02417: Time Series Analysis

Week 1 - Introduction and overview

Peder Bacher
DTU Compute

Based on material previous material from the course

February 5, 2026

Material in the course

- The course webpage `02417.compute.dtu.dk`
- Learn for messages and projects
- Book

Material in the course

- The course webpage `02417.compute.dtu.dk`
- Learn for messages and projects
- Book
- Slides
- Exercises
- Assignments

What to use Time Series Models for

Applications:

- Prediction

What to use Time Series Models for

Applications:

- Prediction
- Estimation and hypothesis testing

What to use Time Series Models for

Applications:

- Prediction
- Estimation and hypothesis testing
- Control and decision making

What to use Time Series Models for

Applications:

- Prediction
- Estimation and hypothesis testing
- Control and decision making

We want a good model!

- Use data to fit a model

What to use Time Series Models for

Applications:

- Prediction
- Estimation and hypothesis testing
- Control and decision making

We want a good model!

- Use data to fit a model
- Basically, any modelling technique can be used, there are no rules!

What to use Time Series Models for

Applications:

- Prediction
- Estimation and hypothesis testing
- Control and decision making

We want a good model!

- Use data to fit a model
- Basically, any modelling technique can be used, there are no rules!
- Pros and cons: robustness, complexity, computation time, man hours to set up, ...

What to use Time Series Models for

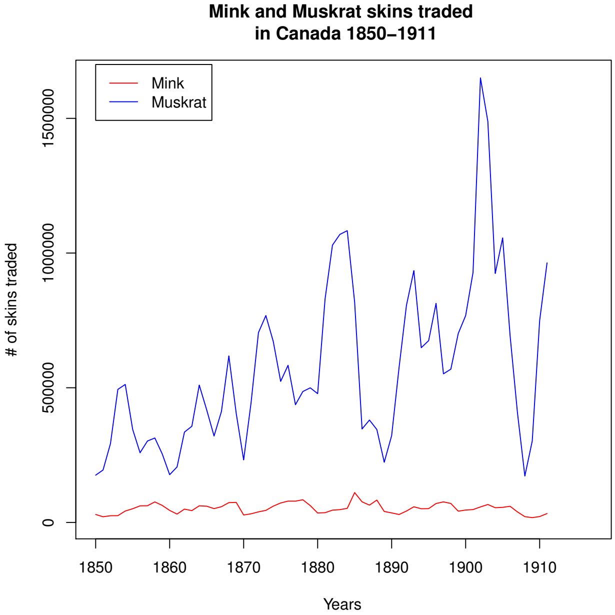
Applications:

- Prediction
- Estimation and hypothesis testing
- Control and decision making

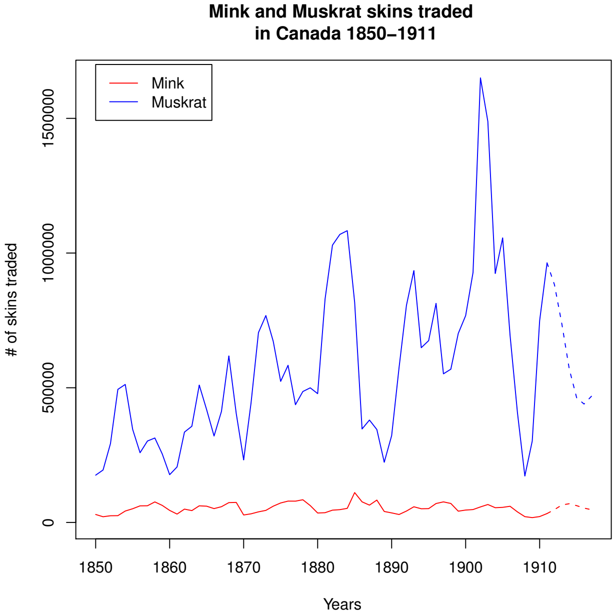
We want a good model!

- Use data to fit a model
- Basically, any modelling technique can be used, there are no rules!
- Pros and cons: robustness, complexity, computation time, man hours to set up, ...
- WE ONLY DO LINEAR MODELS in this course (multiply and add using matrices)! Very fast and reliable, always a good starting points when developing models, can be tweaked later to include non-linear effects...

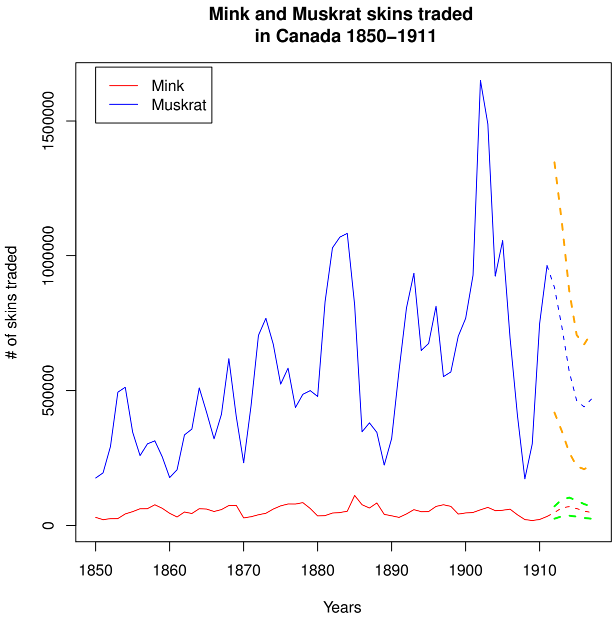
What you should be able to do



What you should be able to do



What you should be able to do



Introductory example – shares (COLO B 1 month)

What do think about the trend here? what would you buy or sell?

Introductory example – shares (COLO B 1 year)

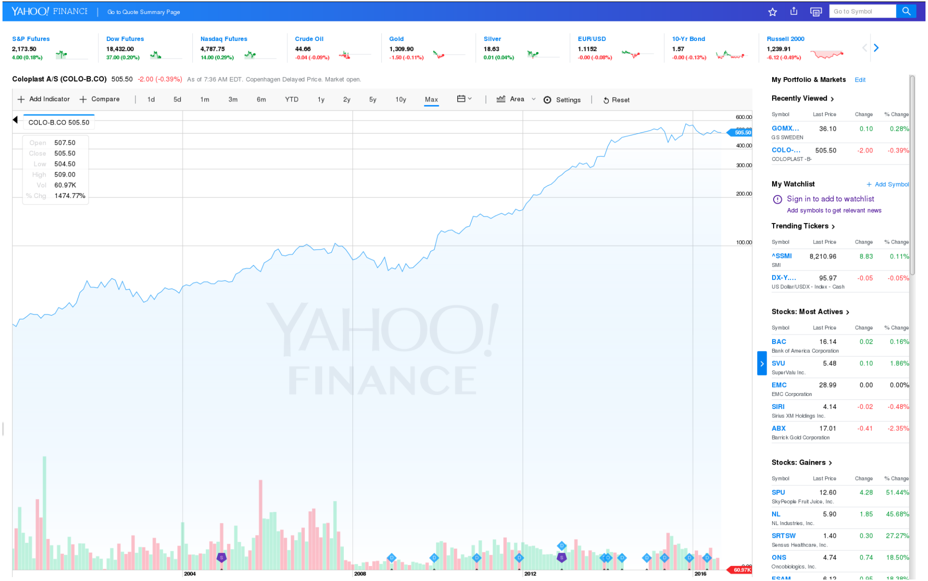
Would you do the same?

Introductory example – shares (COLO B all)

Symbol	Last Price	Change	% Change
GOMX... (G S SWEDEN)	36.10	0.10	0.28%
COLO-... (COLOPLAST -B-)	505.50	-2.00	-0.39%
^SSMI (SMI)	8,210.96	8.83	0.11%
DX-Y.... (US Dollar/USDX - Index - Cash)	95.97	-0.05	-0.05%
BAC (Bank of America Corporation)	16.14	0.02	0.16%
SVU (SuperValu Inc.)	5.48	0.10	1.86%
EMC (EMC Corporation)	28.99	0.00	0.00%
SIRI (Sirius XM Holdings Inc.)	4.14	-0.02	-0.48%
ABX (Barrick Gold Corporation)	17.01	-0.41	-2.35%
SPU (SkyPeople Fruit Juice, Inc.)	12.60	4.28	51.44%
NL (NL Industries, Inc.)	5.90	1.85	45.68%
SRTSW (Sensus Healthcare, Inc.)	1.40	0.30	27.27%
ONS (Oncobiologics, Inc.)	4.74	0.74	18.50%

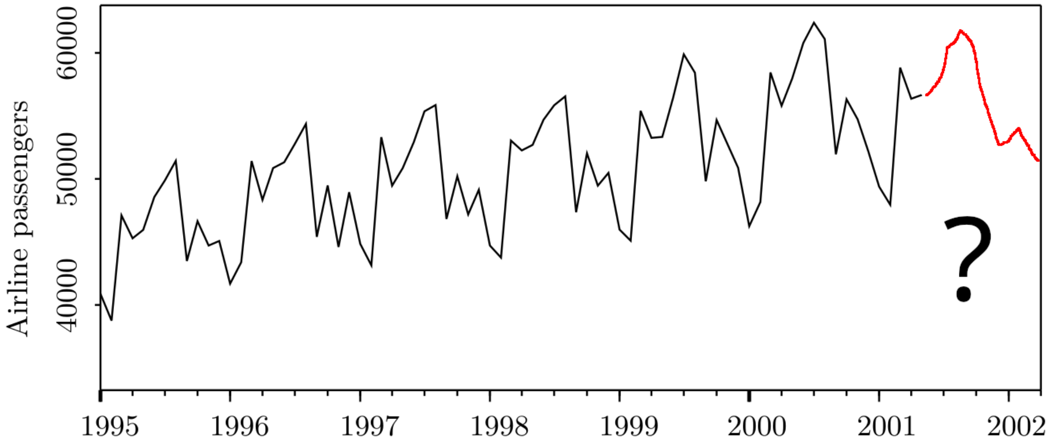
Can we use a linear trend model here?

Introductory example – shares (COLO B log(all))



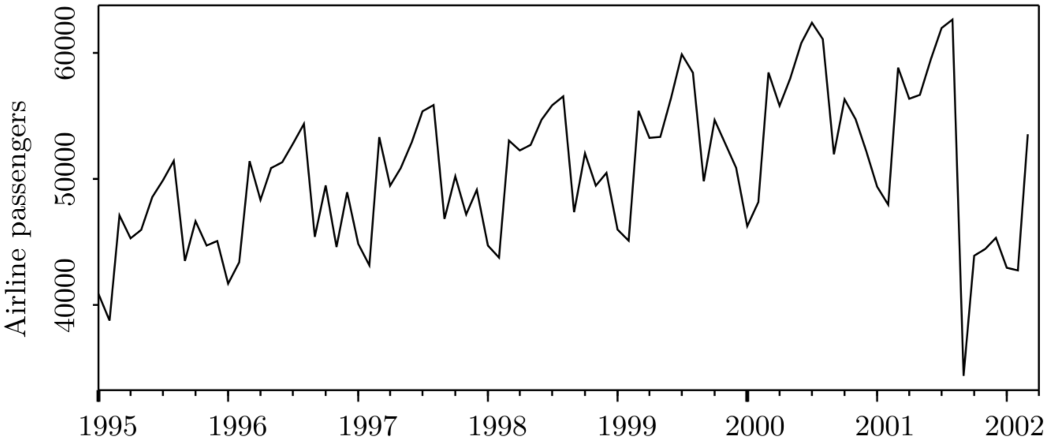
Take log(y): Often we can do non-linear transformations, resolve in cos and sine or splines,...

Number of Monthly Airline Passengers in the US

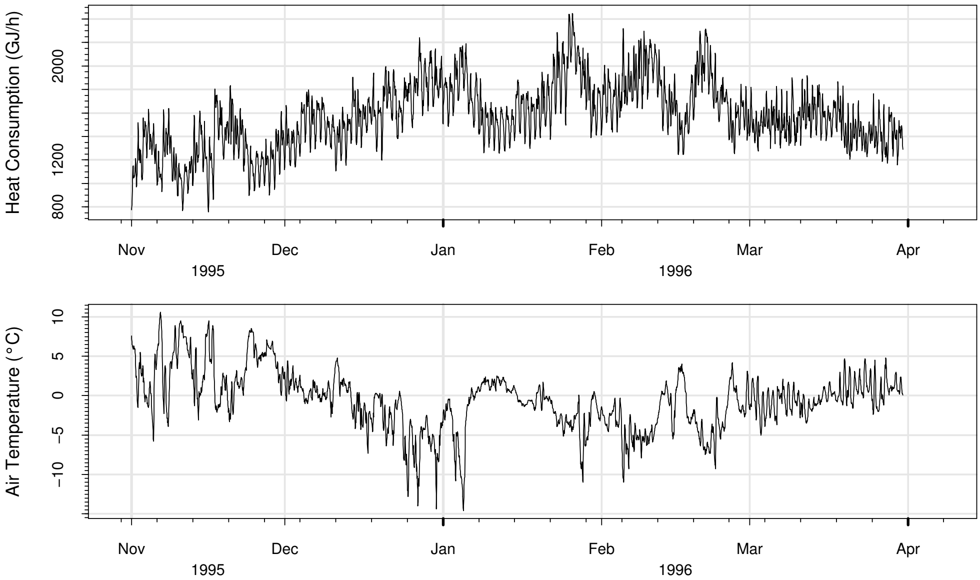


Is this a good prediction?

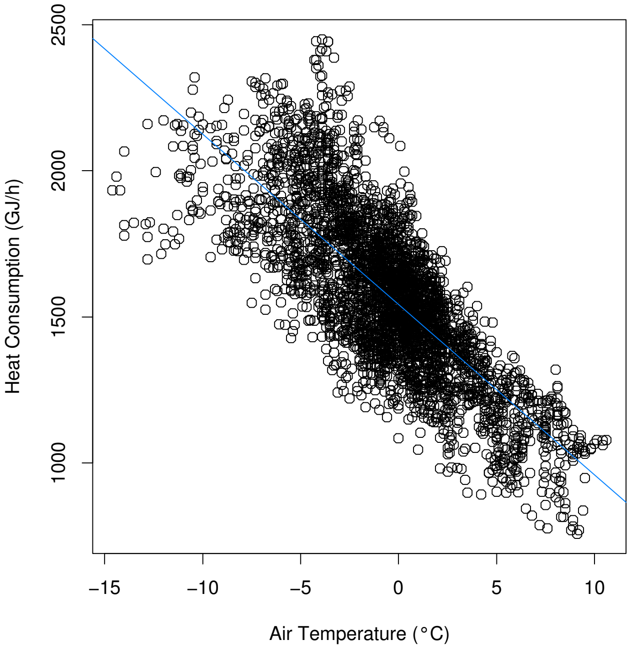
Number of Monthly Airline Passengers in the US



Consumption of District Heating (VEKS) – data

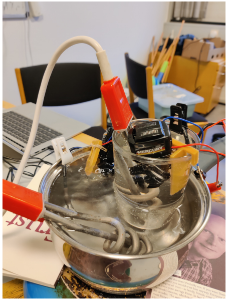


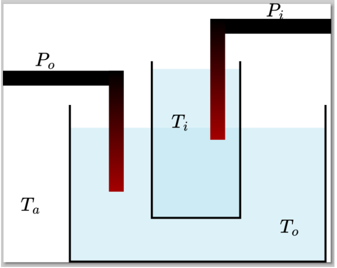
Consumption of DH – simple model



Discussion: What is a dynamical system?

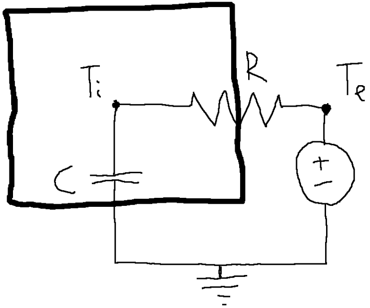
Last year!





Simplest first order RC-system

Single state model of the temperature in a box:



Discretize the ODE

$$\frac{dT_{\mathrm{i}}}{dt} = \frac{1}{RC}(T_{\mathrm{e}} - T_{\mathrm{i}})$$

It has the solution

$$T_{\mathrm{i}}(t + \Delta t) = T_{\mathrm{e}}(t) + e^{-\frac{\Delta t}{RC}}\big(T_{\mathrm{i}}(t) - T_{\mathrm{e}}(t)\big)$$

Discretize the ODE

$$\frac{dT_{\mathrm{i}}}{dt} = \frac{1}{RC}(T_{\mathrm{e}} - T_{\mathrm{i}})$$

It has the solution

$$T_{\mathrm{i}}(t + \Delta t) = T_{\mathrm{e}}(t) + e^{-\frac{\Delta t}{RC}}\big(T_{\mathrm{i}}(t) - T_{\mathrm{e}}(t)\big)$$

if $\Delta t = 1$ and T_{e} is constant between the sample points then

$$T^{\mathrm{i}}_{t+1} = e^{-\frac{1}{RC}}\, T^{\mathrm{i}}_{t} + (1 - e^{-\frac{1}{RC}})\, T^{\mathrm{e}}_{t}$$

Discretize the ODE

$$\frac{dT_{\mathrm{i}}}{dt} = \frac{1}{RC}(T_{\mathrm{e}} - T_{\mathrm{i}})$$

It has the solution

$$T_{\mathrm{i}}(t + \Delta t) = T_{\mathrm{e}}(t) + e^{-\frac{\Delta t}{RC}}\big(T_{\mathrm{i}}(t) - T_{\mathrm{e}}(t)\big)$$

if $\Delta t = 1$ and T_{e} is constant between the sample points then

$$T^{\mathrm{i}}_{t+1} = e^{-\frac{1}{RC}}\, T^{\mathrm{i}}_{t} + (1 - e^{-\frac{1}{RC}})\, T^{\mathrm{e}}_{t}$$

since $e^{-\frac{1}{RC}}$ is between 0 and 1, then write it as

$$T^{\mathrm{i}}_{t+1} = \phi_1\, T^{\mathrm{i}}_{t} + \omega_1\, T^{\mathrm{e}}_{t}$$

where ϕ_1 and ω_1 are between 0 and 1.

Discretize the ODE

$$\frac{dT_{\mathrm{i}}}{dt} = \frac{1}{RC}(T_{\mathrm{e}} - T_{\mathrm{i}})$$

It has the solution

$$T_{\mathrm{i}}(t+\Delta t) = T_{\mathrm{e}}(t) + e^{-\frac{\Delta t}{RC}}\big(T_{\mathrm{i}}(t) - T_{\mathrm{e}}(t)\big)$$

if $\Delta t = 1$ and T_{e} is constant between the sample points then

$$T^{\mathrm{i}}_{t+1} = e^{-\frac{1}{RC}}\,T^{\mathrm{i}}_t + (1 - e^{-\frac{1}{RC}})\,T^{\mathrm{e}}_t$$

since $e^{-\frac{1}{RC}}$ is between 0 and 1, then write it as

$$T^{\mathrm{i}}_{t+1} = \phi_1\,T^{\mathrm{i}}_t + \omega_1\,T^{\mathrm{e}}_t$$

where ϕ_1 and ω_1 are between 0 and 1.

Add a noise term and we have the ARX model

$$T^{\mathrm{i}}_{t+1} = \phi_1\,T^{\mathrm{i}}_t \quad + \omega_1\,T^{\mathrm{e}}_t \quad + \varepsilon_{t+1}$$

Discretize the ODE

$$\frac{dT_{\mathrm{i}}}{dt} = \frac{1}{RC}(T_{\mathrm{e}} - T_{\mathrm{i}})$$

It has the solution

$$T_{\mathrm{i}}(t+\Delta t) = T_{\mathrm{e}}(t) + e^{-\frac{\Delta t}{RC}}\big(T_{\mathrm{i}}(t) - T_{\mathrm{e}}(t)\big)$$

if $\Delta t = 1$ and T_{e} is constant between the sample points then

$$T^{\mathrm{i}}_{t+1} = e^{-\frac{1}{RC}}\, T^{\mathrm{i}}_t + (1 - e^{-\frac{1}{RC}})\, T^{\mathrm{e}}_t$$

since $e^{-\frac{1}{RC}}$ is between 0 and 1, then write it as

$$T^{\mathrm{i}}_{t+1} = \phi_1\, T^{\mathrm{i}}_t + \omega_1\, T^{\mathrm{e}}_t$$

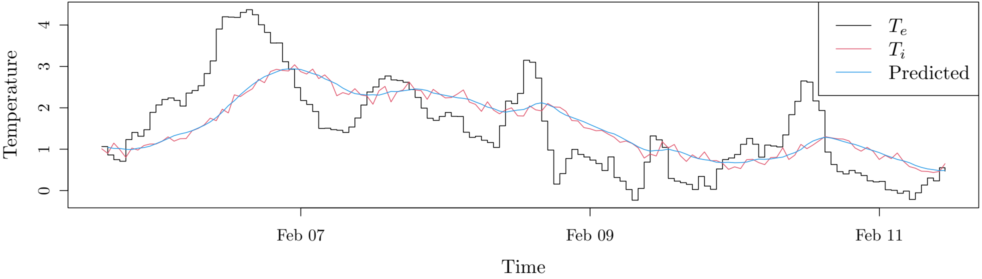
where ϕ_1 and ω_1 are between 0 and 1.

Add a noise term and we have the ARX model

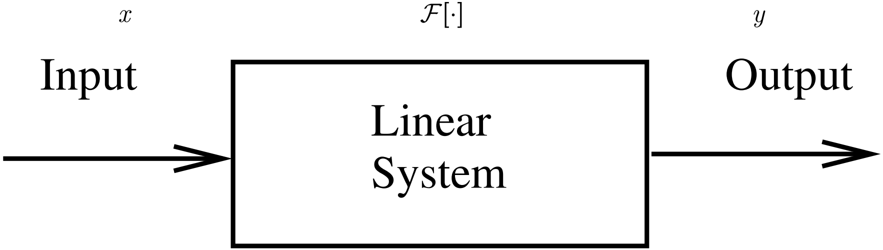
$$T^{\mathrm{i}}_t = \phi_1\, T^{\mathrm{i}}_{t-1} + \omega_1\, T^{\mathrm{e}}_{t-1} + \varepsilon_t$$

An ARMAX model

$$T_t^{\mathrm{i}} = \phi_1 T_{t-1}^{\mathrm{i}} + \omega_1 T_t^{\mathrm{e}} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

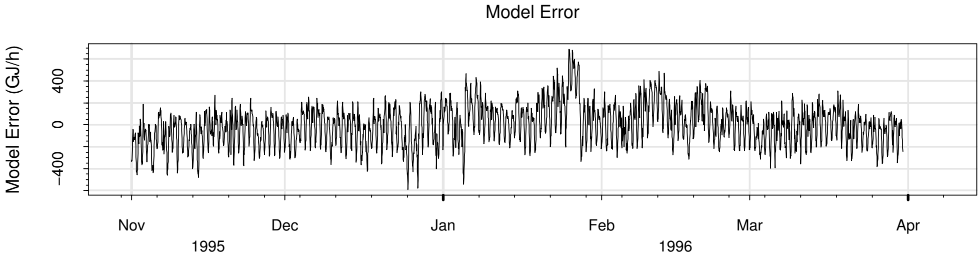


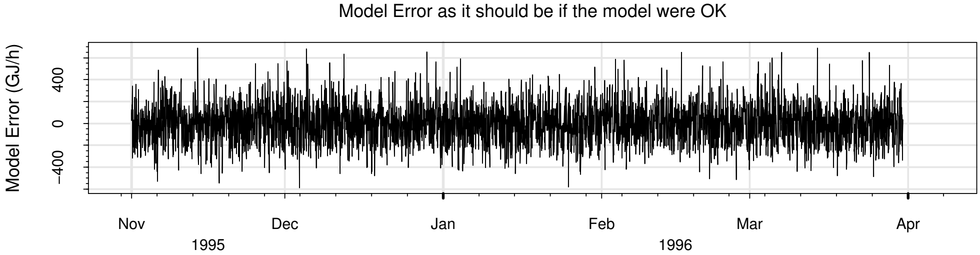
Linear Dynamic Systems – notation



$x(t)$	Differential eq., $h(u)$	$y(t)$
x_t	Difference eq., h_k, $h(B)$	y_t
$X(\omega)$	$\mathcal{H}(\omega)$	$Y(\omega)$
$X(z)$	$H(z)$	$Y(z)$

Consumption of DH – We use the model error to validate the model





Multivariate random variables

- What is a multivariate random variable?

Multivariate random variables

- What is a multivariate random variable?
- Why are multivariate random variables essential in time series analysis?

Multivariate random variables

- Definition (n-dimensional random variable; random vector)

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

- Joint distribution function:

$$F(x_1, \ldots, x_n) = \mathrm{P}\{X_1 \leq x_1, \ldots, X_n \leq x_n\}$$

Multivariate random variables

- Definition (n-dimensional random variable; random vector)

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

- Joint distribution function:

$$F(x_1, \ldots, x_n) = \mathrm{P}\{X_1 \leq x_1, \ldots, X_n \leq x_n\}$$

- When does this simplify to

$$F(x_1, \ldots, x_n) = \mathrm{P}\{X_1 \leq x_1\}\mathrm{P}\{X_2 \leq x_2\} \ldots \mathrm{P}\{X_n \leq x_n\}?$$

Multivariate random variables - joint densities

- Joint distribution function (repeated from last slide):

$$F(x_1, \dots, x_n) = \mathrm{P}\{X_1 \le x_1, \dots, X_n \le x_n\}$$

- Joint density function - continuous case:

$$f(x_1, \dots, x_n) = \frac{\partial^n F(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n}$$

Multivariate random variables - joint densities

- Joint distribution function (repeated from last slide):

$$F(x_1, \ldots, x_n) = \mathrm{P}\{X_1 \le x_1, \ldots, X_n \le x_n\}$$

- Joint density function - continuous case:

$$f(x_1, \ldots, x_n) = \frac{\partial^n F(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n}$$

  - and back to the joint distribution function:

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f(t_1, \ldots, t_n)\, dt_1 \ldots dt_n$$

Multivariate random variables - joint densities

- Joint distribution function (repeated from last slide):

$$F(x_1, \ldots, x_n) = \mathrm{P}\{X_1 \le x_1, \ldots, X_n \le x_n\}$$

- Joint density function - continuous case:

$$f(x_1, \ldots, x_n) = \frac{\partial^n F(x_1, \ldots, x_n)}{\partial x_1 \ldots \partial x_n}$$

  - and back to the joint distribution function:

$$F(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f(t_1, \ldots, t_n)\, dt_1 \ldots dt_n$$

- Joint density function - discrete case:

$$f(x_1, \ldots, x_n) = \mathrm{P}\{X_1 = x_1, X_2 = x_2, \ldots X_n = x_n\}$$

Our Favourite Multivariate Distribution

The Multivariate Normal Distribution

The Multivariate Normal Distribution

- The joint p.d.f.

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{\Sigma}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

The Multivariate Normal Distribution

- The joint p.d.f.

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{\Sigma}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

- $\boldsymbol{\Sigma}$ is symmetric and positive semi-definite
- Notation: $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- Standard multivariate normal: $\boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{I})$

The Multivariate Normal Distribution

- The joint p.d.f.

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{\Sigma}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

- $\boldsymbol{\Sigma}$ is symmetric and positive semi-definite
- Notation: $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- Standard multivariate normal: $\boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{I})$
- If $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{Y} = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{X}$ then $\boldsymbol{Y} \sim \mathrm{N}(\boldsymbol{a} + \boldsymbol{B}\boldsymbol{\mu}, \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^T)$

The Multivariate Normal Distribution

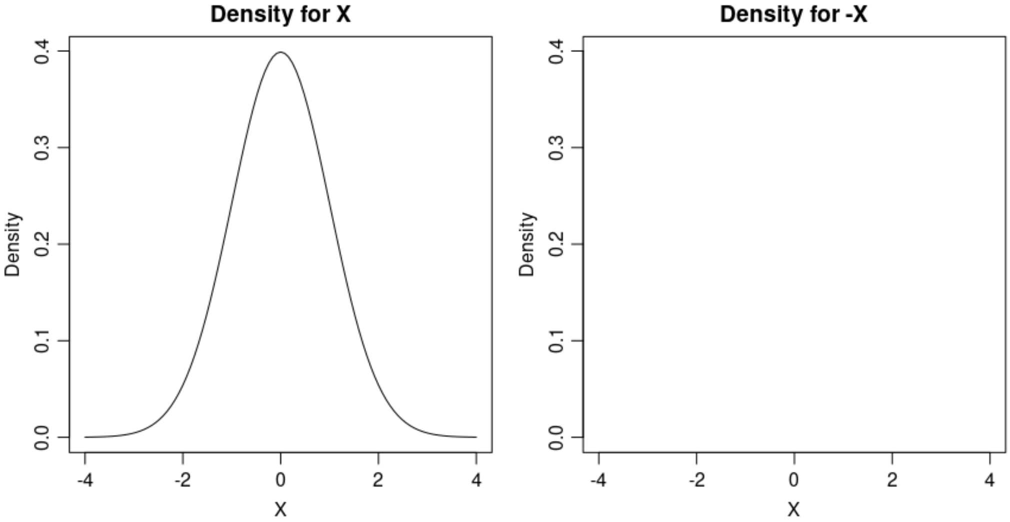
- The joint p.d.f.

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{\Sigma}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

- $\boldsymbol{\Sigma}$ is symmetric and positive semi-definite
- Notation: $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- Standard multivariate normal: $\boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{I})$
- If $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{Y} = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{X}$ then $\boldsymbol{Y} \sim \mathrm{N}(\boldsymbol{a} + \boldsymbol{B}\boldsymbol{\mu}, \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^T)$
- Notice that this implies that linear combinations of normal distributed variables are still normal distributed.

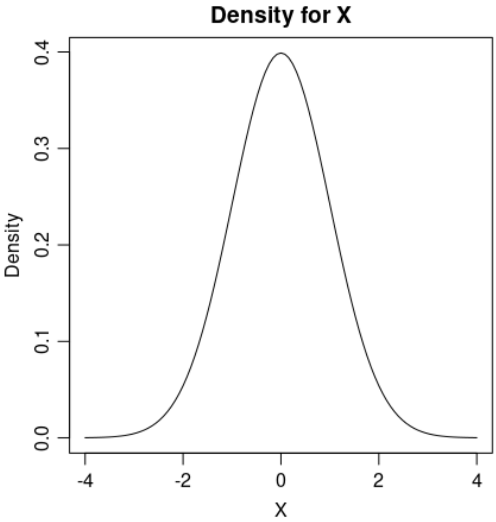
Stochastic variables and distributions

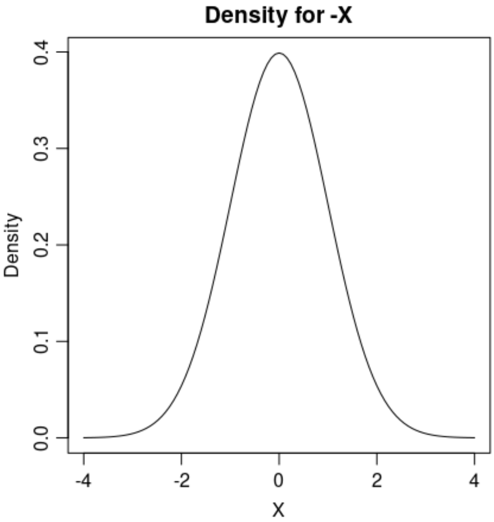
- If $X \sim N(0, 1)$, then what is the distribution of $-X$?



Stochastic variables and distributions

- If $X \sim N(0,1)$, then what is the distribution of $-X$?





- $X, -X$ are different variables but have the same distribution

Marginal density function

When we only consider part of the variables.

- Vector of random variables: $\mathbf{X} = (X_1, \ldots, X_n)^T$

Marginal density function

When we only consider part of the variables.

- Vector of random variables: $\mathbf{X} = (X_1, \ldots, X_n)^T$
- Sub-vector: $(X_1, \ldots, X_k)^T, \quad (k < n)$

Marginal density function

When we only consider part of the variables.

- Vector of random variables: $\mathbf{X} = (X_1, \ldots, X_n)^T$
- Sub-vector: $(X_1, \ldots, X_k)^T, \quad (k < n)$
- Marginal density function:

$$f_S(x_1, \ldots, x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\, dx_{k+1} \ldots dx_n$$

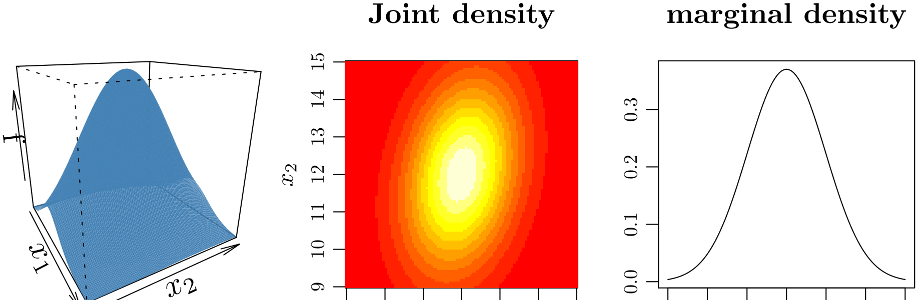
Marginal density function

When we only consider part of the variables.

- Vector of random variables: $\mathbf{X} = (X_1, \ldots, X_n)^T$
- Sub-vector: $(X_1, \ldots, X_k)^T, \quad (k < n)$
- Marginal density function:

$$f_S(x_1, \ldots, x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\, dx_{k+1} \ldots dx_n$$

Two dimensional example:



Conditional distributions

Conditional distributions

- The conditional density of X_1 given $X_2 = x_2$ is given by:

$$f_{X_1|X_2=x_2}(x_1) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)}$$

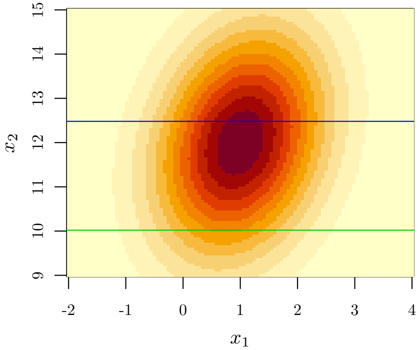
(joint density of (X_1, X_2) divided by the marginal density of X_2 evaluated at x_2)

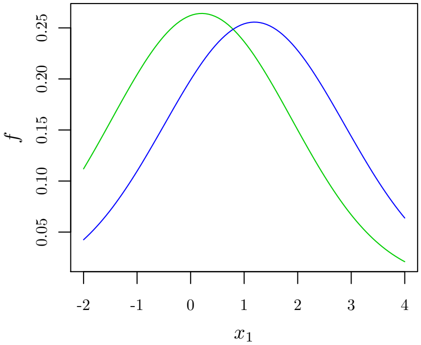
Conditional distributions

- The conditional density of X_1 given $X_2 = x_2$ is given by:

$$f_{X_1|X_2=x_2}(x_1) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)}$$

(joint density of (X_1, X_2) divided by the marginal density of X_2 evaluated at x_2)





Independence

- If knowledge of X does not give information about Y, we get that $f_{Y|X=x}(y) = f_Y(y)$

Independence

- If knowledge of X does not give information about Y, we get that $f_{Y|X=x}(y) = f_Y(y)$
- This leads to the following definition of independence:

$$X, Y \text{ stochastically independent} \overset{def}{\Leftrightarrow}$$

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

Expectation

- The average of unrealised random variables.

Expectation

- The average of unrealised random variables.
- Let X be a univariate random variable with density $f_X(x)$. The expectation of X is then defined as:

$$E[X] = \int_{-\infty}^{\infty} x\, f_X(x) dx \qquad \text{(continuous case)}$$

$$E[X] = \sum_{all\ \ x} x\, P(X = x) \qquad \text{(discrete case)}$$

Expectation

- The average of unrealised random variables.
- Let X be a univariate random variable with density $f_X(x)$. The expectation of X is then defined as:

$$E[X] = \int_{-\infty}^{\infty} x\, f_X(x) dx \qquad \text{(continuous case)}$$

$$E[X] = \sum_{all\ \ x} x\, P(X = x) \qquad \text{(discrete case)}$$

- Expectation is a linear operator:

$$E[a + bX_1 + cX_2] = a + b\, E[X_1] + c\, E[X_2]$$

Moments and Variance

- n'th moment:

$$E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)\, dx$$

Moments and Variance

- n'th moment:

$$E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)\, dx$$

- n'th central moment:

$$E[(X - E[X])^n] = \int_{-\infty}^{\infty} (x - E[X])^n f_X(x)\, dx$$

Moments and Variance

- n'th moment:

$$E[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)\,dx$$

- n'th central moment:

$$E[(X - E[X])^n] = \int_{-\infty}^{\infty} (x - E[X])^n f_X(x)\,dx$$

- The 2'nd central moment is called the variance:

$$V[X] = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

Covariance

- How random variables co-vary.

Covariance

- How random variables co-vary.
- Covariance:

$$\mathrm{Cov}[X_1, X_2] = E[(X_1 - E[X_1])(X_2 - E[X_2])] = E[X_1 X_2] - E[X_1]E[X_2]$$

Covariance

- How random variables co-vary.
- Covariance:

$$\text{Cov}[X_1, X_2] = E[(X_1 - E[X_1])(X_2 - E[X_2])] = E[X_1 X_2] - E[X_1]E[X_2]$$

- Variance and covariance:

  $V[X] = \text{Cov}[X, X]$

Covariance

- How random variables co-vary.
- Covariance:

$$\mathrm{Cov}[X_1, X_2] = E[(X_1 - E[X_1])(X_2 - E[X_2])] = E[X_1 X_2] - E[X_1]E[X_2]$$

- Variance and covariance:

  $V[X] = \mathrm{Cov}[X, X]$

- Calculation rules:

$$\mathrm{Cov}[aX_1 + bX_2, cX_3 + dX_4] =$$
$$ac\,\mathrm{Cov}[X_1, X_3] + ad\,\mathrm{Cov}[X_1, X_4] + bc\,\mathrm{Cov}[X_2, X_3] + bd\,\mathrm{Cov}[X_2, X_4]$$

Covariance

- How random variables co-vary.
- Covariance:

$$\text{Cov}[X_1, X_2] = E[(X_1 - E[X_1])(X_2 - E[X_2])] = E[X_1 X_2] - E[X_1]E[X_2]$$

- Variance and covariance:

  $V[X] = \text{Cov}[X, X]$

- Calculation rules:

$$\begin{aligned}&\text{Cov}[aX_1 + bX_2, cX_3 + dX_4] = \\ &ac\,\text{Cov}[X_1, X_3] + ad\,\text{Cov}[X_1, X_4] + bc\,\text{Cov}[X_2, X_3] + bd\,\text{Cov}[X_2, X_4]\end{aligned}$$

- The calculation rule can be used for the variance as well. For instance:

  $V[a + bX_2] = b^2 V[X_2]$
  $V[aX_1 + bX_2] = a^2 V[X_1] + b^2 V[X_2] + 2ab\text{Cov}[X_1, X_2]$

Moment representation

- All moments up to a given order.
- Second order moment representation:
  - Mean
  - Variance
  - Covariance (If relevant)

Moment representation

- All moments up to a given order.
- Second order moment representation:
    - Mean
    - Variance
    - Covariance (If relevant)

Have you seen something similar to this before?

Moment representation

- All moments up to a given order.
- Second order moment representation:
  - Mean
  - Variance
  - Covariance (If relevant)

Have you seen something similar to this before? Taylor approximations?

Expectation and Variance for Random Vectors

- Expectation: $E[\boldsymbol{X}] = [E[X_1], E[X_2], \ldots, E[X_n]]^T$
- Covariance matrix: $\boldsymbol{\Sigma}_{\boldsymbol{X}} = V[\boldsymbol{X}] = E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T] =$

Expectation and Variance for Random Vectors

- Expectation: $E[\boldsymbol{X}] = [E[X_1], E[X_2], \ldots, E[X_n]]^T$
- Covariance matrix: $\boldsymbol{\Sigma_X} = V[\boldsymbol{X}] = E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T] =$

$$\begin{bmatrix} V[X_1] & \mathrm{Cov}[X_1, X_2] & \cdots & \mathrm{Cov}[X_1, X_n] \\ \mathrm{Cov}[X_2, X_1] & V[X_2] & \cdots & \mathrm{Cov}[X_2, X_n] \\ \vdots & & & \vdots \\ \mathrm{Cov}[X_n, X_1] & \mathrm{Cov}[X_n, X_2] & \cdots & V[X_n] \end{bmatrix}$$

Expectation and Variance for Random Vectors

- Expectation: $E[\boldsymbol{X}] = [E[X_1], E[X_2], \ldots, E[X_n]]^T$
- Covariance matrix: $\boldsymbol{\Sigma}_{\boldsymbol{X}} = V[\boldsymbol{X}] = E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T] =$

$$\begin{bmatrix} V[X_1] & \mathrm{Cov}[X_1, X_2] & \cdots & \mathrm{Cov}[X_1, X_n] \\ \mathrm{Cov}[X_2, X_1] & V[X_2] & \cdots & \mathrm{Cov}[X_2, X_n] \\ \vdots & & & \vdots \\ \mathrm{Cov}[X_n, X_1] & \mathrm{Cov}[X_n, X_2] & \cdots & V[X_n] \end{bmatrix}$$

- How do you think that typical covariance matrices of time series differ from covariance matrices of other multivariate distributions?

Conditional expectation

$$E[Y|X = x] = E[Y|X] = \int_{-\infty}^{\infty} y\, f_{Y|X=x}(y)\, dy$$

Conditional expectation

$$E[Y|X = x] = E[Y|X] = \int_{-\infty}^{\infty} y \, f_{Y|X=x}(y) \, dy$$

1. $E[Y|X] = E[Y]$ if X and Y are independent

Conditional expectation

$$E[Y|X = x] = E[Y|X] = \int_{-\infty}^{\infty} y\, f_{Y|X=x}(y)\, dy$$

1. $E[Y|X] = E[Y]$ if X and Y are independent
2. $E[Y] = E\left[E[Y|X]\right]$

Conditional expectation

$$E[Y|X=x] = E[Y|X] = \int_{-\infty}^{\infty} y\, f_{Y|X=x}(y)\, dy$$

1. $E[Y|X] = E[Y]$ if X and Y are independent
2. $E[Y] = E\left[E[Y|X]\right]$
3. $E[g(X)Y|X] = g(X)E[Y|X]$ - $E[g(X)Y|X=x] = g(x)E[Y|X=x]$

Conditional expectation

$$E[Y|X=x] = E[Y|X] = \int_{-\infty}^{\infty} y \, f_{Y|X=x}(y) \, dy$$

1. $E[Y|X] = E[Y]$ if X and Y are independent
2. $E[Y] = E\left[E[Y|X]\right]$
3. $E[g(X)Y|X] = g(X)E[Y|X]$ - $E[g(X)Y|X=x] = g(x)E[Y|X=x]$
4. $E[g(X)Y] = E\left[g(X)E[Y|X]\right]$

Conditional expectation

$$E[Y|X=x] = E[Y|X] = \int_{-\infty}^{\infty} y \, f_{Y|X=x}(y) \, dy$$

1. $E[Y|X] = E[Y]$ if X and Y are independent
2. $E[Y] = E\left[E[Y|X]\right]$
3. $E[g(X)Y|X] = g(X)E[Y|X]$ - $E[g(X)Y|X=x] = g(x)E[Y|X=x]$
4. $E[g(X)Y] = E\left[g(X)E[Y|X]\right]$
5. $E[a|X] = a$

Conditional expectation

$$E[Y|X=x] = E[Y|X] = \int_{-\infty}^{\infty} y\, f_{Y|X=x}(y)\, dy$$

1. $E[Y|X] = E[Y]$ if X and Y are independent
2. $E[Y] = E\left[E[Y|X]\right]$
3. $E[g(X)Y|X] = g(X)E[Y|X]$ - $E[g(X)Y|X=x] = g(x)E[Y|X=x]$
4. $E[g(X)Y] = E\left[g(X)E[Y|X]\right]$
5. $E[a|X] = a$
6. $E[g(X)|X] = g(X)$ - $E[g(X)|X=x] = g(x)$

Conditional expectation

$$E[Y|X = x] = E[Y|X] = \int_{-\infty}^{\infty} y\, f_{Y|X=x}(y)\, dy$$

1. $E[Y|X] = E[Y]$ if X and Y are independent
2. $E[Y] = E\left[E[Y|X]\right]$
3. $E[g(X)Y|X] = g(X)E[Y|X]$ - $E[g(X)Y|X = x] = g(x)E[Y|X = x]$
4. $E[g(X)Y] = E\left[g(X)E[Y|X]\right]$
5. $E[a|X] = a$
6. $E[g(X)|X] = g(X)$ - $E[g(X)|X = x] = g(x)$
7. $E[cX + dZ|Y] = cE[X|Y] + dE[Z|Y]$

A brief outline of the course

- General aspects of multivariate random variables
- Prediction using the general linear model
- Time series models
- Some theory on linear systems
- Time series models with external input

Some goals:

- Characterization of time series / signals; correlation functions, covariance functions, stationarity, linearity, . . .
- Signal processing; filtering and smoothing
- Modelling; with or without external input
- Prediction with uncertainty