, but(*X*, *Y*) is not bivariate normal
  + e.g.: Let X and Z be independent normal and let Y = Z if XZ > 0 and Y = - Z if XZ < 0

## Estimation of correlation

* Above definition are first introduce for the whole population but in practice most analyses mut be performed using sampled data
* This limitation means the sampled points only are available for estimating the mean, autocovariance, autocorrelation functions
* It brings the issue from a statistical viewpoint of not having iid copies of that are available for estimating the covariance and correlation functions.
* Somehow, we must use average over this single realization to estimate the population means and covariance functions
* ***Sample mean***
* ***Standard error of the estimate***

Thus, after some algebra

* The ***sample autocovariance function*** is defined as

With

* + This is a preferred estimator than the one that would be obtained by dividing by because of the non-negative definite function
  + Both are biased estimator
* The ***sample autocorrelation function*** is defined, analogously to the past
  + The sample autocorrelation function has a sampling distribution that allows us to assess whether the data comes from a completely random or white noise series or whether correlations are statistically significant at some lags
* ***Property: Large sample distribution of the ACF***
  + Under general conditions (aka is iid and fourth moments is finite – all of that verified with is white Gaussian), if is white noise, then for large , the sample ACF,
  + is approximately ***normally distributed with zero means and standard deviation*** given by
* Based on the previous result, we obtain a rough method of assessing whether peaks in are significant by determining whether the observed peak is outside the interval of (or plus/minus two times standard errors which in normal distribution means less 95% change)
* The application of this property develops because many statistical modelling procedures depend on reducing a time series to a white noise series using various kinds of transformations. After such procedure is applied, the plotted ACFs of the residuals should lie roughly within the limits given above ()
* Both sample autocovariance and autocorrelation can be extended to sample cross-covariance and cross-correlation function as originally defined for the theorical case
* Similar property as shown in ***large sample distribution of ACF*** can be extended to ***large sample distribution of cross-correlation*** 
  + Property: the large sample distribution of ***sample cross correlation*** is normal with ***mean zero*** and ***standard error of 1/sqrt(n)*** if at least one of the processes is ***independent white noise.***
* Because of the condition to apply large sample distribution of cross correlation property, we will need to introduce some treatment of whitening a series 🡪 **Prewhitening and Cross Correlation Analysis** 
  + ***Warnings:*** If we don’t pre-white at least one series, the cross-correlation analysis can be misleading
  + Pre-whitening of means we mean that the signal has been removed from the data by running a regression of and , where are the predicted values from the regression.

## Vector-Valued and Multidimensional series

* We frequently encountered situations in which the relationship between a number of jointly measured time series are of interest, hence it would be useful to consider the notion of a ***vector defined timeseries***
* is a vector defined timeseries which contains as its components *p* univariate time series
* We denote column vector of the observed series as , the row vector is its transpose.
* For stationary case,
  + The mean vector is
  + The autocovariance matrix

where elements of the matrix are the cross-covariance function

for . Because , it follows that

* Now the sample autocovariance matrix of the vector series is the matrix of sample cross covariance, defined as follows:

Where the sample mean vector

* The symmetric property of the theoretical autocovariance extends to the sample autocovariance
* The autocovariance function of a stationary multidimensional process can be defined as a function of the multidimensional lag vector, say, , as

Where

Does not depend on the spatial coordinates . In two dimensions,

Which is a function of lag, both in the row and the column

* The multidimensional sample autocovariance function is defined as

Where and the range of the summation for each argument is

For

The mean is computed over r-dimensional array, that is

Where the argument are summed over

* The multidimensional sample autocorrelation function follows, as usual, by taking the scaled ratio
* Opening and issue to consider:
  + The sampling requirements for multidimensional processes are rather severe be- cause values must be available over some uniform grid to compute the ACF.
  + In some areas of application, such as in soil science, we may prefer to sample a limited number of rows or *transects* and hope these are essentially replicates of the basic underlying phenomenon of interest.
  + One-dimensional methods can then be applied. When observations are irregular in time space, modifications to the estimators need to be made.
  + Systematic approaches to the problems introduced by irregularly spaced observations have been developed by Journel and Huijbregts

## Vector-Valued and