Stochastics III

Ulm University
Institute of Stochastics

Lecture Notes
Prof. Dr. Volker Schmidt
Winter Term 2012/2013

Ulm, October 2012

Contents

References

[1] Büning, H., Trenkler, G. (1994)
Nichtparametrische statistische Methoden
de Gruyter, Berlin

[2] Cressie, N.A. (1993)
Statistics for Spatial Data
J. Wiley & Sons, New York

[3] Dobson, A.J. (2002)
An Introduction to Generalized Linear Models
Chapman & Hall, Boca Raton

[4] Falk, M., Marohn, F., Tewes, B. (2002)
Foundations of Statistical Analyses and Applications with SAS
Birkhäuser, Basel

[5] Hastie, T., Tibshirami, R., Friedman, J. (2001)
The Elements of Statistical Learnning
Springer, New York

[6] Koch, K.R. (1997)
Parameterschätzung und Hypothesentests in linearen Modellen
Dümmlers–Verlag, Bonn

[7] Lehmann, E.L. (1999)
Elements of Large–Sample Theory
Springer, New York

[8] Lehmann, E.L., Romano, J.P. (2005)
Testing Statistical Hypotheses
Springer, New York

[9] McCullagh, P., Nelder, J.A. (1989)
Genralized Linear Models
Chapman & Hall, London.

[10] Pruscha, H. (2000)
Vorlesungen über mathematische Statistik
Teubner–Verlag, Stuttgart.

[11] Van der Vaart, A., Wellner, J. (1996)
Weak Convergence and Empirical Processes
Springer–Verlag, New York

[12] Vapnik, V.N. (1998)
Statistical Learning Theory
J. Wiley & Sons, New York

1 Introduction and Mathematical Foundations

These lecture notes are made for students who already have a basic knowledge of mathematical statistics. Estimation and statistical test methods which have been discussed in "Stochastik I" are assumed to be known.

The present lecture notes consist of the following parts:

- multivariate normal distribution (nondegenerate and degenerate normal distribution, linear and quadratic forms)
- linear models (multiple regression, normally distributed disturbance terms, single– and multiple–factor analysis of variance)
- generalized linear models (logistic regression, maximum–likelihood equation, weighted least squares estimator, evaluation of the goodness of fit)
- tests for distribution assumptions (Kolmogorow–Smirnow test, χ^2–goodness–of–fit test of Pearson–Fisher)
- nonparametric location tests (binomial test, iteration tests, linear rank tests)

In particular, we will use notions and results which have been introduced in the lecture notes "Elementare Wahrscheinlichkeitsrechnung und Statistik" and "Stochastik I": we will indicate references to these lecture notes by "WR" and "I" in front of the section number of the cited lemmas, theorems, corollaries and formulas.

1.1 Some Basic Notions and Results of Matrix Algebra

First, we recall some basic notions and results of matrix algebra, which are needed in these lecture notes.

1.1.1 Trace and Rank

- The *trace* $\operatorname{tr}(\mathbf{A})$ of a quadratic $n \times n$ matrix $\mathbf{A} = (a_{ij})$ is given by

$$\operatorname{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii} \,. \tag{1}$$

- Let $\mathbf{A}$ be an arbitrary $n \times m$ matrix. The *rank* $\operatorname{rk}(\mathbf{A})$ is the maximum number of linearly independent rows (or columns) of $\mathbf{A}$.
  - The vectors $\mathbf{a}_1, \ldots, \mathbf{a}_\ell \in \mathbb{R}^m$ are called *linearly dependent* if there exist real numbers $c_1, \ldots, c_\ell \in \mathbb{R}$, which are not all equal to zero and $c_1\mathbf{a}_1 + \ldots + c_\ell\mathbf{a}_\ell = \mathbf{o}$.
  - Otherwise the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_\ell \in \mathbb{R}^m$ are called *linearly independent*.

From the definition of the trace of a matrix in (1) and from the definition of matrix multiplication the next lemma directly follows.

Lemma 1.1 *Let $\mathbf{C}$ be an arbitrary $n \times m$ matrix and $\mathbf{D}$ an arbitrary $m \times n$ matrix. Then* $\operatorname{tr}(\mathbf{CD}) = \operatorname{tr}(\mathbf{DC})$.

It can be proved that a quadratic matrix $\mathbf{A}$ is invertible if and only if $\mathbf{A}$ has full rank or $\det \mathbf{A} \neq 0$, respectively. The following result is also useful in this context.

Lemma 1.2 *Let $\mathbf{A}$ be an $n \times m$ matrix with $n \geq m$ and* $\operatorname{rk}(\mathbf{A}) = m$. *Then* $\operatorname{rk}(\mathbf{A}^\top \mathbf{A}) = m$.

Proof

- It is obvious that the rank $\mathrm{rk}(\mathbf{A}^\top\mathbf{A})$ of the $m \times m$ matrix $\mathbf{A}^\top\mathbf{A}$ cannot exceed m.
- Now, we assume that $\mathrm{rk}(\mathbf{A}^\top\mathbf{A}) < m$. Then, there exists a vector $\mathbf{c} = (c_1, \ldots, c_m)^\top \in \mathbb{R}^m$ with $\mathbf{c} \neq \mathbf{o}$ and $\mathbf{A}^\top\mathbf{A}\mathbf{c} = \mathbf{o}$.
- From this follows that $\mathbf{c}^\top\mathbf{A}^\top\mathbf{A}\mathbf{c} = \mathbf{o}$ and $(\mathbf{A}\mathbf{c})^\top(\mathbf{A}\mathbf{c}) = \mathbf{o}$, i.e., $\mathbf{A}\mathbf{c} = \mathbf{o}$.
- However, this is contradictory to the assumption that $\mathrm{rk}(\mathbf{A}) = m$. □

Furthermore, it can be proved that the following properties of trace and rank are valid.

Lemma 1.3 *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be arbitrary* $n \times n$ *matrices. Then* $\mathrm{tr}(\mathbf{A} - \mathbf{B}) = \mathrm{tr}(\mathbf{A}) - \mathrm{tr}(\mathbf{B})$ *always holds. If* $\mathbf{A}$ *is idempotent and symmetric, i.e.,* $\mathbf{A} = \mathbf{A}^2$ *and* $\mathbf{A} = \mathbf{A}^\top$*, it also holds that* $\mathrm{tr}(\mathbf{A}) = \mathrm{rk}(\mathbf{A})$.

1.1.2 Eigenvalues and Eigenvectors

Definition Let $\mathbf{A}$ be an arbitrary $n \times n$ matrix. Each (complex) number $\lambda \in \mathbb{C}$ is called an *eigenvalue* of the matrix $\mathbf{A}$ if and only if there exists a vector $\mathbf{x} \in \mathbb{C}^n$ with $\mathbf{x} \neq \mathbf{o}$ and

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{o}\,. \tag{2}$$

We call $\mathbf{x}$ an *eigenvector* corresponding to λ.

Remark

- Only if λ is a solution of the so–called *characteristic equation*

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0\,, \tag{3}$$

  there is a solution $\mathbf{x} \in \mathbb{C}^n$ with $\mathbf{x} \neq \mathbf{o}$ for (2). The left–hand side $P(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$ of (3) is called the *characteristic polynomial* of matrix $\mathbf{A}$.
- Let $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ be the real–valued solutions of (3). Then the characteristic polynomial can be written in the form

$$P(\lambda) = (-1)^n(\lambda - \lambda_1)^{a_1} \ldots (\lambda - \lambda_k)^{a_k} q(\lambda)\,, \tag{4}$$

  where $a_1, \ldots, a_k \in \mathbb{N}$ are positive natural numbers, the so–called *algebraic multiplicities* of $\lambda_1, \ldots, \lambda_k$, and $q(\lambda)$ is a polynomial of order $n - \sum_{i=1}^k a_i$ which has no real solutions.

Lemma 1.4 *Let* $\mathbf{A} = (a_{ij})$ *be a symmetric* $n \times n$ *matrix with real–valued entries* a_{ij}*. Then every eigenvalue is a real number and eigenvectors* $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^n$ *which correspond to different eigenvalues* $\lambda_i, \lambda_j \in \mathbb{R}$ *are orthogonal to each other.*

Proof

- The determinant $\det(\mathbf{A} - \lambda\mathbf{I})$ in (3) is given by

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \sum_{\boldsymbol{\pi}} (-1)^{r(\boldsymbol{\pi})} \prod_{i:\, i \neq \pi_i} a_{i\pi_i} \prod_{i:\, i = \pi_i} (a_{i\pi_i} - \lambda)\,, \tag{5}$$

  where the summation extends over all $m!$ permutations $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)$ of the natural numbers $1, \ldots, m$ and $r(\boldsymbol{\pi})$ is the number of pairs in $\boldsymbol{\pi}$, which are not in the natural order.

- Since the elements of $\mathbf{A}$ are real numbers, every solution $\lambda = a + \mathrm{i}\, b$ of (3) implies another solution $\overline{\lambda} = a - \mathrm{i}\, b$ of (3).
- Let $\mathbf{x} = \mathbf{a} + \mathrm{i}\,\mathbf{b}$ and $\overline{\mathbf{x}} = \mathbf{a} - \mathrm{i}\,\mathbf{b}$ be eigenvectors which correspond to λ or $\overline{\lambda}$, respectively. Then $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ and $\mathbf{A}\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$ or
$$\overline{\mathbf{x}}^\top \mathbf{A}\mathbf{x} = \overline{\mathbf{x}}^\top \lambda \mathbf{x} = \lambda \overline{\mathbf{x}}^\top \mathbf{x}$$
and
$$\overline{\mathbf{x}}^\top \mathbf{A}\mathbf{x} = \left(\mathbf{A}^\top \overline{\mathbf{x}}\right)^\top \mathbf{x} = \left(\mathbf{A}\overline{\mathbf{x}}\right)^\top \mathbf{x} = \left(\overline{\lambda}\overline{\mathbf{x}}\right)^\top \mathbf{x} = \overline{\lambda}\overline{\mathbf{x}}^\top \mathbf{x}\,.$$
- From this it follows that $\lambda \overline{\mathbf{x}}^\top \mathbf{x} = \overline{\lambda}\overline{\mathbf{x}}^\top \mathbf{x}$.
- Since $\overline{\mathbf{x}}^\top \mathbf{x} = |\mathbf{a}|^2 + |\mathbf{b}|^2 > 0$, it holds that $\lambda = \overline{\lambda}$, i.e., λ is a real number.
- In a similar way it can be proved that for different eigenvalues $\lambda_i, \lambda_j \in \mathbb{R}$ there exist eigenvectors $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^n$ with real–valued components which are orthogonal to each other.
- Since the matrix $\mathbf{A} - \lambda_i \mathbf{I}$ only contains real–valued elements, it holds that if $\mathbf{x}_i$ is an eigenvector which corresponds to λ_i, then also $\overline{\mathbf{x}}_i$ and $\mathbf{x}_i + \overline{\mathbf{x}}_i \in \mathbb{R}^n$ are eigenvectors that correspond to λ_i .
- Therefore we can (and will) assume w.l.o.g. that $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^n$. Furthermore, if
$$(\mathbf{A} - \lambda_i \mathbf{I})\mathbf{x}_i = \mathbf{o} \qquad \text{and} \qquad (\mathbf{A} - \lambda_j \mathbf{I})\mathbf{x}_j = \mathbf{o},$$
it follows that $\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i$ and $\mathbf{A}\mathbf{x}_j = \lambda_j \mathbf{x}_j$ as well as
$$\mathbf{x}_j^\top \mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_j^\top \mathbf{x}_i \qquad \text{and} \qquad \mathbf{x}_i^\top \mathbf{A}\mathbf{x}_j = \lambda_j \mathbf{x}_i^\top \mathbf{x}_j\,.$$
- On the other hand it is obvious that $\mathbf{x}_j^\top \mathbf{x}_i = \mathbf{x}_i^\top \mathbf{x}_j$ and with the symmetry of $\mathbf{A} = (a_{ij})$ we get the identity $\mathbf{x}_j^\top \mathbf{A}\mathbf{x}_i = \mathbf{x}_i^\top \mathbf{A}\mathbf{x}_j$ since
$$\mathbf{x}_j^\top \mathbf{A}\mathbf{x}_i = \sum_{m=1}^n \sum_{\ell=1}^n x_{\ell j} a_{\ell m} x_{mi} = \sum_{\ell=1}^n \sum_{m=1}^n x_{mi} a_{m\ell} x_{\ell j} = \mathbf{x}_i^\top \mathbf{A}\mathbf{x}_j\,.$$
- Altogether it follows that $\lambda_i \mathbf{x}_j^\top \mathbf{x}_i = \lambda_j \mathbf{x}_i^\top \mathbf{x}_j$ and $(\lambda_i - \lambda_j)\mathbf{x}_j^\top \mathbf{x}_i = 0$.
- As $\lambda_i - \lambda_j \neq 0$, it holds that $\mathbf{x}_j^\top \mathbf{x}_i = 0$. □

1.1.3 Diagonalization Method

- Now, let $\mathbf{A}$ be an invertible symmetric $n \times n$ matrix.
- In Lemma 1.4 we have shown that all eigenvalues $\lambda_1, \ldots, \lambda_n$ of $\mathbf{A}$ are real numbers (where it is possible that one number occurs more than once in this sequence).
- Since $\det \mathbf{A} \neq 0$, we get that $\lambda = 0$ is no solution of (3), i.e., all eigenvalues $\lambda_1, \ldots, \lambda_n$ of $\mathbf{A}$ are different from zero.
- Furthermore, it can be proved that there are *orthonormal* (basis) vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^n$, i.e.,
$$\mathbf{v}_i^\top \mathbf{v}_i = 1\,, \qquad \mathbf{v}_i^\top \mathbf{v}_j = 0\,, \qquad \forall\, i,j \in \{1,\ldots,n\} \text{ with } i \neq j\,, \tag{6}$$
such that $\mathbf{v}_i$ is an eigenvector that corresponds to λ_i; $i = 1, \ldots, n$.
- If all eigenvalues $\lambda_1, \ldots, \lambda_n$ differ from each other, then this is an immediate consequence of part 2 of Lemma 1.4.
- As a consequence, the following *diagonalization method* for invertible symmetric matrices is obtained.

Lemma 1.5

- *Let* $\mathbf{A}$ *be an invertible symmetric* $n \times n$ *matrix and let* $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ *be the* $n \times n$ *matrix that consists of the orthonormal eigenvalues* $\mathbf{v}_1, \ldots, \mathbf{v}_n$.
- *Then*
$$\mathbf{V}^\top \mathbf{A} \mathbf{V} = \mathbf{\Lambda}\,, \tag{7}$$
where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ *denotes the* $n \times n$ *diagonal matrix which consists of the eigenvalues* $\lambda_1, \ldots, \lambda_n$.

Proof

- Equation (2) in the definition of eigenvalues and eigenvectors implies that $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ for each $i = 1, \ldots, n$.
- This means that $\mathbf{A}\mathbf{V} = (\lambda_1 \mathbf{v}_1, \ldots, \lambda_n \mathbf{v}_n)$ and with (6) it follows that $\mathbf{V}^\top \mathbf{A} \mathbf{V} = \mathbf{V}^\top (\lambda_1 \mathbf{v}_1, \ldots, \lambda_n \mathbf{v}_n) = \mathbf{\Lambda}$. □

1.1.4 Symmetric and Definite Matrices; Factorization

Lemma 1.6 *Let* $\mathbf{A}$ *be a symmetric and positive definite* $n \times n$ *matrix, i.e.,* $\mathbf{A} = \mathbf{A}^\top$ *and* $\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0$ *for each vector* $\mathbf{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$ *with* $\mathbf{x} \neq \mathbf{o}$. *Then* $\mathbf{A}$ *is invertible and there is an invertible* $n \times n$ *matrix* $\mathbf{H}$ *such that*
$$\mathbf{A} = \mathbf{H}\mathbf{H}^\top\,. \tag{8}$$

Proof We only prove the second part of Lemma 1.6.

- Lemma 1.5 implies that $\mathbf{V}^\top \mathbf{A} \mathbf{V} = \mathbf{\Lambda}$ and
$$\mathbf{A} = (\mathbf{V}^\top)^{-1} \mathbf{\Lambda} \mathbf{V}^{-1}\,, \tag{9}$$
  - where $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ is the $n \times n$ matrix which consists of the orthonormal eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$,
  - and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ denotes the $n \times n$ diagonal matrix which consists of the (positive) eigenvalues $\lambda_1, \ldots, \lambda_n$.
- Now, let $\mathbf{\Lambda}^{1/2}$ be the $n \times n$ diagonal matrix $\mathbf{\Lambda}^{1/2} = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$ and let
$$\mathbf{H} = (\mathbf{V}^\top)^{-1} \mathbf{\Lambda}^{1/2} \mathbf{V}^\top\,. \tag{10}$$
- It is obvious that the matrix $\mathbf{H}$, given in (10), is invertible. Because of $\mathbf{V}^\top \mathbf{V} = \mathbf{I}$ it also holds that
$$\begin{aligned}\mathbf{H}\mathbf{H}^\top &= (\mathbf{V}^\top)^{-1} \mathbf{\Lambda}^{1/2} \mathbf{V}^\top \Big((\mathbf{V}^\top)^{-1} \mathbf{\Lambda}^{1/2} \mathbf{V}^\top \Big)^\top = (\mathbf{V}^\top)^{-1} \mathbf{\Lambda}^{1/2} \mathbf{V}^\top \mathbf{V} \mathbf{\Lambda}^{1/2} \mathbf{V}^{-1} \\ &= (\mathbf{V}^\top)^{-1} \mathbf{\Lambda}^{1/2} \mathbf{\Lambda}^{1/2} \mathbf{V}^{-1} = (\mathbf{V}^\top)^{-1} \mathbf{\Lambda} \mathbf{V}^{-1} = \mathbf{A}\,,\end{aligned}$$
where the last equality follows from (9). □

Remark

- Each invertible $n \times n$ matrix $\mathbf{H}$ with $\mathbf{A} = \mathbf{H}\mathbf{H}^\top$ is called a *square root* of $\mathbf{A}$ and is denoted by $\mathbf{A}^{1/2}$.
- Using the *Cholesky decomposition* for symmetric and positive definite matrices, one can show that there exists a (uniquely determined) lower triangular matrix $\mathbf{H}$ with $\mathbf{A} = \mathbf{H}\mathbf{H}^\top$.

The following property of symmetric matrices is a generalization of Lemma 1.6.

Lemma 1.7 *Let* $\mathbf{A}$ *be a symmetric and positive semidefinite* $n \times n$ *matrix, i.e., it holds that* $\mathbf{A} = \mathbf{A}^\top$ *and* $\mathbf{x}^\top \mathbf{A}\mathbf{x} \ge 0$ *for each vector* $\mathbf{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$*. Now, let* $\mathrm{rk}(\mathbf{A}) = r \; \bigl(\le n\bigr)$*. Then there exists an* $n \times r$ *matrix* $\mathbf{H}$ *with* $\mathrm{rk}(\mathbf{H}) = r$ *such that* $\mathbf{A} = \mathbf{H}\mathbf{H}^\top$*.*

The *proof* of Lemma 1.7 is similar to the proof of Lemma 1.6.

Lemma 1.8

- *Let* $m, r \in \mathbb{N}$ *be arbitrary natural numbers with* $1 \le r \le m$*. Let* $\mathbf{A}$ *be a symmetric and positive definite* $m \times m$ *matrix and let* $\mathbf{B}$ *be an* $r \times m$ *matrix with full rank* $\mathrm{rk}(\mathbf{B}) = r$*.*
- *Then also the matrices* $\mathbf{B}\mathbf{A}\mathbf{B}^\top$ *and* $\mathbf{A}^{-1}$ *are positive definite.*

Proof

- Because of the full rank of $\mathbf{B}^\top$ it holds that $\mathbf{B}^\top\mathbf{x} \ne \mathbf{o}$ for each $\mathbf{x} \in \mathbb{R}^r$ with $\mathbf{x} \ne \mathbf{o}$.
- Since $\mathbf{A}$ is positive definite, it also holds that

$$\mathbf{x}^\top \bigl(\mathbf{B}\mathbf{A}\mathbf{B}^\top\bigr)\mathbf{x} = (\mathbf{B}^\top\mathbf{x})^\top \mathbf{A} (\mathbf{B}^\top\mathbf{x}) > 0$$

  for each $\mathbf{x} \in \mathbb{R}^r$ with $\mathbf{x} \ne \mathbf{o}$, i.e., $\mathbf{B}\mathbf{A}\mathbf{B}^\top$ is positive definite.
- Therefore, we get for $\mathbf{B} = \mathbf{A}^{-1}$ that

$$\mathbf{A}^{-1} = \mathbf{A}^{-1}\bigl(\mathbf{A}\mathbf{A}^{-1}\bigr) = \mathbf{A}^{-1}\mathbf{A}\bigl(\mathbf{A}^{-1}\bigr)^\top$$

  is positive definite. □

1.2 Multivariate Normal Distribution

In this section we recall the notion of a multivariate normal distribution and discuss some fundamental properties of this family of distributions.

1.2.1 Definition and Fundamental Properties

- Let $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ be independent and (identically) normally distributed random variables, i.e.,

$$X_i \sim \mathrm{N}(\mu, \sigma^2), \qquad \forall i = 1, \ldots, n, \tag{11}$$

  where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.
- The assumption of normality in (11) and the independence of the sample variables $X_1, \ldots, X_n$ mean in vector notation that the distribution of the random sample $\mathbf{X} = (X_1, \ldots, X_n)^\top$ is given by

$$\mathbf{X} \sim \mathrm{N}\bigl(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n\bigr), \tag{12}$$

  where $\boldsymbol{\mu} = (\mu, \ldots, \mu)^\top$ and $\mathrm{N}\bigl(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n\bigr)$ denotes the n–dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\sigma^2 \mathbf{I}_n$.
- *Recall* (cf. Section WR-4.3.4): In general, the n–dimensional normal distribution is defined as follows.
  - Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top \in \mathbb{R}^n$ be an arbitrary vector and let $\mathbf{K}$ be a symmetric and positive definite $n \times n$-matrix.

- Let $\mathbf{Z} = (Z_1, \ldots, Z_n)^\top$ be an absolutely continuous random vector, where the joint density of $\mathbf{Z}$ is given by
$$f(\mathbf{z}) = \Bigl(\frac{1}{\sqrt{2\pi}}\Bigr)^n \frac{1}{\sqrt{\det \mathbf{K}}} \exp\Bigl(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^\top \mathbf{K}^{-1}(\mathbf{z}-\boldsymbol{\mu})\Bigr) \tag{13}$$
for each $\mathbf{z} = (z_1, \ldots, z_n)^\top \in \mathbb{R}^n$.
- Then the random vector $\mathbf{Z} = (Z_1, \ldots, Z_n)^\top$ is called (nondegenerate) *normally distributed.*
- Notation: $\mathbf{Z} \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{K})$

Now, we show that the function given in (13) is an (n–dimensional) probability density.

Theorem 1.1 *Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top \in \mathbb{R}^n$ be an arbitrary vector and let $\mathbf{K}$ be a symmetric and positive definite $n \times n$-matrix. Then it holds that*
$$\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\Bigl(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{K}^{-1}(\mathbf{x}-\boldsymbol{\mu})\Bigr)\, dx_1 \ldots dx_n = (2\pi)^{n/2} (\det \mathbf{K})^{1/2}. \tag{14}$$

Proof

- Since $\mathbf{K}$ is symmetric and positive definite (and therefore invertible), Lemma 1.5 implies that there exists an $n \times n$ matrix $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ consisting of the orthogonal eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathbf{K}$, such that
$$\mathbf{V}^\top \mathbf{K} \mathbf{V} = \boldsymbol{\Lambda}, \tag{15}$$
where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ denotes the $n \times n$ diagonal matrix which is built up of the eigenvalues $\lambda_1, \ldots, \lambda_n$ of $\mathbf{K}$.
- Since $\mathbf{K}$ is positive definite, it holds that $\mathbf{v}_i^\top \mathbf{K} \mathbf{v}_i = \lambda_i > 0$ for each $i = 1, \ldots, n$, i.e., all eigenvalues $\lambda_1, \ldots, \lambda_n$ of $\mathbf{K}$ are positive.
- Because of $\mathbf{V}^\top \mathbf{V} = \mathbf{I}$, it holds $\mathbf{V}^\top = \mathbf{V}^{-1}$ and $\mathbf{V}\mathbf{V}^\top = \mathbf{I}$, respectively.
- Due to the fact that $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ and due to (15), it follows that
$$\bigl(\mathbf{V}^\top \mathbf{K} \mathbf{V}\bigr)^{-1} = \mathbf{V}^\top \mathbf{K}^{-1} \mathbf{V} = \mathrm{diag}\bigl(\lambda_1^{-1}, \ldots, \lambda_n^{-1}\bigr).$$
- The mapping $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ with $\mathbf{y} = \varphi(\mathbf{x}) = \mathbf{V}^\top(\mathbf{x} - \boldsymbol{\mu})$, i.e., $\mathbf{x} - \boldsymbol{\mu} = \mathbf{V}\mathbf{y}$, maps $\mathbb{R}^n$ bijectively onto itself and for the Jacobian determinant of $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ it holds that
$$\det\Bigl(\frac{\partial \varphi_i}{\partial x_j}(x_1, \ldots, x_n)\Bigr) = \det \mathbf{V} = \pm 1,$$
where the last equality follows from the fact that $1 = \det(\mathbf{V}^\top \mathbf{V}) = (\det \mathbf{V})^2$.
- Therefore, the integral on the left–hand side of (14) can be written as
$$\begin{aligned}
&\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\Bigl(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{K}^{-1}(\mathbf{x}-\boldsymbol{\mu})\Bigr)\, dx_1 \ldots dx_n \\
&\quad = \int_{\mathbb{R}^n} \exp\Bigl(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{K}^{-1}(\mathbf{x}-\boldsymbol{\mu})\Bigr)\, d(x_1, \ldots, x_n) = \int_{\mathbb{R}^n} \exp\Biggl(-\frac{1}{2}\sum_{i=1}^n \frac{y_i^2}{\lambda_i}\Biggr)\, d(y_1, \ldots, y_n) \\
&\quad = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\Biggl(-\frac{1}{2}\sum_{i=1}^n \frac{y_i^2}{\lambda_i}\Biggr)\, dy_1 \ldots dy_n = \prod_{i=1}^n (2\pi\lambda_i)^{1/2}.
\end{aligned}$$
- This implies (14) since
$$\prod_{i=1}^n \lambda_i = \det \boldsymbol{\Lambda} = \det\bigl(\mathbf{V}^\top \mathbf{K} \mathbf{V}\bigr) = \det\bigl(\mathbf{V}^\top \mathbf{V}\bigr) \det \mathbf{K} = \det \mathbf{K}. \qquad \square$$

1.2.2 Characteristics of the Multivariate Normal Distribution

- Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top \in \mathbb{R}^n$ be an arbitrary vector and let $\mathbf{K} = (k_{ij})$ be a symmetric and positive definite $n \times n$ matrix.
- First, we determine the characteristic function of a normally distributed random vector.
- *Recall*: The *characteristic function* $\varphi : \mathbb{R}^n \to \mathbb{C}$ of an arbitrary n–dimensional random vector $\mathbf{X} = (X_1, \ldots, X_n)^\top : \Omega \to \mathbb{R}^n$ is given by

$$\varphi(\mathbf{t}) = \mathbb{E} \exp\bigl(\mathrm{i}\,\mathbf{t}^\top \mathbf{X}\bigr) = \mathbb{E} \exp\left(\mathrm{i} \sum_{\ell=1}^{n} t_\ell X_\ell\right), \qquad \forall\, \mathbf{t} = (t_1, \ldots, t_n)^\top \in \mathbb{R}^n . \tag{16}$$

Theorem 1.2

- *Let the random vector* $\mathbf{X} = (X_1, \ldots, X_n)^\top : \Omega \to \mathbb{R}^n$ *be normally distributed with* $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{K})$.
- *Then the characteristic function* $\varphi : \mathbb{R}^n \to \mathbb{C}$ *of* $\mathbf{X}$ *fulfills*

$$\varphi(\mathbf{t}) = \exp\Bigl(\mathrm{i}\,\mathbf{t}^\top \boldsymbol{\mu} - \frac{1}{2}\,\mathbf{t}^\top \mathbf{K} \mathbf{t}\Bigr), \qquad \forall\, \mathbf{t} \in \mathbb{R}^n . \tag{17}$$

Proof

- Equations (13) and (16) imply

$$\begin{aligned}
\varphi(\mathbf{t}) &= \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\left(\mathrm{i} \sum_{\ell=1}^{n} t_\ell x_\ell\right) f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n \\
&= \frac{1}{(2\pi)^{n/2} (\det \mathbf{K})^{1/2}} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\Bigl(\mathrm{i}\,\mathbf{t}^\top \mathbf{x} - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{K}^{-1} (\mathbf{x} - \boldsymbol{\mu})\Bigr)\, dx_1 \ldots dx_n \\
&= \frac{\exp(\mathrm{i}\,\mathbf{t}^\top \boldsymbol{\mu})}{(2\pi)^{n/2} (\det \mathbf{K})^{1/2}} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\Bigl(\mathrm{i}\,\mathbf{t}^\top \mathbf{y} - \frac{1}{2}\,\mathbf{y}^\top \mathbf{K}^{-1} \mathbf{y}\Bigr)\, dy_1 \ldots dy_n ,
\end{aligned}$$

  where the last equality holds due to the substitution $\mathbf{y} = \mathbf{x} - \boldsymbol{\mu}$, for which the matrix of the partial derivatives is the identity matrix and therefore the Jacobian determinant is equal to 1.
- Similar to the proof of Theorem 1.1 it follows with the help of the substitutions $\mathbf{y} = \mathbf{V}\mathbf{x}$ and $\mathbf{t} = \mathbf{V}\mathbf{s}$ that

$$\begin{aligned}
\varphi(\mathbf{t}) &= \frac{\exp(\mathrm{i}\,\mathbf{t}^\top \boldsymbol{\mu})}{(2\pi)^{n/2} (\det \mathbf{K})^{1/2}} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\Bigl(\mathrm{i}\,\mathbf{s}^\top \mathbf{x} - \frac{1}{2}\,\mathbf{x}^\top \mathbf{V}^\top \mathbf{K}^{-1} \mathbf{V} \mathbf{x}\Bigr)\, dx_1 \ldots dx_n \\
&= \frac{\exp(\mathrm{i}\,\mathbf{t}^\top \boldsymbol{\mu})}{(2\pi)^{n/2} (\det \mathbf{K})^{1/2}} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\Bigl(\sum_{\ell=1}^{n} \Bigl(\mathrm{i}\, s_\ell x_\ell - \frac{x_\ell^2}{2\lambda_\ell}\Bigr)\Bigr)\, dx_1 \ldots dx_n
\end{aligned}$$

  and thus

$$\begin{aligned}
\varphi(\mathbf{t}) &= \frac{\exp(\mathrm{i}\,\mathbf{t}^\top \boldsymbol{\mu})}{(2\pi)^{n/2} (\det \mathbf{K})^{1/2}} \prod_{\ell=1}^{n} \int_{-\infty}^{\infty} \exp\Bigl(\mathrm{i}\, s_\ell x_\ell - \frac{x_\ell^2}{2\lambda_\ell}\Bigr)\, dx_\ell \\
&= \exp(\mathrm{i}\,\mathbf{t}^\top \boldsymbol{\mu}) \prod_{\ell=1}^{n} \frac{1}{\sqrt{2\pi\lambda_\ell}} \int_{-\infty}^{\infty} \exp\Bigl(\mathrm{i}\, s_\ell x_\ell - \frac{x_\ell^2}{2\lambda_\ell}\Bigr)\, dx_\ell ,
\end{aligned}$$

  where the matrix $\mathbf{V}$ consists of the orthonormal eigenvectors of $\mathbf{K}$ and $\lambda_1, \ldots, \lambda_n > 0$ are the eigenvalues of $\mathbf{K}$ with $\det \mathbf{K} = \lambda_1 \cdot \ldots \cdot \lambda_n$.

- Now, it is sufficient to consider that $\varphi_\ell : \mathbb{R} \to \mathbb{C}$ with

$$\varphi_\ell(s) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\lambda_\ell}} \exp\Bigl(\mathrm{i}\, sx - \frac{x^2}{2\lambda_\ell}\Bigr)\, dx$$

  is the characteristic function of the (one dimensional) $\mathrm{N}(0, \lambda_\ell)$–distribution.
- In Section WR-5.3.3 we already have seen that $\varphi_\ell(s) = \exp\bigl(-\lambda_\ell s^2/2\bigr)$.
- Hence, we get

$$\begin{aligned} \varphi(\mathbf{t}) &= \exp(\mathrm{i}\,\mathbf{t}^\top \boldsymbol{\mu}) \prod_{\ell=1}^{n} \exp\Bigl(-\frac{\lambda_\ell s_\ell^2}{2}\Bigr) = \exp(\mathrm{i}\,\mathbf{t}^\top \boldsymbol{\mu})\, \exp\Bigl(-\frac{\sum_{\ell=1}^{n} \lambda_\ell s_\ell^2}{2}\Bigr) \\ &= \exp(\mathrm{i}\,\mathbf{t}^\top \boldsymbol{\mu})\, \exp\Bigl(-\frac{\mathbf{t}^\top \mathbf{K} \mathbf{t}}{2}\Bigr). \end{aligned}$$

□

Using (17) for the characteristic function we are able to determine expectation and covariance matrix of a normally distributed random vector.

Corollary 1.1 *If $\mathbf{X} = (X_1, \ldots, X_n)^\top \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{K})$, it holds for arbitrary $i, j = 1, \ldots, n$ that*

$$\mathbb{E}\, X_i = \mu_i\,, \qquad \textit{and} \qquad \mathrm{Cov}\,(X_i, X_j) = k_{ij}\,. \tag{18}$$

Proof

- From (17) it follows that

$$\frac{\partial\, \varphi(\mathbf{t})}{\partial t_i} = \Bigl(\mathrm{i}\, \mu_i - \sum_{\ell=1}^{n} k_{i\ell} t_\ell\Bigr) \varphi(\mathbf{t}) \tag{19}$$

  and

$$\frac{\partial^2 \varphi(\mathbf{t})}{\partial t_i \partial t_j} = -\, k_{ij} \varphi(\mathbf{t}) + \Bigl(\mathrm{i}\, \mu_i - \sum_{\ell=1}^{n} k_{i\ell} t_\ell\Bigr)\Bigl(\mathrm{i}\, \mu_j - \sum_{\ell=1}^{n} k_{j\ell} t_\ell\Bigr) \varphi(\mathbf{t})\,. \tag{20}$$

- It is easy to see that

$$\mathbb{E}\, X_i = \mathrm{i}^{-1}\, \frac{\partial\, \varphi(\mathbf{t})}{\partial t_i}\, \Big|_{\mathbf{t}=\mathbf{o}}\,.$$

  Because of $\varphi(\mathbf{o}) = 1$ and (19), it follows that $\mathbb{E}\, X_i = \mu_i$.
- Furthermore,

$$\mathbb{E}\,(X_i X_j) = -\, \frac{\partial^2 \varphi(\mathbf{t})}{\partial t_i \partial t_j}\, \Big|_{\mathbf{t}=\mathbf{o}}\,.$$

  This equation and (20) imply $\mathrm{Cov}\,(X_i, X_j) = k_{ij}$. □

Remark

- In Theorem WR-4.14 we have shown that the covariance matrix $\mathbf{K} = \mathbf{K}_{\mathbf{X}}$ of an arbitrary random vector $\mathbf{X} = (X_1, \ldots, X_n)^\top$ is always symmetric and positive semidefinite.
- In (13), where the density of the nondegenerate multivariate normal distribution is defined, it is additionally required that the covariance matrix $\mathbf{K}$ is positive definite.
- Here, $\mathbf{K}$ being positive definite is not only sufficient but also necessary to ensure that the matrix $\mathbf{K}$ is invertible, i.e., $\det \mathbf{K} \neq 0$ or $\mathbf{K}$ has full rank.

1.2.3 Marginal Distributions and Independence of Subvectors; Convolution Properties

- In this section it is shown how to derive further interesting properties of the multivariate normal distribution using Theorem 1.2.
- For this purpose we need a *vectorial version* of the uniqueness theorem for characteristic functions (cf. Corollary WR-5.5), which we will state without proof.

Lemma 1.9 *Let $\mathbf{X}, \mathbf{Y} : \Omega \to \mathbb{R}^n$ be arbitrary random vectors; $\mathbf{X} = (X_1, \ldots, X_n)^\top$, $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$. Then it holds that*

$$\mathbf{X} \stackrel{\mathrm{d}}{=} \mathbf{Y} \qquad \text{if and only if} \qquad \varphi_{\mathbf{X}}(\mathbf{t}) = \varphi_{\mathbf{Y}}(\mathbf{t}) \qquad \forall\, \mathbf{t} = (t_1, \ldots, t_n)^\top \in \mathbb{R}^n \,, \tag{21}$$

where

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}\, \exp\Bigl(\mathrm{i} \sum_{j=1}^n t_j X_j\Bigr) \qquad \text{and} \qquad \varphi_{\mathbf{Y}}(\mathbf{t}) = \mathbb{E}\, \exp\Bigl(\mathrm{i} \sum_{j=1}^n t_j Y_j\Bigr)$$

are the characteristic functions of $\mathbf{X}$ and $\mathbf{Y}$, respectively.

First, we show that arbitrary subvectors of normally distributed random vectors are also normally distributed.

- We assume $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top \in \mathbb{R}^n$ to be an arbitrary vector and $\mathbf{K} = (k_{ij})$ to be a symmetric and positive definite $n \times n$-matrix.
- It is obvious that the random vector $(X_{\pi_1}, \ldots, X_{\pi_n})^\top$ is normally distributed for each permutation $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)^\top$ of the natural numbers $1, \ldots, n$ if $\mathbf{X} = (X_1, \ldots, X_n)^\top$ is normally distributed.
- Therefore, we can w.l.o.g restrict the examination of the distribution of subvectors of normally distributed random vectors to the examination of the first components.

Corollary 1.2 *Let $\mathbf{X} = (X_1, \ldots, X_n)^\top \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{K})$, where $\mathbf{K}$ is positive definite. Then it holds that*

$$(X_1, \ldots, X_m)^\top \sim \mathrm{N}(\boldsymbol{\mu}_m, \mathbf{K}_m) \qquad \forall\, m = 1, \ldots, n \,,$$

where $\boldsymbol{\mu}_m = (\mu_1, \ldots, \mu_m)^\top$ and $\mathbf{K}_m$ denotes the $m \times m$ matrix which consists of the first m rows and columns of $\mathbf{K}$.

Proof

- Let $\varphi : \mathbb{R}^n \to \mathbb{C}$ be the characteristic function of $(X_1, \ldots, X_n)^\top$.
- Now, the characteristic function $\varphi_m : \mathbb{R}^m \to \mathbb{C}$ of $(X_1, \ldots, X_m)^\top$ fulfills

$$\varphi_m(\mathbf{t}_m) = \varphi\Bigl((\mathbf{t}_m, \underbrace{0, \ldots, 0}_{n-m})\Bigr)\,, \qquad \forall\, \mathbf{t}_m = (t_1, \ldots, t_m))^\top \in \mathbb{R}^m \,.$$

- This result and (17) imply that

$$\varphi_m(\mathbf{t}_m) = \exp\Bigl(\mathrm{i}\, \mathbf{t}_m^\top \boldsymbol{\mu}_m - \frac{1}{2}\, \mathbf{t}_m^\top \mathbf{K}_m \mathbf{t}_m\Bigr)\,, \qquad \forall\, \mathbf{t}_m \in \mathbb{R}^m \,.$$

- Since $\mathbf{K}$ is symmetric and positive definite, we know that also the $m \times m$ matrix $\mathbf{K}_m$ is symmetric and positive definite. From this fact and from Theorem 1.2 it follows that the characteristic function of the subvector $(X_1, \ldots, X_m)^\top$ is identical with the characteristic function of the $\mathrm{N}(\boldsymbol{\mu}_m, \mathbf{K}_m)$–distribution.
- The statement follows because of the one-to-one correspondence of characteristic functions and distributions of random vectors (cf. Lemma 1.9). □

There is a simple criterion for two subvectors $(X_1, \ldots, X_m)^\top$ and $(X_{m+1}, \ldots, X_n)^\top$, with $1 \le m < n$, of the normally distributed random vector $\mathbf{X} = (X_1, \ldots, X_n)^\top$ being independent.

Corollary 1.3 *Let $\mathbf{X} = (X_1, \ldots, X_n)^\top$ be a normally distributed random vector with $\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{K})$; $\mathbf{K} = (k_{ij})$. The subvectors $(X_1, \ldots, X_m)^\top$ and $(X_{m+1}, \ldots, X_n)^\top$ are independent if and only if $k_{ij} = 0$ for arbitrary $i \in \{1, \ldots, m\}$ and $j \in \{m+1, \ldots, n\}$.*

Proof

- If the subvectors $(X_1, \ldots, X_m)^\top$ and $(X_{m+1}, \ldots, X_n)^\top$ are independent, then the (one–dimensional) random variables X_i and X_j are independent for arbitrary $i \in \{1, \ldots, m\}$ and $j \in \{m+1, \ldots, n\}$.
- Thus, it holds that $\mathrm{Cov}\,(X_i, X_j) = 0$ and Corollary 1.1 implies that $k_{ij} = 0$.
- Let us now assume that $k_{ij} = 0$ for arbitrary $i \in \{1, \ldots, m\}$ and $j \in \{m+1, \ldots, n\}$.
- Then Theorem 1.2 implies that the characteristic function $\varphi(\mathbf{t})$ of $\mathbf{X} = (X_1, \ldots, X_n)^\top$ has the following factorization.
- For each $\mathbf{t} = (t_1, \ldots, t_n)^\top \in \mathbb{R}^n$ it holds that

$$\begin{aligned}\varphi(\mathbf{t}) &= \exp\Bigl(\mathrm{i}\,\mathbf{t}^\top \boldsymbol{\mu} - \frac{1}{2}\mathbf{t}^\top \mathbf{K}\mathbf{t}\Bigr) = \exp\Bigl(\mathrm{i}\sum_{i=1}^n t_i\mu_i - \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n t_i k_{ij} t_j\Bigr)\\ &= \exp\Bigl(\mathrm{i}\sum_{i=1}^m t_i\mu_i - \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m t_i k_{ij} t_j\Bigr)\,\exp\Bigl(\mathrm{i}\sum_{i=m+1}^n t_i\mu_i - \frac{1}{2}\sum_{i=m+1}^n\sum_{j=m+1}^n t_i k_{ij} t_j\Bigr),\end{aligned}$$

  where the factors of the last term are the characteristic functions of $(X_1, \ldots, X_m)^\top$ and $(X_{m+1}, \ldots, X_n)^\top$.
- The statement follows because of the one-to-one correspondence of characteristic functions and distributions of random vectors (cf. Lemma 1.9). □

Remark

- Finally, we show that the family of multivariate normal distributions is closed under convolution. In the following we call this property briefly "convolution stability" of the multivariate normal distribution. In Corollary WR-3.2 we already have proved the convolution stability of one–dimensional normal distributions.
- The following formula for the characteristic function of sums of independent random vectors is useful in this context. The proof is analog to the proof of the one–dimensional case (cf. Theorem WR-5.18).

Lemma 1.10 *Let $\mathbf{Z}_1, \mathbf{Z}_2 : \Omega \to \mathbb{R}^n$ be independent random vectors. The characteristic function $\varphi_{\mathbf{Z}_1+\mathbf{Z}_2} : \mathbb{R}^n \to \mathbb{C}$ of the sum $\mathbf{Z}_1 + \mathbf{Z}_2$ can then be written as*

$$\varphi_{\mathbf{Z}_1+\mathbf{Z}_2}(\mathbf{t}) = \varphi_{\mathbf{Z}_1}(\mathbf{t})\,\varphi_{\mathbf{Z}_2}(\mathbf{t})\,, \qquad \forall\,\mathbf{t} \in \mathbb{R}^n\,, \tag{22}$$

where $\varphi_{\mathbf{Z}_i}$ denotes the characteristic function of $\mathbf{Z}_i$; $i = 1, 2$.

The following statement is called *convolution stability* of the multivariate normal distribution.

Corollary 1.4 *Let $\mathbf{Z}_1, \mathbf{Z}_2 : \Omega \to \mathbb{R}^n$ be independent random vectors with $\mathbf{Z}_i \sim \mathrm{N}(\boldsymbol{\mu}_i, \mathbf{K}_i)$ for $i = 1, 2$. Then it holds that $\mathbf{Z}_1 + \mathbf{Z}_2 \sim \mathrm{N}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2, \mathbf{K}_1 + \mathbf{K}_2)$.*

Proof

- Equations (17) and (22) imply that

$$\begin{aligned}\varphi_{\mathbf{Z}_1+\mathbf{Z}_2}(\mathbf{t}) &= \varphi_{\mathbf{Z}_1}(\mathbf{t})\,\varphi_{\mathbf{Z}_2}(\mathbf{t})\\ &= \exp\Bigl(\mathrm{i}\,\mathbf{t}^\top\boldsymbol{\mu}_1-\frac{1}{2}\,\mathbf{t}^\top\mathbf{K}_1\mathbf{t}\Bigr)\;\exp\Bigl(\mathrm{i}\,\mathbf{t}^\top\boldsymbol{\mu}_2-\frac{1}{2}\,\mathbf{t}^\top\mathbf{K}_2\mathbf{t}\Bigr)\\ &= \exp\Bigl(\mathrm{i}\,\mathbf{t}^\top(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)-\frac{1}{2}\,\mathbf{t}^\top(\mathbf{K}_1+\mathbf{K}_2)\mathbf{t}\Bigr)\,.\end{aligned}$$

- This result and the uniqueness theorem for characteristic functions (cf. Lemma 1.9) imply the statement. □

1.2.4 Linear Transformation of Normally Distributed Random Vectors

Now, we show that the linear transformation of a normally distributed random vector again is a normally distributed random vector.

Theorem 1.3

- *Let* $\mathbf{Y}\sim \mathrm{N}(\boldsymbol{\mu},\mathbf{K})$ *be an* n*–dimensional normally distributed random vector with mean vector* $\boldsymbol{\mu}\in\mathbb{R}^n$ *and (positive definite) covariance matrix* $\mathbf{K}$.
- *Moreover, let* $m\le n$*, let* $\mathbf{A}$ *be an arbitrary* $m\times n$ *matrix having full rank* $\mathrm{rk}(\mathbf{A})=m$ *and let* $\mathbf{c}\in\mathbb{R}^m$ *be an arbitrary* m*–dimensional vector.*
- *Then it holds that* $\mathbf{Z}=\mathbf{A}\mathbf{Y}+\mathbf{c}$ *is an (*m*–dimensional) normally distributed random vector with*

$$\mathbf{Z}\sim \mathrm{N}(\mathbf{A}\boldsymbol{\mu}+\mathbf{c},\,\mathbf{A}\mathbf{K}\mathbf{A}^\top)\,. \tag{23}$$

Proof

- For each $\mathbf{a}\in\mathbb{R}^m$ it holds that

$$\varphi_{\mathbf{Z}}(\mathbf{t})=\exp(\mathrm{i}\,\mathbf{t}^\top\mathbf{a})\varphi_{\mathbf{Z}-\mathbf{a}}(\mathbf{t})\,,\qquad \forall\,\mathbf{t}\in\mathbb{R}^m\,.$$

- From (17) derived in Theorem 1.2 and from the uniqueness theorem for the characteristic function of normally distributed random vectors it follows that

$$\mathbf{Z}\sim \mathrm{N}(\mathbf{A}\boldsymbol{\mu}+\mathbf{c},\,\mathbf{A}\mathbf{K}\mathbf{A}^\top)\quad\text{if and only if}\quad \mathbf{Z}-(\mathbf{A}\boldsymbol{\mu}+\mathbf{c})\sim \mathrm{N}(\mathbf{o},\,\mathbf{A}\mathbf{K}\mathbf{A}^\top)\,.$$

- Therefore, we will w.l.o.g. assume that $\mathbf{Y}\sim \mathrm{N}(\mathbf{o},\mathbf{K})$ and $\mathbf{c}=\mathbf{o}$.
- Then the characteristic function $\varphi_{\mathbf{Z}}(\mathbf{t})$ of $\mathbf{Z}=\mathbf{A}\mathbf{Y}$ fulfills

$$\begin{aligned}\varphi_{\mathbf{Z}}(\mathbf{t}) &= \mathbb{E}\,e^{\mathrm{i}\,\mathbf{t}^\top\mathbf{Z}}\\ &= \mathbb{E}\,e^{\mathrm{i}\,\mathbf{t}^\top\mathbf{A}\mathbf{Y}}=\mathbb{E}\,e^{\mathrm{i}\,(\mathbf{A}^\top\mathbf{t})^\top\mathbf{Y}}\\ &= \varphi_{\mathbf{Y}}(\mathbf{A}^\top\mathbf{t})\,,\end{aligned}$$

  for each $\mathbf{t}\in\mathbb{R}^m$, where $\varphi_{\mathbf{Y}}(\mathbf{A}^\top\mathbf{t})$ denotes the value of the characteristic function of the normally distributed random vector $\mathbf{Y}$ at $\mathbf{A}^\top\mathbf{t}\in\mathbb{R}^n$.

- Now, formula (17) for the characteristic function of normally distributed random vectors implies

$$\begin{aligned}\varphi_{\mathbf{Z}}(\mathbf{t}) &= \varphi_{\mathbf{Y}}(\mathbf{A}^\top\mathbf{t}) \\ &= \exp\Bigl(-\frac{1}{2}\,(\mathbf{A}^\top\mathbf{t})^\top\mathbf{K}(\mathbf{A}^\top\mathbf{t})\Bigr) \\ &= \exp\Bigl(-\frac{1}{2}\,\mathbf{t}^\top(\mathbf{A}\mathbf{K}\mathbf{A}^\top)\mathbf{t}\Bigr)\,.\end{aligned}$$

- In other words: The characteristic function of $\mathbf{Z}$ is equal to the characteristic function of $\mathrm{N}(\mathbf{o},\,\mathbf{A}\mathbf{K}\mathbf{A}^\top)$.
- The uniqueness theorem for characteristic functions of random vectors implies $\mathbf{Z}\sim\mathrm{N}(\mathbf{o},\,\mathbf{A}\mathbf{K}\mathbf{A}^\top)$. □

By using Theorem 1.3 it follows in particular that it is possible to create normally distributed random vectors by a linear transformation of vectors whose components are independent $\mathrm{N}(0,1)$–distributed random variables.

Corollary 1.5

- *Let $Y_1,\ldots,Y_n:\Omega\to\mathbb{R}$ be independent random variables with $Y_i\sim\mathrm{N}(0,1)$ for each $i=1,\ldots,n$, i.e., $\mathbf{Y}=(Y_1,\ldots,Y_n)^\top\sim\mathrm{N}(\mathbf{o},\mathbf{I})$.*
- *Let $\mathbf{K}$ be a symmetric and positive definite $n\times n$ matrix and let $\boldsymbol{\mu}\in\mathbb{R}^n$.*
- *Then the random vector $\mathbf{Z}=\mathbf{K}^{1/2}\mathbf{Y}+\boldsymbol{\mu}$ satisfies $\mathbf{Z}\sim\mathrm{N}(\boldsymbol{\mu},\,\mathbf{K})$, where $\mathbf{K}^{1/2}$ is the square root of $\mathbf{K}$.*

Proof

- With the help of Theorem 1.3 it follows that

$$\mathbf{Z}\sim\mathrm{N}\bigl(\boldsymbol{\mu},\,\mathbf{K}^{1/2}\bigl(\mathbf{K}^{1/2}\bigr)^\top\bigr)\,.$$

- Now, this result and Lemma 1.6 imply the statement. □

1.2.5 Degenerate Multivariate Normal Distribution

In the following, we will give a generalization of the notion of (nondegenerate) multivariate normal distributions, which was introduced in Section 1.2.1.

- A factorization property of covariance matrices which has already been mentioned in Lemma 1.7 is useful in this context.
- *Recall*: Let $\mathbf{K}$ be a symmetric and positive semidefinite $n\times n$ matrix with $\operatorname{rk}(\mathbf{K})=r\le n$. Then there is an $n\times r$ matrix $\mathbf{B}$ with $\operatorname{rk}(\mathbf{B})=r$ such that

$$\mathbf{K}=\mathbf{B}\mathbf{B}^\top\,. \tag{24}$$

Definition

- Let $\mathbf{Y}$ be an n–dimensional random vector with mean vector $\boldsymbol{\mu}=\mathbb{E}\,\mathbf{Y}$ and covariance matrix $\mathbf{K}=\operatorname{Cov}(\mathbf{Y})$ such that $\operatorname{rk}(\mathbf{K})=r$ with $r\le n$.
- Then $\mathbf{Y}$ is called normally distributed if $\mathbf{Y}\overset{\mathrm{d}}{=}\boldsymbol{\mu}+\mathbf{B}\mathbf{Z}$, where $\mathbf{B}$ is an $n\times r$ matrix with $\operatorname{rk}(\mathbf{B})=r$ fulfilling (24) and where $\mathbf{Z}$ is an r–dimensional random vector with $\mathbf{Z}\sim\mathrm{N}(\mathbf{o},\mathbf{I}_r)$.

- We say that $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{K})$ follows a *degenerate normal distribution* if $\mathrm{rk}(\mathbf{K}) < n$. (Notation: $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{K})$)

Remark

- If $\mathrm{rk}(\mathbf{K}) = r < n$, then the random vector $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{K})$ is *not* absolutely continuous
  - because the values of $\mathbf{Y} \overset{\mathrm{d}}{=} \boldsymbol{\mu} + \mathbf{B}\mathbf{Z}$ are almost surely (with probability 1) elements of the r–dimensional subset $\{\boldsymbol{\mu} + \mathbf{B}\mathbf{x} : \mathbf{x} \in \mathbb{R}^r\}$ of $\mathbb{R}^n$,
  - i.e., the distribution of $\mathbf{Y}$ has no density with respect to the n–dimensional Lebesgue measure.
  - An example for this is the random vector $\mathbf{Y} = (Z, Z)^\top = \mathbf{B}Z$ with $Z \sim \mathrm{N}(0, \sigma^2)$ and $\mathbf{B} = (1, 1)^\top$, which only takes values on the diagonal $\{(z_1, z_2) \in \mathbb{R}^2 : z_1 = z_2\}$.
- The distribution of the random vector $\boldsymbol{\mu} + \mathbf{B}\mathbf{Z}$ does *not* depend on the choice of matrix $\mathbf{B}$ of the factorization (24).
- This is an immediate consequence of both of the following criteria for (degenerate and nondegenerate) multivariate normal distributions.

Theorem 1.4

- *Let $\mathbf{Y}$ be an n–dimensional random vector with mean vector $\boldsymbol{\mu} = \mathbb{E}\,\mathbf{Y}$ and covariance matrix $\mathbf{K} = \mathrm{Cov}\,(\mathbf{Y})$ such that $\mathrm{rk}(\mathbf{K}) = r$ with $r \le n$.*
- *The random vector $\mathbf{Y}$ is normally distributed if and only if one of the following conditions is fulfilled:*
  1. *The characteristic function $\varphi(\mathbf{t}) = \mathbb{E}\,\exp\Bigl(\mathrm{i}\sum_{j=1}^n t_j Y_j\Bigr)$ of $\mathbf{Y}$ is given by*
  $$\varphi(\mathbf{t}) = \exp\Bigl(\mathrm{i}\,\mathbf{t}^\top\boldsymbol{\mu} - \frac{1}{2}\,\mathbf{t}^\top\mathbf{K}\mathbf{t}\Bigr), \qquad \forall\,\mathbf{t} = (t_1, \ldots, t_n)^\top \in \mathbb{R}^n . \tag{25}$$
  2. *The linear function $\mathbf{c}^\top\mathbf{Y}$ of $\mathbf{Y}$ is normally distributed for each $\mathbf{c} \in \mathbb{R}^n$ with $\mathbf{c} \ne \mathbf{o}$ and*
  $$\mathbf{c}^\top\mathbf{Y} \sim \mathrm{N}(\mathbf{c}^\top\boldsymbol{\mu}, \mathbf{c}^\top\mathbf{K}\mathbf{c}) .$$

The *proof* of Theorem 1.4 is omitted (and left as an exercise).

1.3 Linear and Quadratic Forms of Normally Distributed Random Vectors

1.3.1 Definition, Expectation and Covariance

Definition

- Let $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ and $\mathbf{Z} = (Z_1, \ldots, Z_n)^\top$ be arbitrary n–dimensional random vectors and let $\mathbf{A}$ be a symmetric $n \times n$ matrix with real–valued entries.
- Then the (real–valued) random variable $\mathbf{Y}^\top\mathbf{A}\mathbf{Y} : \Omega \to \mathbb{R}$ is called a *quadratic form* of $\mathbf{Y}$ (with respect to $\mathbf{A}$).
- The random variable $\mathbf{Y}^\top\mathbf{A}\mathbf{Z} : \Omega \to \mathbb{R}$ is called a *bilinear form* of $\mathbf{Y}$ and $\mathbf{Z}$ (with respect to $\mathbf{A}$).

First, we derive the expectation of quadratic or bilinear forms.

Theorem 1.5 *Let $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ and $\mathbf{Z} = (Z_1, \ldots, Z_n)^\top$ be arbitrary n–dimensional random vectors and let $\mathbf{A}$ be a symmetric $n \times n$ matrix with real–valued entries. Furthermore, let the mean vectors $\boldsymbol{\mu}_\mathbf{Y} = \mathbb{E}\,\mathbf{Y}$ and $\boldsymbol{\mu}_\mathbf{Z} = \mathbb{E}\,\mathbf{Z}$ as well as the covariance matrices $\mathbf{K}_{\mathbf{YY}} = \bigl(\mathrm{Cov}\,(Y_i, Y_j)\bigr)$ and $\mathbf{K}_{\mathbf{ZY}} = \bigl(\mathrm{Cov}\,(Z_i, Y_j)\bigr)$ be well–defined. Then it holds that*

$$\mathbb{E}\,\bigl(\mathbf{Y}^\top\mathbf{A}\mathbf{Y}\bigr) = \mathrm{tr}(\mathbf{A}\mathbf{K}_{\mathbf{YY}}) + \boldsymbol{\mu}_\mathbf{Y}^\top\mathbf{A}\boldsymbol{\mu}_\mathbf{Y} \qquad \textit{and} \qquad \mathbb{E}\,\bigl(\mathbf{Y}^\top\mathbf{A}\mathbf{Z}\bigr) = \mathrm{tr}(\mathbf{A}\mathbf{K}_{\mathbf{ZY}}) + \boldsymbol{\mu}_\mathbf{Y}^\top\mathbf{A}\boldsymbol{\mu}_\mathbf{Z} . \tag{26}$$

Proof

- We only prove the second formula in (26) since the first formula follows as a special case for $\mathbf{Z} = \mathbf{Y}$.
- It obviously holds that $\mathbf{Y}^\top \mathbf{A}\mathbf{Z} = \operatorname{tr}(\mathbf{Y}^\top \mathbf{A}\mathbf{Z})$. Moreover, from Lemma 1.1 it follows that $\operatorname{tr}(\mathbf{Y}^\top \mathbf{A}\mathbf{Z}) = \operatorname{tr}(\mathbf{A}\mathbf{Z}\mathbf{Y}^\top)$.
- Altogether we get

$$\begin{aligned}\mathbb{E}\left(\mathbf{Y}^\top \mathbf{A}\mathbf{Z}\right) &= \mathbb{E}\operatorname{tr}(\mathbf{Y}^\top \mathbf{A}\mathbf{Z}) = \mathbb{E}\operatorname{tr}(\mathbf{A}\mathbf{Z}\mathbf{Y}^\top) = \operatorname{tr}(\mathbf{A}\mathbb{E}\left(\mathbf{Z}\mathbf{Y}^\top\right)) \\ &= \operatorname{tr}(\mathbf{A}(\mathbf{K}_{\mathbf{ZY}} + \boldsymbol{\mu}_{\mathbf{Z}}\boldsymbol{\mu}_{\mathbf{Y}}^\top)) = \operatorname{tr}(\mathbf{A}\mathbf{K}_{\mathbf{ZY}}) + \boldsymbol{\mu}_{\mathbf{Y}}^\top \mathbf{A}\boldsymbol{\mu}_{\mathbf{Z}}\,.\end{aligned}$$

□

In a similar way it is possible to derive a formula for the covariance of quadratic forms of normally distributed random vectors. The following formulas for the third and fourth mixed moments of the components of centered normally distributed random vectors are useful in this context.

Lemma 1.11 *Let* $\mathbf{Z} = (Z_1, \ldots, Z_n)^\top \sim \mathrm{N}(\mathbf{o}, \mathbf{K})$ *be a normally distributed random vector with mean vector* $\boldsymbol{\mu} = \mathbf{o}$ *and with an arbitrary covariance matrix* $\mathbf{K} = (k_{ij})$*. Then it holds that*

$$\mathbb{E}\left(Z_i Z_j Z_\ell\right) = 0 \quad \textit{and} \quad \mathbb{E}\left(Z_i Z_j Z_\ell Z_m\right) = k_{ij}k_{\ell m} + k_{i\ell}k_{jm} + k_{j\ell}k_{im} \qquad \forall\, i, j, \ell, m \in \{1, \ldots, n\}\,. \tag{27}$$

The *proof* of Lemma 1.11 is omitted. It is an immediate consequence of Theorems 1.2 and 1.4, cf. the proof of Corollary 1.1.

Theorem 1.6

- *Let* $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ *be an* n*–dimensional random vector with* $\mathbf{Y} \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{K})$ *and let* $\mathbf{A} = (a_{ij})$, $\mathbf{B} = (b_{ij})$ *be arbitrary symmetric* $n \times n$ *matrices.*
- *Then*

$$\operatorname{Cov}\left(\mathbf{Y}^\top \mathbf{A}\mathbf{Y},\, \mathbf{Y}^\top \mathbf{B}\mathbf{Y}\right) = 2\operatorname{tr}(\mathbf{A}\mathbf{K}\mathbf{B}\mathbf{K}) + 4\boldsymbol{\mu}^\top \mathbf{A}\mathbf{K}\mathbf{B}\boldsymbol{\mu}\,. \tag{28}$$

- *In particular, it holds that* $\operatorname{Var}\left(\mathbf{Y}^\top \mathbf{A}\mathbf{Y}\right) = 2\operatorname{tr}\left((\mathbf{A}\mathbf{K})^2\right) + 4\boldsymbol{\mu}^\top \mathbf{A}\mathbf{K}\mathbf{A}\boldsymbol{\mu}$*.*

Proof

- From the definition of the covariance and from Theorem 1.5 it follows that

$$\begin{aligned}\operatorname{Cov}\left(\mathbf{Y}^\top \mathbf{A}\mathbf{Y},\, \mathbf{Y}^\top \mathbf{B}\mathbf{Y}\right) &= \mathbb{E}\left((\mathbf{Y}^\top \mathbf{A}\mathbf{Y} - \mathbb{E}\left(\mathbf{Y}^\top \mathbf{A}\mathbf{Y}\right))(\mathbf{Y}^\top \mathbf{B}\mathbf{Y} - \mathbb{E}\left(\mathbf{Y}^\top \mathbf{B}\mathbf{Y}\right))\right) \\ &= \mathbb{E}\left((\mathbf{Y}^\top \mathbf{A}\mathbf{Y} - \operatorname{tr}(\mathbf{A}\mathbf{K}) - \boldsymbol{\mu}^\top \mathbf{A}\boldsymbol{\mu})(\mathbf{Y}^\top \mathbf{B}\mathbf{Y} - \operatorname{tr}(\mathbf{B}\mathbf{K}) - \boldsymbol{\mu}^\top \mathbf{B}\boldsymbol{\mu})\right).\end{aligned}$$

- With the substitution $\mathbf{Z} = \mathbf{Y} - \boldsymbol{\mu}$ or $\mathbf{Y} = \mathbf{Z} + \boldsymbol{\mu}$ it follows that

$$\begin{aligned}\operatorname{Cov}\left(\mathbf{Y}^\top \mathbf{A}\mathbf{Y},\, \mathbf{Y}^\top \mathbf{B}\mathbf{Y}\right) &= \mathbb{E}\left((\mathbf{Z}^\top \mathbf{A}\mathbf{Z} + 2\boldsymbol{\mu}^\top \mathbf{A}\mathbf{Z} - \operatorname{tr}(\mathbf{A}\mathbf{K}))(\mathbf{Z}^\top \mathbf{B}\mathbf{Z} + 2\boldsymbol{\mu}^\top \mathbf{B}\mathbf{Z} - \operatorname{tr}(\mathbf{B}\mathbf{K}))\right) \\ &= \mathbb{E}\left(\mathbf{Z}^\top \mathbf{A}\mathbf{Z}\mathbf{Z}^\top \mathbf{B}\mathbf{Z}\right) + 2\boldsymbol{\mu}^\top \mathbf{A}\mathbb{E}\left(\mathbf{Z}\mathbf{Z}^\top \mathbf{B}\mathbf{Z}\right) + 2\boldsymbol{\mu}^\top \mathbf{B}\mathbb{E}\left(\mathbf{Z}\mathbf{Z}^\top \mathbf{A}\mathbf{Z}\right) \\ &\quad - \mathbb{E}\left(\mathbf{Z}^\top \mathbf{A}\mathbf{Z}\right)\operatorname{tr}(\mathbf{B}\mathbf{K}) - \mathbb{E}\left(\mathbf{Z}^\top \mathbf{B}\mathbf{Z}\right)\operatorname{tr}(\mathbf{A}\mathbf{K}) \\ &\quad + 4\boldsymbol{\mu}^\top \mathbf{A}\mathbf{K}\mathbf{B}\boldsymbol{\mu} + \operatorname{tr}(\mathbf{A}\mathbf{K})\operatorname{tr}(\mathbf{B}\mathbf{K}) \\ &= \mathbb{E}\left(\mathbf{Z}^\top \mathbf{A}\mathbf{Z}\mathbf{Z}^\top \mathbf{B}\mathbf{Z}\right) + 2\boldsymbol{\mu}^\top \mathbf{A}\mathbb{E}\left(\mathbf{Z}\mathbf{Z}^\top \mathbf{B}\mathbf{Z}\right) + 2\boldsymbol{\mu}^\top \mathbf{B}\mathbb{E}\left(\mathbf{Z}\mathbf{Z}^\top \mathbf{A}\mathbf{Z}\right) \\ &\quad + 4\boldsymbol{\mu}^\top \mathbf{A}\mathbf{K}\mathbf{B}\boldsymbol{\mu} - \operatorname{tr}(\mathbf{A}\mathbf{K})\operatorname{tr}(\mathbf{B}\mathbf{K})\,,\end{aligned}$$

  where the last equality is a result of Theorem 1.5 because $Z \sim \mathrm{N}(\mathbf{o}, \mathbf{K})$, which implies $\mathbb{E}\left(\mathbf{Z}^\top \mathbf{A}\mathbf{Z}\right) = \operatorname{tr}(\mathbf{A}\mathbf{K})$.

- Since the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{K}$ are symmetric, it follows from Lemma 1.11 that

$$\begin{aligned}\mathbb{E}\left(\mathbf{Z}^\top\mathbf{A}\mathbf{Z}\mathbf{Z}^\top\mathbf{B}\mathbf{Z}\right) &= \mathbb{E}\left(\mathbf{Z}^\top\mathbf{A}\mathbf{Z}\cdot\mathbf{Z}^\top\mathbf{B}\mathbf{Z}\right)\\ &= \sum_{i=1}^n\sum_{j=1}^n\sum_{\ell=1}^n\sum_{m=1}^n a_{ij}b_{\ell m}\mathbb{E}\left(Z_iZ_jZ_\ell Z_m\right)\\ &= \sum_{i=1}^n\sum_{j=1}^n\sum_{\ell=1}^n\sum_{m=1}^n \left(a_{ij}k_{ji}b_{\ell m}k_{m\ell}+a_{ji}k_{i\ell}b_{\ell m}k_{mj}+a_{ij}k_{j\ell}b_{\ell m}k_{mi}\right)\\ &= \operatorname{tr}(\mathbf{AK})\operatorname{tr}(\mathbf{BK})+2\operatorname{tr}(\mathbf{AKBK})\,.\end{aligned}$$

- Furthermore, Lemma 1.11 implies that

$$\mathbb{E}\left(\mathbf{Z}\mathbf{Z}^\top\mathbf{A}\mathbf{Z}\right)=\left(\sum_{i=1}^n\sum_{j=1}^n a_{ij}\mathbb{E}\left(Z_iZ_jZ_\ell\right)\right)_\ell=\mathbf{o}\tag{29}$$

  and analogously $\mathbb{E}\left(\mathbf{Z}\mathbf{Z}^\top\mathbf{B}\mathbf{Z}\right)=\mathbf{o}$.
- This result and the above derived expression for $\operatorname{Cov}\left(\mathbf{Y}^\top\mathbf{A}\mathbf{Y},\,\mathbf{Y}^\top\mathbf{B}\mathbf{Y}\right)$ imply the statement. □

Now, we derive the following formula for the covariance vector of linear or quadratic forms of normally distributed random vectors.

Theorem 1.7 *Let $\mathbf{Y}=(Y_1,\ldots,Y_n)^\top$ be an n–dimensional random vector with $\mathbf{Y}\sim\mathrm{N}(\boldsymbol{\mu},\mathbf{K})$ and let $\mathbf{A}=(a_{ij})$, $\mathbf{B}=(b_{ij})$ be arbitrary symmetric $n\times n$ matrices. Then it holds*

$$\operatorname{Cov}\left(\mathbf{AY},\,\mathbf{Y}^\top\mathbf{B}\mathbf{Y}\right)=2\mathbf{AKB}\boldsymbol{\mu}\,.\tag{30}$$

Proof

- As $\mathbb{E}(\mathbf{AY})=\mathbf{A}\boldsymbol{\mu}$ and as it has been shown in Theorem 1.5 that

$$\mathbb{E}\left(\mathbf{Y}^\top\mathbf{B}\mathbf{Y}\right)=\operatorname{tr}(\mathbf{BK})+\boldsymbol{\mu}^\top\mathbf{B}\boldsymbol{\mu}\,,$$

  it follows that

$$\begin{aligned}\operatorname{Cov}\left(\mathbf{AY},\,\mathbf{Y}^\top\mathbf{B}\mathbf{Y}\right) &= \mathbb{E}\left((\mathbf{AY}-\mathbf{A}\boldsymbol{\mu})(\mathbf{Y}^\top\mathbf{B}\mathbf{Y}-\boldsymbol{\mu}^\top\mathbf{B}\boldsymbol{\mu}-\operatorname{tr}(\mathbf{BK}))\right)\\ &= \mathbb{E}\left((\mathbf{AY}-\mathbf{A}\boldsymbol{\mu})((\mathbf{Y}-\boldsymbol{\mu})^\top\mathbf{B}(\mathbf{Y}-\boldsymbol{\mu})+2(\mathbf{Y}-\boldsymbol{\mu})^\top\mathbf{B}\boldsymbol{\mu}-\operatorname{tr}(\mathbf{BK}))\right)\,.\end{aligned}$$

- Moreover, it holds that $\mathbb{E}(\mathbf{AY}-\mathbf{A}\boldsymbol{\mu})=\mathbf{o}$ and from (29) it follows with $\mathbf{Z}=\mathbf{Y}-\boldsymbol{\mu}$ that

$$\mathbb{E}\left((\mathbf{AY}-\mathbf{A}\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^\top\mathbf{B}(\mathbf{Y}-\boldsymbol{\mu})\right)=\mathbf{A}\mathbb{E}\left((\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^\top\mathbf{B}(\mathbf{Y}-\boldsymbol{\mu})\right)=\mathbf{o}\,.$$

- Therefore, we get

$$\begin{aligned}\operatorname{Cov}\left(\mathbf{AY},\,\mathbf{Y}^\top\mathbf{B}\mathbf{Y}\right) &= 2\mathbb{E}\left((\mathbf{AY}-\mathbf{A}\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^\top\mathbf{B}\boldsymbol{\mu}\right)\\ &= 2\mathbf{A}\mathbb{E}\left((\mathbf{Y}-\boldsymbol{\mu})(\mathbf{Y}-\boldsymbol{\mu})^\top\right)\mathbf{B}\boldsymbol{\mu}\\ &= 2\mathbf{AKB}\boldsymbol{\mu}\,.\end{aligned}$$

□

1.3.2 Noncentral χ^2–Distribution

To determine the distribution of quadratic forms of normally distributed random vectors we introduce the (parametric) family of the noncentral χ^2–distribution.

Definition Let $\boldsymbol{\mu} \in \mathbb{R}^n$ and $(X_1, \ldots, X_n)^\top \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{I})$. Then the random variable

$$Z = (X_1, \ldots, X_n)(X_1, \ldots, X_n)^\top = \sum_{i=1}^n X_i^2$$

is distributed according to a *noncentral χ^2–distribution* with n degrees of freedom and the *noncentrality parameter* $\lambda = \boldsymbol{\mu}^\top \boldsymbol{\mu}$. (Notation: $Z \sim \chi^2_{n,\lambda}$)

Remark

- For $\boldsymbol{\mu} = \mathbf{o}$ we obtain the (central) χ^2–distribution χ^2_n with n degrees of freedom, which has already been introduced in Section I–1.3.1, as a special case.
- To derive a formula for the density of the noncentral χ^2–distribution we consider (in addition to the characteristic function) still another *integral transform* of probability densities.

Definition

- Let $f : \mathbb{R} \to [0, \infty)$ be the density of a real–valued random variable such that the integral $\int_{-\infty}^{\infty} e^{tx} f(x)\, dx$ is well–defined for each $t \in (a, b)$ in a certain interval (a, b) with $a < b$.
- Then the mapping $\psi : (a, b) \to \mathbb{R}$ with

$$\psi(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx\,, \qquad \forall\, t \in (a, b) \tag{31}$$

  is called the *moment generating function* of the density f.

The following *uniqueness theorem* for moment generating functions is true, which we state without proof.

Lemma 1.12

- *Let $f, f' : \mathbb{R} \to [0, \infty)$ be densities of real–valued random variables and let the corresponding moment generating functions $\psi : (a, b) \to \mathbb{R}$ and $\psi' : (a, b) \to \mathbb{R}$ be well–defined in a (common) interval (a, b) with $a < b$.*
- *It holds that $\psi(t) = \psi'(t)$ for each $t \in (a, b)$ if and only if $f(x) = f'(x)$ for almost all $x \in \mathbb{R}$.*

By using Lemma 1.12 we are now able to identify the density of the noncentral χ^2–distribution.

Theorem 1.8

- *Let the random variable $Z_{n,\lambda} : \Omega \to \mathbb{R}$ be distributed according to the $\chi^2_{n,\lambda}$–distribution with n degrees of freedom and noncentrality parameter λ.*
- *Then the density of $Z_{n,\lambda}$ is given by*

$$f_{Z_{n,\lambda}}(z) = \begin{cases} \exp\Big(-\dfrac{\lambda + z}{2}\Big) \displaystyle\sum_{j=0}^{\infty} \dfrac{\Big(\dfrac{\lambda}{2}\Big)^j z^{\frac{n}{2}+j-1}}{j!\, 2^{\frac{n}{2}+j}\, \Gamma\Big(\dfrac{n}{2} + j\Big)}\,, & \text{if } z > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{32}$$

Proof

- Let $\boldsymbol{\mu} \in \mathbb{R}^n$ and $(X_1, \ldots, X_n)^\top \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{I})$.
- The moment generating function $\psi_Z(t)$ of $Z = (X_1, \ldots, X_n)(X_1, \ldots, X_n)^\top = \sum_{j=1}^n X_j^2$ is well–defined for $t \in (-\infty, 1/2)$ and for each $t < 1/2$ it holds that

$$\begin{aligned}\psi_Z(t) &= \mathbb{E}\exp\Big(t\sum_{j=1}^n X_j^2\Big) = \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\exp\Big(t\sum_{j=1}^n x_j^2\Big)\prod_{j=1}^n\frac{1}{\sqrt{2\pi}}\exp\Big(\frac{1}{2}(x_j-\mu_j)^2\Big)\,dx_1\ldots dx_n \\ &= \left(\frac{1}{2\pi}\right)^{n/2}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}\exp\left(t\sum_{j=1}^n x_j^2 - \frac{1}{2}\sum_{j=1}^n(x_j-\mu_j)^2\right)dx_1\ldots dx_n \\ &= \prod_{j=1}^n\int_{-\infty}^{\infty}(2\pi)^{-1/2}\exp\left(tx_j^2 - \frac{1}{2}(x_j-\mu_j)^2\right)dx_j\,.\end{aligned}$$

- It is possible to rewrite the exponent of the last term as follows:

$$\begin{aligned}tx_j^2 - \frac{1}{2}(x_j-\mu_j)^2 &= -\frac{1}{2}\left(-2tx_j^2 + x_j^2 - 2x_j\mu_j + \mu_j^2\right) \\ &= -\frac{1}{2}\Big(x_j^2(1-2t) - 2x_j\mu_j + \mu_j^2(1-2t)^{-1} + \mu_j^2 - \mu_j^2(1-2t)^{-1}\Big) \\ &= -\frac{1}{2}\Big((x_j - \mu_j(1-2t)^{-1})^2(1-2t) + \mu_j^2(1-(1-2t)^{-1})\Big)\,.\end{aligned}$$

- Hence, it holds that

$$\begin{aligned}\psi_Z(t) &= \exp\left(-\frac{1}{2}(1-(1-2t)^{-1})\sum_{j=1}^n\mu_j^2\right)\prod_{j=1}^n\int_{-\infty}^{\infty}(2\pi)^{-1/2}\exp\Big(-\frac{(x_j-\mu_j(1-2t)^{-1})^2}{2(1-2t)^{-1}}\Big)\,dx_j \\ &= (1-2t)^{-n/2}\exp\Big(-\frac{\lambda}{2}(1-(1-2t)^{-1})\Big)\end{aligned}$$

as the integrand represents the density of the one–dimensional normal distribution (except for the constant factor $(1-2t)^{1/2}$); $\lambda = \boldsymbol{\mu}^\top\boldsymbol{\mu}$.

- On the other hand, the moment generating function $\psi(t)$ of the density $f_{Z_{n,\lambda}}(z)$ given in (32) can be written as

$$\psi(t) = \sum_{j=0}^{\infty}\frac{e^{-\lambda/2}(\lambda/2)^j}{j!}\int_0^{\infty}e^{tz}\,\frac{z^{n/2+j-1}e^{-z/2}}{2^{\frac{n}{2}+j}\,\Gamma\Big(\frac{n}{2}+j\Big)}\,dz\,,$$

where the integral is the moment generating function of the (central) χ^2–distribution χ^2_{n+2j} with $n+2j$ degrees of freedom.

- Similar to the way the characteristic function (cf. Theorem I–1.5) is defined, the moment generating function of this distribution is given by

$$\psi_{\chi^2_{n+2j}}(t) = \frac{1}{(1-2t)^{n/2+j}}\,.$$

- Therefore, it holds that

$$\int_0^{\infty}e^{tz}\,\frac{z^{n/2+j-1}e^{-z/2}}{2^{\frac{n}{2}+j}\,\Gamma\Big(\frac{n}{2}+j\Big)}\,dz = \frac{1}{(1-2t)^{n/2+j}}\,,$$

and

$$\begin{aligned}\psi(t) &= e^{-\lambda/2}(1-2t)^{-n/2}\sum_{j=0}^{\infty}\frac{1}{j!}\Big(\frac{\lambda}{2}(1-2t)^{-1}\Big)^j \\ &= (1-2t)^{-n/2}\exp\Big(-\frac{\lambda}{2}\,(1-(1-2t)^{-1})\Big)\,.\end{aligned}$$

- Hence, $\psi(t)=\psi_Z(t)$ for each $t<1/2$ and the statement follows from Lemma 1.12. □

1.3.3 Distributional Properties of Linear and Quadratic Forms

- *Recall*: The definition of the noncentral χ^2–distribution in Section 1.3.2 considers the sum of squares of the components of $\mathrm{N}(\boldsymbol{\mu},\mathbf{I})$–distributed random vectors.
- One can show that the (adequately modified) sum of squares is distributed according to the noncentral χ^2–distribution even if the considered normally distributed random vector has an *arbitrary* positive definite covariance matrix.
- Indeed, let $\boldsymbol{\mu}\in\mathbb{R}^n$ and let $\mathbf{K}$ be a symmetric and positive definite $n\times n$ matrix.
- If $\mathbf{Z}=(Z_1,\ldots,Z_n)^\top\sim\mathrm{N}(\boldsymbol{\mu},\mathbf{K})$, Theorem 1.3 implies that
$$\mathbf{K}^{-1/2}\mathbf{Z}\sim\mathrm{N}(\mathbf{K}^{-1/2}\boldsymbol{\mu},\mathbf{I})\,.$$
- Therefore, by the definition of the noncentral χ^2–distribution it follows that
$$\mathbf{Z}^\top\mathbf{K}^{-1}\mathbf{Z}=\big(\mathbf{K}^{-1/2}\mathbf{Z}\big)^\top\mathbf{K}^{-1/2}\mathbf{Z}\sim\chi^2_{n,\lambda}\,, \tag{33}$$
where $\lambda=(\mathbf{K}^{-1/2}\boldsymbol{\mu})^\top\mathbf{K}^{-1/2}\boldsymbol{\mu}=\boldsymbol{\mu}^\top\mathbf{K}^{-1}\boldsymbol{\mu}$.

The distributional property (33) for quadratic forms of normally distributed random vectors has the following generalization. In this context Lemma 1.7 about the factorization of symmetric and positive semidefinite matrices is useful.

Theorem 1.9

- *Let $\mathbf{Z}=(Z_1,\ldots,Z_n)^\top\sim\mathrm{N}(\boldsymbol{\mu},\mathbf{K})$, where the covariance matrix $\mathbf{K}$ be positive definite. Moreover, let $\mathbf{A}$ be a symmetric $n\times n$ matrix with $\mathrm{rk}(\mathbf{A})=r\le n$.*
- *If the matrix $\mathbf{AK}$ is idempotent, i.e., if $\mathbf{AK}=(\mathbf{AK})^2$, it holds that $\mathbf{Z}^\top\mathbf{AZ}\sim\chi^2_{r,\lambda}$, where $\lambda=\boldsymbol{\mu}^\top\mathbf{A}\boldsymbol{\mu}$.*

Proof

- Let the matrix $\mathbf{AK}$ be idempotent. Then it holds that
$$\mathbf{AK}=\mathbf{AKAK}\,.$$
- Since $\mathbf{K}$ is nondegenerate, it is allowed to multiply both sides of the above equation from the right by $\mathbf{K}^{-1}$. It follows
$$\mathbf{A}=\mathbf{AKA} \tag{34}$$
or
$$\mathbf{x}^\top\mathbf{Ax}=\mathbf{x}^\top\mathbf{AKAx}=(\mathbf{Ax})^\top\mathbf{K}(\mathbf{Ax})\ge 0$$
for each $\mathbf{x}\in\mathbb{R}^n$, i.e., $\mathbf{A}$ is positive semidefinite.

- According to Lemma 1.7 there exists a decomposition
$$\mathbf{A} = \mathbf{H}\mathbf{H}^\top \tag{35}$$
such that the $n \times r$ matrix $\mathbf{H}$ has full column rank r.
- Now, Lemma 1.2 implies that the inverse matrix $(\mathbf{H}^\top\mathbf{H})^{-1}$ exists.
- From Theorem 1.3 about the linear transformation of normally distributed random vectors it follows for the r–dimensional vector $\mathbf{Z}' = \mathbf{H}^\top\mathbf{Z}$ that
$$\mathbf{Z}' \sim \mathrm{N}(\mathbf{H}^\top\boldsymbol{\mu}, \mathbf{I}_r) \tag{36}$$
because
$$\begin{aligned}\mathbf{H}^\top\mathbf{K}\mathbf{H} &= (\mathbf{H}^\top\mathbf{H})^{-1}(\mathbf{H}^\top\mathbf{H})(\mathbf{H}^\top\mathbf{K}\mathbf{H})(\mathbf{H}^\top\mathbf{H})(\mathbf{H}^\top\mathbf{H})^{-1}\\ &= (\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top(\mathbf{A}\mathbf{K}\mathbf{A})\mathbf{H}(\mathbf{H}^\top\mathbf{H})^{-1}\\ &= (\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top\mathbf{A}\mathbf{H}(\mathbf{H}^\top\mathbf{H})^{-1} = \mathbf{I}_r\,,\end{aligned}$$
where the last three equalities follow from (34) and (35).
- As on the other hand
$$\mathbf{Z}^\top\mathbf{A}\mathbf{Z} \;=\; \mathbf{Z}^\top\mathbf{H}\mathbf{H}^\top\mathbf{Z} \;=\; \left(\mathbf{H}^\top\mathbf{Z}\right)^\top\mathbf{H}^\top\mathbf{Z} \;=\; (\mathbf{Z}')^\top\mathbf{Z}'$$
and since
$$\left(\mathbf{H}^\top\boldsymbol{\mu}\right)^\top\mathbf{H}^\top\boldsymbol{\mu} = \boldsymbol{\mu}^\top\mathbf{H}\mathbf{H}^\top\boldsymbol{\mu} = \boldsymbol{\mu}^\top\mathbf{A}\boldsymbol{\mu}\,,$$
the statement follows from (36) and from the definition of the noncentral χ^2–distribution. □

Furthermore, the following criterion for the independence of linear and quadratic forms of normally distributed random vectors is useful. It can be considered as a (vectorial) generalization of Lemma I–5.3.

Theorem 1.10

- *Let* $\mathbf{Z} = (Z_1, \ldots, Z_n)^\top \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{K})$, *where* $\mathbf{K}$ *is an arbitrary (symmetric and positive semidefinite) covariance matrix.*
- *Moreover, let* $\mathbf{A}$, $\mathbf{B}$ *be arbitrary* $r_1 \times n$ *and* $r_2 \times n$ *matrices with* $r_1, r_2 \le n$ *and let* $\mathbf{C}$ *be a symmetric and positive semidefinite* $n \times n$ *matrix.*
- *If the additional condition*
$$\mathbf{A}\mathbf{K}\mathbf{B}^\top = \mathbf{0} \qquad \textit{or} \qquad \mathbf{A}\mathbf{K}\mathbf{C} = \mathbf{0} \tag{37}$$
is fulfilled, the random variables $\mathbf{A}\mathbf{Z}$ *and* $\mathbf{B}\mathbf{Z}$ *or* $\mathbf{A}\mathbf{Z}$ *and* $\mathbf{Z}^\top\mathbf{C}\mathbf{Z}$, *respectively, are independent.*

Proof

- First, we show that (37) implies the independence of the linear forms $\mathbf{A}\mathbf{Z}$ and $\mathbf{B}\mathbf{Z}$.
- Because of the uniqueness theorem for characteristic functions of random vectors (cf. Lemma 1.9), it suffices to show that $\mathbf{t}_2 \in \mathbb{R}^{r_2}$
$$\mathbb{E}\,\exp\bigl(\mathrm{i}\,(\mathbf{t}_1^\top\mathbf{A}\mathbf{Z} + \mathbf{t}_2^\top\mathbf{B}\mathbf{Z})\bigr) = \mathbb{E}\,\exp\bigl(\mathrm{i}\,\mathbf{t}_1^\top\mathbf{A}\mathbf{Z}\bigr)\mathbb{E}\,\exp\bigl(\mathrm{i}\,\mathbf{t}_2^\top\mathbf{B}\mathbf{Z}\bigr)$$
for arbitrary $\mathbf{t}_1 \in \mathbb{R}^{r_1}$.

- From (37) it follows that

$$\mathbf{B}\mathbf{K}\mathbf{A}^\top = \Big((\mathbf{B}\mathbf{K}\mathbf{A}^\top)^\top\Big)^\top = \Big(\mathbf{A}\mathbf{K}\mathbf{B}^\top\Big)^\top = \mathbf{0}.$$

- Therefore, it holds for arbitrary $\mathbf{t}_1 \in \mathbb{R}^{r_1}$, $\mathbf{t}_2 \in \mathbb{R}^{r_2}$ that

$$(\mathbf{t}_1^\top \mathbf{A})\mathbf{K}(\mathbf{t}_2^\top \mathbf{B})^\top = \mathbf{t}_1^\top \mathbf{A}\mathbf{K}\mathbf{B}^\top \mathbf{t}_2 = \mathbf{0}\,, \qquad (\mathbf{t}_2^\top \mathbf{B})\mathbf{K}(\mathbf{t}_1^\top \mathbf{A})^\top = \mathbf{t}_2^\top \mathbf{B}\mathbf{K}\mathbf{A}^\top \mathbf{t}_1 = \mathbf{0}\,. \tag{38}$$

- Then the representation formula (25) for the characteristic function of normally distributed random vectors derived in Theorem 1.4 and (38) imply that

$$\begin{aligned}
&\mathbb{E}\,\exp\big(\mathrm{i}\,(\mathbf{t}_1^\top \mathbf{A}\mathbf{Z} + \mathbf{t}_2^\top \mathbf{B}\mathbf{Z})\big) = \mathbb{E}\,\exp\big(\mathrm{i}\,(\mathbf{t}_1^\top \mathbf{A} + \mathbf{t}_2^\top \mathbf{B})\mathbf{Z}\big)\\
&= \exp\Big(\mathrm{i}\,(\mathbf{t}_1^\top \mathbf{A} + \mathbf{t}_2^\top \mathbf{B})\boldsymbol{\mu} - \frac{1}{2}\,(\mathbf{t}_1^\top \mathbf{A} + \mathbf{t}_2^\top \mathbf{B})\mathbf{K}(\mathbf{t}_1^\top \mathbf{A} + \mathbf{t}_2^\top \mathbf{B})^\top\Big)\\
&= \exp\Big(\mathrm{i}\,(\mathbf{t}_1^\top \mathbf{A} + \mathbf{t}_2^\top \mathbf{B})\boldsymbol{\mu} - \frac{1}{2}\,(\mathbf{t}_1^\top \mathbf{A})\mathbf{K}(\mathbf{t}_1^\top \mathbf{A})^\top - \frac{1}{2}\,(\mathbf{t}_2^\top \mathbf{B})\mathbf{K}(\mathbf{t}_2^\top \mathbf{B})^\top\Big)\\
&= \exp\Big(\mathrm{i}\,(\mathbf{t}_1^\top \mathbf{A})\boldsymbol{\mu} - \frac{1}{2}\,(\mathbf{t}_1^\top \mathbf{A})\mathbf{K}(\mathbf{t}_1^\top \mathbf{A})^\top\Big)\exp\Big(\mathrm{i}\,(\mathbf{t}_2^\top \mathbf{B})\boldsymbol{\mu} - \frac{1}{2}\,(\mathbf{t}_2^\top \mathbf{B})\mathbf{K}(\mathbf{t}_2^\top \mathbf{B})^\top\Big)\\
&= \mathbb{E}\,\exp\big(\mathrm{i}\,\mathbf{t}_1^\top \mathbf{A}\mathbf{Z}\big)\mathbb{E}\,\exp\big(\mathrm{i}\,\mathbf{t}_2^\top \mathbf{B}\mathbf{Z}\big)\,.
\end{aligned}$$

- Now, it remains to show that the independence of $\mathbf{AZ}$ of $\mathbf{Z}^\top\mathbf{CZ}$ is a result of the second condition of (37).
- Let $\mathrm{rk}(\mathbf{C}) = r \le n$. According to Lemma 1.7 there is an $n \times r$ matrix $\mathbf{H}$ with $\mathrm{rk}(\mathbf{H}) = r$ such that $\mathbf{C} = \mathbf{H}\mathbf{H}^\top$.
- Then it follows from (37) that $\mathbf{AKHH}^\top = \mathbf{0}$ and $\mathbf{AKHH}^\top\mathbf{H} = \mathbf{0}$.
- Because of Lemma 1.2, the $r \times r$ matrix $\mathbf{H}^\top\mathbf{H}$ has (full) rank $\mathrm{rk}(\mathbf{H}) = r$. Hence, $\mathbf{H}^\top\mathbf{H}$ is invertible.
- Finally, it follows that $\mathbf{AKH} = \mathbf{0}$.
- Therefore, the first part of the proof implies that the linear forms $\mathbf{AZ}$ and $\mathbf{H}^\top\mathbf{Z}$ are independent.
- Because of

$$\mathbf{Z}^\top\mathbf{C}\mathbf{Z} = \mathbf{Z}^\top\mathbf{H}\mathbf{H}^\top\mathbf{Z} = (\mathbf{H}^\top\mathbf{Z})^\top\mathbf{H}^\top\mathbf{Z}\,,$$

the transformation theorem for independent random vectors (cf. Theorem I–1.8) implies that also $\mathbf{AZ}$ and $\mathbf{Z}^\top\mathbf{CZ}$ are independent. □