Time Series Analysis

Henrik Madsen

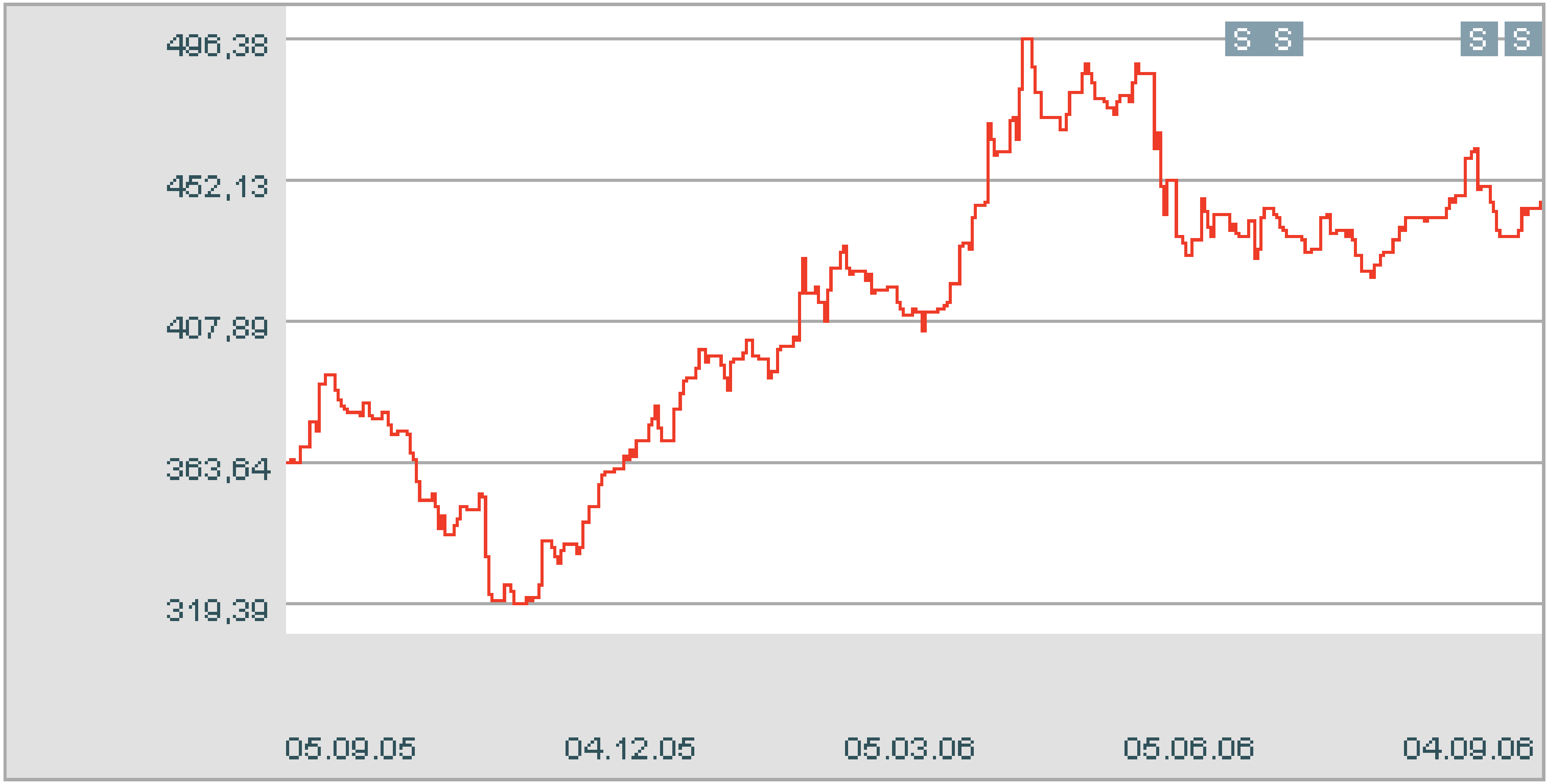
hm@imm.dtu.dk

Informatics and Mathematical Modelling
Technical University of Denmark
DK-2800 Kgs. Lyngby

Outline of the lecture

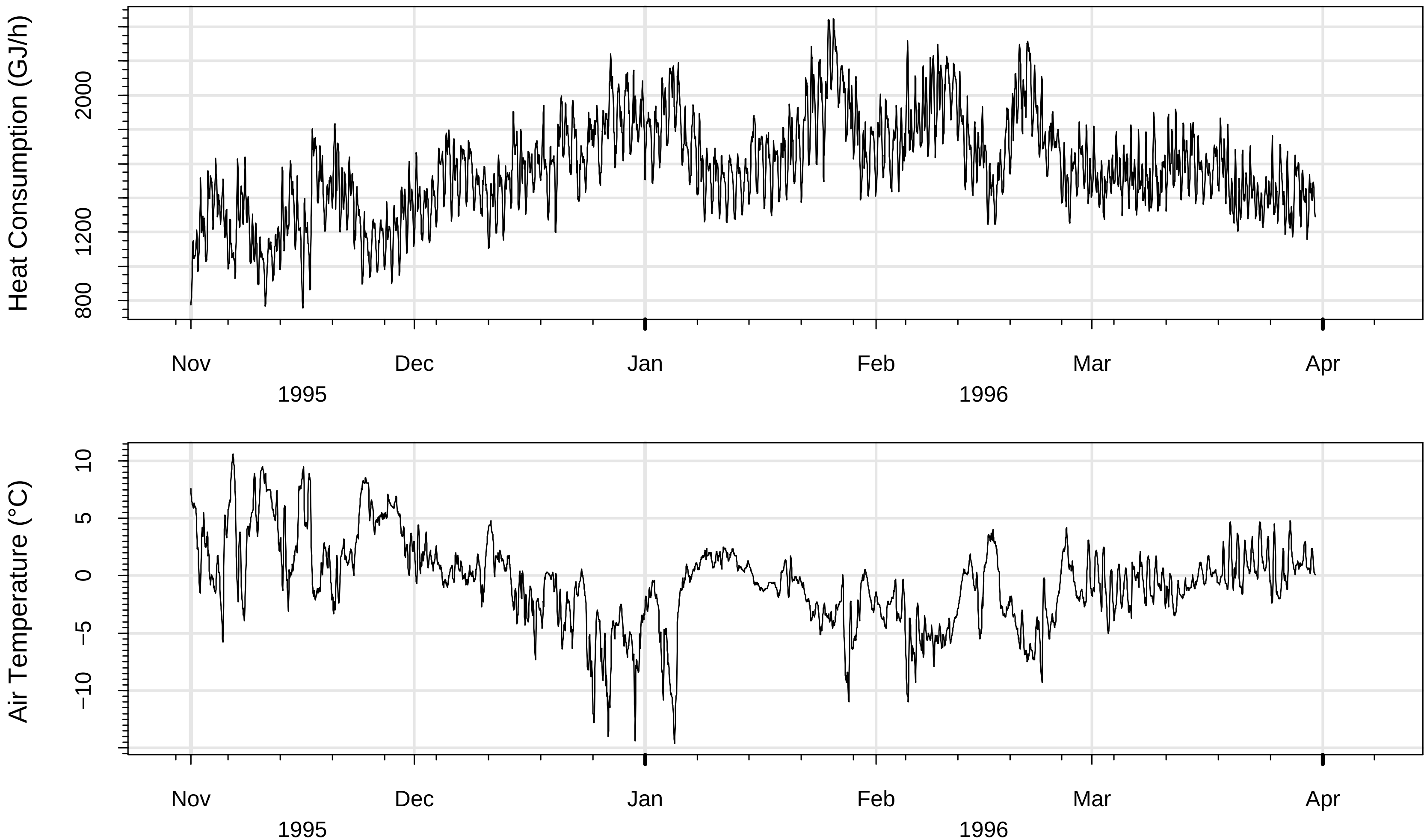
- Practical information
- Introductory examples (See also Chapter 1)
- A brief outline of the course
- Chapter 2:
  - Multivariate random variables
  - The multivariate normal distribution
  - Linear projections
- Example

Introductory example – shares (COLO B 18m)

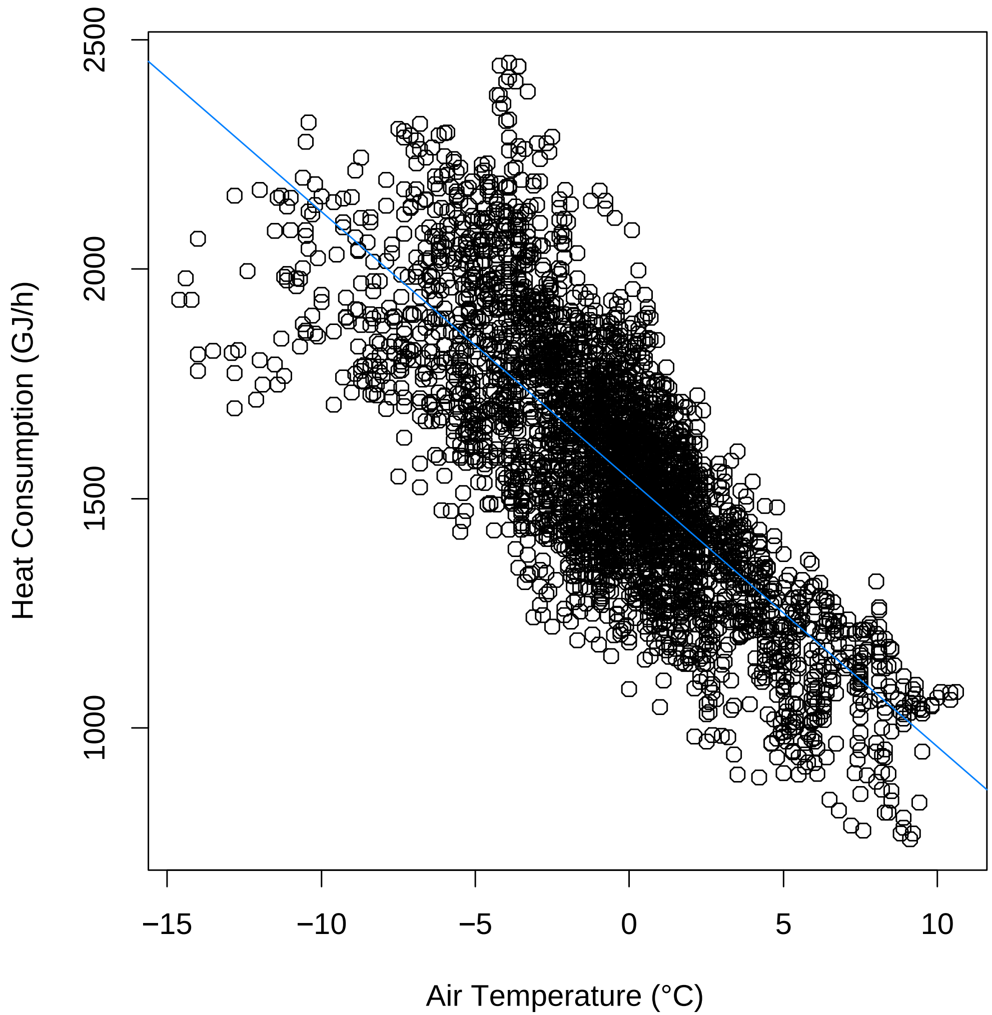


From www.cse.dk

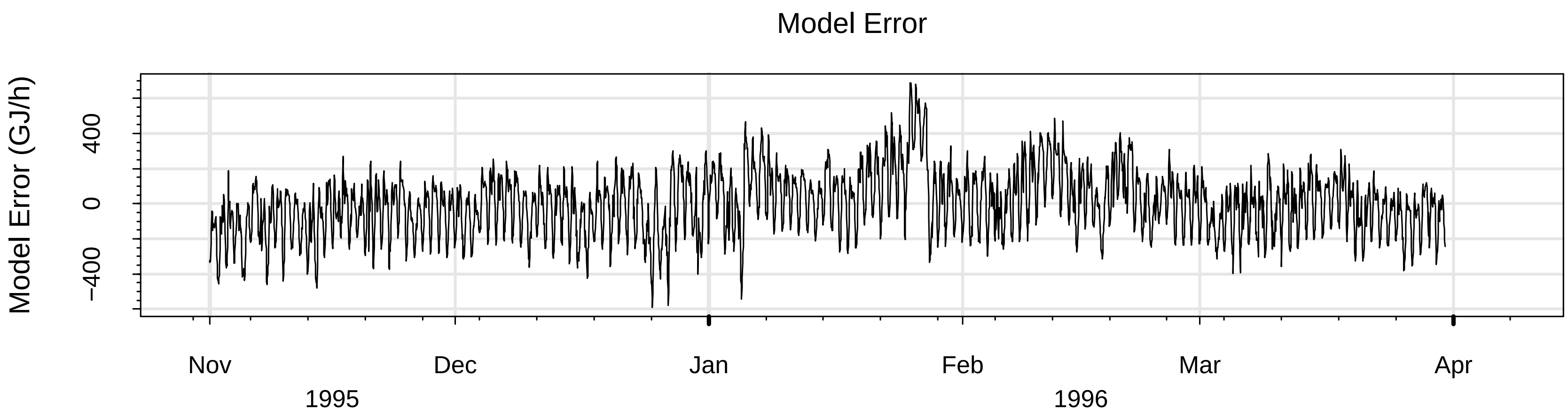
Consumption of District Heating (VEKS) – data

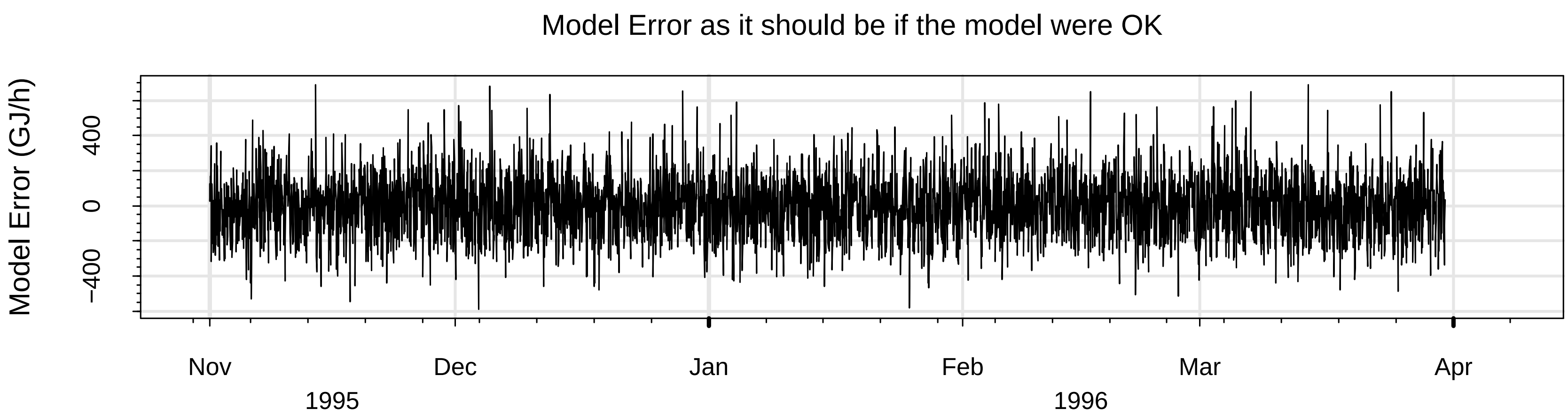


Consumption of DH – simple model



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, spectral distributions, stationarity, ergodicity, linearity, ...
- Signal processing; filtering, sampling, smoothing
- Modelling; with or without external input
- Prediction / Control

Multivariate random variables

- Joint and marginal densities
- Conditional distributions
- Expectations and moments
- Moments of multivariate random variables
- Conditional expectation
- The multivariate normal distribution
- Distributions derived from the normal distribution
- Linear projections

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:

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

Multivariate random variables

- Probability density function (continuous case):

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

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

- Probability density function (discrete case):

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

The Multivariate Normal Distribution

- The joint p.d.f.

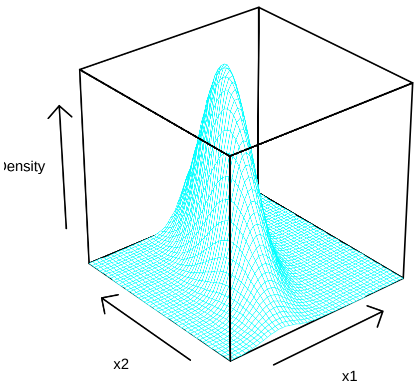
$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \boldsymbol{\Sigma}}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

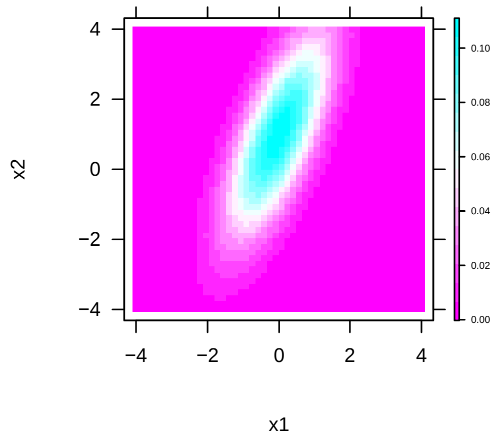
- $\boldsymbol{\Sigma}$ must be positive semidefinite
- Notation: $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- Standardized multivariate normal: $\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{I})$
- $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \boldsymbol{\mu} + \boldsymbol{T}\,\mathrm{N}(\boldsymbol{0}, \boldsymbol{I})$, where $\boldsymbol{\Sigma} = \boldsymbol{T}\boldsymbol{T}^T$
- 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

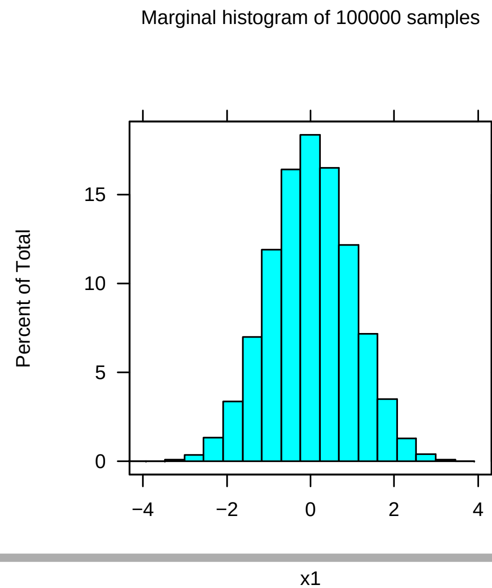
Marginal density function

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

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





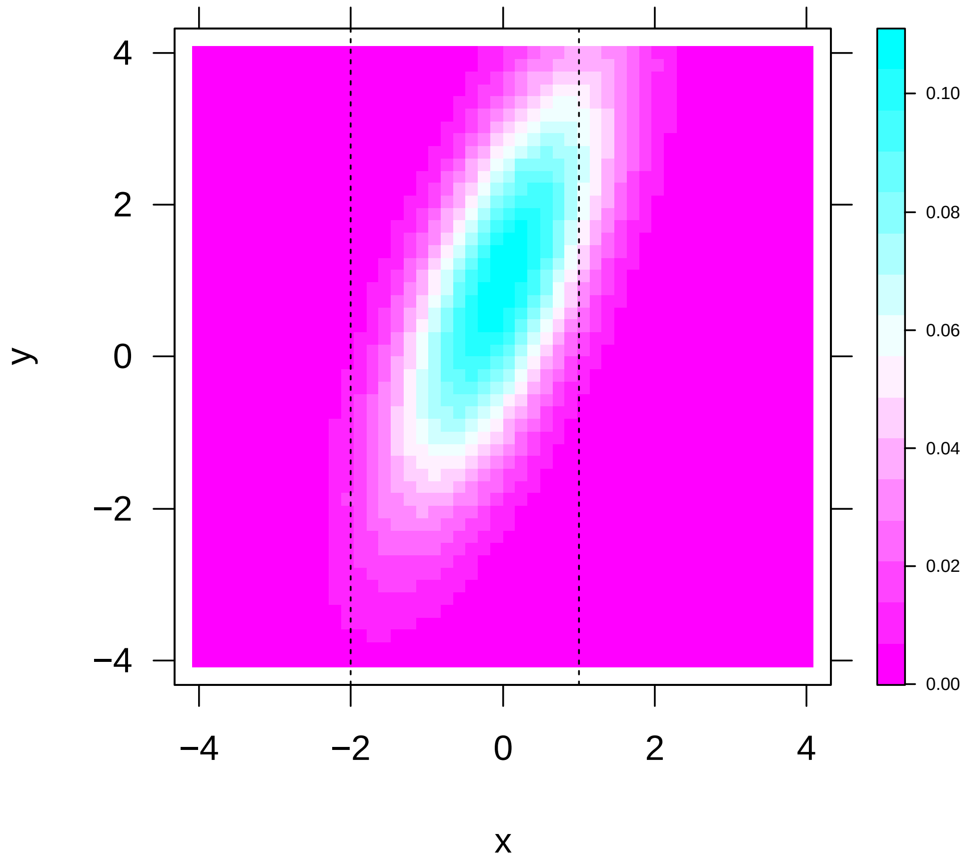


Conditional distributions

- The conditional density of Y given $X = x$ is defined as ($f_X(x) > 0$):

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

(joint density of (X, Y) divided by the marginal density of X evaluated at x)



Independence

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

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

Expectation

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

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

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

- 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:

$$Cov[X_1, X_2] = E[(X_1 - E[X_1])(X_2 - E[X_2])] = E[X_1 X_2] - E[X_1]E[X_2]$$

- Variance and covariance:

$$V[X] = Cov[X, X]$$

- Calculation rule:

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

- The calculation rule can be used for the variance also

Expectation and Variance for Random Vectors

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

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

- Correlation:

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

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):

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

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

$$\boxed{V[\boldsymbol{A}\boldsymbol{X}] = \boldsymbol{A}V[\boldsymbol{X}]\boldsymbol{A}^T}$$

Conditional expectation

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

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

$$E[Y] = E\big[E[Y|X]\big]$$

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

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

$$E[a|X] = a$$

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

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

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

$$E\left[\begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{X} \end{pmatrix}\right] = \begin{pmatrix} \boldsymbol{\mu_Y} \\ \boldsymbol{\mu_X} \end{pmatrix} \text{ and } V\left[\begin{pmatrix} \boldsymbol{Y} \\ \boldsymbol{X} \end{pmatrix}\right] = \begin{pmatrix} \boldsymbol{\Sigma_{YY}} & \boldsymbol{\Sigma_{YX}} \\ \boldsymbol{\Sigma_{XY}} & \boldsymbol{\Sigma_{XX}} \end{pmatrix}$$

- Consider the linear projection: $\boxed{E[\boldsymbol{Y}|\boldsymbol{X}] = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{X}}$
- Then:

$$\begin{aligned} E[\boldsymbol{Y}|\boldsymbol{X}] &= \boldsymbol{\mu_Y} + \boldsymbol{\Sigma_{YX}}\boldsymbol{\Sigma_{XX}}^{-1}(\boldsymbol{X} - \boldsymbol{\mu_X}) \\ V[\boldsymbol{Y} - E[\boldsymbol{Y}|\boldsymbol{X}]] &= \boldsymbol{\Sigma_{YY}} - \boldsymbol{\Sigma_{YX}}\boldsymbol{\Sigma_{XX}}^{-1}\boldsymbol{\Sigma_{YX}}^T \\ C\big[\boldsymbol{Y} - E[\boldsymbol{Y}|\boldsymbol{X}], \boldsymbol{X}\big] &= \boldsymbol{0} \end{aligned}$$

- The linear projection above has minimal variance among all linear projections.

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 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 $93\ \mu g/m^3\ NO$
- Forecast the concentrations at 12:00 ($t+1$)
- What is the variance-covariance of this forecast?

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.
- Estimate the missing NO measurement.
- What is the variance of the error of the estimation?