Time Series Analysis and Modelling

Part 1: Time Series Introduction

Jonathan Mwaura

Khoury College of Computer Sciences

August 1, 2024

The Roux Institute
at Northeastern University

Introduction

Acknowledgements

- These slides have been adapted from Achraf Cohen, University of West Florida - from the class class STA6856
- Lecture Notes from Dewei Wang, Department of Statistics, University of South Carolina (See notes in Canvas)
- C. Chatfield, The Analysis of Time Series: Theory and Practice, Chapman and Hall (1975).

Definitions

A Time Series is a collection of observations x_t made sequentially in time.
A discrete-time time series is a collection of observations x_t in which the set T_0 of times at which observations are made is a discrete set.
A continuous-time time series is a collection of observations x_t are made continuously over some time interval.

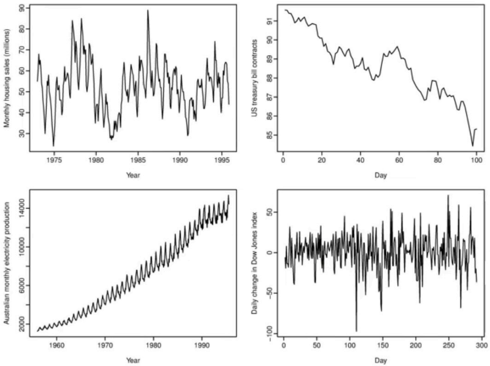
Definitions

A Time Series is a collection of observations x_t made sequentially in time.
A discrete-time time series is a collection of observations x_t in which the set T_0 of times at which observations are made is a discrete set.
A continuous-time time series is a collection of observations x_t are made continuously over some time interval.

We shall be speaking about discrete time series

Definitions

Definition of Time Series: *An <u>ordered</u> sequence of values of a variable at equally spaced time intervals.*



Introduction

Time

We mean by *time*:

- Seconds, hours, years,...

Introduction

Time

We mean by *time*:

- Seconds, hours, years,...
- Spatial: 1^{st} machine in a row, 2^{nd} machine,...

Time

We mean by *time*:

- Seconds, hours, years,...
- Spatial: 1^{st} machine in a row, 2^{nd} machine,...
- Depth: one millimetre down, two millimetre down,...

Introduction

Time

We mean by *time*:

- Seconds, hours, years,...
- Spatial: 1^{st} machine in a row, 2^{nd} machine,...
- Depth: one millimetre down, two millimetre down,...

The important point here is to have an *ordered variable like time* that there is a meaning of direction in its values. Then from a given observation, *past, present and future have a meaning.*

Introduction

Plotting a time series is an important early step in its analysis
In general, a plot can reveal:

- *Trend*: upward or downward pattern that might be extrapolated into the future

Introduction

Plotting a time series is an important early step in its analysis
In general, a plot can reveal:

- *Trend*: upward or downward pattern that might be extrapolated into the future
- *Periodicity*: Repetition of behavior in a regular pattern

Introduction

Plotting a time series is an important early step in its analysis

In general, a plot can reveal:

- *Trend*: upward or downward pattern that might be extrapolated into the future
- *Periodicity*: Repetition of behavior in a regular pattern
- *Seasonality*: Periodic behavior with a known period (hourly, monthly, every 2 months...)

Introduction

Plotting a time series is an important early step in its analysis
In general, a plot can reveal:

- *Trend*: upward or downward pattern that might be extrapolated into the future
- *Periodicity*: Repetition of behavior in a regular pattern
- *Seasonality*: Periodic behavior with a known period (hourly, monthly, every 2 months...)
- *Heteroskedasticity*: changing variance

Introduction

Plotting a time series is an important early step in its analysis
In general, a plot can reveal:

- *Trend*: upward or downward pattern that might be extrapolated into the future
- *Periodicity*: Repetition of behavior in a regular pattern
- *Seasonality*: Periodic behavior with a known period (hourly, monthly, every 2 months...)
- *Heteroskedasticity*: changing variance
- *Dependence*: positive (successive observations are similar) or negative (successive observations are dissimilar)

Introduction

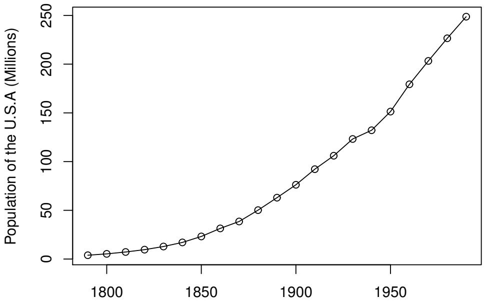
Plotting a time series is an important early step in its analysis
In general, a plot can reveal:

- *Trend*: upward or downward pattern that might be extrapolated into the future
- *Periodicity*: Repetition of behavior in a regular pattern
- *Seasonality*: Periodic behavior with a known period (hourly, monthly, every 2 months...)
- *Heteroskedasticity*: changing variance
- *Dependence*: positive (successive observations are similar) or negative (successive observations are dissimilar)
- Missing data, outliers, breaks...

Introduction

Example 1: U.S.A. population at ten year intervals from 1790-1990

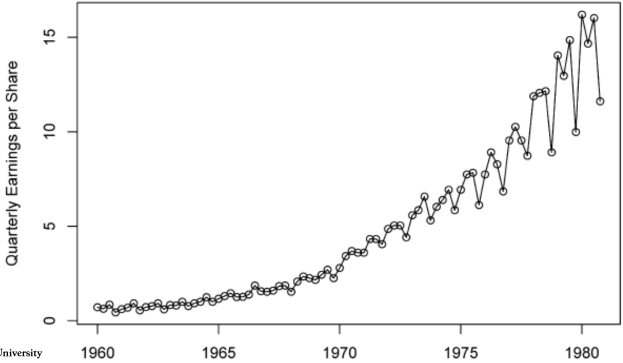
- There is a upward trend
- There is a slight change in shape/structure
- Nonlinear behavior



Introduction

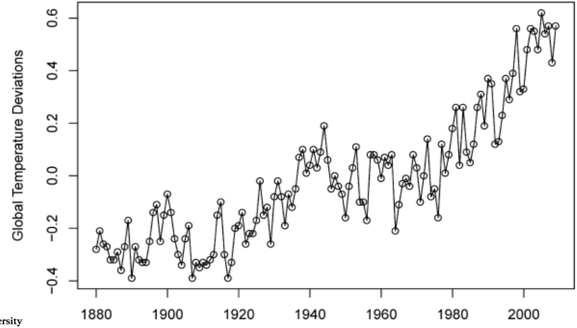
Example 2: Johnson & Johnson Quarterly Earnings

There are 84 quarters (21 years) measured from the 1st quarter of 1960 to the last quarter of 1980.
Note the gradually increasing underlying trend and the rather regular variation superimposed on the trend that seems to repeat over quarters.



Example 3: Global Warming

The data are the global mean land-ocean temperature index from 1880 to 2009. We note an *apparent upward trend* in the series during the latter part of the 20th century that has been used as an argument for the global warming hypothesis (whether the overall trend is natural or whether it is caused by some human-induced interface)

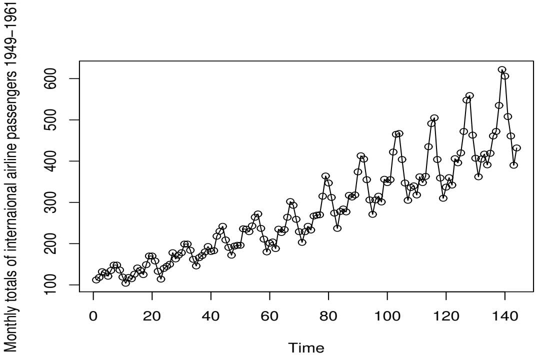


Introduction

Example 4: Airline passengers from 1949-1961

Trend? Seasonality? Heteroskedasticity? ...

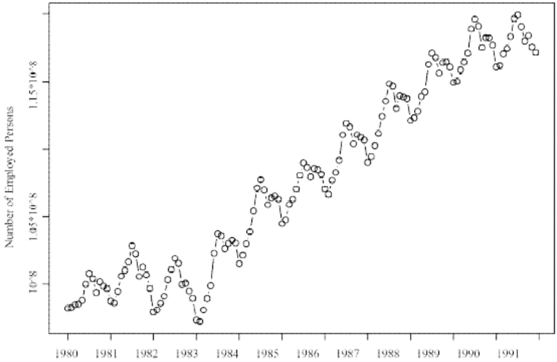
Upward trend, seasonality on a 12 month interval, increasing variability



Example 5: Monthly Employed persons from 1980-1991

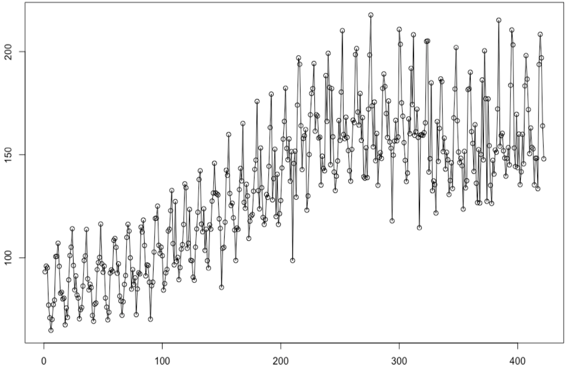
Trend? Seasonality? Heteroskedasticity? ...

Upward trend, seasonality with a structural break



Example 6: Monthly Beer Production in Australia

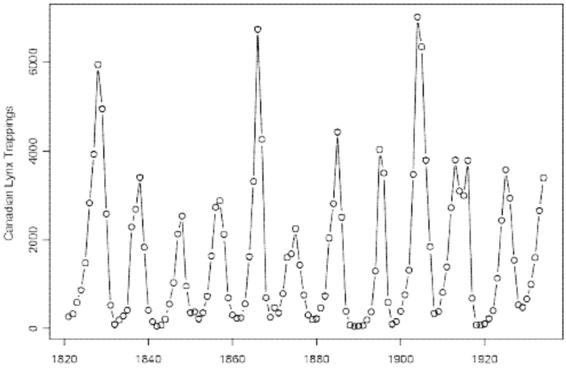
Trend? Seasonality? Heteroskedasticity? breaks?... **no trend in last 100 Months, no clear seasonality**



Introduction

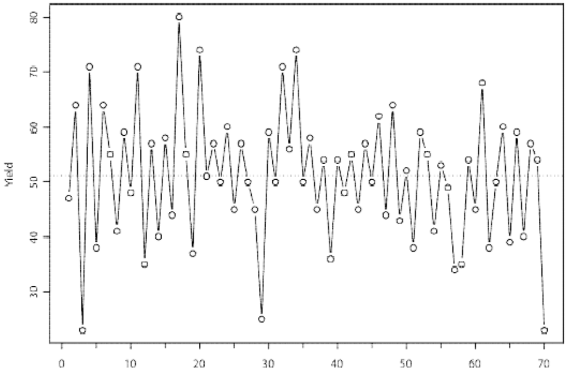
Example 7: Annual number of Candadian Lynx trapped near McKenzie River

Trend? Seasonality? Heteroskedasticity? breaks?... **no trend, no clear seasonality as it does correspond to a known period, periodicity**



Example 8: Yield from a controlled chemical batch process

Trend? Seasonality? Heteroskedasticity? breaks?... **Negative dependence: successive observations tend to lie on opposite sides of the mean.**



Example 9: Monthly real exchange rates between U.S and Canada

Trend? Seasonality? Heteroskedasticity? breaks?...

Introduction

Example 9: Monthly real exchange rates between U.S and Canada

Trend? Seasonality? Heteroskedasticity? breaks?...

No obvious seasonality or trend. Hard to make a long range prediction. Positive Dependence: successive observations tend to lie on the same side of the mean.

Introduction

Remarks

The issue of distinguishing between dependence and trend is difficult: There is no unique decomposition of a series into trend and dependence behaviors.
The issue that tampers this question: we have only one realization. If we had many realizations, we might be able to average to determine trend.

Introduction

Objectives

What do we hope to achieve with time series analysis?

- Provide a model of the data (testing of scientific hypothesis, etc.)

Introduction

Objectives

What do we hope to achieve with time series analysis?

- Provide a model of the data (testing of scientific hypothesis, etc.)
- Predict future values (very common goal of analysis)

Introduction

Objectives

What do we hope to achieve with time series analysis?

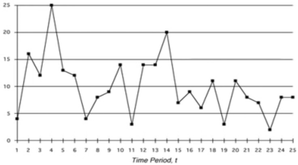
- Provide a model of the data (testing of scientific hypothesis, etc.)
- Predict future values (very common goal of analysis)
- Produce a compact description of the data (a good model can be used for "data compression")

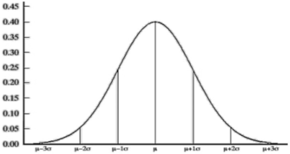
Time Series Analysis vs Statistics / Probability

- Parts of time series processes are described in terms of randomized variable with statistical moments (i.e. mean and variance).
    - In most time series data, mean and/or variance change over time
- Canonical statistical techniques assume that observations are independent and identically distributed (iid)
- The special feature of time series analysis is that successive observations are dependent. Therefore, the analysis must take into account the time order of the observations.
    - Future values can be predicted from past observations

Time Series Analysis vs Statistics / Probability

What's the main difference between a time series and a group of observations of a random variable with a known mean and standard deviation?





We might be able to describe all these data points using a normal (or any other) statistical distribution.

But does this mean that randomly sampling from this distribution gives us a representative time series?

Figure: Time Series Analysis Lecture - Jordan Kern

Time Series Analysis vs Statistics / Probability

Not if the value of x(t) depends in any significant way on the value of x(t-1).

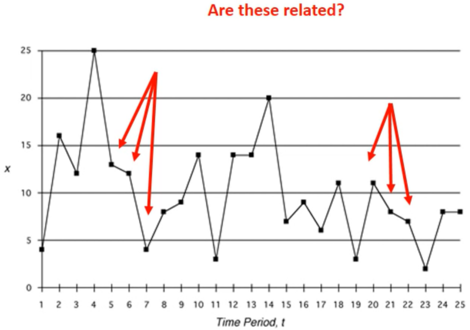


Figure: Time Series Analysis Lecture - Jordan Kern

Time Series Analysis vs Regression

- Time series analysis is focused in identifying trends and patterns, conceptionalizing them mathematically leading to:
  - Models about the data
  - Prediction/ Forecasting
- In this regard, Time series analysis is similar to Regression.
- However, regression aims to quantify specific impacts of underlying independent variables of the form:

$$y = b_1x_1 + b_2x_2 + ... + b_mx_m$$

- Time series decomposes the mathematical process into a combined signals (e.g. yearly agricultural production) and noise (random probabilistic process)

Introduction

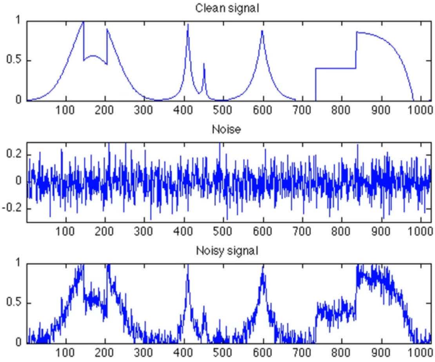


Figure: Time Series Analysis Lecture - Jordan Kern