Time Series Analysis

Lasse Engbo Christiansen

DTU Applied Mathematics and Computer Science
Technical University of Denmark

September 8, 2017

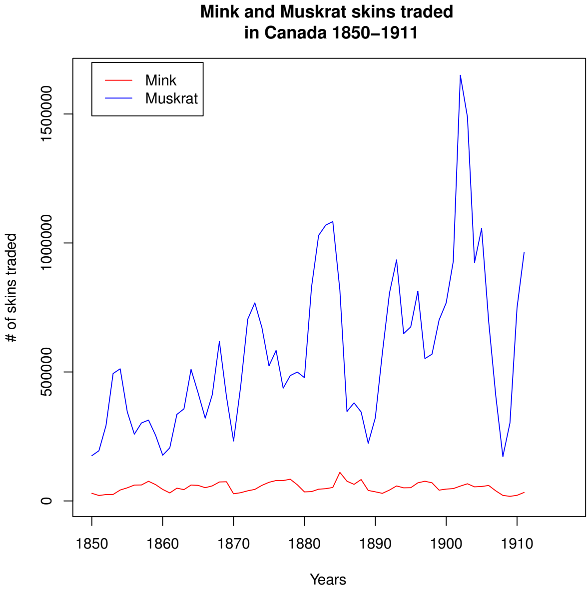
Outline of the lecture

- Practical information
- Introductory examples (see also Chapter 1)
- A brief outline of the course
- Chapter 2:
  - Multivariate random variables
  - The multivariate normal distribution
  - Linear projections
- Example

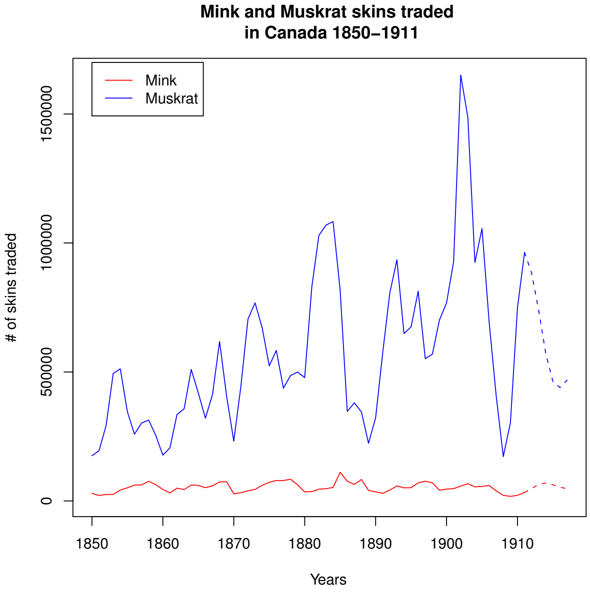
Contact info

- Lasse Engbo Christiansen
  Technical University of Denmark
  Applied Mathematics and Computer Science
  Section for Dynamical Systems
  From October 1st: Building 303B, room 010
  Email `laec@dtu.dk`
- Teaching assistants: Jesper, Sebastian & me.
  For consultation besides that – please drop me an email

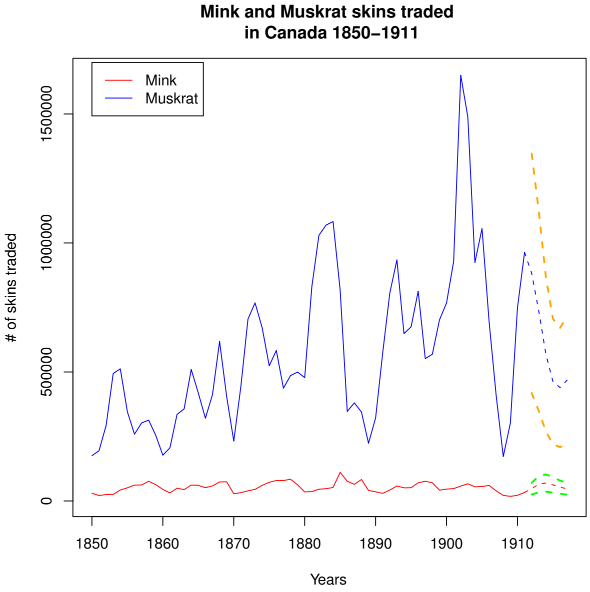
What you should be able to do



What you should be able to do



What you should be able to do



Introductory example – shares (COLO B 1 month)

From `finance.yahoo.com`

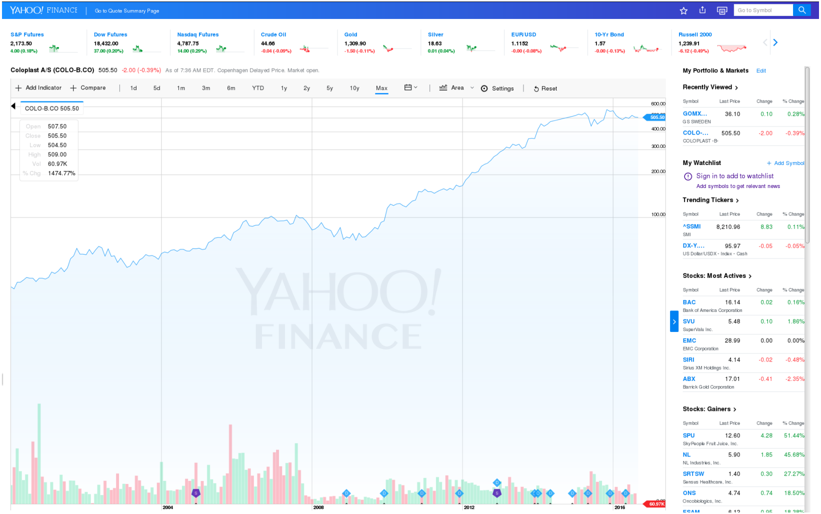
Introductory example – shares (COLO B 1 year)

From finance.yahoo.com

Introductory example – shares (COLO B all)

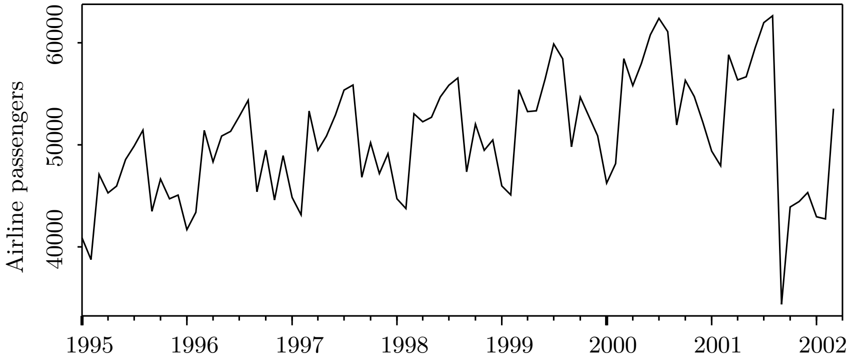
From `finance.yahoo.com`

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

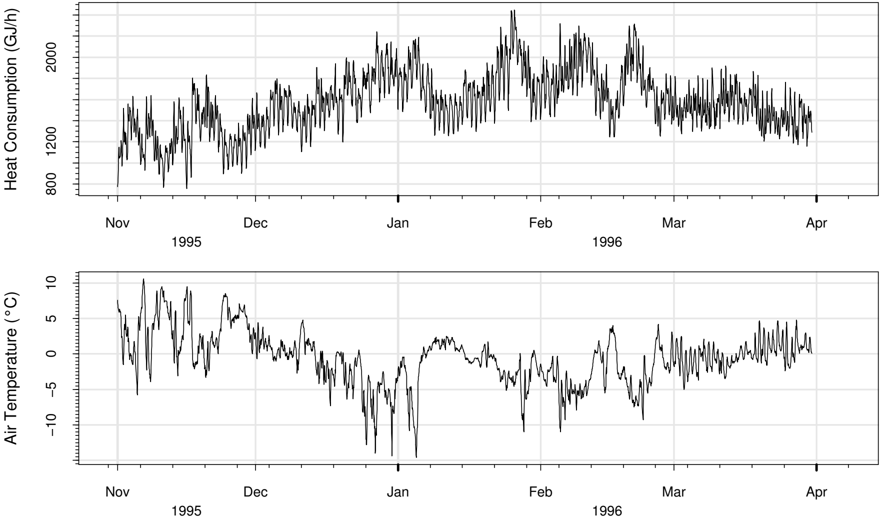


From `finance.yahoo.com`

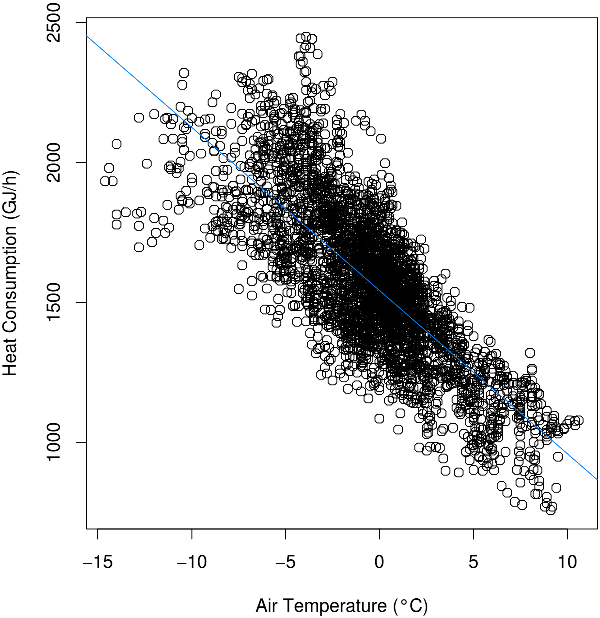
Number of Monthly Airline Passengers in the US



Consumption of District Heating (VEKS) – data

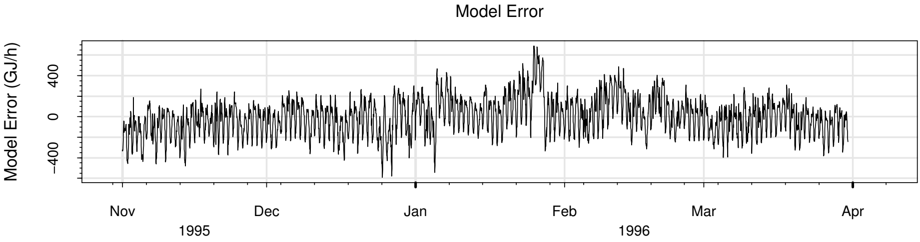


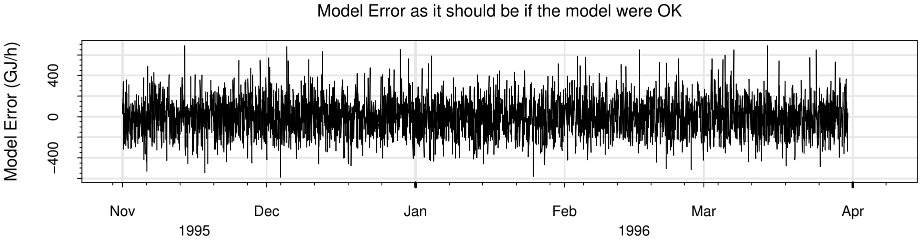
Consumption of DH – simple model



Discussion: What is a dynamical system?

Consumption of DH – model error





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

Today: Multivariate random variables

- Distribution functions
- Density functions
- The multivariate normal distribution
- Marginal densities
- Conditional distributions and independence
- Expectations and moments
- Moments of multivariate random variables
- Conditional expectation
- Distributions derived from the normal distribution
- Linear projections and relations to conditional means

Multivariate random variables – distr. functions

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

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

- Joint distribution function:

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

- Notice notation (lowercase, capital letters, bold font)

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:

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

  - and back to the joint distribution function:

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

- Joint density function - discrete case:

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

The Multivariate Normal Distribution

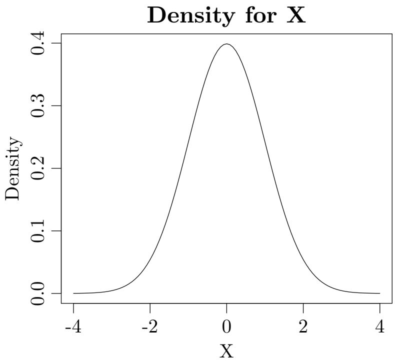
- The joint p.d.f.

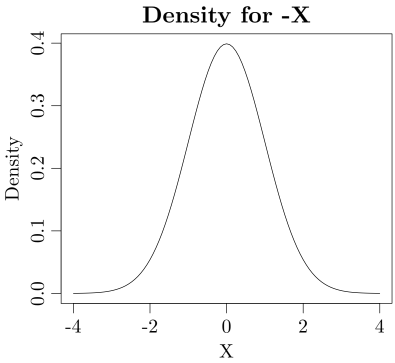
$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{\Sigma}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

- $\boldsymbol{\Sigma}$ 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 $\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{T}\boldsymbol{Z}$, where $\boldsymbol{\Sigma} = \boldsymbol{T}\boldsymbol{T}^T$, then $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- If $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\boldsymbol{Y} = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{X}$ then $\boldsymbol{Y} \sim \mathrm{N}(\boldsymbol{a} + \boldsymbol{B}\boldsymbol{\mu}, \boldsymbol{B}\boldsymbol{\Sigma}\boldsymbol{B}^T)$
- More relations between distributions in Sec. 2.7

Stochastic variables and distributions

- If $X \sim N(0,1)$, then $-X \sim N(0,1)$



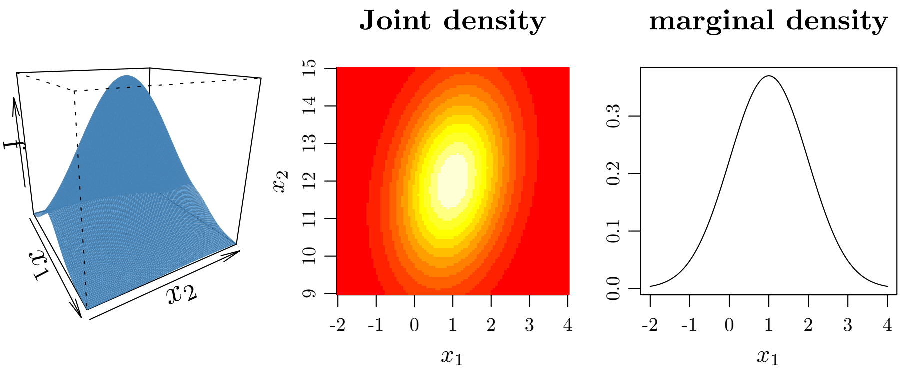


- $X, -X$ are different variables that have the same distribution

Marginal density function

- Sub-vector: $(X_1, \ldots, X_k)^T, \quad (k < n)$
- Marginal density function:

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



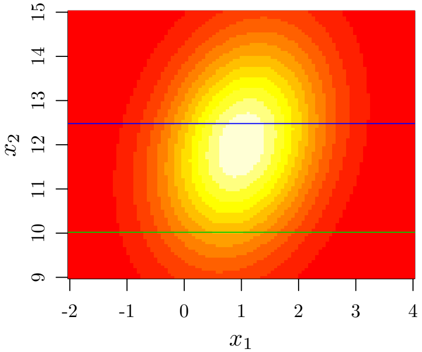
Blackboard - conditional probability

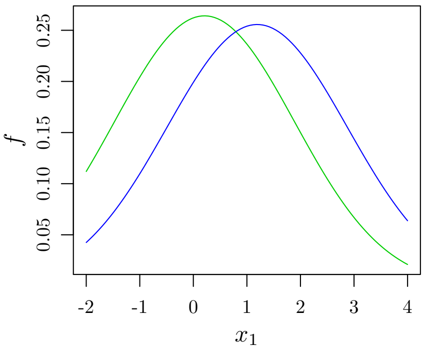
Conditional distributions

- The conditional density of X_1 given $X_2 = x_2$ is defined as ($f_{X_1}(x_1) > 0$):

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

(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 $f_{Y|X=x}(y) = f_Y(y)$
- 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

- 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
- Calculation rule:

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

Moments and Variance

- n'th moment:

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

- 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

- 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:

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

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

  $$V[a + bX_2] = b^2 V[X_2]$$

Moment representation

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

Expectation and Variance for Random Vectors

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

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

- Correlation:

$$\rho_{ij} = \frac{\mathrm{Cov}[X_i, X_j]}{\sqrt{V[X_i]V[X_j]}} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

Correlation and Independence

- If X and Y are independent stochastic variables then $\text{Cov}(X, Y) = 0$ and thus $\text{Corr}(X, Y) = 0$.
- However, if $X \in N(0, 1)$, then

$$\begin{aligned}\text{Cov}(X, X^2) =& E[X \cdot X^2] - E[X] \cdot E[X^2] = E[X^3] \\ &= \int x^3 f_X(x)\, dx = 0\end{aligned}$$

- Thus X and X^2 are uncorrelated, but $E[X^2 | X = x] = x^2$.
- Independence implies no correlation, not the other way around.

Expectation and Variance for Random Vectors

- The correlation matrix $\boldsymbol{R} = \boldsymbol{\rho}$ is an arrangement of ρ_{ij} in a matrix
- Covariance matrix between $\boldsymbol{X}$ (dim. p) and $\boldsymbol{Y}$ (dim. q):

$$\begin{aligned}
\boldsymbol{\Sigma}_{\boldsymbol{XY}} &= C[\boldsymbol{X}, \boldsymbol{Y}] = E\left[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{Y} - \boldsymbol{\nu})^T\right] \\
&= \begin{bmatrix} \mathrm{Cov}[X_1, Y_1] & \cdots & \mathrm{Cov}[X_1, Y_q] \\ \vdots & & \vdots \\ \mathrm{Cov}[X_p, Y_1] & \cdots & \mathrm{Cov}[X_p, Y_q] \end{bmatrix}
\end{aligned}$$

- Calculation rules – see the book.
- The special case of the variance $C[\boldsymbol{X}, \boldsymbol{X}] = V[\boldsymbol{X}]$ results in $\boxed{V[\boldsymbol{AX}] = \boldsymbol{A}V[\boldsymbol{X}]\boldsymbol{A}^T}$

Conditional expectation

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

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

$E[Y] = E\left[E[Y|X]\right]$

$E[g(X)Y|X] = g(X)E[Y|X]$

$E[g(X)Y] = E\left[g(X)E[Y|X]\right]$

$E[a|X] = a$

$E[g(X)|X] = g(X)$

$E[cX + dZ|Y] = cE[X|Y] + dE[Z|Y]$

Variance separation

- Definition of conditional variance and covariance:

$$\begin{aligned} V[\boldsymbol{Y}|\boldsymbol{X}] &= E\left[\left(\boldsymbol{Y}-E[\boldsymbol{Y}|\boldsymbol{X}]\right)\left(\boldsymbol{Y}-E[\boldsymbol{Y}|\boldsymbol{X}]\right)^T|\boldsymbol{X}\right] \\ C[\boldsymbol{Y},\boldsymbol{Z}|\boldsymbol{X}] &= E\left[\left(\boldsymbol{Y}-E[\boldsymbol{Y}|\boldsymbol{X}]\right)\left(\boldsymbol{Z}-E[\boldsymbol{Z}|\boldsymbol{X}]\right)^T|\boldsymbol{X}\right] \end{aligned}$$

- The variance separation theorem:

$$\begin{aligned} V[\boldsymbol{Y}] &= E\left[V[\boldsymbol{Y}|\boldsymbol{X}]\right]+V\left[E[\boldsymbol{Y}|\boldsymbol{X}]\right] \\ C[\boldsymbol{Y},\boldsymbol{Z}] &= E\left[C[\boldsymbol{Y},\boldsymbol{Z}|\boldsymbol{X}]\right]+C\left[E[\boldsymbol{Y}|\boldsymbol{X}],E[\boldsymbol{Z}|\boldsymbol{X}]\right] \end{aligned}$$

Linear Projections

- Consider two random vectors $\boldsymbol{Y}$ and $\boldsymbol{X}$, then

$$E\left[\begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{X} \end{pmatrix}\right] = \begin{pmatrix} \boldsymbol{\mu}_Y \\ \boldsymbol{\mu}_X \end{pmatrix} \text{ and } V\left[\begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{X} \end{pmatrix}\right] = \begin{pmatrix} \boldsymbol{\Sigma}_{YY} & \boldsymbol{\Sigma}_{YX} \\ \boldsymbol{\Sigma}_{XY} & \boldsymbol{\Sigma}_{XX} \end{pmatrix}$$

- Define the *linear projection* $\rho_{\boldsymbol{X}}(\boldsymbol{Y}) \stackrel{def}{=} \boldsymbol{\mu}_Y + \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_X)$
- Then:
  - $\rho_X(\boldsymbol{Y})$ is of the form $\boldsymbol{a} + B\boldsymbol{X}$;
  - $V[\boldsymbol{Y} - \rho_X(\boldsymbol{Y})] = \boldsymbol{\Sigma}_{YY} - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{\Sigma}_{YX}^T$;
  - $\mathrm{Cov}(\boldsymbol{Y} - \rho_X(\boldsymbol{Y}), \boldsymbol{X}) = 0$.

Linear projections and conditional means

- $\rho_X(\boldsymbol{Y})$ minimizes the variance among functions $\boldsymbol{a} + B\boldsymbol{X}$ that gives projection errors uncorrelated with X.
- If $(\boldsymbol{X}, \boldsymbol{Y})$ is multivariate normal, this is a property of $E[\boldsymbol{Y}|\boldsymbol{X}]$.
- Nevertheless, if $\boldsymbol{X} \sim N(0,1)$ then $\rho_X(\boldsymbol{X}^2) = 1$ while $E[\boldsymbol{X}^2|\boldsymbol{X}] = \boldsymbol{X}^2$
- We shall write $E[\boldsymbol{Y}|\boldsymbol{X}]$ for $\rho_X(\boldsymbol{Y})$ anyway.
  BECAUSE:
  1. The book and other time series literature does so
  2. It is true for Normal stochastic variables
  3. $\rho_X(\boldsymbol{Y})$ satisfies the same calculus rules as $E[\boldsymbol{Y}|\boldsymbol{X}]$ for normal distributions
- The differences should be kept in mind.

Air pollution in cities

- Carstensen (1990) has used time series analysis to set up models for NO and NO_2 at Jagtvej in Copenhagen
- Measurements of NO and NO_2 are available every third hour (00, 03, 06, 09, 12, ...)
- We have $\mu_{NO_2} = 48\mu g/m^3$ and $\mu_{NO} = 79\mu g/m^3$
- In the model $X_{1,t} = NO_{2,t} - \mu_{NO_2}$ and $X_{2,t} = NO_t - \mu_{NO}$ is used

Air pollution in cities – model and forecast

$$\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} 0.9 & -0.1 \\ 0.4 & 0.8 \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} \xi_{1,t} \\ \xi_{2,t} \end{pmatrix}$$

$$\boldsymbol{X}_t = \boldsymbol{\Phi} \boldsymbol{X}_{t-1} + \boldsymbol{\xi}_t$$

$$V[\boldsymbol{\xi}_t] = \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} 30 & 21 \\ 21 & 23 \end{pmatrix} (\mu g/m^3)^2$$

- Assume that t corresponds to 09:00 today and we have measurements 64 $\mu g/m^3$ NO_2 and 93 $\mu g/m^3$ NO
- $\boxed{\text{Forecast the concentrations at 12:00 } (t+1)}$
- $\boxed{\text{What is the variance-covariance of the forecast error?}}$
- The best predictor is the conditional expectation

Air pollution in cities – model and forecast

The forecast:

$$\begin{aligned} E(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t) &= E(\boldsymbol{\Phi}\boldsymbol{X}_t + \boldsymbol{\xi}_{t+1}|\boldsymbol{X}_t) \\ &= \boldsymbol{\Phi}\boldsymbol{X}_t \end{aligned}$$

Variance-covariance of the forecast error.

$$\begin{aligned} V(\boldsymbol{X}_{t+1} - E(\boldsymbol{X}_{t+1}|\boldsymbol{X}_t)|\boldsymbol{X}_t) &= V(\boldsymbol{\Phi}\boldsymbol{X}_t + \boldsymbol{\xi}_{t+1} - \boldsymbol{\Phi}\boldsymbol{X}_t|\boldsymbol{X_t}) \\ &= V(\xi_{t+1}|\boldsymbol{X}_t) \\ &= \boldsymbol{\Sigma} \end{aligned}$$

Air pollution in cities – forecast with R

```r
## The system
mu <- matrix(c(48,79),nrow=2)
Phi <- matrix(c(.9,.4,-.1,.8),nrow=2)
Sigma <- matrix(c(20,21,21,23),nrow=2)
## The forecast of the concentrations
Xt <- matrix(c(64,93),nrow=2)-mu
Xtp1.hat <- Phi%*%Xt
Xtp1.hat + mu

##      [,1]
## [1,] 61.0
## [2,] 96.6

## Variance of the error is trivial
```

Air pollution in cities – linear projection

- At 12:00 ($t+1$) we now assume that NO_2 is measured with $67\ \mu g/m^3$ as the result, **but** NO cannot be measured due to some trouble with the equipment.
- $\boxed{\text{Estimate the missing } NO \text{ measurement.}}$
- $\boxed{\text{What is the variance of the error of the estimation?}}$

Air pollution in cities – linear projection

$$E[X_{2,t+1}|X_{1,t+1},\boldsymbol{X}_t] = \overbrace{E[(X_{2,t+1}|X_{1,t+1})|\boldsymbol{X_t}]}^{\text{compare with (2.65)}} =$$

$$E[X_{2,t+1}|\boldsymbol{X}_t]+\mathrm{Cov}(X_{1,t+1},X_{2,t+1}|\boldsymbol{X}_t)V[X_{1,t+1}|\boldsymbol{X}_t]^{-1}(X_{1,t+1}-E(X_{1,t+1}|\boldsymbol{X}_t))$$

$$= (\Phi_{21}X_{1,t} + \Phi_{22}X_{2,t}) + \Sigma_{12}\Sigma_{11}^{-1}(X_{1,t+1} - (\Phi_{11}X_{1,t} + \Phi_{12}X_{2,t}))$$

The variance of the projection error is (2.66)

$$\begin{aligned} &E(V(X_{2,t+1}|X_{1,t+1},\boldsymbol{X}_t)) \\ &\quad = V(X_{2,t+1}|\boldsymbol{X}_t) - \mathrm{Cov}(X_{2,t+1},X_{1,t+1}|\boldsymbol{X}_t)^2 V(X_{1,t+1}|\boldsymbol{X}_t)^{-1} \\ &\quad = \Sigma_{22} - \Sigma_{12}^2/\Sigma_{11} \end{aligned}$$

Air pollution in cities – linear projection with R

```
## The new observation of X_{1,t+1}
X1tp1 <- 67 - mu[1]
## The projection
Xtp1.hat[2] + mu[2] +
    Sigma[1,2]/Sigma[1,1] * (X1tp1 - Xtp1.hat[1])

## [1] 102.9

## The variance of the projection error
Sigma[2,2] - Sigma[1,2]^2/Sigma[1,1]

## [1] 0.95
```

Highlights

- Covariance calculation rule

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

- The variance separation theorem:

$$
\begin{aligned}
V[\boldsymbol{Y}] &= E\left[V[\boldsymbol{Y}|\boldsymbol{X}]\right] + V\left[E[\boldsymbol{Y}|\boldsymbol{X}]\right] \\
C[\boldsymbol{Y}, \boldsymbol{Z}] &= E\left[C[\boldsymbol{Y}, \boldsymbol{Z}|\boldsymbol{X}]\right] + C\left[E[\boldsymbol{Y}|\boldsymbol{X}], E[\boldsymbol{Z}|\boldsymbol{X}]\right]
\end{aligned}
$$

- Linear projection:

$$
(E[\boldsymbol{Y}|\boldsymbol{X}] =)\rho_{\boldsymbol{X}}(\boldsymbol{Y}) \stackrel{def}{=} \boldsymbol{\mu}_{\boldsymbol{Y}} + \boldsymbol{\Sigma}_{\boldsymbol{YX}}\boldsymbol{\Sigma}_{\boldsymbol{XX}}^{-1}(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})
$$

- Second order moment representation:
  All moments up to second order
  (mean, variance and covariance).

Exercises

Exercises 2.1, 2.2, 2.3
Correction in exercise 2.2: X and ε are mutually independent.