NTNU
Norwegian University of
Science and Technology

TMA4267 Linear Statistical Models V2014 (26)

Learning outcome, exam, course map, key concepts, reading list

Mette Langaas

To be lectured: April 28, 2014
wiki.math.ntnu.no/emner/tma4267/2014v/start/

Outline

— Learning outcomes.

— Exam.

— Course map - connecting the 7 parts.

— Key concepts.

— Final reading list.

— Acitivites before the exam.

TMA4267 Linear statistical methods Learning outcome, Knowledge

- The student has strong theoretical knowledge about the most popular statistical models and methods that are used in science and technology, with emphasis on regression-type statistical models.
- The statistical properties of the multivariate normal distribution are well known to the student, and the student is familiar with the role of the multivariate normal distribution within linear statistical models.

TMA4267 Linear statistical methods Learning outcome, Skills

- The student knows how to design an experiment and
- how to collect informative data of high quality to study a phenomenon of interest.
- Subsequently, the student is able to choose a suitable statistical model,
- apply sound statistical methods, and
- perform the analyses using statistical software.
- The student knows how to present the results from the statistical analyses, and how to draw conclusions about the phenomenon under study.

Exam

— 9.00-13.00, May 22, 2014.

— Written.

— Makes up 80% of the final grade, the remaining 20 % from the compulsory DOE project.

— Permitted aids: (Code C). One yellow A5 with own handwritten notes, Rottmann: Matematisk formelsamling, Tabeller og formler i statistikk, specified calculator.

Why one yellow A5 sheet?

- Force you to structure the course key concepts?
- Memorizing not needed?
- Security blanket.

TMA4267 – summing up

P7
- Model selection
- Shrinkage; ridge & lasso
- Dimension reduction PCA

P1
Simple linear regression
Correlation ρ vs β
Bivariate normal

P2
N, χ^2, t, F
Analysis of variance ANOVA

P6
Design of Experiments DOE
X orthogonal
2^k, blocking, fraction

Central model
$$Y = X\beta + \varepsilon$$
$n\times 1 \quad n\times p \quad p\times 1 \quad n\times 1$

$\varepsilon \sim N(0, \sigma^2 I)$

multiple linear regression model

P3
E and Cov random vectors
Σ properties, spectral theorem
multivariate Normal
MVN $N(\mu, \Sigma)$

P4
Projection matrices
Quadratic form
$\hat{\beta}$, SSE, SST, SSR (LS)

P5
Model assessment
$\varepsilon \sim N(0, \sigma^2 I)$
Taylor approx
transformation

Some key concepts

E = exercise
L = lecture
p = part

Random vectors:

$\underset{p\times 1}{\underline{X}}$ random vector with $\underset{p\times 1}{E(\underline{X})} = \mu$

and $\underset{p\times p}{Cov(\underline{X})} = \Sigma$ ← (real) positive definite symmetric

$f(x_1, \dots, x_p)$ joint distribution

variances on the diagonal
covariances off-diagonal

$$Cov(\underline{X}) = E\left((X-\mu)(X-\mu)^T\right)$$
$$= E(XX^T) - \mu\mu^T \qquad [E3]$$

Linear combinations : $\underset{k\times p}{A}\ \underset{p\times 1}{\underline{X}}$

$$E(AX) = A\mu$$
$$Cov(AX) = A\Sigma A^T$$

Spectral theorem:

(λ_i, e_i) eigenvalue/eigenvector pairs for Σ

$$\Sigma = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T + \dots + \lambda_p e_p e_p^T$$

$$= P \Lambda P^T$$

$P = [e_1\ e_2 \cdots e_p]$, $\Lambda = \mathrm{diag}(\lambda_i)$

$\Rightarrow$ used much

1) DEF: $\Sigma^{1/2} = P \Lambda^{\frac{1}{2}} P^T$

$\Lambda^{\frac{1}{2}} = \mathrm{diag}(\sqrt{\lambda_i})$

2) Define principal components

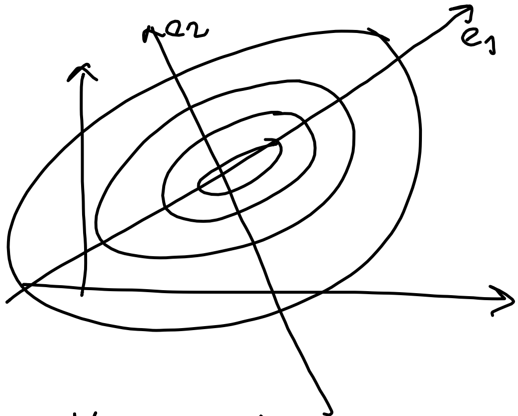
$Z_i = e_i^T X$ to be linear combinations

that are uncorrelated with maximal variance

$E(Z_i) = e_i^T \mu$, $\mathrm{Cov}(Z_i) = e_i^T \Sigma e_i = \lambda_i$

mvN [P3]

$$f(\underline{x}; \mu, \Sigma) = (2\pi)^{-p/2} |\Sigma|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)\right\}$$



[1, 3]

e_1, e_2 of Σ

$Z = \Sigma^{-1/2}(X - \mu)$ ← standard normal and independent components

Tools: MGF, mv transformation formula

$(X-\mu)^T \Sigma^{-1}(X-\mu) \sim \chi^2_p$

$$\Sigma = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 4 & 5 \\ 0 & 5 & 1 \end{bmatrix}$$

Zero correlation implies independence

AX and BX are independent iff $A \Sigma B^T = 0$

Subsets, condition distribution, linear combinations of X are all mvN.

N, χ^2, t, F

E2 ⇐ give the relationship between these.

Exercise 2, Problem 1

The LF for E2, P1 may be too elaborate and technical for the exam.

↑ last parts of

The MLR model

$$\underset{n \times 1}{Y} = \underset{n \times p}{X} \underset{p \times 1}{\beta} + \varepsilon$$

$E(\varepsilon) = 0$, $Cov(\varepsilon) = \sigma^2 I$

and often also $\varepsilon \sim mvN$

β: Part 1-5 : β_0 ($\nearrow \beta_1$) in β ← p covariates with intercept included

6+7 + old exams : β_0 + p addition covariates

What p means differ throughout the course

(you may be confused with formulas with $n-p$ or $n-p-1$)

you may be confused by n-p and n-p-1

LS & ML : $\hat{\beta} = (X^T X)^{-1} X^T Y$

$$E(\hat{\beta}) = \beta \,, \quad Cov(\hat{\beta}) = (X^T X)^{-1} \sigma^2$$

$$\hat{Y} = X\hat{\beta} = \underbrace{X (X^T X)^{-1} X^T}_{\substack{H \\ n \times n}} Y = HY$$

$H \rightarrow$ projects Y down into the space spanned by the column vectors of X.

$(I - H) \rightarrow$ projects to the space orthogonal to the column space of X.

Projection matrices and quadratic forms [P4]

P = projection matrix

↓

idempotent $P^2 = P$

P also symmetric : $P^T = P$ } P is orthogonal

rank(P) ← eigenvalues that were 0 and 1

||

tr(P)

E4 gives many such matrices: H, $I-H$, J, ...

* Let P have rank r and $Y \sim N_n(\mu, \sigma^2 I)$

↑ $X\beta$

then $(Y-\mu)^T P (Y-\mu) \sim \sigma^2 \chi^2_r$

* A, B symmetric projection matrices and $AB = 0$:

$(Y-\mu)^T A (Y-\mu)$ and $(Y-\mu)^T B (Y-\mu)$

are independent

MLR and mvN and symmetric projection matrices

$\varepsilon \sim N_n(0, \sigma^2 I)$

$$\hat{\beta} \sim \underline{N}_p\left(\beta, (X^TX)^{-1}\sigma^2\right)$$

$$S^2 = \frac{SSE}{n-p}$$

$$SSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = Y^T(I-H)Y \sim \sigma^2 \chi^2_{\underbrace{\text{rank}(I-H)}_{n-p}}$$

$$SST = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = Y^T\left(I - \tfrac{1}{n}J\right)Y$$

$$\sim \sigma^2 \chi^2_{\overbrace{\text{rank}(I-\frac{1}{n}J)}^{n-1}}$$

$$SSR = SST - SSE = Y^T\left(H - \tfrac{1}{n}J\right)Y \sim \sigma^2 \chi^2_{\text{rank}\underbrace{(H-\frac{1}{n}J)}_{p-1}}$$

$H_0: \beta_j = 0 \rightarrow$ t-test

$H_1: \beta_1 = \beta_2 = \dots = \beta_p = 0 \rightarrow$ F-test

This was all I had time for on April 28. Some comments:

More on ANOVA will be done in connection with the lecture on the Exam August 2011, April 29, 12.15 in S7.

Part 1 is just a special case of the mvN and the MLR, so no special care is needed.

Parts 5 and 6: I assume you have worked a lot with on your DOE compulsory project.

Part 7: we did some wrt PCA under "Random vectors", and the model selection and shrinkage part is covered on Exercise 7.

Final reading list

- Bingham and Fry (2010) chapters 1-4, and 7.
- James, Witten, Hastie and Tibshirani (2013) chapter 6 and 10.2.
- DOEnote by John Tyssedal.
- The class notes - mainly to cover SSR and SST (defined differently from Ch 3.4 in Bingham and Fry (L15), and the derivation of principal components (L25).
- The 7 exercises.

Comparison with reading list earlier years

Not on the reading list V2014, but on before:

Analysis of contingency tables.

The most complex parts of Design of experiments (folding, combining blocking and fractionating).

Random effects ANOVA.

Multiple testing (Bonferroni, Tukey).

More effort made earlier with quadratic forms and projection matrices in ANOVA (we did ANOVA before projection matrices).

Hotelling T^2.

New on the reading list:

New strategies to Model selection (best subset and forward selection),

Ridge regression and Lasso regression is new.

Activities before the exam

The exam is Thursday May 22 at 9.00.
Exam problems from earlier year is available from the course www-page.
Should we schedule supervision 10.15-12 on Monday May 19 and Tuesday May 20? Or other times? Other activities?