Time Series Data Analysis

Knowledge Discovery and Data Mining 2 (VU) (706.715)

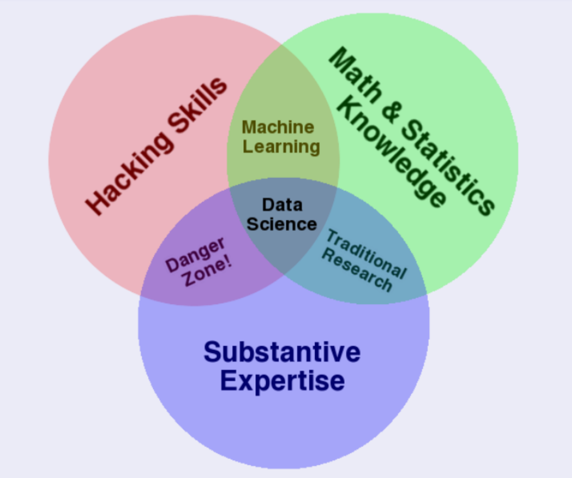
Maximilian Toller

Know-Center GmbH

2019-03-28

Recall from earlier

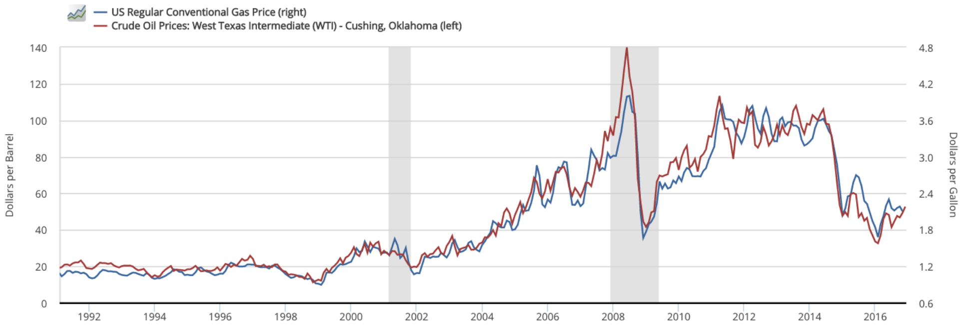
Why KDDM2?



What are *time series*?

What are time series?

Data observed over time



What are time series?

Stochastic processes indexed by integers

- $\{X_t | t \in T\}\ T = \mathbb{Z}$
- Confirmatory data analysis
- Goal: See if model is sound
- Mainly about: theorems, models, proofs
- Pros: Provably correct, theoretically sound
- Cons: "*All models are wrong*" - George Box

What are time series?

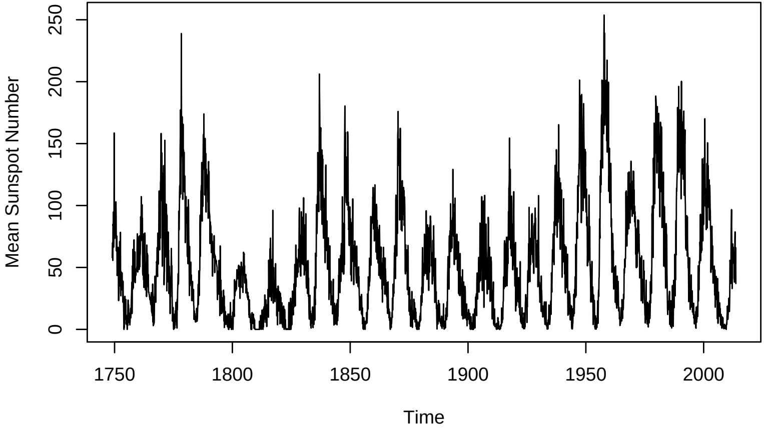
Data vs Process

- $x_t = \{114, 117, 104, \dots\}$
- Exploratory data analysis
- Work with data
- Pros: fast, domain specific
- Cons: possibly unsound

- $\{X_t | t \in T\}$ $T = \mathbb{Z}$
- Confirmatory data analysis
- Work with models
- Pros: theoretically sound
- Cons: slow, simplification

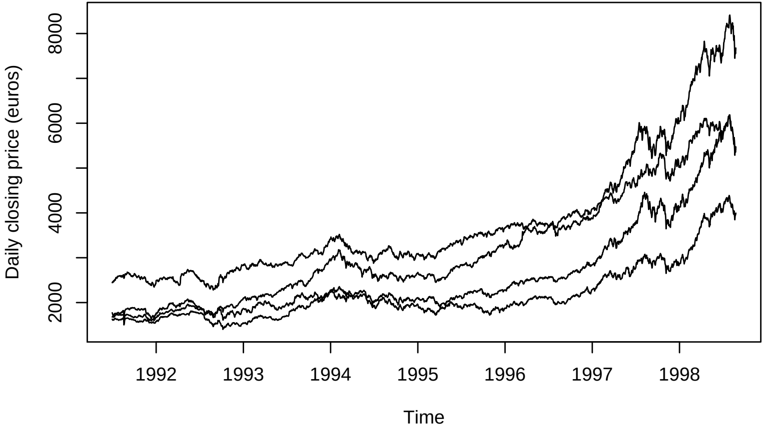
What are time series data?

Sunspot counts (monthly)



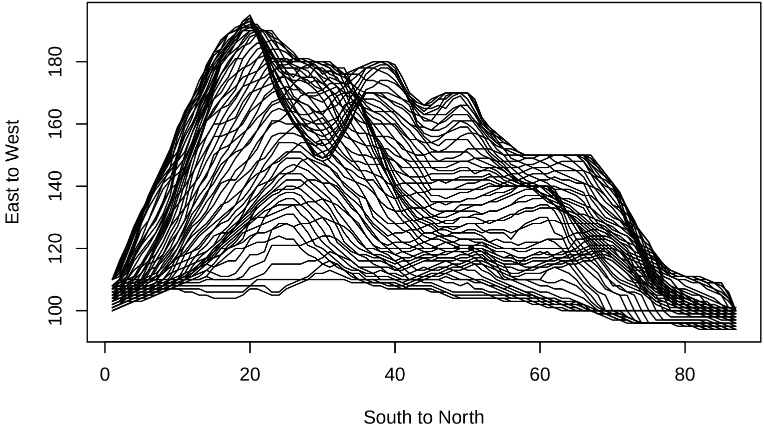
What are time series data?

EU stock market prices (daily)?



What are time series data?

Volcano topography?



Obtaining time series data

- Stream data mining repository
  http://www.cse.fau.edu/~xqzhu/stream.html
- UCI machine learning repository
  https://archive.ics.uci.edu/ml/datasets.html
- UEA & UCR time series classification repository
  http://timeseriesclassification.com/

Obtaining time series data

Unevenly spaced & incomplete data

- Often data are not evenly sampled
- Time series theory requires $t \in \mathbb{Z}$
- Linear interpolation: $x_t = x_0 + t\frac{x_1 - x_0}{r_1 - r_0} \quad r \in \mathbb{R}$
- Missing value imputation

γ and ρ

Some math we'll need later

Some math we'll need later

Autocovariance

- $\mu_t = \mathbb{E}[X_t]$
- $\gamma(\tau, k) = \mathbb{E}[(X_\tau - \mu_\tau)(X_k - \mu_k)]$
- $\hat{\mu} = \text{undefined}, \quad \hat{\mu}_t = \frac{1}{m}\sum_{i=1}^{m} x_t^{(i)}$
- $\hat{\gamma}(\tau, k) = \frac{1}{n-1}\sum_{i=1}^{N}(x_{i\tau} - \mu_\tau)(x_{ik} - \mu_k)$

Some math we'll need later

Autocorrelation

- $\rho(\tau) = \frac{\gamma(\tau,k)}{\sqrt{\gamma(\tau,\tau)\gamma(k,k)}}$
- $\hat{\rho}(\tau, k) = \frac{\hat{\gamma}(\tau,k)}{\sqrt{\hat{\gamma}(\tau,\tau)\gamma(k,k)}}$
- *With only one realization x_t, we can't compute this*

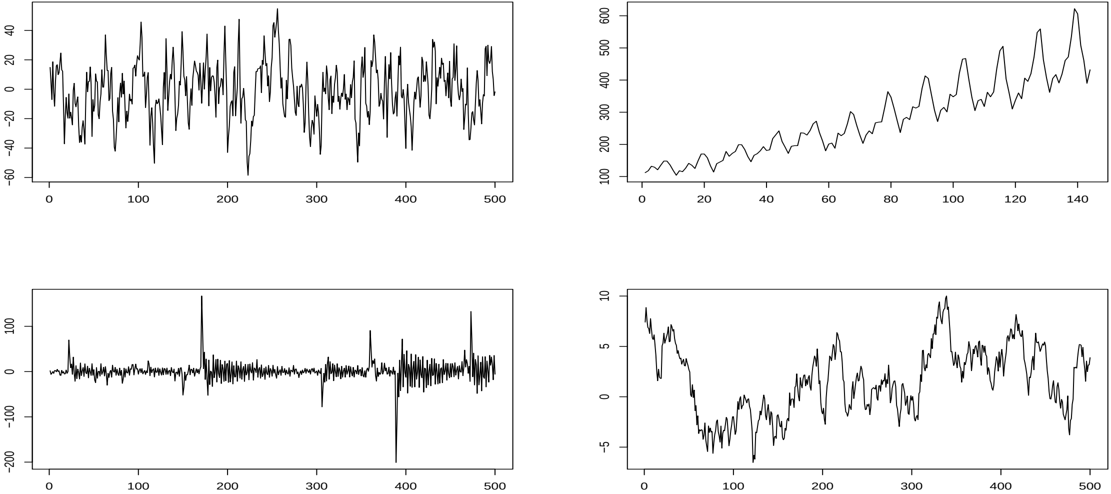
Stationarity

What, why and how?

What? - Theoretical definition

- Strict stationarity
  - $F_X(X_t, \dots, X_{t+k}) = F_X(X_{t+\tau}, \dots, X_{t+\tau+k}) \quad \text{for all } t, \tau, k \in \mathbb{Z}$
  - Time and order do not matter

- Weak stationarity
  - $E[X_t] = \mu \quad \text{for all } t$
  - $E[X^2] < \infty$
  - $E[(X_t - \mu)(X_{t+\tau} - \mu)] = \gamma(\tau) \quad \text{for all } t \text{ and any } \tau$



40
20
0
−20
−40
−60
0
100
200
300
400
500
600
500
400
300
200
100
0
20
40
60
80
100
120
140
100
0
−100
−200
0
100
200
300
400
500
10
5
0
−5
0
100
200
300
400
500

Stationarity

For ~~dummies~~ non-statisticians

- Data can't be stationary or non-stationary
- Stationarity is a property of processes
- Correct question: "*Was my data generated by a stationary process?*"
- Roughly: "no change over time"

Stationarity

Why?

- Classical statistics require strict stationarity
- Most models require at least weak stationarity
- Transformation to stationary form often possible
- Non-stationary theory is complex
- We can estimate autocorrelation

Stationarity
How?

- Augmented Dickey-Fuller test
- Priestley-Subba Rao test
- Hyndman's suggestion
- ~~Visual inspection~~

γ and ρ

Revisited

Autocovariance

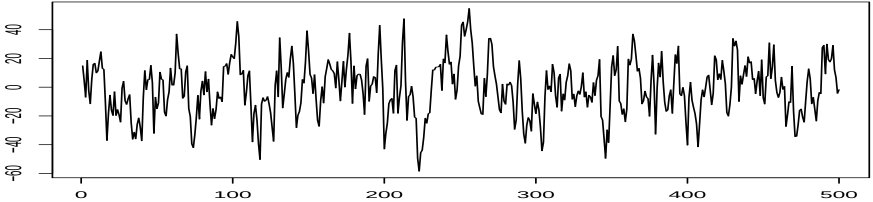
This time with only one parameter

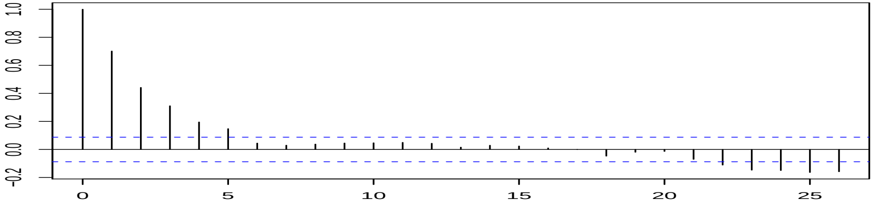
- $\mu = \mathbb{E}[X_t]$
- $\gamma(\tau) = \mathbb{E}[(X_t - \mu)(X_{t+\tau} - \mu)] \quad \text{for all } t, \tau \in \mathbb{Z}$
- $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$
- $\hat{\gamma}(\tau) = \frac{1}{n} \sum_{i=1}^{n-\tau} (x_i - \hat{\mu})(x_{i+\tau} - \hat{\mu})$

Autocorrelation

This time with only one parameter

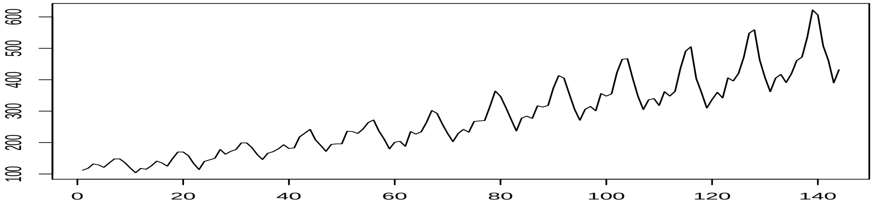
- $\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$
- $\hat{\rho}(\tau) = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)}$

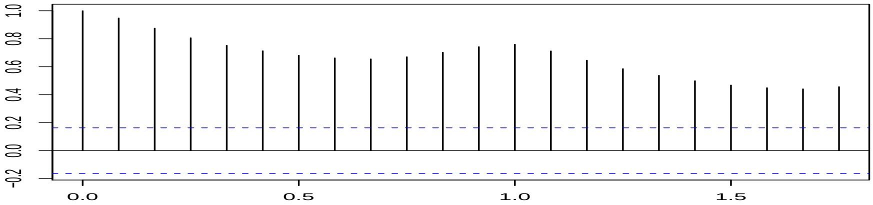

40
20
0
-20
-40
-60
0
100
200
300
400
500


1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
20
25

Autocorrelation

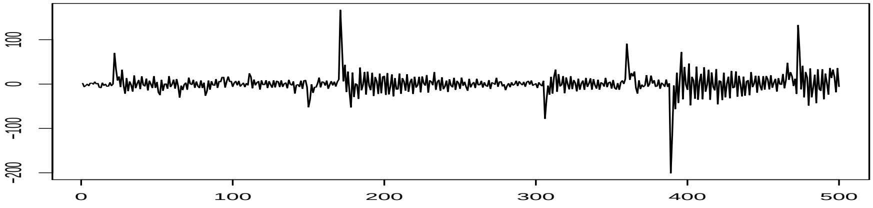
Examples

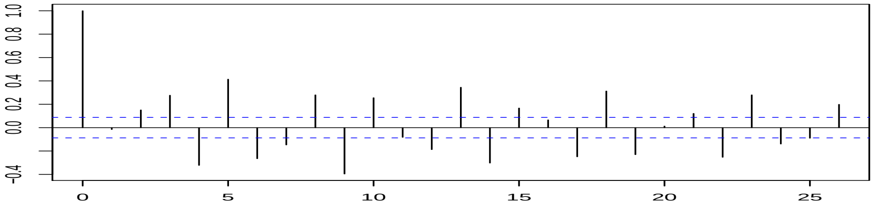


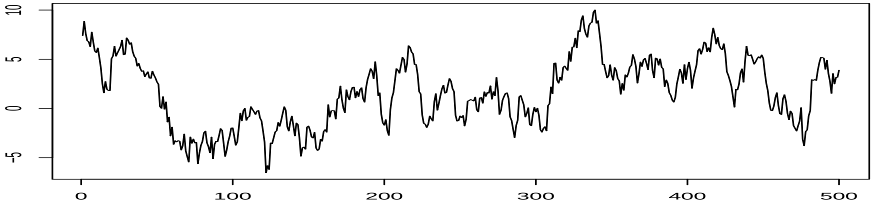


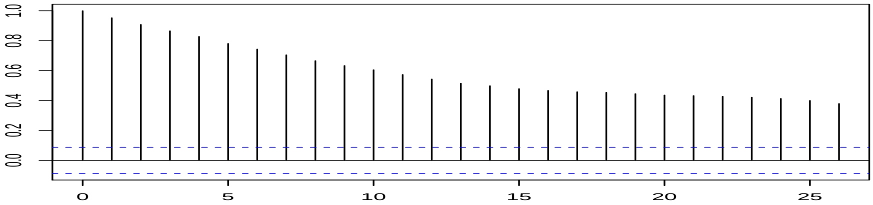
Autocorrelation

Examples






10
5
0
-5
0
100
200
300
400
500


1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25

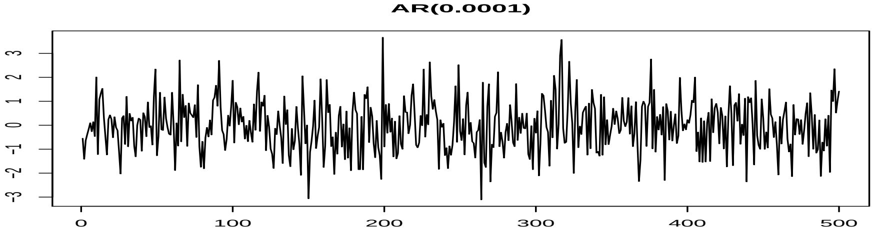
Time Series Models

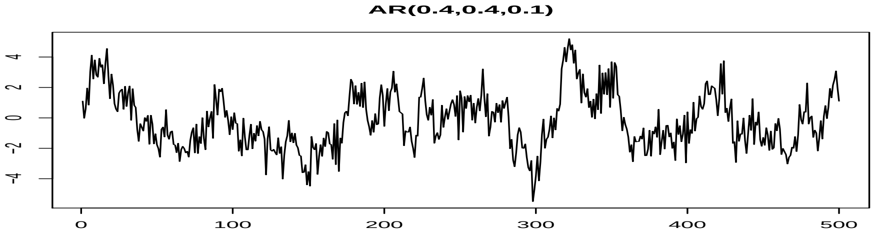
AR, MA, ARMA,...

- $\mathrm{AR}(1): \ X_t = c + \theta X_{t-1} + \epsilon_t$
- $\mathrm{AR}(p): \ X_t = c + \theta_1 X_{t-1} + \theta_2 X_{t-2} + \ldots + \theta_p X_{t_p} + \epsilon_t$
- Simple linear model of past
- Stationary if $\sum \theta$ is small
- Least squares parameter fitting

AR-Model

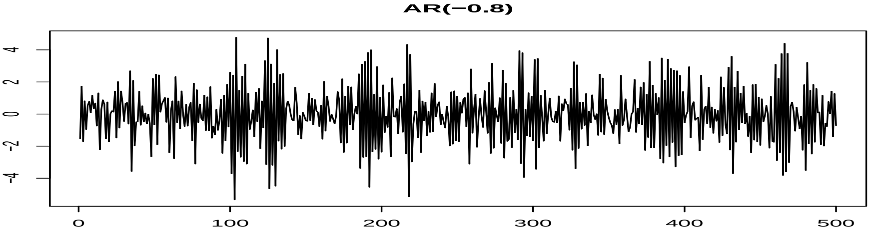
Examples

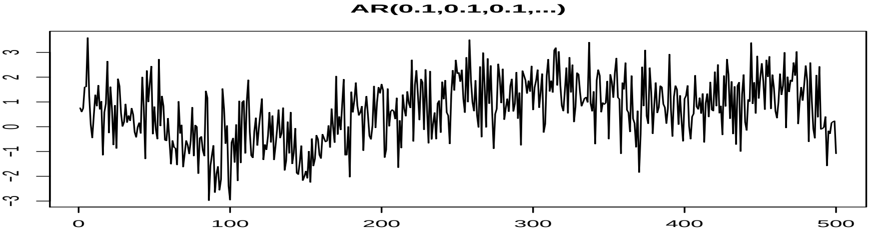




AR-Model

Examples



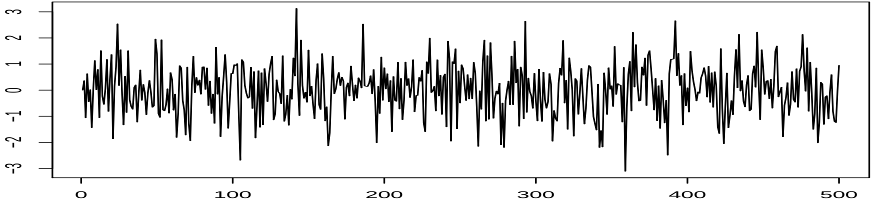


- $\mathrm{MA}(1): \; X_t = c + \epsilon_t + \phi\epsilon_{t-1}$
- $\mathrm{MA}(q): \; X_t = c + \epsilon_t + \phi_1\epsilon_{t-1} + \phi_2\epsilon_{t-2} + \ldots + \phi_q\epsilon_{t-q}$
- Don't confuse with rolling average
- Always weakly-stationary
- Assume distribution and maximize likelihood

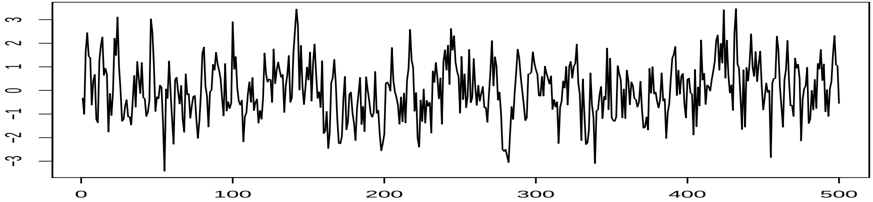
MA-Model

Examples

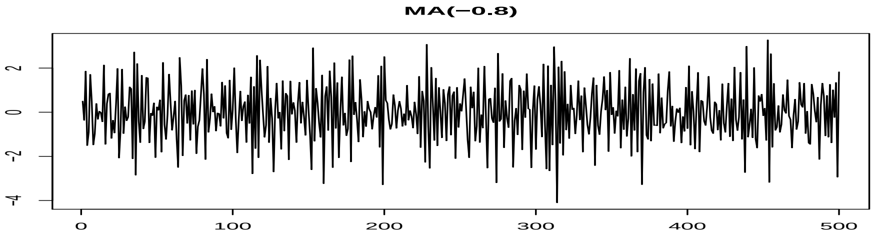
MA(0.0001)

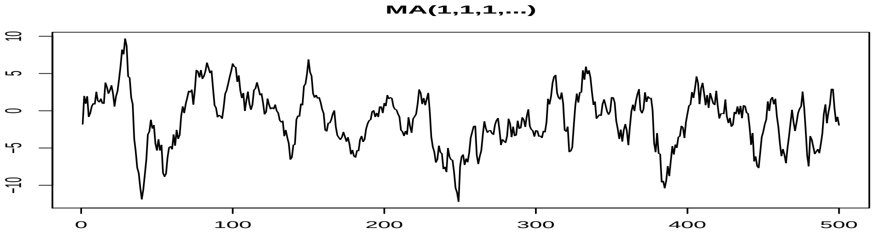


MA(0.4,0.4,0.1)



MA-Model
Examples





- $\mathrm{ARMA}(p, q): \; X_t = c + \sum_{i=1}^{p} \theta_i X_{t-i} + \sum_{j=1}^{q} \phi_j \epsilon_{t-j} + \epsilon_t$
- $\mathrm{ARMA}(p, q): \; x_t = \mathrm{AR}(p) + \mathrm{MA}(q) - c - \epsilon_t$
- Approximates large p or q
- Stationary if AR part stationary
- Parameter fitting as above

Other models

- Exponential Smoothing
- Hidden Markov Models
- NARX
- GARCH

How can we choose p and q?

ARMA order estimation

- $\alpha(1) = \rho(1)$
- $\alpha(\tau) = \frac{\mathbb{E}[(X_{\tau+1} - P_{\overline{\text{sp}}\{1,X_2,\ldots,X_\tau\}}(X_{\tau+1}) - \mu)(X_1 - P_{\overline{\text{sp}}\{1,X_2,\ldots,X_\tau\}}(X_1) - \mu)]}{\sqrt{\mathbb{E}[(X_{\tau+1} - P_{\overline{\text{sp}}\{1,X_2,\ldots,X_\tau\}}(X_{\tau+1}) - \mu)^2]\mathbb{E}[(X_1 - P_{\overline{\text{sp}}\{1,X_2,\ldots,X_\tau\}}(X_1) - \mu)^2]}}$
- ACF with lagged values estimated by linear model
- Usually Yule-Walker equations or OLS

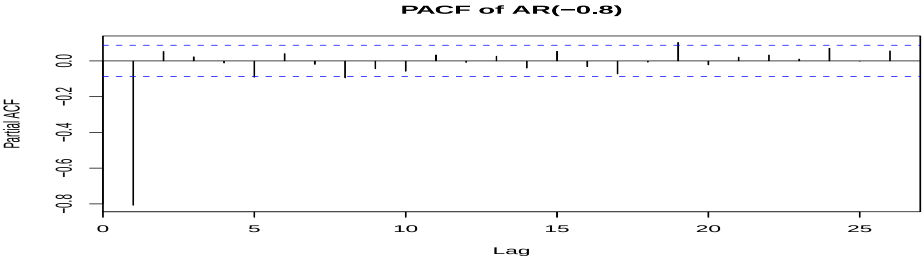
ARMA order estimation

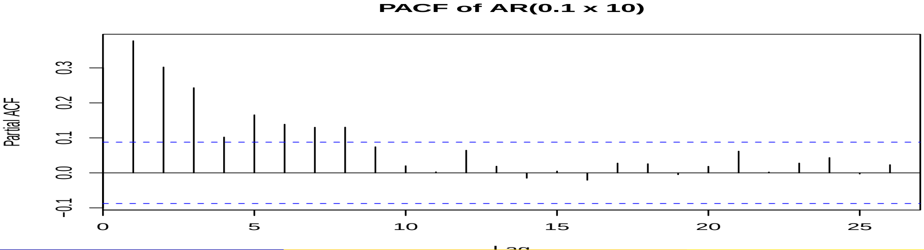
Estimating AR order p

- $\alpha(\tau \leq p)$ will be non-zero
- $\alpha(\tau > p)$ will be zero
- Compute $\hat{\alpha}$
- p is lag where $\hat{\alpha}$ enters confidence borders

ARMA order estimation

Estimating AR order p





ARMA order estimation

Estimating MA order q

- Plot ACF
- q is lag where ACF becomes zero
- Hyndman's method for stationary

ARMA order estimation

The Box-Jenkins Method

ACF Shape	**Indication**
Some spikes, almost zero	MA model, q = time to first zero
Exponential decay to zero	AR model, plot PACF to find p
Alternating exp. decay to zero	AR model, plot PACF to find p
Delayed decay	ARMA model
Peaks at fixed intervals	Data are seasonal, use SARMA
Never reaches zero	Probably not stationary, detrend
Everything almost zero	Data are independent, noise

D_t and S_t

Trend and Seasonality

Trend and Seasonality

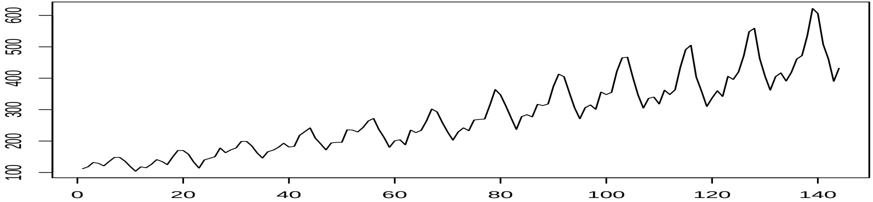
The additive model

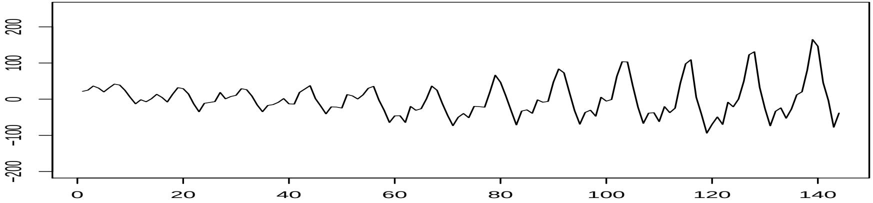
- $X_t = D_t + S_t + Y_t \quad D_t = f(t),\ S_t = g(t), S_t = S_{t+k}$
- $Y_t \ldots$ stochastic residual
- Estimate $\hat{D}_t$ and $\hat{S}_t$
- Subtract and analyze residual

- Filters
  - Assume $S_t = 0 \ \forall t$
  - Remove arbitrary polynomial
- Regression
  - Linear
  - Non-isotonic
  - Isotonic
- Differencing
  - Stochastic trend
  - $\nabla(X_t) = X_t - X_{t-1}$
- $\log(X_t)$

Trend and Seasonality

Detrending: Example



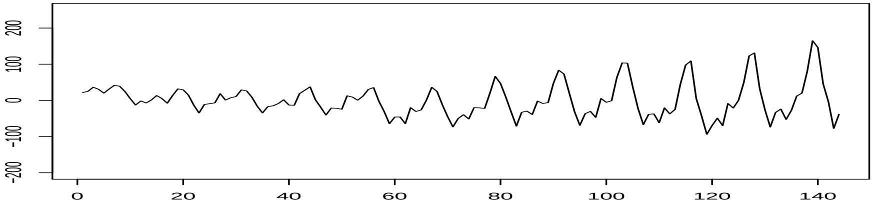


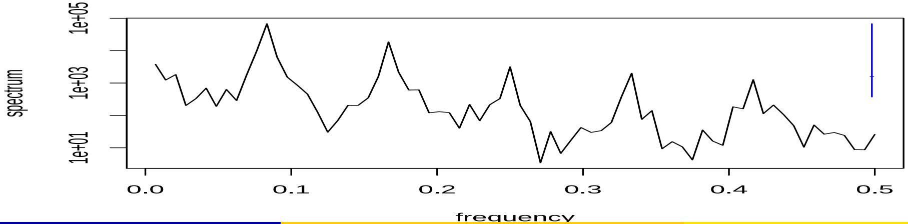
- Discrete integration $\int_{-\infty}^{\infty} X_t dt \approx \sum_{i=1}^{t-1} X_i$
- Idea: Model integrated data
- $\mathrm{ARIMA}(p, d, q)$: Integrate $\mathrm{AR}(p) + \mathrm{MA}(q)$ d times
- Actually ∇x_t computed

- Repeating events $\rightarrow$ Fourier Analysis
- Periodogram:
  - Fourier Sequence $\mathcal{F}_n(\omega)$
  - Fast Fourier Transform of ACF
- Peak Analysis: $s = \frac{1}{\arg\max_{\omega}(F_n)}$
- $\mathrm{SARIMA}(p, d, q)(P, \mathcal{D}, Q)_s$

Trend and Seasonality

Periodogram: Example





Time Series Forecasting

Estimating x_{t+k} from $x_1, \ldots, x_t$

Time Series Forecasting

Facts

- In pure theory, we are done: Set $s = t + 1$
- Maximum likelihood estimator
- Models have forecast function
- Residual analysis

Time Series Forecasting

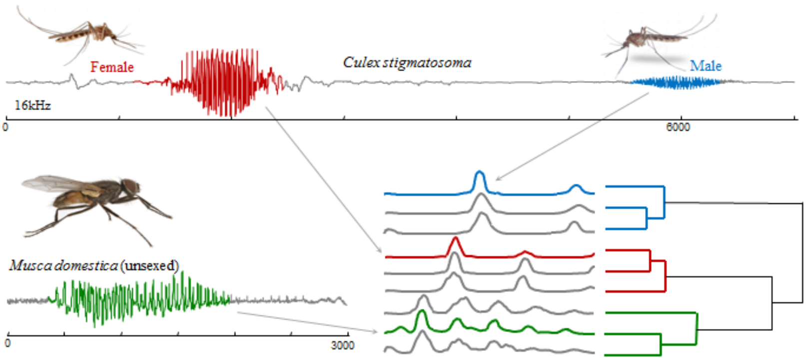
Applying what we learned so far

Live Demo

- Time Series Database
- Identify class
- Distance/Similarity Measures
  - Euclidean distance
  - Cosine similarity
  - Dynamic time warping
  - Edit distance
  - ...

Other Time Series Data Mining

Classification



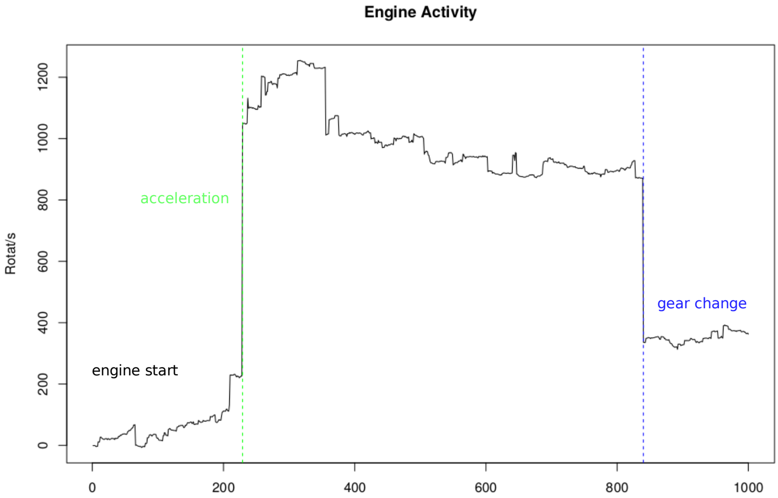
Insect classification by clustering audio snippet time series. Adapted from *Insect Detection and Classification Based on Wingbeat Sound* by Yanping Chen 2014, retrieved from `http://alumni.cs.ucr.edu/~ychen053/`. Copyright 2014 by Yanping Chen.

- Discretization: $x_t = a, b, a, c, a, c, d, c, \ldots$
- Piecewise Aggregate Approximation
- Breakpoints
- Symbolic time series

- Time series segmentation
- Change points/novelties
- Sliding windows
- CUSUM
- Detection-threshold problem

Other Time Series Data Mining

Event Detection



Tools

Some help for the practicals

- R
  - http://www.statmethods.net/advstats/timeseries.html
  - https://cran.r-project.org/web/views/TimeSeries.html
  - https://github.com/robjhyndman/
- Python
  - Prophet
  - TS-Fresh
  - Pandas, NumPy, scikit-learn, Statsmodels
- MatLab/Octave
  - TSA
  - Signal
  - ...
- Java
  - JMotif
  - Weka
  - ...

One last thing. . .

Remarks about artificial neural networks

- Feedforward ANN simulates nonlinear-$\mathrm{MA}(q)$
- Recurrent ANN simulates nonlinear -$\mathrm{ARMA}(p, q)$
- Autoregressive ANN $\neq \mathrm{AR}(p)$
- Long Short-Term Memory

The End

Next: Information Retrieval

```r
library(forecast)

ts_data <- AirPassengers %>% c() %>% as.ts()

ts_data %>% plot.ts()
ts_data %>% acf()
ts_data %>% pacf()

model1 <- Arima(y = ts_data, order = c(2,0,0))
model1 %>% forecast %>% plot(showgap=F)
model1$sigma2
model1$aic

detrended_data <- ts_data %>% diff()
detrended_data %>% plot()

model2 <- Arima(y = ts_data, order = c(2,1,0))
model2 %>% forecast %>% plot(showgap=F)
model2$sigma2
model2$aic

detrended_data %>% plot()
detrended_data %>% acf()
detrended_data %>% pacf()

model3 <- Arima(y = ts_data, order = c(2,1,1))
model3 %>% forecast() %>% plot(showgap=F)
model3$sigma2
model3$aic

detrended_data %>% acf(lag.max = 100)
pgram <- ts_data %>% spec.pgram()
{pgram$spec} %>% which.max() %>% {1/pgram$freq[.]}

model4 <- Arima(y = ts_data, order = c(2,1,1),seasonal = list(order=c(0,1,0),period=12))
model4 %>% forecast() %>% plot(showgap=F)
model4$sigma2
model4$aic

#short version
model5 <- auto.arima(ts(ts_data,frequency = 12))
model5 %>% forecast() %>% plot(showgap=F)
model5$sigma2
model5$aic
```