TMA4267 Linear Statistical Models V2017 [L4]

Part 1: Multivariate RVs, and the multivariate normal distribution
The multivariate normal distribution (pdf and mgf) [H:4.2-4.4]

Mette Langaas

Department of Mathematical Sciences, NTNU

To be lectured: January 20, 2017

What we know, and the plan for this lecture

- A random vector $\boldsymbol{X}$ can be described by the joint pdf $f(\boldsymbol{x})$.
- Mean: $\boldsymbol{\mu} = E(\boldsymbol{X}) = \{\mathrm{E}(X_j)\}$
- Covariance matrix: $\mathrm{Cov}(\boldsymbol{X}) = \mathrm{E}((\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T)$, symmetric and we often require the matrix to be positive definite.
- Linear combinations $\boldsymbol{CX}$: $\mathrm{E}(\boldsymbol{CX}) = \boldsymbol{C}\boldsymbol{\mu}_X$ and $\mathrm{Cov}(\boldsymbol{CX}) = \boldsymbol{C\Sigma C}^T$.
- Now: derive the joint pdf and the moment generating function for the multivariate normal distribution.

Why is the mulitivariate normal distribution so important in statistics?

- Many natural phenomena may be modelled using this distribution (just as in the univariate case).
- Multivariate version of the central limit theorem- the sample mean will be approximately multivariate normal for large samples.
- Good interpretability of the covariance.
- Mathematically tractable.
- Building block in many models and methods.

Cramer-Wold and moment generating functions

$\boldsymbol{X}_{(p\times 1)}$ is a random vector. The distribution of $\boldsymbol{X}$ is completely determined by the set of all one-dimensional distributions of the linear combinations $Y = \boldsymbol{t}^T\boldsymbol{X} = \sum_{i=1}^{p} t_i X_i$ where $\boldsymbol{t}$ ranges over all fixed p-vectors.

- $Y = \boldsymbol{t}^T\boldsymbol{X}$ has MGF $M_Y(s) = \mathrm{E}(\exp(sY)) = \mathrm{E}(\exp(s\boldsymbol{t}^T\boldsymbol{X}))$.
- If we choose $s = 1$ $M_Y(1) = \mathrm{E}(\exp(\boldsymbol{t}^T\boldsymbol{X})) = M_{\boldsymbol{X}}(\boldsymbol{t})$, which is the MGF of $\boldsymbol{X}$ and thus determines the distribution of $\boldsymbol{X}$.

Härdle and Simes (2015) use characteristic functions, $\mathrm{E}(e^{i\boldsymbol{t}^T\boldsymbol{X}})$ but we stick with moment generating functions $\mathrm{E}(e^{\boldsymbol{t}^T\boldsymbol{X}})$. Why: we will only work with nice distributions and do not have problems with integrals not existing, and we know MGFs from previous course.

Multivariate transformation formula [H:4.3]

$$X = u(Y) \tag{4.43}$$

for a one-to-one transformation $u: \mathbb{R}^p \to \mathbb{R}^p$. Define the Jacobian of u as

$$\mathcal{J} = \left(\frac{\partial x_i}{\partial y_j}\right) = \left(\frac{\partial u_i(y)}{\partial y_j}\right)$$

and let $\text{abs}(|\mathcal{J}|)$ be the absolute value of the determinant of this Jacobian. The pdf of Y is given by

$$f_Y(y) = \text{abs}(|\mathcal{J}|) \cdot f_X\{u(y)\}. \tag{4.44}$$

The Chi-square distribution

pdf χ^2_p:

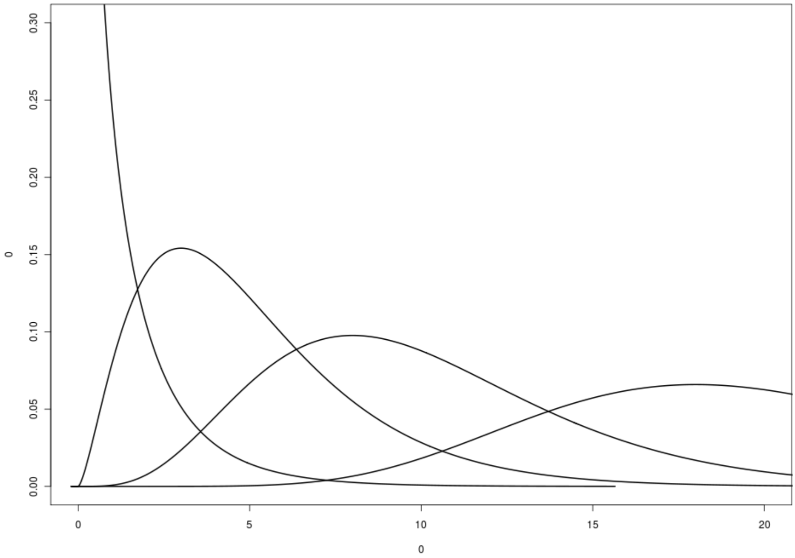
$$f(y) = \frac{1}{2^{p/2}\Gamma(p/2)} y^{p/2-1} e^{(-y/2)} \text{ for } y > 0$$

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



This lecture: derived the MGF and pdf of the multivariate normal distribution

1. $Z \sim N_1(0, 1)$
   - MGF: $M_Z(t) = \mathrm{E}(e^{tz}) = e^{\frac{1}{2}t^2}$
2. $Z_1, Z_2, \ldots, Z_p$ iid $N_1(0, 1) \rightarrow \boldsymbol{Z}_{p\times 1} \sim N_p(\boldsymbol{0}, \boldsymbol{I})$
   - MGF: $M_{\boldsymbol{Z}}(\boldsymbol{t}) = \mathrm{E}(e^{\boldsymbol{t}^T\boldsymbol{z}}) = e^{\frac{1}{2}\boldsymbol{t}^T\boldsymbol{t}}$
3. $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{\mu}$, $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{\Sigma}$ gives $\boldsymbol{X}_{p\times 1} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
   - MGF: $M_{\boldsymbol{X}}(\boldsymbol{t}) = \mathrm{E}(e^{\boldsymbol{t}^T\boldsymbol{x}}) = e^{\boldsymbol{t}^T\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{t}}$
   - pdf (invertible):

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} exp\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\}$$

Properties of the mvN - plan for L5

Let $\boldsymbol{X}_{(p\times 1)}$ be a random vector from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

1. Probability density function $f(\boldsymbol{x})$ (both when $\boldsymbol{\Sigma}$ is invertible and not).
2. Moment generating function: $M_X(\boldsymbol{t}) = \exp(\boldsymbol{t}^T\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t})$
3. Graphical display, contours (ellipsoids), and chisq-distributed $(\boldsymbol{X}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})$.
4. Linear combinations of components of $\boldsymbol{X}$ are (multivariate) normal.
5. All subsets of the components of $\boldsymbol{X}$ are (multivariate) normal.
6. Zero covariance implies that the corresponding components are independently distributed.
7. $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}^T = \boldsymbol{0} \Leftrightarrow \boldsymbol{A}\boldsymbol{X}$ and $\boldsymbol{B}\boldsymbol{X}$ are independent.
8. The conditional distributions of the components are (multivariate) normal. $\boldsymbol{X}_2 \mid (\boldsymbol{X}_1 = \boldsymbol{x}_1) \sim N_{p2}(\boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})$

And then remains estimators for parameters and properties of quadratic forms in L6.