Mathematics for Economists: linear algebra, matrixalgebra, stochastic matrices

Stefan Lang, Janette Walde

Department of Statistics, University of Innsbruck

October 2014

Chapter 1

Real valued vectors

Real valued vectors

- The set of all n–tupels

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

  of real valued numbers $x_1, \ldots, x_n$ equipped with a vector addition and scalar multiplication (see below) is called the n–dimensional vector space over $\mathbb{R}$, short $\mathbb{R}^n$.

- The numbers $x_1, \ldots, x_n$ are also called scalars.

Real valued vectors

- The null vector

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

is denoted by $\mathbf{0}$.

- The one vector

$$\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

is denoted by $\mathbf{1}$.

Real valued vectors

- **Vector addition**

$$x + y = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

- **Scalar multiplication**

$$\lambda \cdot x = \lambda \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda \cdot x_1 \\ \lambda \cdot x_2 \\ \vdots \\ \lambda \cdot x_n \end{pmatrix}$$

Real valued vectors

Example

$$x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \qquad y = \begin{pmatrix} -1 \\ 2 \\ -3 \end{pmatrix}$$

Geometrical properties of real valued vectors

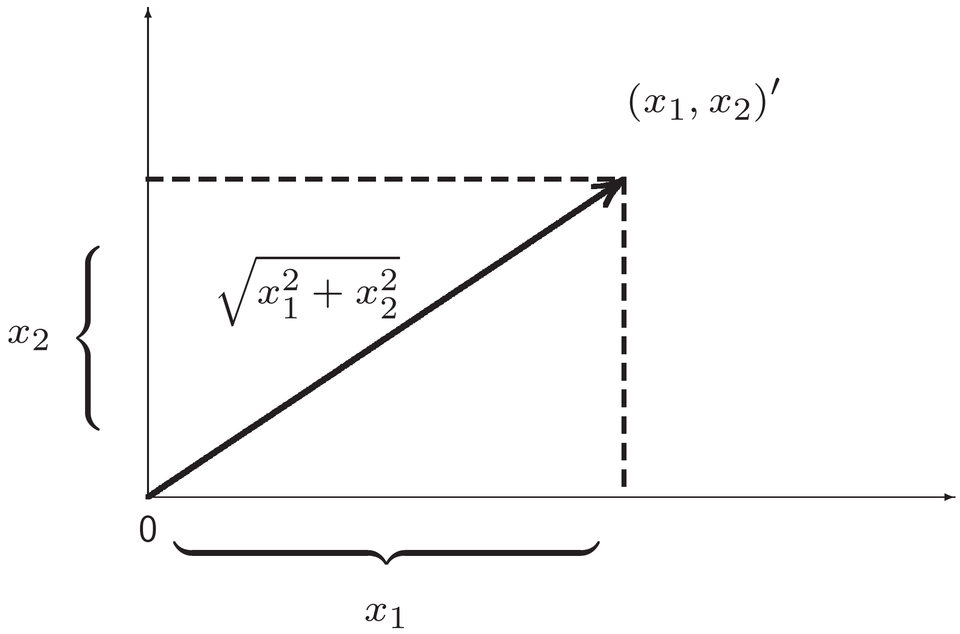

Figure 1: Geometrical visualization of a vector in $\mathbb{R}^2$.

Geometrical properties of real valued vectors

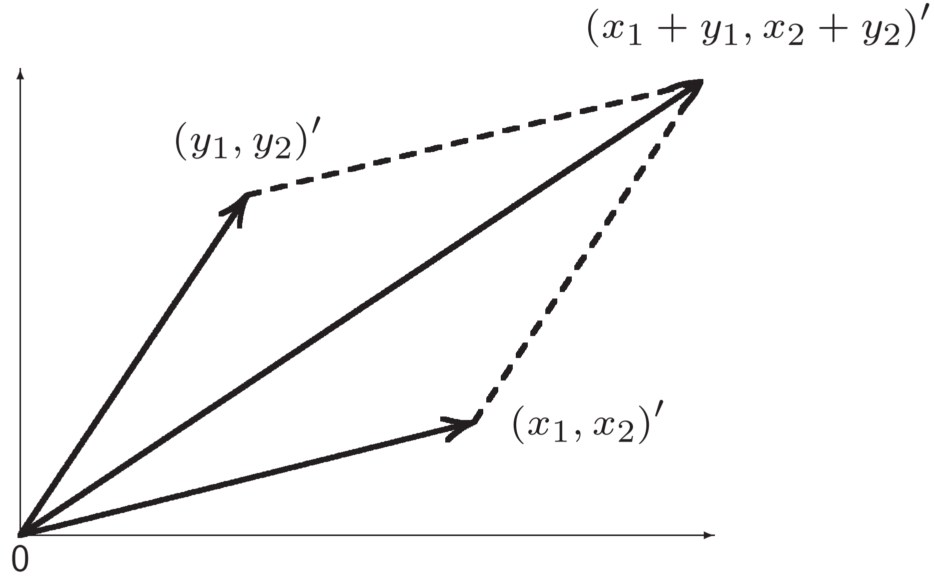

Figure 2: vector addition in $\mathbb{R}^2$.

Geometrical properties of real valued vectors

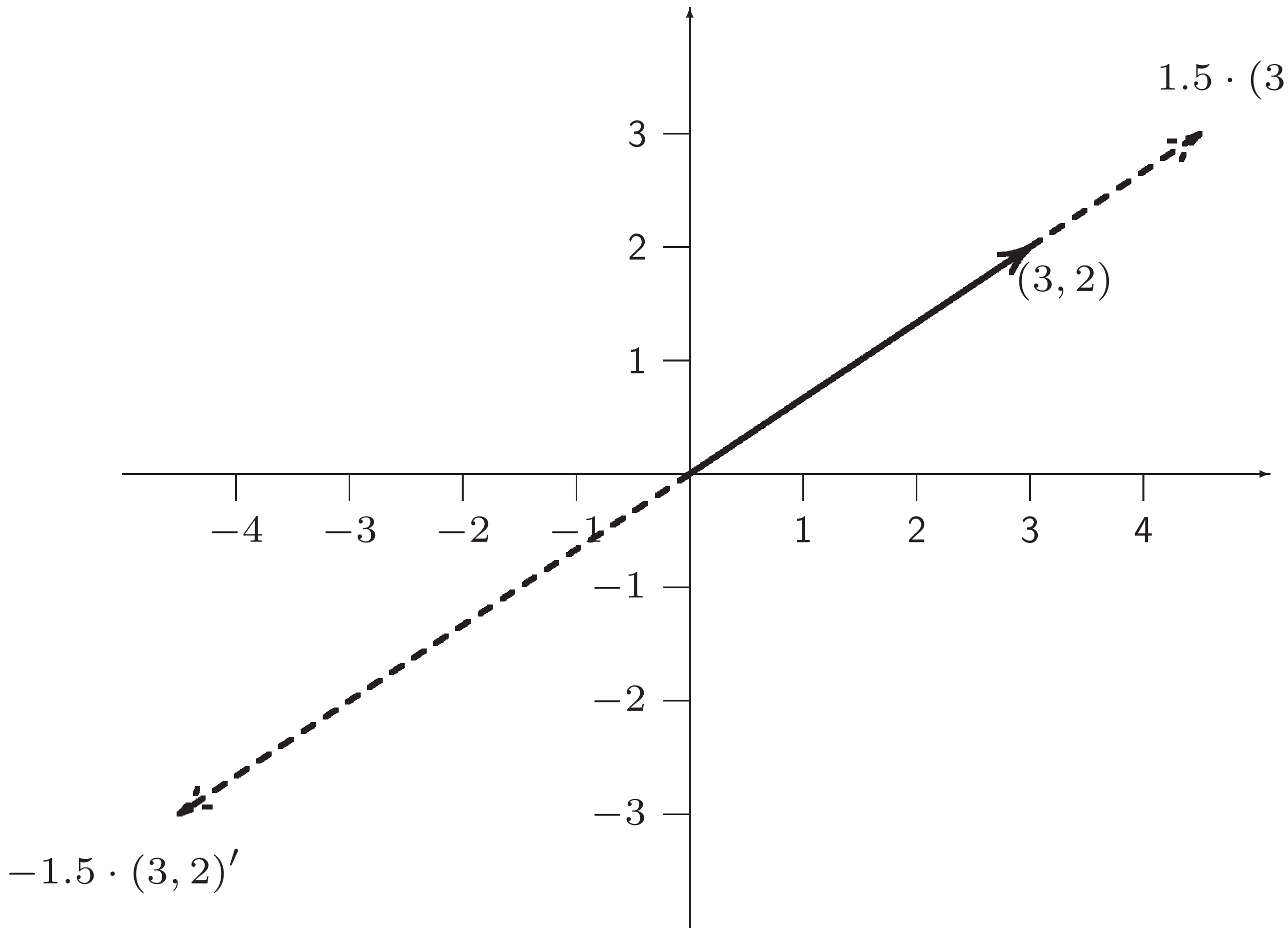

Figure 3: Scalar multiplication.

Properties of real valued vectors

For arbitrary vectors $x, y, z \in \mathbb{R}^n$ and scalars $\lambda, \mu \in \mathbb{R}$ the following properties hold:

1. Associative law: $x + (y + z) = (x + y) + z$
2. Commutative law: $x + y = y + x$
3. $x + \mathbf{0} = x$
4. $x + (-x) = \mathbf{0}$
5. Distributive law for scalar multiplication: $(\lambda + \mu)x = \lambda x + \mu x$ respectively $\lambda(x + y) = \lambda x + \lambda y$
6. Associative law for scalar multiplication : $(\lambda\mu)x = \lambda(\mu x)$
7. $1 \cdot x = x$

Scalar product

- The scalar product (or inner product or dot product) $\langle x, y\rangle$ of the vectors $x, y \in \mathbb{R}^n$ is defined as

$$\langle x, y\rangle = x_1 \cdot y_1 + x_2 \cdot y_2 + \cdots + x_n \cdot y_n.$$

- Tow vectors are called orthogonal, if

$$\langle x, y\rangle = 0$$

  holds.

- The space $\mathbb{R}^n$ equipped with a vector addition, scalar multiplication and the scalar product is called Euclidian space.

Scalar product

Example

$$x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \qquad y = \begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix} \qquad z = \begin{pmatrix} -1 \\ 0 \\ \frac{1}{3} \end{pmatrix}$$

Scalar product

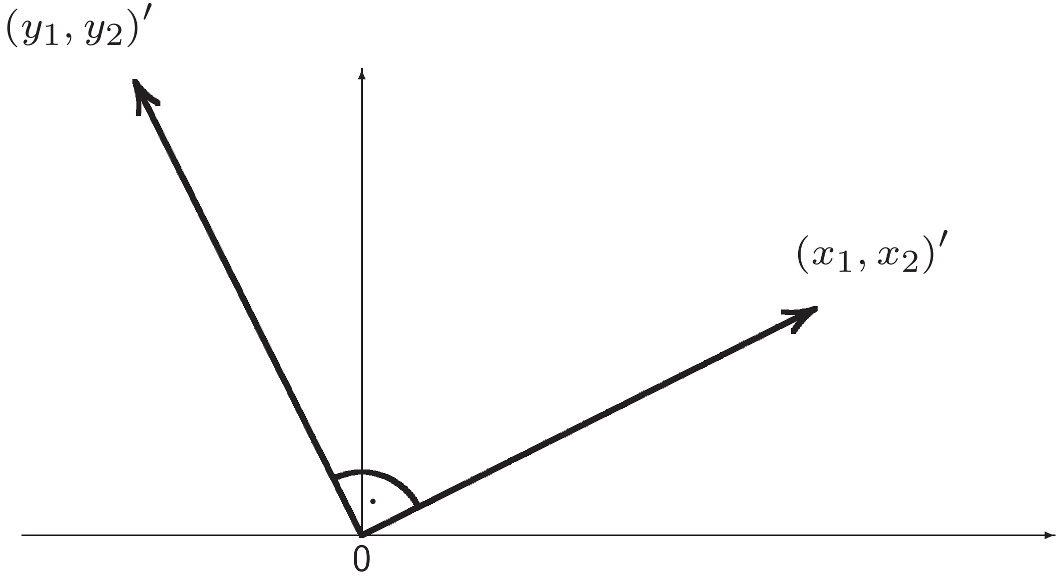

Figure 4: Two orthogonal vectors x and y.

Distance between vectors, length of a vector

- The (euclidian) distance $d(x, y)$ between the vectors x and y is defined as

$$\begin{aligned} d(x,y) &= \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \\ &= \sqrt{\langle x - y, x - y \rangle}. \end{aligned}$$

- The (euclidian) length $\|x\|$ of a vector $x \in \mathbb{R}^n$ is defined as

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{\langle x, x \rangle}.$$

Distance between vectors, length of a vector

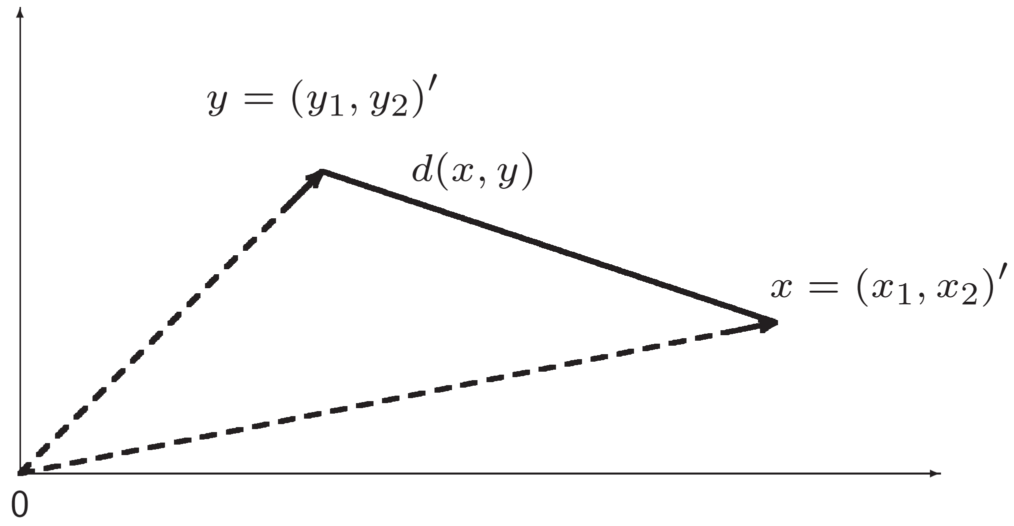

Figure 5: Euclidian distance in R^2

Distance between vectors, length of a vector

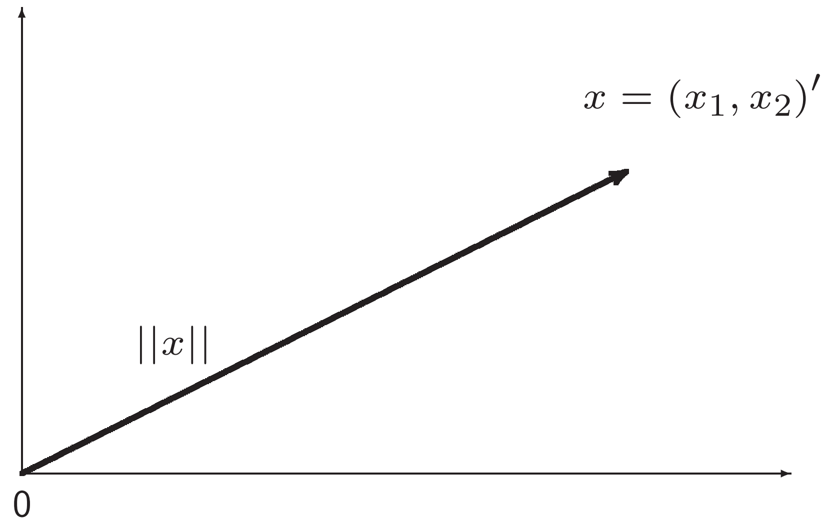

Figure 6: Length $||x||$ in $\mathbb{R}^2$.

Distance between vectors, length of a vector

Example

$$x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \qquad y = \begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix}$$

Chapter 2

Real valued matrices

Real valued matrices

- A in m rows and n columns ordered scheme $\mathbf{A}$ of mn elements $a_{ij} \in \mathbb{R}$

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

  is called real valued matrix of order $m \times n$ or short $m \times n$ matrix. Short: $\mathbf{A} = (a_{ij}), i = 1, \ldots, m, j = 1, \ldots, n$.

- The rows of $\mathbf{A}$ can be seen as vectors in $\mathbb{R}^n$ (so called row vectors) and the columns as vectors in $\mathbb{R}^m$ (so called column vectors). The j–th row vector of $\mathbf{A}$ is denoted by $a^j = (a_{j1}, \ldots, a_{jn})$ and the j–th column vector by $a_j = (a_{1j}, \ldots, a_{nj})'$.

- Two $m \times n$ matrices $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ are equal, if and only if for all i, j $a_{ij} = b_{ij}$ holds.

Some special matrices

- **Quadratic matrix:** A matrix $\mathbf{A}$ of order $n \times n$ is called quadratic. The diagonal consisting of the elements $a_{11}, \ldots, a_{nn}$ is called main diagonal.

- **Identity matrix:**

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & \ldots & 1 \end{pmatrix}.$$

- **Diagonal matrix:**

$$\mathbf{D} = \begin{pmatrix} d_1 & 0 & \ldots & 0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & \ldots & d_n \end{pmatrix}$$

  Short: $\mathbf{D} = diag(d_1, \ldots, d_n)$

- **Symmetric matrix:** A quadratic matrix $\mathbf{A}$ is called symmetric, if $\mathbf{A} = \mathbf{A}'$.

Submatrices

- Partition a matrix $\mathbf{A}$ in submatrices $\mathbf{A}_{ij}$ to obtain:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1c} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \mathbf{A}_{r1} & \mathbf{A}_{r2} & \cdots & \mathbf{A}_{rc} \end{pmatrix} = (\mathbf{A}_{ij})$$

- The submatrices $\mathbf{A}_{i1}, \ldots, \mathbf{A}_{ic}$, $i = 1, \ldots, r$ have the same number of rows, the submatrices $\mathbf{A}_{1j}, \ldots, \mathbf{A}_{rj}$, $j = 1, \ldots, c$ the same number of columns.

Transpose of a matrix

- The transpose $\mathbf{A}'$ of a matrix $\mathbf{A}$ is defined as the $n \times m$ matrix, that is obtained by exchanging the rows and columns of $\mathbf{A}$, i.e.

$$\mathbf{A}' = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}$$

- The transpose of a partitioned matrix is given as the transpose of the transposed submatrices, i.e.

$$\mathbf{A}' = \begin{pmatrix} \mathbf{A}'_{11} & \mathbf{A}'_{21} & \cdots & \mathbf{A}'_{r1} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \mathbf{A}'_{1c} & \mathbf{A}'_{2c} & \cdots & \mathbf{A}'_{rc} \end{pmatrix}.$$

Transpose of a matrix

Example

$$\mathbf{A} = \begin{pmatrix} 2 & 4 & 1 & 6 \\ 1 & 0 & 3 & 2 \\ 9 & 3 & 4 & 3 \end{pmatrix} \quad \mathbf{B} = \left(\begin{array}{cc|cc} 1 & 2 & -1 & 3 \\ 2 & -2 & 1 & 0 \\ \hline 1 & -2 & 3 & 4 \\ -2 & 4 & 5 & 1 \end{array}\right)$$

Matrix addition and scalar multiplication

- The sum $\mathbf{A}+\mathbf{B}$ of two $m \times n$ matrices $\mathbf{A}=(a_{ij})$ and $\mathbf{B}=(b_{ij})$ ist defined as

$$\mathbf{A}+\mathbf{B} := (a_{ij}+b_{ij}).$$

- The scalar multiplication of $\mathbf{A}$ with a scalar $\lambda \in \mathbb{R}$ is defined as

$$\lambda\mathbf{A} := (\lambda a_{ij}).$$

Matrix addition and scalar multiplication

Example

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 5 & 2 \\ 1 & 2 & 2 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 1 & 4 & 2 \\ 3 & 1 & 0 \\ -1 & 2 & -4 \end{pmatrix}$$

Matrix multiplication

The product between the $m \times n$ matrix $\mathbf{A} = (a_{ij})$ and the $n \times p$ matrix $\mathbf{B} = (b_{ij})$ is the $m \times p$ matrix

$$\mathbf{AB} = \mathbf{C} = (c_{ik}) \qquad \text{with} \qquad c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}.$$

We have

$$\mathbf{A} \cdot \mathbf{B} = \begin{pmatrix} \sum_{j=1}^{n} a_{1j} b_{j1} & \sum_{j=1}^{n} a_{1j} b_{j2} & \cdots & \sum_{j=1}^{n} a_{1j} b_{jp} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \sum_{j=1}^{n} a_{mj} b_{j1} & \sum_{j=1}^{n} a_{mj} b_{j2} & \cdots & \sum_{j=1}^{n} a_{mj} b_{jp} \end{pmatrix}.$$

Matrix multiplication

Example

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} -1 & 2 \\ 1 & 2 \end{pmatrix}$$

Matrix multiplication

Example

$$\mathbf{A} = \begin{pmatrix} 2 & 4 & 16 \\ 1 & -3 & -7 \\ -2 & 2 & 2 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} -2 & -4 & -8 \\ -3 & -6 & -12 \\ 1 & 2 & 4 \end{pmatrix}$$

Multiplication of partitioned matrices

- For partitioned matrices $\mathbf{A}$ and $\mathbf{B}$ with submatrices $\mathbf{A}_{ij}$, $i = 1, \ldots, r$, $j = 1 \ldots, c$, and $\mathbf{B}_{lk}$, $l = 1, \ldots, c$, $k = 1, \ldots, d$, we obtain

$$\mathbf{AB} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \cdots & \mathbf{C}_{1d} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \mathbf{C}_{r1} & \mathbf{C}_{r2} & \cdots & \mathbf{C}_{rd} \end{pmatrix},$$

mit

$$\mathbf{C}_{ik} = \sum_{j=1}^{c} \mathbf{A}_{ij}\mathbf{B}_{jk} \qquad i = 1, \ldots, r \qquad k = 1, \ldots, d.$$

- Partitioned matrices can only be multiplied in partitioned form, if the corresponding submatrices have proper order for matrix multiplication.

Multiplication of partitioned matrices

Example

$$\mathbf{A} = \left(\begin{array}{rr|rr} 1 & 2 & -1 & 3 \\ 2 & -2 & 1 & 0 \\ \hline 1 & -2 & 3 & 4 \\ -2 & 4 & 5 & 1 \end{array}\right) \qquad \mathbf{B} = \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}$$

with

$$\mathbf{B}_{11} = \begin{pmatrix} 2 & 3 \\ -1 & 4 \end{pmatrix}, \qquad \mathbf{B}_{12} = \begin{pmatrix} 2 & 4 \\ 1 & 1 \end{pmatrix}$$

$$\mathbf{B}_{21} = \begin{pmatrix} -1 & -2 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{B}_{22} = \begin{pmatrix} 1 & 2 \\ 0 & -2 \end{pmatrix}.$$

Properties matrix addition and multiplication

1. $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$

2. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$

3. $\mathbf{A} + \mathbf{0} = \mathbf{A}$

4. $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$

5. $(k + r)\mathbf{A} = k\mathbf{A} + r\mathbf{A}$ respectively $k(\mathbf{A} + \mathbf{B}) = k\mathbf{A} + k\mathbf{B}$

6. $(kr)\mathbf{A} = k(r\mathbf{A})$

7. $1 \cdot \mathbf{A} = \mathbf{A}$

8. $0 \cdot \mathbf{A} = \mathbf{0}$.

9. $(k\mathbf{A})' = k\mathbf{A}'$

10. $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$

11. $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$

12. $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$

13. $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$

14. $\mathbf{AI}_n = \mathbf{A}$ respectively $\mathbf{I}_n\mathbf{A} = \mathbf{A}$

Matrices and linear operators

Consider a function $f : \mathbb{R}^n \to \mathbb{R}^m$ of the form

$$f(x) = f\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n \end{pmatrix}$$

f is called a linear operator or linear mapping.

Defining the $m \times n$ matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & & a_{mn} \end{pmatrix}$$

the linear operator can be expressed as

$$f(x) = \mathbf{A}x.$$

Matrices and linear operators

Sum of two linear operators

Consider the linear operators $f(x) = \mathbf{A}x$ and $g(x) = \mathbf{B}x$ where $\mathbf{A}$ and $\mathbf{B}$ are of order $m \times n$. Then

$$(f + g)(x) = (\mathbf{A} + \mathbf{B})x.$$

Composition of two linear operators

Consider the linear operators $f(x) = \mathbf{A}x$ and $g(y) = \mathbf{B}y$ where $\mathbf{A}$ and $\mathbf{B}$ are of order $m \times n$ and $n \times p$, respectively. The vectors x and y are $n \times 1$ and $p \times 1$ dimensional. Then

$$f(g(y)) = \mathbf{A}\mathbf{B}y.$$

Matrices and linear operators

Example

$$f(x) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} x$$

defines a rotation by 90 degrees.

Chapter 3

Vector Spaces

Definition vector space

A vector space over the field of real numbers $\mathbb{R}$ is a set of elements $v \in V$ equipped with a vector addition

$$\begin{array}{rll} + : & V \times V & \mapsto V \\ & (x, y) & \mapsto x + y \end{array}$$

and a scalar multiplication

$$\begin{array}{rll} \cdot : & \mathbb{R} \times V & \mapsto V \\ & (a, x) & \mapsto a \cdot x \end{array}$$

such that the following properties hold:

I. vector addition

1. Associative law: $x + (y + z) = (x + y) + z$ for all $x, y, z \in V$.

2. Commutative law: $x + y = y + x$ for all $x, y \in V$.

3. There is a vector $0 \in V$, called the zero vector such that $x + 0 = x$ for all $x \in V$.

4. For each $x \in V$ there exists a vector $-x \in V$ such that $x + (-x) = 0$.

II. Scalar multiplication

1. Distributive law $(a + b)x = ax + bx$ respectively $a(x + y) = ax + ay$ for all $x, y \in V$, $a, b \in \mathbb{R}$.

2. Associative law: $(ab)x = a(bx)$ for all $x \in V$, $a, b \in \mathbb{R}$.

3. There exists an element $1 \in \mathbb{R}$ called the unit element such that $1 \cdot x = x$ for all $x \in V$.

Examples of vector spaces

- **Vector space $\mathbb{R}^n$**

  V is the set of all vectors in $\mathbb{R}^n$ equipped with the vector addition and scalar multiplication defined in chapter 1.

- **Vector space of all $m \times n$ matrices**

  V is the set of all $m \times n$ matrices equipped with matrix addition and scalar multiplication defined in chapter 2.

Examples of vector spaces

- **Vector space of polynomials of degree n**

  V is the set of all polynomials of degree n, i.e.

$$P(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n.$$

Define for

$$P_1(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n$$

and

$$P_2(t) = b_0 + b_1 t + b_2 t^2 + \cdots + b_n t^n$$

the polynomial addition through

$$P_1(t) + P_2(t) = (a_0 + b_0) + (a_1 + b_1)t + (a_2 + b_2)t^2 + \cdots + (a_n + b_n)t^n.$$

Multiplication with a Scalar $b \in \mathbb{R}$ is defined through

$$bP(t) = ba_0 + ba_1 t + ba_2 t^2 + \cdots + ba_n t^n.$$

Subspaces

Let U be a subset of V. U is called a subspace of V if U itself is a vector space.

Theorem

$U \subseteq V$ is a subspace, if and only if

1. U is not empty,
2. U is closed with respect to vector addition, i.e. for $u_1, u_2 \in U$ we have $u_1 + u_2 \in U$,
3. U is closed with respect to scalar multiplication, i.e. for $u \in U$ we have $k \cdot u \in U$ for $k \in \mathbb{R}$.

Subspaces

The theorem provides us with a recipe to show that U is a subspace:

- Show $0 \in U$.
- Show, that for arbitrary $k \in \mathbb{R}$ and $u, u_1, u_2 \in U$ the vectors ku and $u_1 + u_2$ are contained in U.

Remark:

Let V be an arbitrary vector space. Then the set $\{0\}$ and V are subspaces.

Subspaces

Examples

- $V = \mathbb{R}^3, U := \{(0, a, b) : a, b \in \mathbb{R}\}$
- $V = \mathbb{R}^2, U := \{(y, x) : y = a + bx,\ a, b \in \mathbb{R}\}$.
- Let V be the vector space of polynomials of degree n. U is the subset of polynomials of degree p with $p \leq n$.
- $V = \mathbb{R}^n, U := \{(x_1, \ldots, x_n)' : x_1 = 0\}$.
- $V = \mathbb{R}^n, U := \{(x_1, \ldots, x_n)' : \sum_{i=1}^n x_i = 0\}$.

Linear dependence

A set of n vectors $x_1, x_2, \ldots, x_n \in V$ is called linear independent, if the equation

$$a_1 x_1 + \cdots + a_n x_n = 0$$

with $a_i \in \mathbb{R}$ is true only for

$$a_1 = a_2 = \cdots = a_n = 0.$$

Otherwise $x_1, \ldots, x_n$ are called linear dependent.

Linear dependence

Example

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

Basis and dimension

- A vector space V is finite dimensional or n–dimensional, short $dim(V) = n$, if linear independent vectors $b_1, \ldots, b_n$ exist, such that every vector $x \in V$ can be written as a linear combination of the b_i, i.e.

$$x = a_1 b_1 + \cdots + a_n b_n.$$

- The set $B := \{b_1, \ldots, b_n\}$ is called basis of V.

Basis and dimension

Examples

- Basis of $\mathbb{R}^n$
  The canonical basis is given by

$$E := \{e_i \in \mathbb{R}^n : e_i = (\delta_{i1}, \ldots, \delta_{in})', \quad i = 1, \ldots, n\}$$

  where

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

- Basis of polynomials of degree n

$$f(x) = a_0 \cdot 1 + a_1 \cdot x + \cdots + a_n \cdot x^n,$$

  i.e. $B_0(x) = 1, B_1(x) = x, \cdots, B_n(x) = x^n$.

- Vector space of m times continuously differentiable functions.

Basis and dimension

Some facts

Let V be a n-dimensional vector space.

- The basis of a vector space is not unique.
- For a given basis the representation of a vector through the basis is unique.
- Every basis of V has the same number of elements.
- An arbitrary set of $n + 1$ vectors is linear dependent.
- An arbitrary set of linear independent vectors can be expanded to a basis of V.
- A set of n linear independent vectors forms a basis.

Chapter 4

Rank and determinant

Rank of a matrix

Column and row rank

- The maximum number of linear independent column vectors of a $m \times n$ matrix is called column rank of $\mathbf{A}$.
- The maximum number of linear independent row vectors of $\mathbf{A}$ is called row rank.
- It can be shown that the row and column rank of a matrix is equal.

Rank

The rank of a $m \times n$ matrix $\mathbf{A}$, short $rank(\mathbf{A})$ or $rk(\mathbf{A})$, is defined as the maximum number of linear independent columns or rows of $\mathbf{A}$.

Rank of a matrix

Example

$$\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 2 & 5 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$$

Properties of the rank of a matrix

1. $rk(\mathbf{A}) = rk(-\mathbf{A})$

2. $rk(\mathbf{A}') = rk(\mathbf{A})$

3. $rk(\mathbf{A}) - rk(\mathbf{B}) \leq rk(\mathbf{A}+\mathbf{B}) \leq rk(\mathbf{A}) + rk(\mathbf{B})$

4. $rk(\mathbf{AB}) \leq \min\{rk(\mathbf{A}), rk(\mathbf{B})\}$

5. $rk(\mathbf{I}_n) = n$

Inverse of a matrix

Definition

Let $\mathbf{A}$ be a $n \times n$ matrix. The matrix $\mathbf{A}^{-1}$ is called the inverse of $\mathbf{A}$ if

$$\mathbf{A} \cdot \mathbf{A}^{-1} = \mathbf{A}^{-1} \cdot \mathbf{A} = \mathbf{I}.$$

Existence

The inverse of a $n \times n$ matrix $\mathbf{A}$ exists if and only if $rk(\mathbf{A}) = n$. In this case $\mathbf{A}$ is called regular. In case of existence $\mathbf{A}^{-1}$ is unique.

Inverse of a matrix

Example

$$\mathbf{A} = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 0 & 1 \\ 3 & 5 & 1 \end{pmatrix} \qquad \mathbf{A}^{-1} = \begin{pmatrix} -5 & 2 & 3 \\ 2 & -1 & -1 \\ 5 & -1 & -3 \end{pmatrix}$$

Properties of the matrix inverse

1. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$

2. $(k\mathbf{A})^{-1} = k^{-1}\mathbf{A}^{-1} = \dfrac{1}{k}\mathbf{A}^{-1}$

3. $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$

4. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

5. $(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$

6. $\mathbf{A}$ symmetric $\Longrightarrow \mathbf{A}^{-1}$ symmetric.

7. For $\mathbf{A} = diag(a_1, \ldots, a_n)$ the inverse is given by $\mathbf{A}^{-1} = diag(a_1^{-1}, \ldots, a_n^{-1})$.

8. If $\mathbf{A}$ is orthogonal, then $\mathbf{A}^{-1} = \mathbf{A}'$.

The determinant of a matrix

Consider the matrix

$$\mathbf{A} = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}.$$

The columns a_1 and a_2 of $\mathbf{A}$ form a parallelogram The determinant of $\mathbf{A}$ is defined as the area under the parallelogram

$$det(\mathbf{A}) = 4 \cdot 3 - 2 \cdot 1.$$

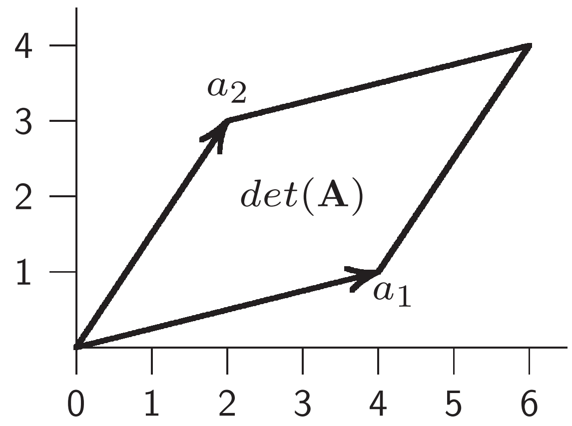

Figure 7: Geometrical visualization of the determinant of a 2×2 matrix.

The determinant of a matrix

The following properties are obvious:

- If the columns of $\mathbf{A}$ are linearly dependent the parallelogram would collapse to a line and would have zero area.
- Hence, if the columns of $\mathbf{A}$ are linearly dependent the determinant of $\mathbf{A}$ is zero.
- The determinant of $\mathbf{A}$ is nonzero if it has full rank.

General definition of the determinant

The determinant $det(\mathbf{A})$ of a $n \times n$ matrix $\mathbf{A}$ is the volume of the hyperparallelogram spanned by the columns of $\mathbf{A}$.

The determinant of a matrix

Example

The columns of the diagonal matrix

$$\mathbf{D} = \text{diag}(d_1, \ldots, d_n)$$

define a box in $\mathbb{R}^n$. Its volume is the product of the length of the sides, i.e.

$$det(\mathbf{D}) = d_1 \cdot d_2 \cdots d_n.$$

Properties of the determinant

1. $det(k\mathbf{A}) = k^n det(\mathbf{A})$
2. $det(\mathbf{A}) \neq 0 \iff rg(\mathbf{A}) = n$
3. $det(\mathbf{AB}) = det(\mathbf{A}) det(\mathbf{B})$
4. $det(A^{-1}) = \dfrac{1}{det(\mathbf{A})}$
5. $\mathbf{A}$ orthogonal $\Longrightarrow det(\mathbf{A}) = \pm 1$
6. $det(\mathbf{A}) = det(\mathbf{A}')$

Computation of the determinant

2×2 matrix

For a 2×2 matrix

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

we have $det(\mathbf{A}) = a \cdot d - b \cdot c$.

Diagonal matrix

For a diagonal matrix $\mathbf{D} = diag(d_1, \dots, d_n)$ we have $det(\mathbf{D}) = d_1 \cdot d_2 \cdots d_n$.

Computation of the determinant

General $n \times n$ matrix

The determinant $det(\mathbf{A})$ of a $n \times n$ matrix is given by $det(\mathbf{A}) = a_{11}$ for $n = 1$ and

$$\begin{aligned} det(\mathbf{A}) &= (-1)^{i+1} a_{i1} det(\mathbf{A}_{i1}) + \cdots + (-1)^{i+n} a_{in} det(\mathbf{A}_{in}) \\ &= (-1)^{1+j} a_{1j} det(\mathbf{A}_{1j}) + \cdots + (-1)^{n+j} a_{nj} det(\mathbf{A}_{nj}) \end{aligned}$$

for $n > 1$ where $\mathbf{A}_{ij}$ is obtained from $\mathbf{A}$ by deleting the i-th row and j-th column.

Computation of the determinant

Example

$$\mathbf{A} = \begin{pmatrix} 3 & 2 & 1 \\ 0 & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix}$$

Chapter 5

Eigenvalues and eigenvectors

Complex numbers

- A complex number x is an ordered pair $x = (x_1, x_2)$ of real numbers x_1 and x_2.
- Two complex numbers x and y are equal if $x_1 = y_1$ and $x_2 = y_2$.
- For complex numbers the following addition and multiplication is defined:

$$\begin{aligned} x + y &= (x_1, x_2) + (y_1, y_2) &= (x_1 + y_1, x_2 + y_2) \\ x \cdot y &= (x_1, x_2) \cdot (y_1, y_2) &= (x_1 y_1 - x_2 y_2, x_1 y_2 + x_2 y_1) \end{aligned}$$

- The set of complex numbers is denoted by the symbol $\mathbb{C}$.

Complex numbers

Some remarks

- Because of
$$(x_1, 0) + (y_1, 0) = (x_1 + y_1, 0)$$
and
$$(x_1, 0)(y_1, 0) = (x_1 y_1, 0)$$
the real number x can be identified with the complex number $(x, 0)$, i.e. $\mathbb{R}$ is a subset of $\mathbb{C}$.

- The complex number $i = (0, 1)$ is of particular importance. We have:
$$i^2 = i \cdot i = (0, 1)(0, 1) = (-1, 0) = -1$$
and the representation of $x = (x_1, x_2)$ as
$$x = (x_1, x_2) = (x_1, 0) + (x_2, 0) \cdot (0, 1) = x_1 + x_2 \cdot i.$$

Using this representation complex numbers can be handled just like real numbers

- The complex number

$$\bar{x} = x_1 - x_2 \cdot i$$

is called the complex conjugate to

$$x = x_1 + x_2 \cdot i$$

We have

$$x \cdot \bar{x} = x_1^2 + x_2^2.$$

Complex numbers

Example

$$(3,2) = 3 + 2i \qquad (2,1) = 2 + 1i \qquad \frac{1}{3+2i}$$

Complex numbers

Example

Quadrat equation

$$x^2 + p = 0.$$

Complex numbers

Absolute value

The absolute value of the complex number

$$x = (x_1, x_2) = x_1 + x_2 \cdot i$$

is defined as

$$|x| = \sqrt{x_1^2 + x_2^2}.$$

Example

$$x = 4 + 3i$$

Eigenvalues

Let $\mathbf{A}$ be a quadratic $n \times n$ matrix.

- The (possibly) complex number $\lambda \in \mathbb{C}$ is called eigenvalue of $\mathbf{A}$, if a vector $x \in \mathbb{C}^n$ with $x \neq 0$ exists, such that:

$$\mathbf{A}x = \lambda x \text{ resp. } (\mathbf{A} - \lambda \mathbf{I})x = 0$$

- The vector x is called the eigenvector with respect to the eigenvalue λ.

- The eigenvalues of a square matrix $\mathbf{A}$ are the roots of the characteristic polynomial

$$q(\lambda) := det(\mathbf{A} - \lambda \mathbf{I}).$$

Eigenvalues

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 2 & -2 \end{pmatrix} \qquad \mathbf{B} = \begin{pmatrix} 2 & -1 \\ 8 & -2 \end{pmatrix}$$

Properties of the eigenvalues of a matrix

Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of the square matrix $\mathbf{A}$.

1. $det(\mathbf{A}) = \prod_{i=1}^{n} \lambda_i$

2. $tr(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$

3. $\mathbf{A}$ is regular if and only if $\lambda_i \neq 0$ for all i=1,. . . ,n.

4. The matrices $\mathbf{A}$ und $\mathbf{A}'$ share the same characteristic polynomial. Therefore the eigenvalues of the two matrices are identical.

5. If λ is an eigenvalue of the regular matrix $\mathbf{A}$, then $\dfrac{1}{\lambda}$ is an eigenvalue of $\mathbf{A}^{-1}$.

6. The eigenvalues of a diagonal matrix $\mathbf{D}$ are given by the elements of the main diagonal.

7. The eigenvalues λ_i of an orthogonal matrix $\mathbf{A}$ are either 1 or -1, i.e. $\lambda_i = \pm 1$.

8. The eigenvalues of an idempotent matrix $\mathbf{A}$ are either 1 or 0.

9. The eigenvalues of a symmetric matrix are real valued.

Chapter 6

Quadratic forms and definite matrices

Quadratic forms

Let $\mathbf{A}$ be a symmetric $n \times n$ matrix. A quadratic Form in $x \in \mathbb{R}^n$ is defined as definiert as:

$$Q(x) = x'\mathbf{A}x = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j =$$

$$= \sum_{i=1}^{n} a_{ii}x_i^2 + 2 \cdot \sum_{i=1}^{n}\sum_{j>i} a_{ij}x_i x_j$$

Quadratic forms

The quadratic form $x'\mathbf{A}x$ and the matrix $\mathbf{A}$ are called

1. positive definite, if $x'\mathbf{A}x > 0$ for all $x \neq 0$. Short: $\mathbf{A} > 0$.

2. positive semidefinite, if $x'\mathbf{A}x \geq 0$ and $x'\mathbf{A}x = 0$ for at least one $x \neq 0$.

3. nonnegative definite, if $x'\mathbf{A}x$ respectively $\mathbf{A}$ are either positive or positive semidefinite. Short: $\mathbf{A} \geq 0$.

4. negative definite, if $-\mathbf{A}$ is positive definite.

5. negativ semidefinite, if $-\mathbf{A}$ positive semidefinite.

6. indefinite in all other cases.

Eigenvalues and definite matrices

Let $\mathbf{A}$ be a symmetric matrix with (real valued) eigenvalues $\lambda_1, \ldots, \lambda_n$. Then $\mathbf{A}$

1. is positive definite , if and only if $\lambda_i > 0$ for $i = 1, \ldots, n$,
2. positive semi definite if and only if $\lambda_i \geq 0$ for $i = 1, \ldots, n$ and $\lambda_i = 0$ for some λ_i,
3. negative definite, if and only if $\lambda_i < 0$ for alle $i = 1 \ldots, n$,
4. negative semi definite, if and only if $\lambda_i \leq 0$ for $i = 1, \ldots, n$ and $\lambda_i = 0$ for some λ_i,
5. indefinite, if and only if the matrix $\mathbf{A}$ has positive and negative eigenvalues.

Properties of positive definite matrices

Let $\mathbf{A}$ be positive definite.

1. $\mathbf{A}$ is regular.

2. The diagonal elements a_{ii} are positive, i.e. $a_{ii} > 0$ for $i = 1, \ldots, n$.

3. $tr(\mathbf{A}) > 0$

4. Let $\mathbf{B}$ be positive semi definite. Then $\mathbf{A} + \mathbf{B}$ is positive definite.

Chapter 7

Linear equation systems

Linear equations

A linear equation system with unknowns $x_1, \ldots, x_n \in \mathbb{R}$ is a system of m equations of the form

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & c_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & c_2 \\ \vdots & & \vdots & & & & \vdots & = & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & c_m \end{array}$$

where the scalars $a_{ij}, c_i \in \mathbb{R}$ are known coefficients.

Combining the scalars a_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, n$, to the $m \times n$ matrix $\mathbf{A}$ and x_i and c_i to the $n \times 1$ respectively $m \times 1$ column vectors x and c, the equation system can be written in matrix notation as

$$\mathbf{A}x = c.$$

Linear equations

Some facts

- The system $\mathbf{A}x = c$ is solvable if and only if $rk(\mathbf{A}\, c) = rk(\mathbf{A})$.
- The set of solutions is of the general form

  $$L = x_0 + L_0,$$

  where x_0 is a particular solution of $\mathbf{A}x = c$ and L_0 is the set of solutions of the homogenous system $\mathbf{A}x = \mathbf{0}$.
- The set of solutions of $\mathbf{A}x = \mathbf{0}$ is called the nullspace of $\mathbf{A}$:

  $$N(\mathbf{A}) = \{x \in \mathbb{R}^n : \mathbf{A}x = \mathbf{0}\}$$

  The nullspace is a subspace of $\mathbf{A}$.

Cholesky decomposition

A symmetric and positive definite $n \times n$ matrix $\mathbf{A}$ can be uniquely represented as

$$\mathbf{A} = \mathbf{L}\mathbf{L}',$$

where $\mathbf{L}$ is a lower triangular matrix with positive diagonal elements.

For $j = 1, \ldots, n$ and $i = j+1, \ldots, n$ we have

$$l_{jj} = \left(a_{jj} - \sum_{k=1}^{j-1} l_{jk}^2 \right)^{\frac{1}{2}}$$

$$l_{ij} = \frac{1}{l_{jj}} \left(a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk} \right).$$

Cholesky decomposition

Example

$$\mathbf{A} = \begin{pmatrix} 4 & 6 & 6 \\ 6 & 13 & 11 \\ 6 & 11 & 14 \end{pmatrix}$$

Cholesky decomposition

Example

$$\mathbf{A} = \begin{pmatrix} 4 & 2 & 4 & 4 \\ 2 & 10 & 17 & 11 \\ 4 & 17 & 33 & 29 \\ 4 & 11 & 29 & 39 \end{pmatrix}$$

Computing the determinant

Using the Cholesky decomposition the determinant of a matrix $\mathbf{A}$ can be computed as

$$det(\mathbf{A}) = det(\mathbf{L}\mathbf{L}') = det(\mathbf{L})det(\mathbf{L}') = (l_{11} \cdot l_{22} \cdots l_{nn})^2.$$

Example

Solving linear equation systems

To solve the linear equation system $\mathbf{A}x = b$ with $\mathbf{A} > 0$ we proceed as follows:

1. Compute the Cholesky decomposition $\mathbf{A} = \mathbf{L}\mathbf{L}'$.
2. Solve recursively the linear equation system $\mathbf{L}y = b$ starting with y_1. Proceed with $y_1, y_2, \ldots$.
3. Solve recursively the linear equation system $\mathbf{L}'x = y$ starting with x_n. Proceed with $x_{n-1}, x_{n-2}, \ldots$

Solving linear equation systems

Example

$$\begin{pmatrix} 4 & 2 & 4 & 4 \\ 2 & 10 & 17 & 11 \\ 4 & 17 & 33 & 29 \\ 4 & 11 & 29 & 39 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 44 \\ 133 \\ 269 \\ 257 \end{pmatrix}$$

Chapter 8

Matrix calculus

Differentiation with respect to a vector

Let $x = (x_1, \ldots, x_n)'$ be a $(n \times 1)$-vector and $f(x)$ a real function differentiable with respect to the elements x_i of x.

The $(n \times 1)$-vector

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

is then called differential of f with respect to x. We denote by

$$\frac{\partial f}{\partial x'} = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right)$$

the transpose of $\frac{\partial f}{\partial x}$.

Differentiation with respect to a vector

Example

$$f(\mathbf{x}) = \mathbf{y}'\mathbf{x} = \sum_{i=1}^{n} y_i x_i,$$

where $\mathbf{y} = (y_1, \ldots, y_n)'$ is constant.

Differentiation of a Vector Function with Respect to a Vector

Let $x = (x_1, \ldots, x_n)'$ be a $(n \times 1)$-vector and $f(x) = (f_1(x), \ldots, f_m(x))'$ a $(m \times 1)$ vector function differentiable with respect to the elements x_i of x. The $(n \times m)$-matrix

$$\frac{\partial f}{\partial x} = \left(\frac{\partial f_j}{\partial x_i}\right) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial f_1}{\partial x_n} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

is then called differential of f with respect to x. We denote by

$$\frac{\partial f}{\partial x'} = \left(\frac{\partial f}{\partial x}\right)' = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

the transpose of $\frac{\partial f}{\partial x}$.

Differentiation of a Vector Function with Respect to a Vector

Example

$$\frac{\partial \mathbf{A}x}{\partial x} = \mathbf{A}'$$

Differentiation Rules

Assume that $\mathbf{A}$ is a matrix and a, x and y are vectors.

1. $\dfrac{\partial y'x}{\partial x} = y$

2. $\dfrac{\partial x'\mathbf{A}x}{\partial x} = (\mathbf{A} + \mathbf{A}')x$

3. If $\mathbf{A}$ is symmetric, then

$$\frac{\partial x'\mathbf{A}x}{\partial x} = 2\mathbf{A}x = 2\mathbf{A}'x.$$

4. $\dfrac{\partial \mathbf{A}x}{\partial x} = \mathbf{A}'$

5. $\dfrac{\partial \mathbf{A}x}{\partial x'} = \mathbf{A}$

Local Extremes

Let $x = (x_1, \ldots, x_n)'$ be a $(n \times 1)$-vector and $f(x)$ a real function differentiable with respect to the elements x_i of x. Define the vector

$$s(x) = \frac{\partial f(x)}{\partial x}$$

of first derivatives and the matrix

$$\mathbf{H}(x) = \frac{\partial s(x)}{\partial x'} = \begin{pmatrix} \frac{\partial s_1(x)}{\partial x_1} & \cdots & \frac{\partial s_1(x)}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial s_n(x)}{\partial x_1} & \cdots & \frac{\partial s_n(x)}{\partial x_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \end{pmatrix}$$

of second derivatives. $\mathbf{H}(x)$ is also called *Hessian matrix* or simply the *Hessian*.

A necessary condition for $x = x_0$ being a local extreme of f is

$$s(x_0) = \mathbf{0}. \tag{1}$$

If (1) is true, the following sufficient condition holds:

- If $\mathbf{H}(x_0)$ is positive definite x_0 is a local minimum.
- If $\mathbf{H}(x_0)$ is negative definite x_0 is a local maximum.

Local Extremes

Example

$$f(x) = (y - \mathbf{Z}x)'(y - \mathbf{Z}x)$$

where the $n \times p$ matrix $\mathbf{Z}$ has full column rank, x and y are $p \times 1$ and $n \times 1$ vectors.

Chapter 9

Stochastic vectors and matrices

Random vectors

- The vector $X = (X_1, \ldots, X_p)'$ is called a random vector or p-dimensional random variable, if the components $X_1, \ldots, X_p$ are one dimensional random variables.

- The vector X is called continuous if there is a function $f(x) = f(x_1, \ldots, x_p) \geq 0$ such that

$$P(a_1 \leq X_1 \leq b_1, \ldots, a_p \leq X_p \leq b_p) = \int_{a_p}^{b_p} \cdots \int_{a_1}^{b_1} f(x_1, \ldots, x_p)\, dx_1 \ldots dx_p.$$

  The function f is called (joint) probability density function (p.d.f.) of X.

Random vectors

- The random vector X is called discrete, if X has only values in a finite or countable set $\{x_1, x_2, \ldots\} \subset \mathbb{R}^p$. The function f with

$$f(x) = \begin{cases} P(X = x) & x \in \{x_1, x_2, \ldots\} \\ 0 & \text{else} \end{cases}$$

is called probability function or discrete p.d.f. of X.

Random vectors

Example

Consider the 2-dimensional continuous random vector $x = (x_1, x_2)'$ with pdf

$$f(x_1, x_2) = \begin{cases} 0.8(x_1 + x_2 + x_1 x_2) & \begin{array}{l} 0 \le x_1 \le 1 \\ 0 \le x_2 \le 1 \end{array} \\ 0 & \text{else.} \end{cases}$$

Random vectors

Example

Consider the zwo dimensional random vector $(X, Y)'$ with

$$X = \text{proseminar grade} \qquad Y = \text{final exam grade}$$

We have the following distribution:

Y/X	1	2	3	4
1	$\frac{50}{1007}$	$\frac{40}{1007}$	$\frac{12}{1007}$	$\frac{6}{1007}$
2	$\frac{55}{1007}$	$\frac{97}{1007}$	$\frac{54}{1007}$	$\frac{10}{1007}$
3	$\frac{28}{1007}$	$\frac{100}{1007}$	$\frac{68}{1007}$	$\frac{33}{1007}$
4	$\frac{13}{1007}$	$\frac{79}{1007}$	$\frac{75}{1007}$	$\frac{36}{1007}$
5	$\frac{9}{1007}$	$\frac{44}{1007}$	$\frac{119}{1007}$	$\frac{79}{1007}$

Marginal and conditional distribution

- Let the p-dimensional random vector $X = (X_1, \ldots, X_p)'$ be partitioned into the p_1-dimensional vector X_1 and the p_2-dimensional Vector X_2, i.e. $X = (X_1', X_2')'$.

- The p_1-dimensional p.d.f. or probability function $f_{X_1}(x_1)$ of X_1 is then called marginal p.d.f. or marginal probability function of X. It is given by

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_{p_1+1} \ldots dx_p$$

for continuous random vectors, and

$$f_{X_1}(x_1) = \sum_{x_2} f(x_1, x_2)$$

for discrete random vectors.

Marginal and conditional distribution

- The conditional p.d.f. or probability function of X_1 given $X_2 = x_2$ is defined as

$$f(x_1|x_2) = \begin{cases} \dfrac{f(x_1, x_2)}{f_{X_2}(x_2)} & \text{for } f_{X_2}(x_2) > 0 \\ 0 & \text{else.} \end{cases}$$

The marginal and conditional p.d.f.'s or probability functions for X_2 are defined in complete analogy.

Marginal and conditional distribution

Example

Consider the 2-dimensional continuous random vector $x = (x_1, x_2)'$ with pdf

$$f(x_1, x_2) = \begin{cases} 0.8(x_1 + x_2 + x_1 x_2) & \begin{array}{l} 0 \leq x_1 \leq 1 \\ 0 \leq x_2 \leq 1 \end{array} \\ 0 & \text{else.} \end{cases}$$

Marginal and conditional distribution

Example

Y/X	1	2	3	4
1	$\frac{50}{1007}$	$\frac{40}{1007}$	$\frac{12}{1007}$	$\frac{6}{1007}$
2	$\frac{55}{1007}$	$\frac{97}{1007}$	$\frac{54}{1007}$	$\frac{10}{1007}$
3	$\frac{28}{1007}$	$\frac{100}{1007}$	$\frac{68}{1007}$	$\frac{33}{1007}$
4	$\frac{13}{1007}$	$\frac{79}{1007}$	$\frac{75}{1007}$	$\frac{36}{1007}$
5	$\frac{9}{1007}$	$\frac{44}{1007}$	$\frac{119}{1007}$	$\frac{79}{1007}$

Expectation or mean vector

Let $X = (X_1, \dots, X_p)'$ be a p-dimensional random vector. Then

$$E(X) = \mu = (\mu_1, \dots, \mu_p)' = (E(X_1), \dots, E(X_p))'$$

is called mean vector of X.

Example

Covariance and correlation matrix

The covariance matrix $Cov(X) = \mathbf{\Sigma}$ of a p-dimensional random vector X is defined as

$$Cov(X) = \mathbf{\Sigma} \quad = \quad E(X-\mu)(X-\mu)' = \begin{pmatrix} \sigma_{11} & \dots & \sigma_{1p} \\ \vdots & & \vdots \\ \sigma_{p1} & \dots & \sigma_{pp} \end{pmatrix},$$

where $\sigma_{ij} = Cov(X_i, X_j)$, $i \neq j$, is the covariance between X_i and X_j, and $\sigma_{ii} = \sigma_i^2 = Var(X_i)$ is the variance of X_i.

Covariance and correlation matrix

The correlation matrix $\mathbf{R}$ of X is defined as

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \vdots & & & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & 1 \end{pmatrix},$$

where

$$\rho_{ij} = \frac{\mathsf{Cov}(X_i, X_j)}{\sqrt{\mathsf{Var}(X_i) \cdot \mathsf{Var}(X_j)}}.$$

Covariance and correlation matrix

Example

Properties of expectations and covariance matrices

Let X and Y be random vectors and $\mathbf{A}, \mathbf{B}, a, b$ matrices and vectors.

1. $E(X+Y) = E(X) + E(Y)$

2. $E(\mathbf{A}X + b) = \mathbf{A} \cdot E(X) + b$

3. $Cov(X) = E(XX') - \mu\mu'$

4. $Var(a'X) = a'Cov(X)a = \sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j \sigma_{ij}$

5. The covariance matrix is symmetric and positive semi definite.

6. $Cov(\mathbf{A}X + b) = \mathbf{A}Cov(X)\mathbf{A}'$

Multivariate Normal Distribution

- A continuous p-dimensional random vector $X = (X_1, X_2, \ldots, X_p)'$ is said to have a multivariate normal (or Gaussian) distribution if it has p.d.f.

$$f(x) = (2\pi)^{-\frac{p}{2}}|\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(x-\mu)'\mathbf{\Sigma}^{-1}(x-\mu)\right]$$

  with $\mu \in \mathbb{R}^p$ and positive definite $(p \times p)$-matrix $\mathbf{\Sigma}$.

- It can be shown that $E(X) = \mu$ and Cov$(X) = \mathbf{\Sigma}$.

- We write

$$X \sim N_p(\mu, \mathbf{\Sigma}),$$

  The special case $\mu = \mathbf{0}$ and $\mathbf{\Sigma} = \mathbf{I}$ is called the (multivariate) standard normal distribution.

Multivariate Normal Distribution

Example

$X = (X_1, X_2, X_3, X_4)' \sim N(\mu, \mathbf{\Sigma})$ with

$$\mu = (1, 2, 3, 4)' \qquad \mathbf{\Sigma} = \begin{pmatrix} 3 & 1 & 0 & 1 \\ 1 & 4 & 2 & 0 \\ 0 & 2 & 5 & 2 \\ 1 & 0 & 2 & 4 \end{pmatrix}$$

Marginal and conditional distributions

- Let the multivariate normal random variable $X \sim N(\mu, \mathbf{\Sigma})$ be partitioned into the subvectors $Y = (X_1, \ldots, X_r)'$ and $Z = (X_{r+1}, \ldots, X_p)'$, i.e.

$$X = \begin{pmatrix} Y \\ Z \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_Y \\ \mu_Z \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_Y & \mathbf{\Sigma}_{YZ} \\ \mathbf{\Sigma}_{ZY} & \mathbf{\Sigma}_Z \end{pmatrix}.$$

- Then Y has an r-dimensional normal distribution $Y \sim N(\mu_Y, \mathbf{\Sigma}_Y)$.

- The conditional distribution of Y given Z is again multivariate normal with mean
$$\mu_{Y|Z} = \mu_Y + \mathbf{\Sigma}_{YZ} \cdot \mathbf{\Sigma}_Z^{-1}(Z - \mu_Z)$$
and covariance matrix
$$\mathbf{\Sigma}_{Y|Z} = \mathbf{\Sigma}_Y - \mathbf{\Sigma}_{YZ}\mathbf{\Sigma}_Z^{-1}\mathbf{\Sigma}_{ZY}.$$

Marginal and conditional distributions

- Furthermore, $\mathbf{Y}$ and $\mathbf{Z}$ are independent if and only if $\mathbf{Y}$ and $\mathbf{Z}$ are uncorrelated, i.e. if $\mathbf{\Sigma}_{YZ} = \mathbf{\Sigma}_{ZY} = \mathbf{0}$.

- The equivalence is generally not true for non-normal random vectors: If $\mathbf{Y}$ and $\mathbf{Z}$ are independent they are also uncorrelated, but in general $\mathbf{\Sigma}_{ZY} = \mathbf{0}$ does not imply independence.

Marginal and conditional distributions

Example

$Y = (X_1, X_2)'$, $Z = (Z_1, Z_2)'$

Linear transformations

Assume $X \sim N_p(\mu, \mathbf{\Sigma})$ is multivariate normal. Then the linear transformation

$$Y = \mathbf{D}X + d$$

with the $m \times p$ matrix $\mathbf{D}$ and the $m \times 1$ vector d is again multivariate normal

$$Y \sim N_m(\mathbf{D}\mu + d, \mathbf{D\Sigma D}').$$

Example (simulation)

Chapter 10

Convergence of random variables

Random variables as mappings

- Consider a random experiment with possible outcome in the space Ω and a corresponding probability measure P.
- A random variable X is a mapping that assigns every $\omega \in \Omega$ a real value X, i.e. $X(\omega){=}x$. More specifically

$$\begin{aligned} X : \Omega &\to \mathbb{R} \\ \omega &\mapsto X(\omega) = x. \end{aligned}$$

Random variables as mappings

Example (rolling a device twice)

Tschebyschov inequality

Let X be a random variable with expected value μ and variance σ^2. For $\epsilon > 0$ the inequality

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

or equivalently

$$P(|X - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2}$$

holds.

Example

Let X_1,X_2,. . . be a sequence of i.i.d. random variables with finite expected value μ and finite variance σ^2.

Tschebyschov inequality

Proof of the inequality:

Convergence of random variables

- We consider in the following a sequence of random variables X_1,X_2,X_3,. . . and investigate the limit behavior for $n \to \infty$.
- Consider e.g. the estimator $\hat{\mu} = \bar{X}$ for the expected value μ of a distribution in dependence of n:

$$
\begin{aligned}
\bar{X}_1 &= \frac{1}{1} X_1 \\
\bar{X}_2 &= \frac{1}{2}(X_1 + X_2) \\
\bar{X}_3 &= \frac{1}{3}(X_1 + X_2 + X_3) \\
&\vdots \\
\bar{X}_n &= \frac{1}{n}(X_1 + \cdots + X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i
\end{aligned}
$$

Almost sure convergence

- A sequence X_n of random variables converges almost surely to a random variable X if, for every $\epsilon > 0$

$$P(\lim_{n\to 0}|X_n - X| < \epsilon) = 1.$$

- The definition stats that $X_n(\omega)$ converges to $X(\omega)$ for all $\omega \in \Omega$, except perhaps for $\omega \in N$ where $N \subset \Omega$ and $P(N)=0$.

- Notation: $X_n \xrightarrow{\text{a.s.}} X$

Almost sure convergence

Example

Ω=[0,1] and P the uniform distribution. Define

$$X_n(\omega) = \omega + \omega^n$$
$$X(\omega) = \omega.$$

Convergence in probability

- X_n converges to X in probability, if for $\epsilon > 0$

$$\lim_{n\to\infty} P(\left|X_n - X\right| \geq \epsilon) = 0$$

or

$$\lim_{n\to\infty} P(\left|X_n - X\right| < \epsilon) = 1$$

- Notation: $X_n \xrightarrow{\text{p}} X$.

Convergence in probability

Example (weak law of large numbers)

- Let X_1,X_2,. . . be a sequence of i.i.d. random variables with finite expected value μ and finite variance σ^2. Then

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{p}} \mu \qquad \text{for} \quad n \to \infty$$

- $\bar{X}_n \xrightarrow{\text{a.s.}} \mu$ can be established as well. Then we speak of the strong law of large numbers.

- The law says, that for large n

$$\left|\bar{X}_n - \mu\right| < \epsilon$$

with high probability.

Convergence in probability

Proof of the weak law of large numbers:

Convergence in probability

Example (convergence in probability, not almost sure)

Let Ω=[0,1] and P is the uniform distribution. Define X_1,X_2,. . . as follows:

$$\begin{array}{rcl} X_1(\omega) & = & \omega + I_{[0,1]}(\omega) \\ X_2(\omega) & = & \omega + I_{[0,\frac{1}{2}]}(\omega) \\ X_3(\omega) & = & \omega + I_{[\frac{1}{2},1]}(\omega) \\ X_4(\omega) & = & \omega + I_{[0,\frac{1}{3}]}(\omega) \\ X_5(\omega) & = & \omega + I_{[\frac{1}{3},\frac{2}{3}]}(\omega) \\ X_6(\omega) & = & \omega + I_{[\frac{2}{3},1]}(\omega) \\ & \vdots & \end{array}$$

Define $X(\omega)$=ω.

Convergence in probability

Example (convergence in probability, not almost sure)

Convergence in the r-th mean

- X_n converges to X in the r-th mean, if

$$E(|X_n^r|) < \infty \qquad \text{for all } n$$

  and

$$\lim_{n\to\infty} E(|X_n - X|^r) = 0.$$

- Notation: $X_n \xrightarrow{\text{r}} X$.

- For r=2 we say that X_n converges in mean square to X.

Convergence in the r-th mean

Example (arbitrary distribution)

Let X_1,X_2,. . . be a sequence of i.i.d. random variables with finite expected value μ and finite variance σ^2.

Convergence in distribution

- X_n converges to X in distribution, if

$$\lim_{n\to\infty} P(X_n \leq x) = P(X \leq x)$$

or

$$\lim_{n\to\infty} F_{X_n}(x) = F_x(x)$$

at all points x where $F_x(x)$ is continuous.

- Notation: $X_n \xrightarrow{\mathsf{d}} X$.

Convergence in distribution

- Note that if

$$f_n \to f \qquad n \to \infty,$$

  where f_n, f are the probability functions (or densities), then the distributions defined through f_n converge to the distribution defined through f.

- The reverse is not correct in general: Convergence in distribution does not imply that the densities converge.

Convergence in distribution

Example (convergence of the Binomial to the Poisson distribution)

- Let $X_n \sim B(n,\pi)$ with probability function

$$f(x) = P(X = x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}$$

- For $n \to \infty$, $n\pi = \lambda$ or $\frac{\lambda}{n} = \pi$ the probability function $f(x)$ converges to the probability function of the Poisson distribution, i.e.

$$\lim_{n\to\infty} f(x) = \frac{\lambda^x \exp(-\lambda)}{x!}$$

Convergence in distribution

Proof:

Convergence in distribution

Application of the limit theorem

- If $X \sim B(n, \pi)$, n large, π small, i.e. $\lambda = n\pi$ moderate (rule of thump $n > 30$, $\pi \leq 0.05$) then X can be approximated by the Poisson distribution with parameter $\lambda = n\pi$.

- E.g. $X \sim B(40, 0.01)$, then $\lambda = 40 \cdot 0.01 = 0.4$ and

$$P(X = 2) \approx \frac{0.4^2}{2!} \exp(-0.4) = 0.0536$$

Exact: $P(Y = 2) = 0.0532$

Convergence in distribution

Example

The sentence “at all x for which $F_X(x)$ is continuous” matters! Let

$$X_n \sim N\left(0, \frac{1}{n}\right)$$

and X a degenerated distribution at 0, i.e. $P(X=0)=1$.

Relationships among modes of convergence