02417: Time Series Analysis

Week 2 - Regression based methods, 1st part

Peder Bacher

DTU Compute

Based on material previous material from the course

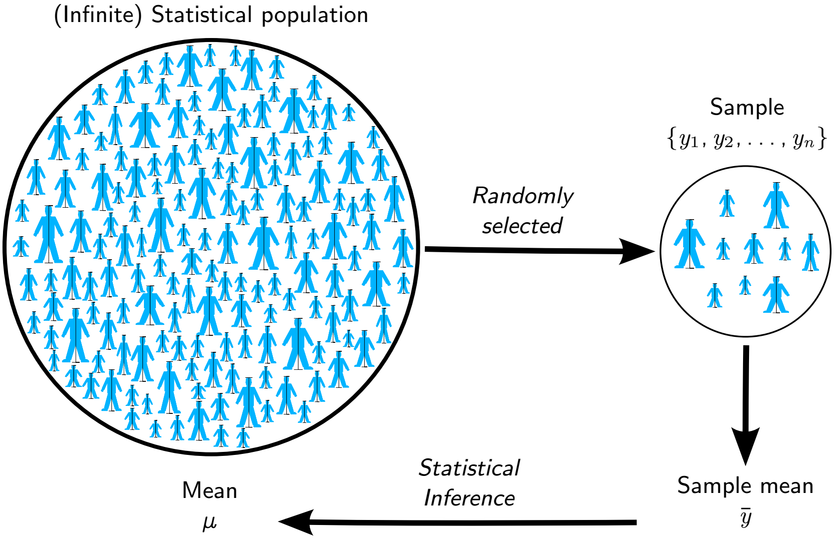
February 24, 2025

Outline

Overview

1. Estimation in the General Linear Model (GLM):
   - Ordinary Least Squares (OLS)
   - Maximum Likelihood (ML) estimation
2. Recursive Least Squares (RLS)
3. Global trend models
4. Essential principles needed to find a good model:
   - Residual analysis and model validation
   - Model selection

Population and sample



Ordinary least squares example

The General Linear Model (GLM), see the ModellingReference.pdf

```r
# see the script example_OLS_RLS.R
```

Parameter estimation with example

Simplest example: a constant model for the mean

- Model

$$Y_i = \theta_1 + \varepsilon_i \quad \text{, where } \varepsilon_i \sim N(0, \sigma^2) \text{ and i.i.d.}$$

- i.i.d.: identically and independent distributed
- The parameters are: the mean μ and the standard deviation σ
- We take a sample of $n = 100$ observations

$$(y_1, y_2, \ldots, y_{100})$$

Parameter estimation with example

Simplest example: a constant model for the mean

- Model

$$Y_i = \mu + \varepsilon_i \quad \text{, where } \varepsilon_i \sim N(0, \sigma^2) \text{ and i.i.d.}$$

- i.i.d.: identically and independent distributed
- The parameters are: the mean μ and the standard deviation σ
- We take a sample of $n = 100$ observations

$$(y_1, y_2, \ldots, y_{100})$$

Likelihood

The likelihood is defined by the joint probability function of the data

$$L(\mu, \sigma) \equiv p(y_1, y_2, \ldots, y_{100}|\mu, \sigma)$$

Hence, it's a function of the two parameters (the sample is observed, so it is not varying).

Due to independence

$$= \prod_{i=1}^{100} p(y_i|\mu, \sigma)$$

In our model the error $\varepsilon_i = Y_i - \mu$ is normal distributed (Gaussian), so

$$p(y_i|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y_i - \mu)^2}{2\sigma^2}\right) \tag{1}$$

Maximum likelihood estimation

Parameter estimation

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \Big(-\ln\big(L(\theta)\big)\Big)$$

where $\theta = (\mu, \sigma)$

Maximum likelihood estimation

Parameter estimation

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \Big(-\ln\big(L(\theta)\big)\Big)$$

where $\theta = (\mu, \sigma)$

```
# see the example_likelihood.R script
```

Likelihood for time correlated data

Given a sequence of measurements $\mathcal{Y}_N$

$$
\begin{aligned}
L(\theta) = p(\mathcal{Y}_N|\theta) &= p(y_N, y_{N-1}, \ldots, y_0|\theta) \\
&= \left(\prod_{k=1}^{N} p(y_k|\mathcal{Y}_{k-1}, \theta) \right) p(y_0|\theta)
\end{aligned}
$$

Parameter estimation

$$
\hat{\theta} = \arg\min_{\theta \in \Theta} \left(-\ln(L(\theta)) \right)
$$

Likelihood for time correlated data

If Gaussian

$$
\begin{aligned}
\hat{y}_{k|k-1} &= E[y_k|\mathcal{Y}_{k-1}, \theta] \\
R_{k|k-1} &= V[y_k|\mathcal{Y}_{k-1}, \theta] \\
\varepsilon_k &= y_k - \hat{y}_{k|k-1}
\end{aligned}
$$

then the likelihood is

$$
L(\theta) = \left(\prod_{k=1}^{N} \frac{\exp(-\frac{1}{2}\varepsilon_k^T R_{k|k-1}^{-1}\varepsilon_k)}{\sqrt{|R_{k|k-1}|}\sqrt{2\pi}^{l}} \right)
$$

Maximised using quasi Newton (in practise always minimize the negative log-likelihood)

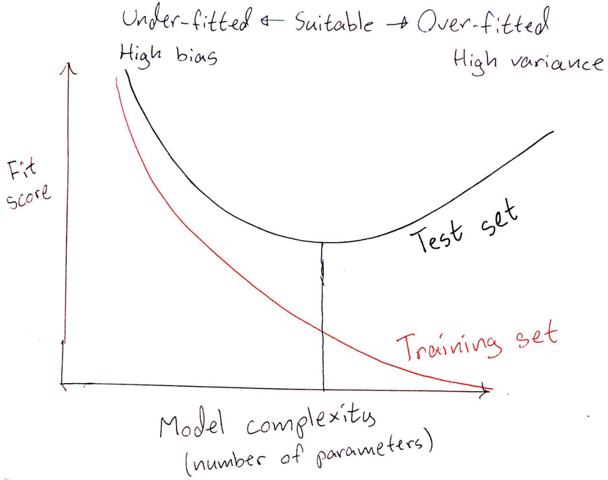
The global trend model, see the ModelExamples.pdf
https://gml.noaa.gov/ccgg/trends/data.html

Model selection and the bias-variance tradeoff

We want to find a *suitable model* – the model which is neither too simple (under-fitted) nor too complex (over-fitted):

- Divide the data into a training set and a test set
- Define a fit score (smaller is better e.g. summed squared error)

How do we find the balance?



Model selection

How do we find the balance?

Model selection

How do we find the balance?

- Information criteria (AIC, BIC)
- Goodness of fit test (likelihood-ratio test, F-test)
- Cross-validation technique (n-fold CV)

Model selection

How do we find the balance?

- Information criteria (AIC, BIC)
- Goodness of fit test (likelihood-ratio test, F-test)
- Cross-validation technique (n-fold CV)

Model selection procedure

- Forward selection (start with a small model and extend)
- Backward selection (start with a large model and remove)

Model validation (use also while finding a model)

After fitting the model then analyse the residuals

If no patterns are left, hence we have white noise residual, then we are done!

Residuals analysis

Plot, plot, plot!

- Time series plots of residuals aligned with input series
- Scatter plots of residuals vs. inputs
- ACF and CCF (We will get back to them!)

Model validation (use also while finding a model)

After fitting the model then analyse the residuals

If no patterns are left, hence we have white noise residual, then we are done!

Residuals analysis

Plot, plot, plot!

- Time series plots of residuals aligned with input series
- Scatter plots of residuals vs. inputs
- ACF and CCF (We will get back to them!)

Forward selection

Fit a simple model, analyse the residuals: Can you see some patterns left related to some inputs? Improve the model and repeat...

- Good approach for modelling with new data
- Good approach for articles (you get a story)

Training and Test Set Techniques for Time Series

- When data is time dependent, we can't just select some random points as test set and for cross-validation
- We have to take time into account
- **Holdout Method:**
  - Split data into training and test sets
  - **Illustration:**

  - Problems often occur: if the underlying process changed from training to test set
- **Cross-Validation:**
  - K-fold, Leave-One-Out
  - **Illustration:**

  - This is often a good compromise (a burn-in period can be needed).

Rolling Horizon Test Sets

- **Concept:** Use a moving window to create multiple training and test sets
- **Illustration:**
- **Steps:**
  1. Start with an initial training set
  2. Train the model
  3. Test on the next time point
  4. Expand the training set to include the test point
  5. Repeat the process
- A burn-in period is needed
- Recursive estimation is really nice, it always does this!