TMA4267 Linear Statistical Models V2017 [L6]

Part 1: Multivariate RVs, and the multivariate normal distribution
Estimators for mean and covariance
Quadratic forms [H:3.3,4.5,5.1,F:AppB3]

Mette Langaas

Department of Mathematical Sciences, NTNU

To be lectured: January 27, 2017

Plan for today

- estimators for mean and covariance
- quadratic forms and rules for quadratic forms
- idempotent matrices
- more rules for quadratic forms - with idempotent matrices

Maximum likelihood estimators

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ be a random sample of size n from the multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The maximum likelihood estimators for are found by maximizing the likelihood:

$$
\begin{aligned}
L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{j=1}^{n} f(\boldsymbol{x}_j; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= \prod_{j=1}^{n} (\frac{1}{2\pi})^{\frac{p}{2}} \det(\boldsymbol{\Sigma})^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\boldsymbol{x}_j - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}_j - \boldsymbol{\mu})\}
\end{aligned}
$$

Could take ln and then partial derivatives, but easier to add and subtract the mean $\bar{\boldsymbol{x}}$ and rewrite (using trace-formulas)

$$
\begin{aligned}
&L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\frac{1}{2\pi})^{\frac{np}{2}} \det(\boldsymbol{\Sigma})^{-\frac{n}{2}} \\
&\exp\{-\frac{1}{2}[\mathrm{tr}(\boldsymbol{\Sigma}^{-1} \sum_{j=1}^{n} (\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})^T) + n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{\mu})]\}
\end{aligned}
$$

Maximum likelihood estimators: first for $\boldsymbol{\mu}$

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\frac{1}{2\pi})^{\frac{np}{2}} \det(\boldsymbol{\Sigma})^{-\frac{n}{2}}$$

$$\exp\{-\frac{1}{2}[\mathrm{tr}(\boldsymbol{\Sigma}^{-1} \sum_{j=1}^{n} (\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})^T) + n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{\mu})]\}$$

and see directly for SPD $\boldsymbol{\Sigma}$ that the maximum is achieved for $\boldsymbol{\mu} = \bar{\boldsymbol{x}}$, so that the MLE for $\boldsymbol{\mu}$ is

$$\bar{\boldsymbol{X}} = \frac{1}{n} \sum_{j=1}^{n} \boldsymbol{X}_j$$

Maximum likelihood estimators: then for $\boldsymbol{\Sigma}$

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\frac{1}{2\pi})^{\frac{np}{2}} \det(\boldsymbol{\Sigma})^{\frac{n}{2}}$$

$$\exp\{-\frac{1}{2}[\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\sum_{j=1}^{n}(\boldsymbol{x}_j - \bar{\boldsymbol{x}})(\boldsymbol{x}_j - \bar{\boldsymbol{x}})^T) + n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})]\}$$

A maximization theorem for matrices it used to find that the MLE for $\boldsymbol{\Sigma}$ is

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{j=1}^{n}(\boldsymbol{X}_j - \bar{\boldsymbol{X}})(\boldsymbol{X}_j - \bar{\boldsymbol{X}})^T$$

Properties of the ML estimators

- $\bar{\boldsymbol{X}}$ is distributed as $N_p(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma})$
- $n\boldsymbol{S}$ is distributed as a Wishart random matrix with $n-1$ degrees of freedom.
- $\bar{\boldsymbol{X}}$ and $n\boldsymbol{S}$ are independent.

The Wishart distribution is not on the reading list for TMA4267. General properties of maximum likelihood estimation is covered in detail in TMA4295 Statistical Inference.

Quadratic forms - first results [F:B3.3, Theorem B.2]

We stay with our random vector $\boldsymbol{X}$ with $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and a symmetric constant matrix $\boldsymbol{A}$.

- What is a quadratic form? $\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}$
- The "trace-formula": $\mathrm{E}(\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X})$.

Exam V2014: Problem 1a

Let $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$ be a random vector with mean $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{X}) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

and covariance matrix $\boldsymbol{\Sigma} = \mathrm{Cov}(\boldsymbol{X}) = \boldsymbol{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Further, let

$\boldsymbol{A} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix}$ be a matrix of constants.

Define $\boldsymbol{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \boldsymbol{AX}$.

Find $\mathrm{E}(\boldsymbol{Y})$ and $\mathrm{Cov}(\boldsymbol{Y})$. Are X_1 and X_2 independent? Are Y_1 and Y_2 independent? Justify your answers.
Find the mean of $\boldsymbol{X}^T\boldsymbol{AX}$.

Useful facts about the trace [H:2.1] and [F:Theorem A.18]

Let A, B and C be conformable matrices

$$\begin{aligned}
\mathrm{tr}(A+B) &= \mathrm{tr}(A)+\mathrm{tr}(B) \\
\mathrm{tr}(AB) &= \mathrm{tr}(BA) \\
\mathrm{tr}(ABC) &= \mathrm{tr}(CAB) = \mathrm{tr}(BCA)
\end{aligned}$$

Quadratic forms - last results [F:B3.3, Theorem B.2]

Now: $\boldsymbol{X}$ is multivariate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{I}$, and we also have a symmetric and idempotent matrix $\boldsymbol{R}_{(p \times p)}$ with rank r.

- Properties of an idempotent matrix.
- Distribution of $\boldsymbol{X}^T \boldsymbol{R} \boldsymbol{X} \sim \chi^2_r$.
- Distribution of a ratio of two quadratic forms and the Fisher distribution.

Properties of symmetric idempotent matrices

A symmetric matrix $\boldsymbol{A}$ is idempotent, $\boldsymbol{A}^2 = \boldsymbol{A}$, and has the following properties (to be proven in RecEx1.P7).

1. The eigenvalues are 0 and 1.
2. The rank of a symmetric matrix (actually: a diagonalizable quadratic matrix) equals the number of nonero eigenvaluse of the matrix. Should be known from previous courses.
3. (Combining 1+2). If a $(n \times n)$ symmetric idempotent matrix $\boldsymbol{A}$ has rank r then r eigenvalues are 1 and $n - r$ are 0.
4. The trace and rank of a symmetric projection matrix are equal: $tr(\boldsymbol{A}) = \text{rank}(\boldsymbol{A})$.
5. The matrix $\boldsymbol{I} - \boldsymbol{A}$ is also idempotent, and $\boldsymbol{A}(\boldsymbol{I} - \boldsymbol{A}) = 0$.

The Chi-square distribution

pdf χ^2_p:

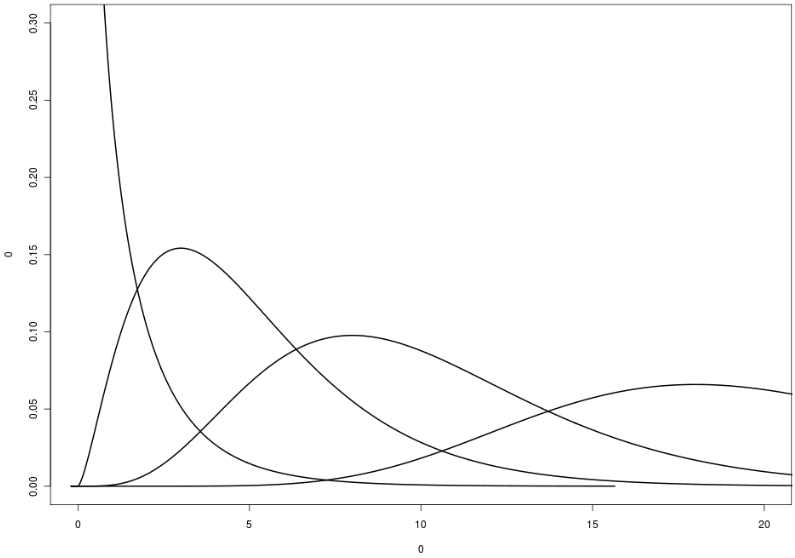
$$f(y) = \frac{1}{2^{p/2}\Gamma(p/2)} y^{p/2-1} e^{(-y/2)} \text{ for } y > 0$$

MGF χ^2_p:

$$M_Y(t) = \frac{1}{(1-2t)^{p/2}}$$

Addition property:
Let $X_1 \sim \chi^2_p$ and $X_2 \sim \chi^2_q$, and let X_1 and X_2 be independent.
Then $X_1 + X_2 \sim \chi^2_{p+q}$.
Subtraction property:
Let $X = X_1 + X_2$ with $X_1 \sim \chi^2_p$ and $X \sim \chi^2_{p+q}$. Assume that X_1 and X_2 are independent. Then $X_2 \sim \chi^2_q$.

Kjikvadrat 1,5,10,20



The Fisher distribution [F: B.1 Def 8.14], RecEx2.P5+6

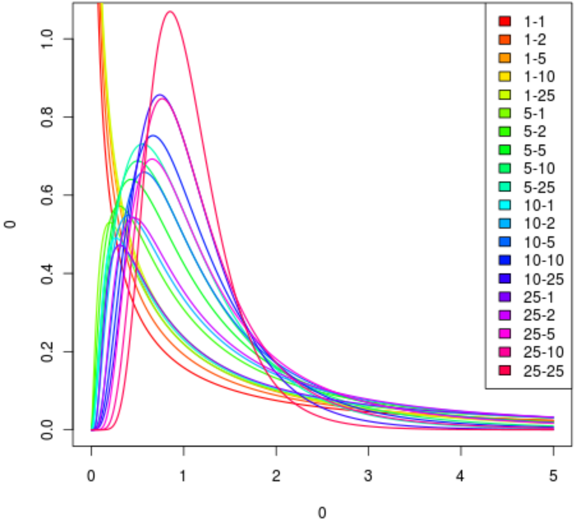
"Tabeller og formeler i statistikk":
If Z_1 and Z_2 are independent and χ^2-distributed with ν_1 and ν_2 degrees of freedom, then

$$F = \frac{Z_1/\nu_1}{Z_2/\nu_2}$$

is F(isher)-distributed with ν_1 and ν_2 degrees of freedom.

- The expected value of F is $\mathrm{E}(F) = \frac{\nu_2}{\nu_2 - 2}$.
- The mode is at $\frac{\nu_1 - 2}{\nu_1} \frac{\nu_2}{\nu_2 + 2}$.
- Identity:

$$f_{1-\alpha,\nu_1,\nu_2} = \frac{1}{f_{\alpha,\nu_2,\nu_1}}$$



The Fisher distribution with different degrees of freedom ν_1 and ν_2 (given in the legend).

Quadratic forms [F:B3.3, Theorem B.2]

Random vector $\boldsymbol{X}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, symmetric constant matrix $\boldsymbol{A}$.

- Quadratic form: $\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}$.
- The "trace-formula": $\mathrm{E}(\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}) = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) - \boldsymbol{\mu}^T\boldsymbol{A}\boldsymbol{\mu}$.

Then, let $\boldsymbol{X} \sim N_p(\boldsymbol{0}, \boldsymbol{I})$, and $\boldsymbol{R}$ is a symmetric and idempotent matrix with rank r.

$$\boldsymbol{X}^T\boldsymbol{R}\boldsymbol{X} \sim \chi^2_r$$

Now, also $\boldsymbol{S}$ is a symmetric and idempotent matrix with rank s, and $\boldsymbol{R}\boldsymbol{S} = \boldsymbol{0}$.

$$\frac{s\boldsymbol{X}^T\boldsymbol{R}\boldsymbol{X}}{r\boldsymbol{X}^T\boldsymbol{S}\boldsymbol{X}} \sim F_{r,s}$$

Plan for the last week of Part 1

Supervision in lecture times.
See Blackboard: Part 1: dates and places for supervision.

Kahoot!

Summing up the last three lectures with a few multiple choice questions.