Time Series Analysis and Modelling

Part 2: Time Series - Preprocessing and Filtering

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).

Introduction

Modeling

We take the approach that the data is a realization of random variable. However, many statistical tools are based on assuming any R.V. are IID.

In Times Series:

- R.V. are usually not independent (affected by trend and seasonality)

Introduction

Modeling

We take the approach that the data is a realization of random variable. However, many statistical tools are based on assuming any R.V. are IID.

In Times Series:

- R.V. are usually not independent (affected by trend and seasonality)
- Variance may change significantly

Modeling

We take the approach that the data is a realization of random variable. However, many statistical tools are based on assuming any R.V. are IID.

In Times Series:

- R.V. are usually not independent (affected by trend and seasonality)
- Variance may change significantly
- R.V. are usually not identically distributed

The first goal in time series modeling **is to reduce the analysis needed to a simpler case**: Eliminate Trend, Seasonality, and heteroskedasticity then we model the remainder as dependent but Identically distributed

Introduction

The probabilistic model

A complete probabilistic model/description of a time series X_t observed as a collection of n random variables at times $t_1, t_2, \ldots, t_n$ for any positive integer n is provided by the joint probability distribution,

$$F(C_1, C_2, ..., C_n) = P(X_1 \leq C_1, ..., X_n \leq C_n)$$

- This is generally difficult to write, unless the case the variables are jointly normal.

Introduction

The probabilistic model

A complete probabilistic model/description of a time series X_t observed as a collection of n random variables at times t_1,t_2,. . . , t_n for any positive integer n is provided by the joint probability distribution,

$$F(C_1, C_2, ..., C_n) = P(X_1 \leq C_1, ..., X_n \leq C_n)$$

- This is generally difficult to write, unless the case the variables are jointly normal.
- Thus, we look for other statistical tools => quantifying dependencies

Introduction

Recall the basic concepts

X and Y are r.v.'s with finite variance

$$Cov(X, Y) = E((X - E(X))(Y - E(Y)))$$

and correlation

$$corr(X, Y) = \frac{Cov(X, Y)}{S_X S_Y}$$

- r.v.'s with zero correlation are uncorrelated
- Uncorrelated does not imply Independence
-

Recall the equation for describing the correlation between two random variables, X and Y.

$$r_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

$$r_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \ \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Some properties of Expectation and Variances/Covariances

X, Y, W, and Z are r.v.'s

- Cov(Y,Z) = E(YZ) - E(Y)E(Z)
- Var(X) = Cov(X,X) = E(X^2) - $(E(X))^2$
- Var(a+bX)= b^2 Var(X)
- Cov(aX+bY,cZ+dW)= ac Cov(X,Z) + ad Cov(X,W) + bc Cov(Y,Z) +bd Cov(Y,W)
- $E(\sum X_i) = \sum E(X_i)$

These are important! Remember them :)

Introduction

A time series model for the observed data x_t

- The mean function $\mu_X = E(X_t)$
- The Covariance function
  $\gamma_X(r, s) = E((X_r - \mu_X(r))(X_s - \mu_X(s)))$ for all integers r and s

The focus will be to determine the mean function and the Covariance function to define the time series model.

Introduction

The equation for autocorrelation looks pretty similar – we're just measuring the correlation between a time series and itself, at different "lags" (shifts in time).

Given measurements, $Y_1, Y_2, \ldots, Y_N$ at time $t_1, t_2, \ldots, t_N$, the autocorrelation for a lag of k is:

$$r_k = \frac{\sum_{i=1}^{N-k}(Y_i - \bar{Y})(Y_{i+k} - \bar{Y})}{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}$$

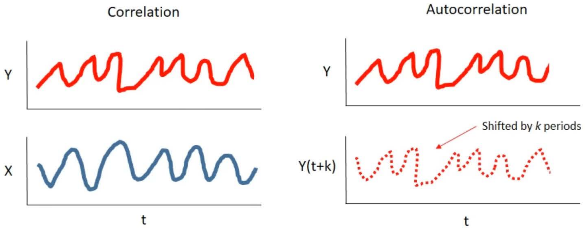


Figure: Time Series Analysis Lecture - Jordan Kern

Some zero-Mean Models

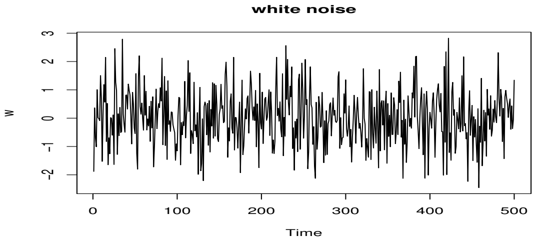
iid Noise

The simplest model for a times series: no trend or seasonal component and in which the observations are IID with zero mean.

- We can write, for any integer n and real numbers x_1, x_2,...,x_n,

$$P(X_1 \leq x_1, ..., X_n \leq x_n) = P(X_1 \leq x_1)...P(X_n \leq x_n)$$

It plays an important role as a building block for more complicated time series models



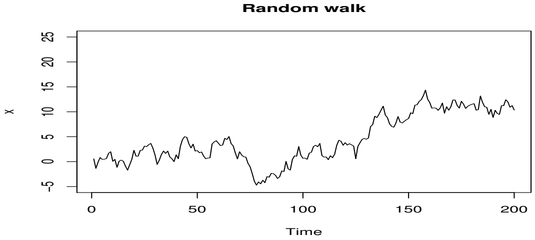
Some zero-Mean Models

Random Walk

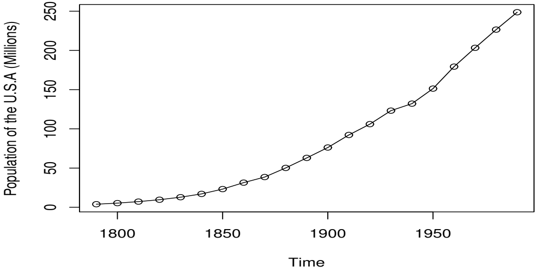
The random walk $\{S_t\}, t = 0, 1, 2, ...$ is obtained by cumulatively summing iid random variables, $S_0 = 0$

$$S_t = X_1 + X_2 + \cdots + X_t, \qquad t = 1, 2, ...$$

where X_t is iid noise. It plays an important role as a building block for more complicated time series models



Models with Trend



In this case a zero-mean model for the data is clearly inappropriate. The graph suggests trying a model of the form:

$$X_t = m_t + Y_t$$

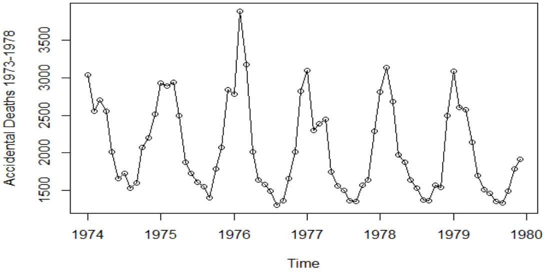
where m_t is a function known as the trend component and Y_t has a zero mean. Estimating m_t?

Models with Trend

m_t can be estimated using a least squares regression procedure (quadratic regression)
The estimated trend component $\hat{m}_t$ provides a natural predictor of future values of X_t

$$\hat{X}_t = \hat{m}_t + Y_t$$

Models with Seasonality



In this case a zero-mean model for the data is clearly inappropriate. The graph suggests trying a model of the form:

$$X_t = S_t + Y_t$$

where S_t is a function known as the season component and Y_t has a zero mean. Estimating S_t?

Time series Modeling

- Plot the series => examine the main characteristics (trend, seasonality, ...)
- Remove the trend and seasonal components to get stationary residuals/models
- Choose a model to fit the residuals using sample statistics (sample autocorrelation function)
- Forecasting will be given by forecasting the residuals to arrive at forecasts of the original series X_t

Stationary and Autocorrelation function

Let X_t be a time series:

- The mean function

$$\mu_X(t) = E(X_t)$$

- The Covariance function

$$\gamma_X(r, s) = Cov(X_r, X_s) = E((X_r - \mu_X(r))(X_s - \mu_X(s)))$$

for all integers r and s

Stationary and Autocorrelation function

Definitions

1. X_t is **strictly** stationary if $\{X_1, \ldots X_n\}$ and $\{X_{1+h}, \ldots X_{n+h}\}$ have the same joint distributions for all integers h and $n > 0$.

Stationary and Autocorrelation function

Definitions

1. X_t is **strictly** stationary if $\{X_1, \ldots X_n\}$ and $\{X_{1+h}, \ldots X_{n+h}\}$ have the same joint distributions for all integers h and $n > 0$.
2. X_t is **weakly** stationary if
   - $\mu_X(t)$ is independent of t.
   - $\gamma_X(t+h, t)$ is independent of t for each h.

Stationary and Autocorrelation function

Definitions

1. X_t is **strictly** stationary if $\{X_1, \ldots X_n\}$ and $\{X_{1+h}, \ldots X_{n+h}\}$ have the same joint distributions for all integers h and $n > 0$.
2. X_t is **weakly** stationary if
   - $\mu_X(t)$ is independent of t.
   - $\gamma_X(t+h, t)$ is independent of t for each h.
3. Let X_t be a stationary time series. **The autocovariance function** (ACVF) of X_t at lag h is

$$\gamma_X(h) = Cov(X_{t+h}, X_t)$$

The **autocorrelation function** (ACF) of X_t at lag h is

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = Cor(X_{t+h}, X_t)$$

Stationary and Autocorrelation function

iid Noise

If X_t is iid noise and $E(X_t^2) = \sigma^2 < \infty$, then The process X_t is strictly stationary since the joint distribution can be written, for any integer n and real numbers c_1, c_2,...,c_n, as follows:

$$\begin{aligned} P(X_1 \leq c_1, ..., X_n \leq c_n) &= P(X_1 \leq c_1)...P(X_n \leq c_n) \\ &= F(c_1)\dots F(c_n) \\ &= \prod_{i=1}^{n} F(c_i) \end{aligned}$$

This does not depend on t. $X_t \sim IID(0, \sigma^2)$

The autocovariance function

$$\gamma_X(t+h, h) = \begin{cases} \sigma^2, & \text{if } h = 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Stationary and Autocorrelation function

White Noise

If X_t is a sequence of uncorrelated random variables, each with zero mean and variance σ^2, then clearly X_t is stationary with the same autocovariance function as the iid noise. We can write

$$X_t \sim WN(0, \sigma^2)$$

Clearly, every $IID(0, \sigma^2)$ sequence is $WN(0, \sigma^2)$ but not conversely

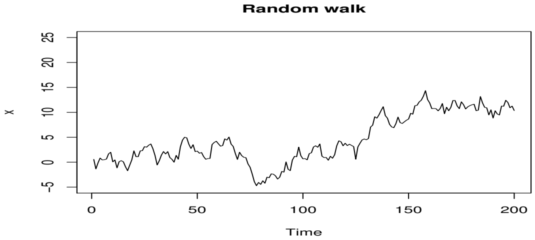
Stationary and Autocorrelation function

Random Walk

If $\{S_t\}$ is the random walk with X_t as a $IID(0, \sigma^2)$ sequence, then The random walk $\{S_t\}$, $t = 0, 1, 2, ...$ is obtained by cumulatively summing iid random variables, $S_0 = 0$

$$S_t = X_1 + X_2 + \cdots + X_t, \qquad t = 1, 2, ...$$

where X_t is iid noise. It plays an important role as a building block for more complicated time series models



First-Order Moving Average or MA(1) Process

Consider the series defined by the equation

$$X_t = Z_t + \theta Z_{t-1}, \qquad t = 0, \pm 1, \ldots.$$

where Z_t is $WN(0, \sigma^2)$ noise and θ is a real-valued constant.

- $E(X_t)$?
- $\gamma_X(t+h, h)$?

First-Order AutoRegressive AR(1) Process

Let us assume now that X_t is a stationary series satisfying the equation

$$X_t = \Phi X_{t-1} + Z_t, \qquad t = 0, \pm 1, \ldots.$$

where Z_t is $WN(0, \sigma^2)$ noise, $|\Phi| < 1$, and Z_t is uncorrelated with X_s for each $s < t$.

- $E(X_t)$?
- $\gamma_X(t+h, h)$?

The Sample Autocorrelation function

In practical problems, we do not start with a model, but with observed data $(x_1, x_2, \ldots, x_n)$. **To assess the degree of dependence** in the data and to **select a model** for the data, one of the important tools we use is the sample autocorrelation function (Sample ACF).

Definition

Let $x_1, x_2, \ldots, x_n$ be observations of a time series. The sample mean of $x_1, x_2, \ldots, x_n$ is

$$\overline{x} = 1/n \sum_{t=1}^{n} x_i$$

The sample autocovariance function is

$$\hat{\gamma}(h) := 1/n \sum_{t=1}^{n-|h|} (x_{t+|h|} - \overline{x})(x_t - \overline{x}), \quad -n < h < n$$

The sample autocorrelation function is

$$\hat{\rho} = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

The Sample Autocorrelation function

Remarks

1. The sample autocorrelation function (ACF) can be computed for any data set and is not restricted to observations from a stationary time series.
2. For data containing a Trend, $|\hat{\rho}(h)|$ will display slow decay as h increases.
3. For data containing a substantial deterministic periodic component, $|\hat{\rho}(h)|$ will exhibit similar behavior with the same periodicity.

The Sample Autocorrelation function

We may recognize the sample autocorrelation function of many time series:

Remarks

1. White Noise => Zero
2. Trend => Slow decay
3. Periodic => Periodic
4. Moving Average (q) => Zero for $|h| > q$
5. AutoRegression (p) => Decay to zero exponentially

Estimation and Elimination of Trend and Seasonal Components

The first step in the analysis of any time series is to plot the data. Inspection of the graph may suggest the possibility of representing the data as follows (the classical decomposition):

$$X_t = m_t + s_t + Y_t$$

where

- m_t is the trend component
- s_t is the seasonal component
- Y_t random noise component / Residuals

if seasonal and noise fluctuations appear to increase with the level of the process => Eliminate by using a preliminary transformation of the data (natural log,...).

Estimation and Elimination of Trend and Seasonal Components

Approaches

1. Estimate and eliminate the trend and the seasonal components in the hope that the residual Y_t will turn out to be a stationary time series => Find a Model using stationary process theory.
2. Box and Jenkins (1976) proposed to apply differencing operators to the series until the differenced observations resemble a realization of some stationary time series.

Estimation and Elimination of Trend and Seasonal Components

Trend Estimation

Moving average and spectral smoothing are an essentially nonparametric methods for trend (or signal) estimation

$$X_t = m_t + Y_t, \quad E(Y_t) = 0$$

1. Smoothing with a finite moving average filter
2. Exponential smoothing
3. Smoothing by eliminating the high-frequency components (Fourier series)
4. Polynomial fitting (Regression)

Estimation and Elimination of Trend and Seasonal Components

Trend Estimation: Smoothing Moving Average

Let q be a nonnegative integer and consider the two-sided moving average

$$W_t = \frac{\sum_{j=-q}^{q} X_{t-j}}{2q+1}$$

It is useful to think about $\hat{m}_t$ as a process obtained from X_t by application of a linear operator or linear filter $\hat{m}_t = \sum_{j=-\infty}^{\infty} a_j X_{t-j}$ with weights $a_j = \frac{1}{2q+1}$.

This particular filter is a low-pass filter in the sense that it takes the data and removes from it the rapidly fluctuating (high frequency) components $\hat{Y}_t$ to leave slowly varying estimated trend term $\hat{m}_t$.

Estimation and Elimination of Trend and Seasonal Components

Trend Estimation: Exponential Smoothing

For any fixed $\alpha \in [0, 1]$, the one-sided moving averages $\hat{m}_t$ defined by:

$$\hat{m}_t = \alpha X_t + (1 - \alpha)\hat{m}_{t-1}; t = 2, ...n$$

and $\hat{m}_1 = X_1$

Estimation and Elimination of Trend and Seasonal Components

Trend Estimation: Smoothing by eliminating of high frequency

Using Fourier Transform we can delete some high frequency.

Trend Estimation: Polynomial fitting

Using regression procedures

Nonseasonal Model With Trend: Estimation

Remark: Smoothing Moving Average

- There are many filters that could be used for smoothing!
- Large q will allow linear trend function $m_t = c_0 + c_1 t$ to pass
- We must be aware of choosing q to be too large if m_t is not linear

- Clever choice of the weights a_j can design a filter that will not only be effective in attenuating noise in the data, but that will also allow a larger class of trend functions (for example all polynomials of degree ≤ 3) to pass.

- Spencer 15-point moving average is a filter that passes polynomials of degree ≤ 3 without distortion. Its weights are:
$1/320[-3, -6, -5, 3, 21, 46, 67, 74, 67, 46, 21, 3, -5, -6, -3]$

Estimation and Elimination of Trend and Seasonal Components

Trend Elimination: Differencing

Instead of attempting to remove the noise by smoothing as in Method 1, we now attempt to eliminate the trend term by differencing. We define the lag-1 difference operator ∇ by:

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t$$

where B is the backward shift operator : $BX_t = X_{t-1}$

Estimation and Elimination of Trend and Seasonal Components

Trend Elimination: Differencing

Powers of the operators B and ∇ are defined as follows:

$$\nabla^j(X_t) = \nabla(\nabla^{j-1}(X_t))$$

where $j >= 1, \nabla^0(X_t) = X_t$ and

$$B^j(X_t) = X_{t-j}$$

Example: $\nabla^2 X_t$?

If the operator ∇ is applied to a linear trend function $m_t = c_0 + c_1 t$, then we obtain the constant function $\nabla m_t = m_t - m_{t-1} = c_1$. In the same way we can show any polynomial trend of degree k can be reduced to a constant by application of the operator ∇^k.

It is found in practice that the order k of differencing required is quite small; frequently 1 or 2.

Estimation and Elimination of Trend and Seasonal Components

Both Trend and Seasonal

The classical Decomposition model

$$X_t = m_t + s_t + Y_t$$

where $E(Y_t) = 0$, $s_{t+d} = s_t$, and $\sum_{j=1}^{d} s_j = 0$

d is the period of the seasonal component.
The methods described previously (for the trend) can be adapted in a natural way to eliminate both the trend and seasonality components.

Estimation: Both Trend and Seasonal [Method 1]

Suppose we have observations $x_1, x_2, ... x_n$

1. The trend is first estimated by applying a moving average filter specially to eliminate the seasonal component of period d
   - if the period is even, say d=2q then $\hat{m}_t = (0.5x_{t-q} + x_{t-q+1}, \ldots, 0.5x_{t+q})/d \;\; q < t \leq n - q$
   - if the period is odd, say d=2q+1 then we use the simple moving average
2. Estimate the seasonal component; for each k=1,..., d, we compute the average w_k of the deviation $x_{k+jd} - \hat{m}_{k+jd}$, $q < k + jd \leq n - q$; and we estimate the seasonal component as follows:

$$\hat{s}_k = w_k - \frac{\sum_{i=1}^{d} w_i}{d}; \quad k = 1, \ldots, d$$

   and $\hat{s}_k = \hat{s}_{k-d}; \quad k > d$
3. The *deseasonalized* data is then $d_t = x_t - \hat{s}_t$ t=1,...n
4. Reestimate the trend from the *deseasonalized* data using one of the methods already described.

Eliminating: Both Trend and Seasonal [Method 2]

We can use the Differencing operator to eliminate the trend and the seasonal component

1. Eliminate the seasonality of period d using $\nabla_d X_t = X_t - X_{t-d}$
   $X_t = m_t + s_t + Y_t$ Applying $\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d}$
2. Eliminate the trend $m_t - m_{t-d}$ using $\nabla^k X_t$

Testing Noise squence

IID Null Hypothesis

1. Q_{LB} Ljung-Box Test (H_0 : data are Independent vs. H_1 : data are not independent)
2. Q_{ML} McLeod- Li Test (autocorrelations of squared data)
3. Turning point Test (IID vs. not IID)
4. The difference-sign Test (randomness)
5. The rank Test (detecting a linear trend)

Remember that as you increase the number of tests, the probability of at least one rejects the null hypothesis when it is true increases.

Summary

Trend and Seasonality

- Estimation and Elimination of Trend and Seasonal Components
  1. Smoothing methods
  2. Differencing Operator ∇
  3. Properties of the operator ∇
  4. Procedure to estimate Both Trend and Seasonality
- Testing Noise sequence