

Formelsammlung zur Einführung in die angewandte Stochastik

Website: [RWTHmoodle](#)

Kurs: Einführung in die angewandte Stochastik (VO) [23ss-
11.02.117]

Buch: Formelsammlung zur Einführung in die angewandte
Stochastik

Gedruckt von: Mause, Nils

Datum: Dienstag, 25. April 2023, 12:47

Inhaltsverzeichnis

1. Mathematik

- 1.1. Folgen, Summen, Reihen
- 1.2. Finanzmathematik
- 1.3. Funktionen
- 1.4. Differentialrechnung
- 1.5. Taylorpolynom, Taylorentwicklung
- 1.6. Integration
- 1.7. Vektoren
- 1.8. Matrizen
- 1.9. Lineare Gleichungssysteme
- 1.10. Differenzengleichungen, Eigenwert, Eigenvektor
- 1.11. Funktionen mit mehreren Variablen
- 1.12. Optimierung von Funktionen
- 1.13. Mehrdimensionale Integration

2. Wahrscheinlichkeitsrechnung

- 2.1. Grundbegriffe
- 2.2. Zufallsvariablen und ihre Verteilung
- 2.3. Erwartungswert, Varianz, Momente und Entropie
- 2.4. Diskrete Verteilungsmodelle
- 2.5. Stetige Verteilungsmodelle
- 2.6. Zufallszahlen
- 2.7. Zufallsvektoren
- 2.8. Grenzwertsätze
- 2.9. Mehrdimensionale Verteilungen
- 2.10. (Erzeugende Funktionen)
- 2.11. (Markov-Ketten)

3. Inferenzstatistik

- 3.1. Statistische Schätzer
- 3.2. Gütekriterien für statistische Schätzer
- 3.3. Testverteilungen
- 3.4. Konfidenzintervalle
- 3.5. Einführung in die statistische Testtheorie
- 3.6. 1-Stichproben-Tests
- 3.7. 2-Stichproben-Tests
- 3.8. Korrelationstests
- 3.9. Analyse von Kontingenztafeln
- 3.10. Lineares Regressionsmodell
- 3.11. (Multiple Lineare Regression)
- 3.12. (Elemente der Bayes-Statistik)

4. Deskriptive Statistik

- 4.1. Grundbegriffe der deskriptive Statistik
- 4.2. Aufbereitung von univariaten Daten
- 4.3. Lagemaße
- 4.4. Streuung
- 4.5. Schiefe und Symmetrie
- 4.6. Quantile
- 4.7. Konzentrationsmessung
- 4.8. Kontingenztafeln, Randverteilungen und Kontingenzkoeffizienten
- 4.9. Empirische Kovarianz, Korrelationskoeffizient und Rangkorrelationskoeffizient
- 4.10. Einfache lineare Regression und KQ-Methode

1. Mathematik

1.1. Folgen, Summen, Reihen

Folgen

Eine Folge $(a_n)_{n \in \mathbb{N}}$,

$$a_1, a_2, \dots,$$

von reellen Zahlen heißt **Folge**. Der **Index** n durchläuft die **Indexmenge** der natürlichen Zahlen \mathbb{N} . Oft tritt auch \mathbb{N}_0 als Indexmenge auf. Allgemeiner betrachtet man Indexmengen $I \subset \mathbb{N}_0$. Ist die Indexmenge endlich, dann heißt die Folge **endlich**.

Eigenschaften einer Folge $(a_n)_{n \in \mathbb{N}}$:

- monoton wachsend: $a_n \leq a_{n+1}$ für alle $n \in \mathbb{N}$
- streng monoton wachsend: $a_n < a_{n+1}$ für alle $n \in \mathbb{N}$
- monoton fallend: $a_n \geq a_{n+1}$ für alle $n \in \mathbb{N}$
- streng monoton fallend: $a_n > a_{n+1}$ für alle $n \in \mathbb{N}$
- beschränkt nach oben: Es gibt eine Zahl $C > 0$ mit $a_n \leq C$ für alle $n \in \mathbb{N}$
- beschränkt nach unten: Es gibt eine Zahl $C < 0$ mit $a_n \geq C$ für alle $n \in \mathbb{N}$
- beschränkt: Es gibt eine Zahl $C > 0$ mit: $|a_n| \leq C$ für alle $n \in \mathbb{N}$

Jede monoton wachsende oder monoton fallende Folge, die zusätzlich beschränkt ist, konvergiert.

Rechnen mit konvergenten Folgen:

•

$$\lim_{n \rightarrow \infty} (a_n + b_n) = \lim_{n \rightarrow \infty} a_n + \lim_{n \rightarrow \infty} b_n$$

•

$$\lim_{n \rightarrow \infty} (a_n - b_n) = \lim_{n \rightarrow \infty} a_n - \lim_{n \rightarrow \infty} b_n$$

•

$$\lim_{n \rightarrow \infty} (a_n \cdot b_n) = \lim_{n \rightarrow \infty} a_n \cdot \lim_{n \rightarrow \infty} b_n$$

- Falls $\lim_{n \rightarrow \infty} b_n \neq 0$, dann gilt

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \frac{\lim_{n \rightarrow \infty} a_n}{\lim_{n \rightarrow \infty} b_n}$$

Summen

Die Summe von n Zahlen x_1, \dots, x_n wird mit

$$\sum_{i=1}^n x_i = x_1 + \dots + x_n$$

notiert.

Endliche geometrische Summe:

$$\sum_{i=0}^n x^i = \frac{1 - x^{n+1}}{1 - x}$$

Weitere Summen:

$$\sum_{i=1}^n i = \frac{n(n+1)}{2}$$

$$\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$$

Reihen

Die Folge der sukzessiven Summen der Zahlen a_0, a_1, \dots , d.h. der **Partialsummen**

$$s_n = \sum_{k=0}^n a_k, \quad n = 0, 1, 2, \dots,$$

heißt **Reihe**. Notation: $s_n, n \in \mathbb{N}_0$ oder auch $\sum_{k=0}^{\infty} a_k$. Die Summation kann auch bei 1 beginnen.

Die Reihe konvergiert gegen s , wenn $s_n \rightarrow s, n \rightarrow \infty$. Man schreibt dann: $\sum_{k=0}^{\infty} a_k = s$. Wenn $\sum_{k=0}^{\infty} |a_k|$ konvergiert, heißt die Reihe **absolut konvergent**. Wenn die Reihe nicht konvergiert, heißt sie **divergent**.

Quotientenkriterium

Ab $n_0 \in \mathbb{N}_0$ gelte $a_k \neq 0$ für $k \geq n_0$. Gilt für ein $q \in (0, 1)$

$$\left| \frac{a_{k+1}}{a_k} \right| \leq q, \quad \text{für alle } k \geq n_0$$

oder

$$\lim_{k \rightarrow \infty} \left| \frac{a_{k+1}}{a_k} \right| = q,$$

dann konvergiert die Reihe $s_n = \sum_{k=0}^n a_k, n = 0, 1, 2, \dots$

Gilt

$$\left| \frac{a_{k+1}}{a_k} \right| \geq 1, \quad \text{für } k \geq n_0,$$

dann konvergiert die Reihe nicht gegen eine reelle Zahl.

Exponentialreihe

$$e^x = \exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}, \quad x \in \mathbb{R}.$$

Geometrische Reihe

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}, \quad \text{für alle reellen Zahlen } x \text{ mit } |x| < 1$$

1.2. Finanzmathematik

Betrachtet wird eine Zahlungsreihe über die n Perioden ausgehend von einem Kapitalstand (Anfangskapital) in $t = 0$ bis zu einem Vermögensendwert in $t = n$. Diese kann ein normales Spar- oder Girokonto darstellen oder die Abwicklung eines Finanzierungsobjekts über ein zugeordnetes Konto, bspw. eine Investition mit anschließenden Ein- und Auszahlungen oder ein Kredit/Darlehen. Zahlungen an verschiedenen Zeitpunkten werden durch Auf- oder Abzinsen finanzmathematisch bzw. ökonomisch gleichwertig gemacht.

Notation:

K_0 : Anfangskapital (Startkapital)

Vorzeichenkonvention:

$K_0 > 0$: Investition

$K_0 < 0$: Darlehen/Kredit

p, i : Zinssatz (Zinsfuß) (in Prozent)

$q = 1 + i$: Aufzinsungsfaktor

Z_k : Zahlung mit der Vorzeichenkonvention (auf ein Konto)

$Z_k > 0$: Einzahlung (aus Sicht des Verrechnungskontos des Finanzierungsobjekts)

$Z_k < 0$: Auszahlung ("")

Man muss unterscheiden, ob eine Zahlung zu Beginn oder am Ende einer Periode stattfindet.

vorschüssige Zahlung: zu Beginn der Periode

nachschüssige Zahlung: am Ende der Periode

Voraussetzungen:

(1) Auf positive und negative Guthaben wird derselbe Zinssatz angewendet (Habenzins = Sollzins).

(2) Der Zinssatz ist in allen Perioden gleich.

Sonderformen:

$K_0 > 0$ und $Z_k < 0$ für alle k : Investition mit jährlichen Auszahlungen (Entnahmen)

$K_0 < 0$ und $Z_k > 0$ für alle k : Darlehen mit jährlichen Rückzahlungen

Barwert und Vermögensendwert:

Fall 1) Nachschüssigen Zahlungen Z_1, \dots, Z_n

Z_0, Z_1, \dots, Z_n Zahlungsreihe. Kalkulationszinssatz: i

Barwert:

$$B_0 = Z_0 + \sum_{k=1}^n q^{-k} Z_k$$

Fall 2) Vorschüssige Zahlungen Z_1, \dots, Z_n :

$$B_0 = Z_0 + \sum_{k=1}^n q^{-k+1} Z_k$$

Voraussetzung:

(3) Die Verzinsung erfolgt nachschüssig. Anfallende Zinsen werden am Periodenende gutgeschrieben

Kapitalstand (Vermögensendwert) bei nachschüssigen Zahlungen:

$$K_n = K_0 q^n + \sum_{k=0}^{n-1} Z_{n-k} q^k$$

Spezialfälle:

1) Konstante Zahlungen: $Z_k = Z$ für $k = 1, \dots, n$.

Dann ist

$$\sum_{k=0}^{n-1} Z_{n-k} q^k = Z \sum_{k=0} = \frac{1 - q^n}{1 - q}$$

Also:

$$K_n = K_0 q^n + Z \frac{q^n - 1}{q - 1}$$

2) Keine Zahlungen: $Z_k = 0$ für $k = 1, \dots, n$.

$$K_n = K_0 q^n$$

Auflösen nach i :

$$i = \sqrt[n]{K_n/K_0} - 1$$

Auflösen nach n :

$$n = \log_q(K_n/K_0) = \frac{\ln(\frac{K_n}{K_0})}{\ln(q)}$$

(Hinweis: Aufrunden!)

Tilgungsrechnung bei Darlehen/Ratenkrediten:

Ein Darlehen i.H.v. S_0 ($= K_0$) wird durch nachschüssige Zahlungen Z_1, Z_2, \dots, Z_n in n Jahren getilgt. n heißt Laufzeit.

S_j sei die Restschuld nach j Jahren.

Die Zahlung im Jahr j (Annuität, Rate) setzt sich wie folgt zusammen:

Zinsen für die Restschuld S_{j-1} : $i \cdot S_{j-1}$

Tilgungsbetrag: T_j

Zahlung:

$$Z_j = i \cdot S_{j-1} + T_j$$

Fall 1): Bei konstanter Tilgung $T_j = T$, $j = 1, \dots, n$, ist

$$Z_j = i \cdot S_{j-1} + T$$

Restschuld nach n Jahren:

$$S_n = S_{n-1} - T = S_0 - nT$$

Laufzeit des Kredits:

$$n = \frac{S_0}{T}$$

Fall 2): Konstante Annuitäten (Raten): $Z_k = Z > 0$, $k = 1, \dots, n$.

Restschuld nach n Jahren:

$$S_n = S_0 q^n - Z \frac{q^n - 1}{q - 1}$$

Annuität bei einer Laufzeit von n Jahren:

$$Z = S_0 q^n \frac{q - 1}{q^n - 1}$$

betrugen.

Tilgung im k ten Jahr: $T_k = Z - i S_{k-1} = q^{k-1} T_1$

3) Ratenkredit bei monatlicher Abwicklung. r : Zinssatz p.a., R : Monatsrate

Monatlicher Zinssatz: $i = \frac{r}{12}$. Aufzinsungsfaktor $q = 1 + \frac{r}{12}$.

Restschuld nach n Monaten:

$$S_n = S_0 \left(1 + \frac{r}{12}\right)^n - R \frac{\left(1 + \frac{r}{12}\right)^n - 1}{\frac{r}{12}}$$

1.3. Funktionen

Funktionen

Eine **Funktion**

$$f : D \rightarrow W$$

ordnet jedem Element $x \in D$ eine Zahl $y = f(x)$ zu. Die Menge $D \subset \mathbb{R}$ heißt **Definitionsbereich**, die Menge $W = \{f(x) | x \in D\} \subset \mathbb{R}$ heißt **Wertebereich**.

Wenn

$$f : D \rightarrow W \quad \text{und} \quad g : E \rightarrow V$$

Funktionen mit $W \subset E$ sind, dann ist die Funktion $y = g(f(x))$ für alle $x \in D$ definiert und heißt **Komposition** oder **Verkettung** von f und g .

Eine Funktion

$$f : D \rightarrow W$$

heißt **umkehrbar**, wenn es zu jedem $y \in W$ genau ein $x \in D$ mit $y = f(x)$ gibt. Die Umkehrfunktion $f^{-1} : W \rightarrow D$ wird dann durch $f^{-1}(y) = x$ definiert.

- Es gilt $f(f^{-1}(y)) = y$ und $f^{-1}(f(x)) = x$.
- Jede streng monotone Funktion ist umkehrbar.

Polynome und Potenzfunktionen

Eine Funktion $p : \mathbb{R} \rightarrow \mathbb{R}$ mit

$$p(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \dots + a_n \cdot x^n$$

heißt **Polynom vom Grad n** oder **ganz-rationale Funktion**.

- $a_1, \dots, a_n \in \mathbb{R}$ heißen **Koeffizienten**.
- Zwei Polynome sind gleich, wenn ihre Koeffizienten gleich sind.
- Wenn x_1 eine Nullstelle eines Polynoms $f(x)$ ist, dann gilt $f(x) = (x - x_1)g(x)$ für ein Polynom $g(x)$ vom Grad $n - 1$.

Gebrochen-rationale Funktionen

Sind $p(x)$ und $q(x)$ zwei Polynome, wobei $q(x)$ keine Nullstellen in der Menge D hat, dann ist die Funktion

$$f(x) = \frac{p(x)}{q(x)}, \quad x \in D,$$

wohldefiniert und wird **gebrochen-rationale Funktion** genannt. Die Nullstellen des Polynoms $q(x)$ sind dann **Polstellen** (senkrechte Asymptoten) der Funktion $f(x)$.

Wurzelfunktion

Für $x \in [0, \infty)$ und $n \in \mathbb{N}$ ist die Funktion $f(x) = x^n$ streng monoton steigend und daher umkehrbar. Die Umkehrfunktion $f^{-1}(y) = \sqrt[n]{y}$ heißt **n -te Wurzelfunktion** und ist die eindeutige Lösung der Gleichung $y = x^n$ auf dem Wertebereich $[0, \infty)$.

Potenzfunktion

Für $a \neq 0$ heißt die Funktion $f(x) = x^a$ **Potenzfunktion**. Der maximale Definitionsbereich ist $[0, \infty)$ für $a > 0$, und $(0, \infty)$ für $a < 0$.

Exponentialfunktion

Für $b > 0$ und $x \in \mathbb{R}$ heißt die durch

$$f(x) = b^x$$

gegebene Funktion **allgemeine Exponentialfunktion zur Basis b** .

Exponentialfunktion und Logarithmus

Durch $e \approx 2.718282$ erhält man die streng monoton wachsende **Exponentialfunktion** $f(x) = e^x$ mit Wertebereich $(0, \infty)$. Die Umkehrfunktion $y = \ln(x)$ heißt **natürlicher Logarithmus** und hat $(0, \infty)$ als Definitionsbereich.

Beziehungen zwischen (allgemeiner) Exponentialfunktion und Logarithmus

- Es gilt $y = e^x \Leftrightarrow x = \ln(y)$.
- Für $b > 0$ und $x \in \mathbb{R}$ gilt

$$b^x = e^{x \cdot \ln(b)}.$$

- Für $y > 0$ und $b \neq 1$ ist die Umkehrfunktion von $y = b^x$ daher $x = \log_b(y) = \frac{\ln(y)}{\ln(b)}$.

Rechenregeln für die Exponentialfunktion

- $e^0 = 1, e^x > 1$ für $x > 0$ und $0 < e^x < 1$ für $x < 0$.
- $e^{-x} = 1/e^x$.
- $e^{x+y} = e^x \cdot e^y$ und $e^{x-y} = e^x / e^y$.
- $(e^x)^y = e^{xy}$.

Rechenregeln für den Logarithmus

- $\ln(1) = 0$.
- $\ln(x \cdot y) = \ln(x) + \ln(y), \ln(x/y) = \ln(x) - \ln(y)$.
- $\ln(x^y) = y \ln(x)$ für $x > 0$ und $y \in \mathbb{R}$.

Sinus und Kosinus

Der Einheitskreis im \mathbb{R}^2 ist der Kreis mit Radius 1 und Mittelpunkt $(0, 0)$. Zu jeder Zahl $t \in [0, 2\pi]$ gibt es einen Punkt (x, y) auf dem Einheitskreis, sodass die Länge der Kreislinie zwischen dem Punkt $(1, 0)$ und dem Punkt (x, y) - gegen den Uhrzeigersinn laufend - genau t beträgt. Die Koordinaten werden mit $x = \sin(t)$ und $y = \cos(t)$ bezeichnet. Dadurch werden für $x \in [0, 2\pi]$ zwei Funktionen $\sin(x)$ und $\cos(x)$ definiert, die **Sinus** und **Kosinus** heißen.

Eigenschaften von Sinus und Kosinus

- Sinus und Kosinus sind 2π -periodisch: $\sin(x) = \sin(x + 2\pi), \cos(x) = \cos(x + 2\pi)$.
- $\sin(x + \pi) = -\sin(x), \cos(x + \pi) = -\cos(x)$.
- Nullstellen Sinus: $\sin(x) = 0$ für Vielfache von π , d.h. $x = k\pi$, mit $k \in \mathbb{Z}$.
- Nullstellen Kosinus: $\cos(x) = 0$ für $x = k\pi + \pi/2$ für $k \in \mathbb{Z}$.
- Sinus ist ungerade: $\sin(x) = -\sin(-x)$, Kosinus ist gerade: $\cos(x) = \cos(-x)$.
- Satz des Pythagoras: $(\sin(x))^2 + (\cos(x))^2 = 1$.
- $|\sin(x)| \leq 1, |\cos(x)| \leq 1$.
- Halber Winkel: $(\sin(x))^2 = \frac{1}{2}(1 - \cos(2x)), (\cos(x))^2 = \frac{1}{2}(1 + \cos(2x))$.

Additionstheoreme

- $\cos(x + y) = \cos(x)\cos(y) - \sin(x)\sin(y)$.
- $\cos(x - y) = \cos(x)\cos(y) + \sin(x)\sin(y)$.
- $\sin(x + y) = \sin(x)\cos(y) + \cos(x)\sin(y)$.
- $\sin(x - y) = \sin(x)\cos(y) - \cos(x)\sin(y)$.

Sonstige Funktionen

Die **Indikatorfunktion** oder **charakteristische Funktion** 1_I einer Menge I ist gegeben durch

$$1_I = 1(x \in I) = \begin{cases} 1, & x \in I \\ 0, & x \notin I \end{cases}.$$

Konvergenz von Funktionen

Eine Funktion $f : D \rightarrow \mathbb{R}$ hat in einem Punkt $a \in \mathbb{R}$ den **Grenzwert** c , wenn für jede Folge $(x_n)_n$ mit $x_n \in D$ für alle n und $\lim_{n \rightarrow \infty} x_n = a$ gilt, dass $\lim_{n \rightarrow \infty} f(x_n) = c$. Man sagt dann, dass die Funktion f für $x \rightarrow a$ gegen c **konvergiert** und schreibt

$$\lim_{x \rightarrow a} f(x) = c.$$

Einseitiger Grenzwert

c heißt **linksseitiger Grenzwert im Punkt** a , wenn für alle Folgen $(x_n)_n$ mit $x_n \in D$ und $x_n \leq a$ für alle n sowie $\lim_{n \rightarrow \infty} x_n = a$ gilt, dass $\lim_{n \rightarrow \infty} f(x_n) = c$. Notation:

$$f(a-) = \lim_{x \uparrow a} f(x) = c.$$

c heißt **rechtsseitiger Grenzwert im Punkt** a , wenn für alle Folgen $(x_n)_n$ mit $x_n \in D$ und $x_n \geq a$ für alle n sowie $\lim_{n \rightarrow \infty} x_n = a$ gilt, dass $\lim_{n \rightarrow \infty} f(x_n) = c$. Notation:

$$f(a+) = \lim_{x \downarrow a} f(x) = c.$$

$\pm\infty$ ist als Grenzwert jeweils zugelassen.

Sind $f(a-)$ und $f(a+)$ endlich mit $f(a-) \neq f(a+)$, dann hat $f(x)$ in a einen Sprung der Höhe $f(a+) - f(a-)$.

Stetigkeit von Funktionen

Eine Funktion $f : D \rightarrow \mathbb{R}$ heißt **stetig im Punkt** $x \in D$, wenn für alle Folgen $(x_n)_n$ mit $x_n \rightarrow x$, $n \rightarrow \infty$, gilt, dass $f(x_n) \rightarrow f(x)$, $n \rightarrow \infty$.
 $f(x)$ heißt **stetig**, wenn f in allen Punkten $x \in D$ stetig ist.

Eigenschaften

- Eine Funktion f ist genau dann stetig in x , wenn $f(x-) = f(x+) = f(x)$.
- Seien $f(x)$ und $g(x)$ stetige Funktionen. Dann sind $f(x) \pm g(x)$, $f(x) \cdot g(x)$ und $f(x)/g(x)$ für $g(x) \neq 0$ stetig. Wenn $f(g(x))$ definiert ist, dann ist auch $f(g(x))$ stetig.
- Alle Polynome und gebrochen-rationalen Funktionen sind stetig.
- e^x und $\ln(x)$ sowie $|x|$ sind stetig.

Potenzreihen

Seien $x \in \mathbb{R}$ und $a_k \in \mathbb{R}$, $k \in \mathbb{N}_0$. Dann heißt

$$f(x) = \sum_{k=0}^{\infty} a_k (x - x_0)^k$$

formale Potenzreihe mit Entwicklungspunkt x_0 .

- $f(x)$ konvergiert nur für $x = 0$, auf einem ganzen Intervall $I \subset \mathbb{R}$ oder auf ganz \mathbb{R} .
- Wenn es eine Zahl $R > 0$ gibt, sodass $f(x)$ für alle $|x - x_0| < R$ absolut konvergiert und für $|x - x_0| > R$ divergiert, dann heißt R **Konvergenzradius**, und es gilt

$$R = \lim_{n \rightarrow \infty} \frac{a_n}{a_{n+1}}.$$

1.4. Differentialrechnung

Ableitung

Eine Funktion $f : D \rightarrow \mathbb{R}$ ist im Punkt $x \in D$ **differenzierbar**, wenn der Differenzenquotient

$$\frac{f(x+h) - f(x)}{h}$$

für $h \rightarrow 0$ gegen eine reelle Zahl konvergiert. Der Grenzwert wird mit $f'(x) = \frac{df(x)}{dx}$ bezeichnet und heißt **Ableitung von f an der Stelle x** . Die Funktion f ist **differenzierbar**, wenn $f(x)$ in jedem Punkt $x \in D$ differenzierbar ist. f ist dann eine **Stammfunktion** von f' .

Die **linksseitige Ableitung** ist durch $f'(x-) = \lim_{h \uparrow 0} \frac{f(x+h) - f(x)}{h}$ gegeben, die **rechtsseitige Ableitung** ist durch $f'(x+) = \lim_{h \downarrow 0} \frac{f(x+h) - f(x)}{h}$ definiert.

Rechenregeln

- Wenn zwei Funktionen $f(x)$ und $g(x)$ in einem Punkt x differenzierbar sind, dann sind auch $f(x) \pm g(x)$, $f(x) \cdot g(x)$ und $f(x)/g(x)$ (für $g(x) \neq 0$) in x differenzierbar.
- Faktorregel: $(cf(x))' = cf'(x)$ für alle $c \in \mathbb{R}$.
- Summenregel: $(f(x) + g(x))' = f'(x) + g'(x)$.
- Produktregel: $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$.
- Quotientenregel: $\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$.
- Kettenregel: $(f(g(x)))' = f'(g(x))g'(x)$.
- Umkehrregel: $(f^{-1}(y))' = \frac{1}{f'(x)} = \frac{1}{f(f^{-1}(y))}$, $y = f(x)$, $x = f^{-1}(y)$.
- Regel von L'Hospital: Wenn zwei Funktionen $f(x)$ und $g(x)$ für $x \rightarrow x_0$ beide gegen 0 oder $\pm\infty$ konvergieren und $\frac{f'(x)}{g'(x)} \rightarrow c \in \mathbb{R}$ für $x \rightarrow x_0$ gilt, dann gilt auch $\frac{f(x)}{g(x)} \rightarrow c$ für $x \rightarrow x_0$.

Ableitungen und Stammfunktionen einiger Funktionen

Stammfunktion F(x)	Funktion f(x)	Ableitung f'(x)
ax	a	0
$\frac{ax^2}{2} + bx$	$ax + b$	a
$\frac{x^{n+1}}{n+1}$	x^n ($n \in \mathbb{N}$)	nx^{n-1}
$\frac{x^{r+1}}{r+1}$	x^r ($r \in \mathbb{R}$)	rx^{r-1}
$a_0x + a_1\frac{x^2}{2} + \dots + a_n\frac{x^{n+1}}{n+1}$	$a_0 + a_1x + \dots + a_nx^n$	$a_1 + 2a_2x + \dots + na_nx^{n-1}$
$\frac{b^x}{\ln(b)}$	b^x ($b > 0$)	$\ln(b)b^x$
e^x	e^x	e^x
$x \ln(x) - x$	$\ln(x)$	$1/x$
$-\cos(x)$	$\sin(x)$	$\cos(x)$
$\sin(x)$	$\cos(x)$	$-\sin(x)$

Höhere Ableitungen

Wenn $f'(x)$ in x differenzierbar ist, dann heißt

$$f''(x) = f^{(2)}(x) = \frac{d^2f(x)}{dx^2} = (f'(x))'$$

zweite Ableitung von f an der Stelle x . Rekursiv wird die n -te Ableitung von f an der Stelle x für $n \geq 3$ durch $f^{(n)}(x) = (f^{(n-1)}(x))'$ definiert, wenn $f^{(n-1)}(x)$ an der Stelle x differenzierbar ist.

1.5. Taylorpolynom, Taylorentwicklung

Taylorpolynom mit Restglied

Die Funktion $f(x)$ sei in x_0 n -mal differenzierbar. Dann heißt

$$P_n(f, x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$$

Taylorpolynom von $f(x)$ an der Stelle x_0 . Der Term $R_n(f, x) = |f(x) - P_n(f, x)|$ wird **Restglied** genannt und stellt den Approximationsfehler dar.

Für x -Werte mit $|x - x_0| \leq c, c > 0$, gilt

$$R_n(f, x) = |f(x) - P_n(f, x)| \leq \frac{c^{n+1}}{(n+1)!} \max_{t \in [x_0 - c, x_0 + c]} |f^{(n+1)}(t)|,$$

wenn f $(n+1)$ -mal differenzierbar ist.

Taylorreihe

Sei $f : (a, b) \rightarrow \mathbb{R}$ eine Funktion, die für alle x mit $|x - x_0| < R$ ($R > 0$) durch eine Potenzreihe

$$f(x) = \sum_{k=0}^{\infty} a_k (x - x_0)^k$$

darstellbar ist. Dann heißt diese Potenzreihe **Taylorreihe von f mit Entwicklungspunkt x_0** , und es gilt $a_k = \frac{f^{(k)}(x_0)}{k!}$.

1.6. Integration

Sei $f : [a, b] \rightarrow \mathbb{R}$ eine Funktion und $a = x_0 < \dots < x_n = b$ eine **Partition** von $[a, b]$. Mit $x_i^* \in (x_{i-1}, x_i]$ ist durch

$$R_n(f) = \sum_{i=1}^n f(x_i^*)(x_i - x_{i-1})$$

die **Riemann-Summe** von $f(x)$ zu den Stützstellen x_1^*, \dots, x_n^* definiert. Die **Obersumme** erhält man, in dem man die x_i^* jeweils als Maximum der Funktion $f(x)$ auf dem Intervall $(x_{i-1}, x_i]$ wählt, die **Untersumme** ergibt sich, wenn die x_i^* jeweils als Minimum der Funktion $f(x)$ auf dem Intervall $(x_{i-1}, x_i]$ gewählt werden.

Die Funktion f heißt **(Riemann-) integrierbar auf $[a, b]$** , wenn Obersumme und Untersumme für $\max_{i=1, \dots, n} |x_{i-1} - x_i| \rightarrow 0, n \rightarrow \infty$, gegen den gleichen Wert I konvergieren. Das **(Riemann-) Integral** ist dann definiert durch

$$\int_a^b f(x) dx = I.$$

Insbesondere ist jede (stückweise) stetige Funktion $f : [a, b] \rightarrow \mathbb{R}$ integrierbar.

Rechenregeln:

$f, g : D \rightarrow \mathbb{R}$ seien integrierbare Funktionen, $a, b \in D$ und $c \in \mathbb{R}$.

(1) Konstante Funktionen des Integranden können vor das Integral gezogen werden:

$$\int_a^b cf(x) dx = c \int_a^b f(x) dx$$

(2) Linearität: Es gilt:

$$\int_a^b (f(x) + g(x)) dx = \int_a^b f(x) dx + \int_a^b g(x) dx$$

(3) Regel (1) und (2) liefern: Sind c, d Konstanten, so gilt:

$$\int_a^b (cf(x) + dg(x)) dx = c \int_a^b f(x) dx + d \int_a^b g(x) dx$$

(4) Für jedes c mit $a \leq c \leq b$ gilt:

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx$$

(Die Flächen addieren sich).

(5) Für $f(x) \leq g(x)$ für alle $x \in [a, b]$, dann folgt:

$$\int_a^b f(x) dx \leq \int_a^b g(x) dx$$

(6) Ist $f(x)$ beschränkt auf $[a, b]$, d.h. es gilt für eine Konstante K

$$|f(x)| \leq K, \quad \text{für alle } x \in [a, b],$$

dann folgt:

$$\left| \int_a^b f(x) dx \right| \leq K(b - a)$$

(7) Es gilt

$$\int_a^b f(x) dx = - \int_b^a f(x) dx$$

Stammfunktionen

Eine Funktion $F(x)$ heißt **Stammfunktion** von $f(x)$ auf $[a, b]$, wenn für alle $x \in [a, b]$ gilt, dass $F'(x) = f(x)$.

- Wenn $F(x)$ eine Stammfunktion von $f(x)$ ist, dann ist auch $F(x) + c$ mit $c \in \mathbb{R}$ eine Stammfunktion von $f(x)$.
- $F(x) = \int_a^x f(t) dt$ ist eine Stammfunktion von $f(x)$.
- Ist $F(x)$ eine Stammfunktion von $f(x)$, dann gilt

$$\int_a^b f(x)dx = F(x)|_a^b = F(b) - F(a).$$

Integrationsregeln

- Partielle Integration:

$$\int_a^b f'(x)g(x)dx = f(x)g(x)|_a^b - \int_a^b f(x)g'(x)dx.$$

- Substitutionsregel:

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(y)dy, \quad y = g(x).$$

Uneigentliches Integral

Sei $b \in \mathbb{R} \cup \{\infty\}$ und $f : [a, b] \rightarrow \mathbb{R}$ auf jedem Teilintervall $[a, c] \subset [a, b]$ integrierbar. Wenn der Grenzwert

$$I = \lim_{c \uparrow b} \int_a^c f(x)dx$$

existiert (oder $\pm\infty$ ist), ist $f(x)$ (**uneigentlich**) **integrierbar auf** $[a, b]$ und $I = \int_a^b f(x)dx$ heißt (**uneigentliches**) **Integral von** f . Analog definiert man für $a \in \mathbb{R} \cup \{-\infty\}$

$$\int_a^b f(x)dx = \lim_{c \downarrow a} \int_c^b f(x)dx.$$

1.7. Vektoren

Ein **Spaltenvektor** hat die Form

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad x_1, \dots, x_n \in \mathbb{R},$$

ein **Zeilenvektor** ist durch (x_1, \dots, x_n) gegeben. Ist \mathbf{x} ein Spaltenvektor, dann bezeichnet $\mathbf{x}' = \mathbf{x}^t = (x_1, \dots, x_n)$ den gleichen Vektor als Zeilenvektor und heißt **transponierter Vektor** von \mathbf{x} . Analog ist für einen Zeilenvektor (x_1, \dots, x_n) der transponierte Vektor durch $(x_1, \dots, x_n)'$ definiert. Alle (Spalten-) Vektoren \mathbf{x} mit n Einträgen bilden den **n -dimensionalen Vektorraum \mathbb{R}^n** .

Spezielle Vektoren

- $\mathbf{0} = \mathbf{0}_n = (0, \dots, 0)' \in \mathbb{R}$ heißt **Nullvektor**.
- Der Vektor e_i hat eine 1 im i -ten Eintrag und sonst Nullen, und heißt **i -ter Einheitsvektor**:

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \dots, \quad e_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

Rechenregeln

Vektoraddition: Zwei n -dimensionale Vektoren \mathbf{x} und \mathbf{y} werden koordinatenweise addiert:

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

Skalare Multiplikation: Ein Skalar, d.h. eine Zahl $c \in \mathbb{R}$, wird koordinatenweise mit einem Vektor \mathbf{x} multipliziert:

$$c\mathbf{x} = \begin{pmatrix} cx_1 \\ \vdots \\ cx_n \end{pmatrix}.$$

Für Skalare $c, d \in \mathbb{R}$ und Vektoren $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ gilt:

1. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.
2. $c(\mathbf{x} + \mathbf{y}) = c\mathbf{x} + c\mathbf{y}$.
3. $(c + d)\mathbf{x} = c\mathbf{x} + d\mathbf{x}$.

Lineare Unabhängigkeit

Für Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{R}^n$ und Skalare $c_1, \dots, c_k \in \mathbb{R}$ heißt

$$c_1\mathbf{x}_1 + \dots + c_k\mathbf{x}_k$$

Linearkombination von $\mathbf{x}_1, \dots, \mathbf{x}_k$. Ein Vektor $\mathbf{y} \in \mathbb{R}^n$ ist **linear kombinierbar** aus den Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_k$, wenn es Skalare $c_1, \dots, c_k \in \mathbb{R}$ gibt, sodass

$$\mathbf{y} = c_1\mathbf{x}_1 + \dots + c_k\mathbf{x}_k.$$

Die Vektoren $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{R}^n$ heißen **linear unabhängig**, wenn

$$\mathbf{0} = c_1\mathbf{x}_1 + \dots + c_k\mathbf{x}_k$$

nur für $c_1 = \dots = c_k = 0$ gilt, sonst heißen die Vektoren **linear abhängig**.

Skalarprodukt

Das **Skalarprodukt** zweier Vektoren $\mathbf{x} = (x_1, \dots, x_n)'$ und $\mathbf{y} = (y_1, \dots, y_n)'$ ist durch die Zahl

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

gegeben.

Rechenregeln

Für Vektoren $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ und $c \in \mathbb{R}$ gilt:

1. $\mathbf{x}'\mathbf{x} = \sum_{i=1}^n x_i^2$.
2. $\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x}$.
3. $(\mathbf{x} + \mathbf{y})'\mathbf{z} = \mathbf{x}'\mathbf{z} + \mathbf{y}'\mathbf{z}$.
4. $(c\mathbf{x})'\mathbf{y} = c\mathbf{x}'\mathbf{y} = \mathbf{x}'(c\mathbf{y})$.

Zwei Vektoren \mathbf{x} und \mathbf{y} mit $\mathbf{x}'\mathbf{y} = 0$ heißen **orthogonal** (senkrecht) zueinander.

Norm

Die (**euklidische**) **Norm** des Vektors $\mathbf{x} \in \mathbb{R}^n$ ist gegeben durch

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}'\mathbf{x}}$$

Der Vektor \mathbf{x} ist **normiert**, wenn $\|\mathbf{x}\| = 1$.

Rechenregeln

Für Vektoren $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ und $c \in \mathbb{R}$ gilt:

1. $\|\mathbf{x}\| = 0$ genau dann, wenn $\mathbf{x} = \mathbf{0}$.
2. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.
3. $\|c\mathbf{x}\| = |c|\|\mathbf{x}\|$.
4. Jeder Vektor $\mathbf{x} \neq \mathbf{0}$ kann normiert werden: $\mathbf{x}^* = \frac{\mathbf{x}}{\|\mathbf{x}\|}$ hat Norm 1.
5. Cauchy-Schwarz-Ungleichung: $|\mathbf{x}'\mathbf{y}| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$.

Jede Abbildung, $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}$, die die Rechenregeln 1. - 3. erfüllt, heißt **Norm**.

Winkel

Der **Winkel** zwischen zwei Vektoren $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ist gegeben durch

$$(\mathbf{x}, \mathbf{y}) = \arccos\left(\frac{\mathbf{x}}{\|\mathbf{x}\|} \cdot \frac{\mathbf{y}}{\|\mathbf{y}\|}\right)$$

Satz des Pythagoras

Für zwei orthogonale Vektoren $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ gilt $\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$.

1.8. Matrizen

Eine $(m \times n)$ -**Matrix** ist eine Anordnung von $m \cdot n$ Zahlen $a_{ij} \in \mathbb{R}$, $i = 1, \dots, m$, $j = 1, \dots, n$, der Form

$$A = (a_{ij})_{i,j} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

(m, n) heißt **Dimension** der Matrix.

Spezielle Matrizen

1. Die Matrix $\mathbf{0} = \mathbf{0}_{m \times n}$, deren Einträge alle Null sind, heißt **Nullmatrix**.
2. Eine $(n \times n)$ -Matrix mit $a_{ij} = 0$ für alle $i \neq j$ heißt **Diagonalmatrix**.
3. Die Diagonalmatrix $\mathbf{I} = \mathbf{I}_{n \times n} = \text{diag}(1, \dots, 1)$, deren Diagonaleinträge alle 1 sind, heißt **Einheitsmatrix**.

Rechenregeln

Matrixaddition: Die Matrix $\mathbf{C} = \mathbf{A} + \mathbf{B}$ ergibt sich aus zwei Matrizen \mathbf{A} und \mathbf{B} mit gleicher Dimension durch $c_{ij} = a_{ij} + b_{ij}$.

Skalare Multiplikation: Ein Skalar $c \in \mathbb{R}$ wird mit einer Matrix \mathbf{A} elementweise multipliziert: $c\mathbf{A} = (ca_{ij})_{i,j}$.

Für Matrizen $\mathbf{A}, \mathbf{B}, \mathbf{C}$ gleicher Dimension und Skalare $c, d \in \mathbb{R}$ gilt:

1. $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$.
2. $c(\mathbf{A} + \mathbf{B}) = c\mathbf{A} + c\mathbf{B}$.
3. $(c + d)\mathbf{A} = c\mathbf{A} + d\mathbf{A}$.

Matrix-Vektor-/Matrix-Multiplikation

Sei \mathbf{A} eine $(m \times n)$ -Matrix. Dann werden mit $(\mathbf{a}_1, \dots, \mathbf{a}_m)$ die Zeilen und mit $(\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(n)})$ die Spalten von \mathbf{A} bezeichnet.

Matrix-Vektor-Multiplikation

Sei $\mathbf{A} = (a_{ij})_{i,j}$ eine $(m \times n)$ -Matrix und \mathbf{x} ein n -dimensionaler Vektor. Dann ist die **Multiplikation von \mathbf{A} mit \mathbf{x}** definiert durch

$$\mathbf{y} = \mathbf{Ax} = \begin{pmatrix} \mathbf{a}'_1 \mathbf{x} \\ \vdots \\ \mathbf{a}'_m \mathbf{x} \end{pmatrix},$$

wobei $\mathbf{a}'_i \mathbf{x}$ das Skalarprodukt der i -ten Zeile von \mathbf{A} mit \mathbf{x} ist.

Rechenregeln

Für $(m \times n)$ -Matrizen \mathbf{A}, \mathbf{B} , Vektoren $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ und $c \in \mathbb{R}$ gilt:

1. $(\mathbf{A} + \mathbf{B})\mathbf{x} = \mathbf{Ax} + \mathbf{Bx}$.
2. $\mathbf{A}(\mathbf{x} + \mathbf{y}) = \mathbf{Ax} + \mathbf{Ay}$.
3. $\mathbf{A}(c \cdot \mathbf{x}) = c \cdot \mathbf{Ax}$ (Linearität von $\mathbf{x} \mapsto \mathbf{Ax}$).

Matrizenmultiplikation

Sei \mathbf{A} eine $(m \times n)$ -Matrix und \mathbf{B} eine $(n \times r)$ -Matrix. Dann ist das **Produkt** von \mathbf{A} und \mathbf{B} definiert durch die $(m \times r)$ -Matrix

$$\mathbf{C} = \mathbf{AB} = (c_{ij})_{i,j} \in \mathbb{R}^{m \times r},$$

wobei

$$c_{ij} = \sum_{k=1}^n a_{i,k} b_{k,j}$$

das Skalarprodukt der i -ten Zeile von \mathbf{A} mit der j -ten Spalte von \mathbf{B} ist.

Zwei Matrizen heißen **multiplikations-kompatibel**, wenn die Spaltenzahl von \mathbf{A} und die Zeilenzahl von \mathbf{B} übereinstimmen, dann kann das Produkt \mathbf{AB} gebildet werden. Daraus folgt aber nicht, dass auch das Produkt \mathbf{BA} existiert!

Rechenregeln

Für Matrizen $\mathbf{A}, \mathbf{B}, \mathbf{C}$, einen Vektor $\mathbf{x} \in \mathbb{R}^n$ und $c \in \mathbb{R}$, gilt, sofern die Matrixprodukte gebildet werden können:

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

2. $\mathbf{A}(\mathbf{Bx}) = (\mathbf{AB})\mathbf{x}$.
3. $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$.
4. In der Regel gilt: $\mathbf{AB} \neq \mathbf{BA}$ (selbst wenn beide Produkte existieren).

Rang

Der **Zeilenrang** einer Matrix ist die maximale Anzahl linear unabhängiger Zeilenvektoren der Matrix, analog dazu ist der **Spaltenrang** einer Matrix als die maximale Anzahl linear unabhängiger Spaltenvektoren der Matrix definiert. Da Zeilen- und Spaltenrang bei jeder Matrix übereinstimmen, spricht man meistens nur vom **Rang** $\text{rg}(\mathbf{A})$ der Matrix \mathbf{A} .

Determinanten

Die **Determinante** einer (2×2) -Matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

ist gegeben durch

$$\det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21}.$$

Die **Determinante** einer (3×3) -Matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

ist gegeben durch

$$\det(\mathbf{A}) = a_{11}\det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{12}\det \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} + a_{13}\det \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}.$$

Allgemein ist die **Determinante** (entwickelt nach der i -ten Zeile) einer $n \times n$ -Matrix \mathbf{A} gegeben durch

$$\det(\mathbf{A}) = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det(\mathbf{A}_{ij}),$$

wobei sich \mathbf{A}_{ij} durch Streichen der i -ten Zeile und der j -ten Spalte ergibt. Insbesondere ist

$$\det(\mathbf{A}) = a_{11}\det(\mathbf{A}_{11}) - a_{12}\det(\mathbf{A}_{12}) \pm \dots \pm (-1)^{n+1}\det(\mathbf{A}_{1n}).$$

Rechenregeln

Für multiplikations-kompatible Matrizen \mathbf{A}, \mathbf{B} und $c \in \mathbb{R}$ gilt:

1. Das Vertauschen zweier Zeilen oder Spalten ändert das Vorzeichen der Determinante.
2. $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$.
3. $\det(c\mathbf{A}) = c^n \det(\mathbf{A})$.
4. $\det(\mathbf{A}) = \det(\mathbf{A}')$.
5. $\det(\mathbf{A}) = 0$ genau dann, wenn $\text{rg}(\mathbf{A}) < n$.
6. $\det(\mathbf{A}) \neq 0$ genau dann, wenn die Zeilen (Spalten) von \mathbf{A} linear unabhängig sind.
7. Die Determinante ist in jeder Zeile bzw. Spalte linear.
8. Haben alle Elemente unterhalb der Hauptdiagonalen den Wert 0, dann gilt

$$\det(\mathbf{A}) = a_{11}a_{22} \cdot \dots \cdot a_{nn}.$$

Wegen Regel 4 kann man auch nach der j -ten Spalte entwickeln:

$$\det(\mathbf{A}) = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det(\mathbf{A}_{ij}).$$

Generell entwickelt man nach derjenigen Zeile oder Spalte, die die meisten Nullen enthält.

Inverse Matrix

Sei \mathbf{A} eine $(n \times n)$ -Matrix. Dann heißt eine Matrix \mathbf{B} mit

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}$$

inverse Matrix von \mathbf{A} und wird mit \mathbf{A}^{-1} bezeichnet. Die Matrix \mathbf{A} heißt dann **invertierbar**.

Eine Matrix \mathbf{A} ist genau dann invertierbar, wenn $\det(\mathbf{A}) \neq 0$.

Für eine invertierbare $(n \times n)$ -Matrix \mathbf{A} und $c \in \mathbb{R}^n$ gilt:

1. Ist $\mathbf{AB} = \mathbf{I}$ oder $\mathbf{BA} = \mathbf{I}$, dann folgt $\mathbf{B} = \mathbf{A}^{-1}$.
2. $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$.
3. $(c\mathbf{A})^{-1} = \frac{1}{c}\mathbf{A}^{-1}$.
4. Ist \mathbf{A} symmetrisch (d.h. $\mathbf{A} = \mathbf{A}'$), dann ist auch \mathbf{A}^{-1} symmetrisch.
5. Mit \mathbf{A} und \mathbf{B} sind auch \mathbf{AB} und \mathbf{BA} invertierbar, wobei

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}, \quad (\mathbf{BA})^{-1} = \mathbf{A}^{-1}\mathbf{B}^{-1}.$$

1.9. Lineare Gleichungssysteme

Sei \mathbf{A} eine $(m \times n)$ -Matrix und $\mathbf{b} \in \mathbb{R}^m$. Gesucht werde ein Vektor $\mathbf{x} \in \mathbb{R}^n$, sodass

$$\mathbf{Ax} = \mathbf{b}.$$

Dies wird als **Lineares Gleichungssystem** (LGS) mit m Gleichungen und n Unbekannten bezeichnet. Hierbei sind:

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

Das LGS lautet in ausführlicher Form:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

- Das LGS ist genau lösbar, wenn die rechte Seite \mathbf{b} durch die Spalten der Koeffizientenmatrix \mathbf{A} kombiniert werden kann. Die Koeffizienten der Linearkombinationen sind dann eine Lösung. D.h.: Das LGS ist genau dann lösbar, wenn die rechte Seite im Spaltenraum von \mathbf{A} liegt.
- Das LGS $\mathbf{Ax} = \mathbf{b}$ besitzt genau dann eine Lösung, wenn $\text{rg}(\mathbf{A}) = \text{rg}(\mathbf{A}|\mathbf{b})$.
- Ist \mathbf{A} eine $(n \times n)$ -Matrix, dann hat das LGS $\mathbf{Ax} = \mathbf{b}$ genau dann eine Lösung, wenn $\det(\mathbf{A}) \neq 0$.
- Ist \mathbf{A} eine reguläre (d.h. invertierbare) Matrix und ist die inverse Matrix \mathbf{A}^{-1} bekannt, so ist der eindeutig Lösungsvektor durch

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

gegeben.

Gauß-Verfahren

Bringe die erweiterte Koeffizientenmatrix $(\mathbf{A}|\mathbf{b})$ durch (ggf. mehrmaliges)

1. Vertauschen von Zeilen,
2. Addition eines Vielfachen einer Zeile mit einer anderen Zeile,
3. Multiplikation einer Zeile mit einer Zahl $c \neq 0$

in die Gestalt

$$\begin{pmatrix} T & \mathbf{d} \\ \mathbf{0} & \mathbf{e} \end{pmatrix},$$

wobei $\mathbf{T} \in \mathbb{R}^{k \times n}$ in Treppenform ist. Wenn \mathbf{e} nicht der Nullvektor ist, dann besitzt das LGS keine Lösung. Der Rang der Matrix ist k .

\mathbf{T} habe in den Spalten mit Indizes s_1, \dots, s_k Stufen, dann können diese Gleichungen nach den Variablen x_{s_1}, \dots, x_{s_k} aufgelöst werden.

Die übrigen Variablen x_j mit $j \notin \{s_1, \dots, s_k\}$ sind **freie Parameter**. Löse die k -te Zeile des obigen Schemas nach x_{s_k} auf:

$$x_{s_k} = \frac{d_k}{t_{k,s_k}} - \frac{t_{k,s_{k+1}}}{t_{k,s_k}}x_{s_{k+1}} - \cdots - \frac{d_k}{t_{k,s_k}}x_n.$$

Dann ist x_{s_k} eine Funktion der freien Variablen $x_{s_{k+1}}, \dots, x_n$, die beliebig gewählt werden können. Löse dann schrittweise nach den Variablen x_{s_k}, \dots, x_{s_1} auf.

Gauß-Verfahren für mehrere rechte Seiten

Sind mehrere lineare Gleichungssysteme zu lösen, bei denen sich jeweils nur die rechte Seite unterscheidet, die Koeffizientenmatrix jedoch stets dieselbe ist, so geht man folgendermaßen vor:

1. Ergänze \mathbf{A} um die rechten Seiten $\mathbf{b}_1, \dots, \mathbf{b}_r$.
2. Führe das Gauß-Verfahren wie gewohnt durch, bis Treppengestalt erreicht ist bzw. links die Einheitsmatrix erscheint.
3. Berechne die r Lösungen durch Rückwärtseinsetzen bzw. lese die Lösung nun ab.

$$(\mathbf{A}|\mathbf{b}_1 \dots \mathbf{b}_r) \rightarrow (\mathbf{I}|\mathbf{x}_1 \dots \mathbf{x}_r)$$

Berechnung der inversen Matrix

Die inverse Matrix einer regulären $n \times n$ -Matrix \mathbf{A} kann mit dem Gauß-Verfahren so berechnet werden:

1. Ergänze \mathbf{A} um die Einheitsmatrix \mathbf{I} , also die rechten Seiten $\mathbf{e}_1, \dots, \mathbf{e}_n$.
2. Führe das Gauß-Verfahren durch, bis links die Einheitsmatrix erscheint.

3. Rechts steht nun die inverse Matrix \mathbf{A}^{-1} .

$$(\mathbf{A} | \mathbf{e}_1 \dots \mathbf{e}_n) = (\mathbf{A} | \mathbf{I}) \rightarrow (\mathbf{I} | \mathbf{A}^{-1})$$

Cramersche Regel

Wenn \mathbf{A} invertierbar ist, dann ist die i -te Koordinate der eindeutigen Lösung \mathbf{x} des LGS $\mathbf{Ax} = \mathbf{b}$ gegeben durch

$$x_i = \frac{\det(\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(i-1)}, \mathbf{b}, \mathbf{a}^{(i+1)}, \dots, \mathbf{a}^{(n)})}{\det(\mathbf{A})}.$$

Hierbei bezeichnet

$$(\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(i-1)}, \mathbf{b}, \mathbf{a}^{(i+1)}, \dots, \mathbf{a}^{(n)})$$

diejenige Matrix, die man durch Ersetzen der i -ten Spalte von A durch die Spalte b erhält.

1.10. Differenzengleichungen, Eigenwert, Eigenvektor

Homogene lineare Differenzengleichungen $x_{n+1} = (1-p)x_n + d$, $n > 0$. $q = 1-p$, d : Strukturparameter $x_{n+1} = f(x_n)$, $f(x) = qx + d$ x^* ist ein Gleichgewicht, wenn $f(x^*) = x^*$. Es gilt

$$x^* = f(x^*) \Leftrightarrow x^* = qx^* + d \Leftrightarrow x^* = \frac{d}{1-q}.$$

Lösungsfolge

$$x_1 = qx_0 + d$$

$$x_2 = q(qx_0 + d) + d = q^2x_0 + q^1d + q^0d$$

usw.

$$x_n = q^n x_0 + d \sum_{i=0}^{n-1} q^i = q^n x_0 + d \frac{1 - q^n}{1 - q}, \quad n = 0, 1, \dots$$

Asymptotisches Verhalten

$$\lim_{n \rightarrow \infty} x_n = \frac{d}{1-q} = x^* \text{ für } |q| < 1$$

$$\lim_{n \rightarrow \infty} x_n = \infty \text{ für } q > 1.$$

Systeme linearer Entwicklungsmodelle x und y seien zwei Variablen an Zeitpunkten t_1, t_2, \dots . Weiter sei $x_n = x(t_n)$ der Wert von x am Ende der n -ten Periode $(t_{n-1}, t_n]$ und $y_n = y(t_n)$ der Wert von y am Ende der n -ten Periode $(t_{n-1}, t_n]$. (x_n, y_n) heißt Zustand des Systems zur Zeit t_n und

$$\mathbf{z}_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix} \in \mathbb{R}^2$$

Zustandsvektor. Es sei $\mathbf{z}_{n+1} = \mathbf{A}\mathbf{z}_n + \mathbf{b}$ für $n = 1, 2, \dots$. Ein Zustandsvektor \mathbf{z}^* heißt Gleichgewichtsvektor, wenn

$$\mathbf{z}^* = \mathbf{A}\mathbf{z}^* + \mathbf{b}.$$

Auflösen nach \mathbf{z}^* ergibt

$$(\mathbf{I} - \mathbf{A})\mathbf{z}^* = \mathbf{b},$$

d.h. \mathbf{z}^* ist Lösung des Gleichungssystems $(\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{b}$. Ist $(\mathbf{I} - \mathbf{A})$ invertierbar, dann ist $\mathbf{z}^* = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} = -(\mathbf{A} - \mathbf{I})^{-1}\mathbf{b}$. Lösungsfolge
Homogener Fall: $\mathbf{z}_{n+1} = \mathbf{A}\mathbf{z}_n$ ($\mathbf{b} = \mathbf{0}$). Sei

$$\mathbf{z}_n = c\mathbf{v}\lambda^n.$$

Dann ist $\mathbf{A}\mathbf{z}_n = \mathbf{A}(c\mathbf{v}\lambda^n) = c\lambda^n \mathbf{A}\mathbf{v}$ und $\mathbf{z}_{n+1} = c\mathbf{v}\lambda^{n+1}$. Gleichsetzen führt auf $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Gilt für einen Vektor $\mathbf{v} \neq \mathbf{0}$ und eine Zahl $\lambda \in \mathbb{R}$

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v},$$

dann heißt \mathbf{v} Eigenvektor und λ Eigenwert von \mathbf{A} . $\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \Leftrightarrow (\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0} \Rightarrow \underbrace{\det(\mathbf{A} - \lambda\mathbf{I})}_{\text{Polynom in } \lambda} = 0$. Mit diesem λ erhält man ein lineares Gleichungssystem, das auf \mathbf{v} führt.

1.11. Funktionen mit mehreren Variablen

Eine Zuordnung $f: D \rightarrow \mathbb{R}$ mit $D \subset \mathbb{R}^n$, die jedem Punkt $\mathbf{x} = (x_1, \dots, x_n) \in D$ genau einen Wert $y = f(x_1, \dots, x_n) \in \mathbb{R}$ zuordnet, wird **Funktion von x_1, \dots, x_n** genannt. x_1, \dots, x_n heißen (**unabhängige, exogene, Argument-**) **Variablen** aus dem **Definitionsbereich D** . Die Menge $W = \{f(\mathbf{x}) : \mathbf{x} \in D\}$ heißt **Wertebereich**, deren Elemente $y = f(x_1, \dots, x_n)$ auch als **endogene Variablen** bezeichnet werden.

Konvergenz von Punktfolgen

Eine Folge $(\mathbf{x}_k)_{k \in \mathbb{N}}$, $\mathbf{x}_k = (x_{k_1}, \dots, x_{k_n})$ konvergiert gegen $\mathbf{x} = (x_1, \dots, x_n)$, wenn alle n Koordinatenfolgen gegen die entsprechenden Koordinaten von \mathbf{x} konvergieren, d.h.

$$x_{k_1} \rightarrow x_1, \dots, x_{k_n} \rightarrow x_n, k \rightarrow \infty.$$

Stetigkeit

Eine Funktion $f(\mathbf{x}), \mathbf{x} \in D$, heißt **stetig im Punkt \mathbf{a}** , wenn für alle Folgen $(\mathbf{x}_k)_k$ mit $\mathbf{x}_k \rightarrow \mathbf{a}$ folgt, dass

$$f(\mathbf{x}_k) \rightarrow f(\mathbf{a}), k \rightarrow \infty.$$

f ist **stetig**, wenn $f(\mathbf{x})$ in jedem Punkt \mathbf{a} stetig ist.

Wie auch im eindimensionalen Fall sind alle Polynome in n Variablen stetig, sowie alle Summen, Differenzen, Produkte und Divisionen (ohne Nullstellen der Funktion im Nenner) stetiger Funktionen. Auch die Verkettung $f(g_1(\mathbf{x}), \dots, g_n(\mathbf{x}))$ von stetigen Funktionen $f, g_1(\mathbf{x}), \dots, g_n(\mathbf{x})$ ist stetig.

Partielle Differenzierbarkeit

Für eine Funktion $f(\mathbf{x}) = f(x_1, \dots, x_n)$ ist die (*i-te*) **partielle Ableitung nach x_i im Punkt \mathbf{x}** definiert durch

$$\frac{\partial f(\mathbf{x})}{\partial x_i} := \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{e}_i) - f(\mathbf{x})}{h},$$

wenn der Grenzwert (in \mathbb{R}) existiert. f heißt **partiell differenzierbar** (im Punkt \mathbf{x}), wenn alle n partiellen Ableitungen (im Punkt \mathbf{x}) existieren. f heißt **stetig partiell differenzierbar**, wenn alle n partiellen Ableitungen stetig sind.

Der Vektor der n partiellen Ableitungen

$$\text{grad } f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{pmatrix}$$

wird als **Gradient von $f(\mathbf{x})$** bezeichnet.

Mehrmaliges Ableiten

Wenn die Funktion $\frac{\partial f(x_1, \dots, x_n)}{\partial x_i}$ partiell differenzierbar nach x_j ist, dann wird die resultierende **partielle Ableitung zweiter Ordnung** mit $\frac{\partial^2 f(x_1, \dots, x_n)}{\partial x_j \partial x_i}$ bezeichnet. Analog wird die **partielle Ableitung k -ter Ordnung** mit $\frac{\partial^k f(x_1, \dots, x_n)}{\partial x_{i_k} \dots \partial x_{i_1}}$ notiert, falls sie existiert.

Ist f zweimal stetig differenzierbar im Punkt \mathbf{x} , dann wird die symmetrische $(n \times n)$ -Matrix

$$\mathbf{H}_f(\mathbf{x}) = \left(\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} \right)_{i,j}$$

Hesse-Matrix von $f(\mathbf{x})$ an der Stelle \mathbf{x} genannt.

Vertauschbarkeitsregel

Wenn alle partiellen Ableitungen 2. Ordnung $\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$ der Funktion f existieren und stetig sind, dann gilt

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i} \left(\frac{\partial f(\mathbf{x})}{\partial x_j} \right) = \frac{\partial}{\partial x_j} \left(\frac{\partial f(\mathbf{x})}{\partial x_i} \right) = \frac{\partial^2 f(\mathbf{x})}{\partial x_j \partial x_i}.$$

Kettenregel

Für differenzierbare Funktionen $f(x_1, \dots, x_n), g_1(t), \dots, g_n(t)$ gilt

$$\frac{df(g_1(t), \dots, g_n(t))}{dt} = (\text{grad}f(g_1(t), \dots, g_n(t)))' \begin{pmatrix} \frac{\partial g_1(t)}{\partial t} \\ \vdots \\ \frac{\partial g_n(t)}{\partial t} \end{pmatrix}.$$

Lineare und quadratische Approximation

Für eine Funktion $f(\mathbf{x})$ von n Variablen ist die lineare Approximation im Punkt \mathbf{x}_0 gegeben durch

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + (\text{grad}f(\mathbf{x}_0))'(\mathbf{x} - \mathbf{x}_0).$$

Für eine Funktion $f(\mathbf{x})$ ist eine quadratische Approximation in \mathbf{x}_0 gegeben durch

$$Q(\mathbf{x}) = f(\mathbf{x}_0) + (\text{grad}f(\mathbf{x}_0))'(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)' \mathbf{H}_f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0).$$

1.12. Optimierung von Funktionen

Optimierung von Funktionen mit einer Variablen

Lokale und globale Extrema

Eine Funktion $f : (a, b) \rightarrow \mathbb{R}$ besitzt im Punkt $x_0 \in (a, b)$ ein **lokales Minimum**, wenn es ein $c > 0$ gibt, sodass

$$f(x_0) \leq f(x) \text{ für alle } x \text{ mit } |x - x_0| < c.$$

x_0 ist ein **globales Minimum**, wenn $f(x_0) \leq f(x)$ für alle $x \in (a, b)$. Der Punkt $x_0 \in (a, b)$ ist ein **lokales Maximum**, wenn es ein $c > 0$ gibt, sodass

$$f(x_0) \geq f(x) \text{ für alle } x \text{ mit } |x - x_0| < c.$$

f hat in x_0 ein **globales Maximum**, wenn $f(x_0) \geq f(x)$ für alle $x \in (a, b)$.

Notwendiges Kriterium

Wenn $x_0 \in (a, b)$ ein lokales Extremum ist, dann gilt $f'(x_0) = 0$. Ein Punkt x mit $f'(x) = 0$ heißt **stationärer Punkt**.

Hinreichendes Kriterium 1. Ordnung

Sei $x_0 \in (a, b)$ ein stationärer Punkt von $f(x)$. Wechselt das Vorzeichen von $f'(x)$ in x_0 ...

1. ... von $-$ nach $+$, dann ist x_0 ein lokales Minimum.
2. ... von $+$ nach $-$, dann ist x_0 ein lokales Maximum.

Hinreichendes Kriterium 2. Ordnung

Sei $x_0 \in (a, b)$ ein stationärer Punkt von $f(x)$.

1. Ist $f''(x_0) > 0$, dann ist x_0 ein lokales Minimum.
2. Ist $f''(x_0) < 0$, dann ist x_0 ein lokales Maximum.

Eine Funktion $f(x)$ ist **konvex** auf (a, b) , wenn alle Verbindungsstrecken von zwei Punkten auf dem Graphen mit x -Koordinaten in (a, b) oberhalb des Graphen verlaufen. Verlaufen diese immer unterhalb des Graphen, nennt man $f(x)$ **konkav**.

Die Funktion $f(x)$ sei zweimal differenzierbar. Gilt $f''(x) > 0$ für alle $x \in (a, b)$, dann ist $f(x)$ in (a, b) konvex. Gilt $f''(x) < 0$ für alle $x \in (a, b)$, dann ist $f(x)$ in (a, b) konkav.

Optimierung von Funktionen mit mehreren Variablen

Sei $f : D \rightarrow \mathbb{R}$, $D \subset \mathbb{R}^n$, eine Funktion. Ein Punkt \mathbf{x}_0 heißt **lokales Minimum** von f , wenn es ein $c > 0$ gibt, sodass $f(\mathbf{x}_0) \leq f(\mathbf{x})$ für alle \mathbf{x} mit $\|\mathbf{x} - \mathbf{x}_0\| \leq c$. Ist der Punkt \mathbf{x}_0 ein lokales Minimum der Funktion $-f(\mathbf{x})$, dann heißt \mathbf{x}_0 **lokales Maximum** von f . In beiden Fällen wird \mathbf{x}_0 **lokales Extremum** von f genannt.

Ein Punkt $\mathbf{x} \in \mathbb{R}^n$ mit $\text{grad } f(\mathbf{x}) = \mathbf{0}$ heißt **stationärer Punkt**.

Ein Punkt $\mathbf{x}_0 \in D$ heißt **innerer Punkt** der Menge D , wenn ein $c > 0$ existiert, sodass alle Punkte \mathbf{x} mit $\|\mathbf{x} - \mathbf{x}_0\| < c$ auch in der Menge D liegen.

Notwendiges und hinreichendes Kriterium für ein Extremum

Wenn $\mathbf{x}_0 \in D$ ein innerer Punkt von D und ein lokales Extremum der Funktion $f : D \rightarrow \mathbb{R}$ ist, dann gilt $\text{grad } f(\mathbf{x}_0) = \mathbf{0}$.

Eine symmetrische $(n \times n)$ -Matrix \mathbf{A} heißt **positiv definit**, wenn $\mathbf{x}' \mathbf{A} \mathbf{x} > 0$ für alle $\mathbf{x} \neq 0$. Gilt nur $\mathbf{x}' \mathbf{A} \mathbf{x} \geq 0$ für alle $\mathbf{x} \neq 0$, dann nennt man \mathbf{A} **positiv semidefinit**. Wenn $-\mathbf{A}$ positiv (semi-)definit ist, dann heißt \mathbf{A} **negativ (semi-)definit**. Ist \mathbf{A} weder positiv noch negativ definit, nennt man \mathbf{A} **indefinit**.

Wie überprüft man eine gegebene Matrix?

1. Eine $(n \times n)$ -Matrix \mathbf{A} ist genau dann positiv definit, wenn alle Determinanten $\det(\mathbf{A}_i)$ der Teilmatrizen \mathbf{A}_i , die aus den ersten i Zeilen und Spalten bestehen, positiv sind.
2. Eine (2×2) -Matrix $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ ist insbesondere genau dann positiv definit, wenn $a > 0$ und $\det(\mathbf{A}) = ad - bc > 0$ gilt.
3. Eine $(n \times n)$ -Matrix \mathbf{A} ist genau dann negativ definit, wenn für alle Teilmatrizen die Determinanten mit alternierenden Vorzeichen positiv sind, d.h. wenn $(-1)^i \det(\mathbf{A}_i) > 0$. Also wenn $\det(\mathbf{A}_1) < 0$ und $\det(\mathbf{A}_2) > 0$ und $\det(\mathbf{A}_3) < 0$ usw.
4. Gilt $\det(\mathbf{A}_i) > 0$ für $i = 1, \dots, n-1$ und $\det(\mathbf{A}) = 0$, dann ist eine $(n \times n)$ -Matrix \mathbf{A} positiv semidefinit.
5. Gilt $(-1)^i \det(\mathbf{A}_i) > 0$ für $i = 1, \dots, n-1$ und $\det(\mathbf{A}) = 0$, dann ist eine $(n \times n)$ -Matrix \mathbf{A} negativ semidefinit.
6. Gilt zwar $\det(\mathbf{A}) \neq 0$ aber weder 1. noch 3. gelten, dann ist die Matrix indefinit.

Sei $f : D \rightarrow \mathbb{R}$ zweimal stetig differenzierbar und \mathbf{x}_0 ein stationärer Punkt, der innerer Punkt von D ist. Dann gilt:

1. \mathbf{x}_0 ist ein lokales Minimum, wenn $\mathbf{H}_f(\mathbf{x}_0)$ positiv definit ist.
2. \mathbf{x}_0 ist ein lokales Maximum, wenn $\mathbf{H}_f(\mathbf{x}_0)$ negativ definit ist. (Man prüfe die negative Matrix $-\mathbf{H}_f(\mathbf{x}_0)$ auf positive Definitheit.)
3. \mathbf{x}_0 ist ein Sattelpunkt, wenn $\mathbf{H}_f(\mathbf{x}_0)$ indefinit ist.

Optimierung unter Nebenbedingungen

In einem **restriktiven Optimierungsproblem** sollen die Extrema einer Funktion $f : D \rightarrow \mathbb{R}$, $D \subset \mathbb{R}^n$, unter den m Nebenbedingungen

$$g_1(\mathbf{x}) = 0, g_2(\mathbf{x}) = 0, \dots, g_m(\mathbf{x}) = 0$$

gefunden werden.

Seien die Funktionen $f, g_1, \dots, g_m : D \rightarrow \mathbb{R}$ stetig differenzierbar und \mathbf{x}_0 ein lokales Extremum von $f(\mathbf{x})$ unter den Nebenbedingungen $g_i(\mathbf{x}) = 0$, $i = 1, \dots, m$. Weiter habe die $(m \times n)$ -Jakobi-Matrix

$$g'(\mathbf{x}_0) = \begin{pmatrix} \frac{\partial g_1(\mathbf{x}_0)}{\partial x_1} & \dots & \frac{\partial g_1(\mathbf{x}_0)}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial g_m(\mathbf{x}_0)}{\partial x_1} & \dots & \frac{\partial g_m(\mathbf{x}_0)}{\partial x_n} \end{pmatrix}$$

der partiellen Ableitungen der g_i nach x_1, \dots, x_n vollen Rang m . Dann existieren eindeutige Zahlen $\lambda_1, \dots, \lambda_m \in \mathbb{R}$, sodass für die Funktion $F : D \rightarrow \mathbb{R}$,

$$F(\mathbf{x}, \lambda_1, \dots, \lambda_m) = f(\mathbf{x}) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x})$$

gilt, dass

$$\text{grad } F(\mathbf{x}_0, \lambda_1, \dots, \lambda_m) = \text{grad } f(\mathbf{x}_0) + \sum_{i=1}^m \lambda_i \text{grad } g_i(\mathbf{x}_0) = \mathbf{0}.$$

Die Funktion F heißt **Lagrange-Funktion** und die λ_i **Lagrange-Multiplikatoren**.

Satz (Lagrange): Die Zielfunktion $f(x, y)$, $(x, y) \in D \subset \mathbb{R}^2$, und $g(x, y)$ seien stetig differenzierbar. Ist $(x_0, y_0) \in D$ eine lokale Extremalstelle von $f(x, y)$ unter der Restriktion

$$g(x, y) = 0$$

und gilt

$$(*) \quad \text{grad } g(x_0, y_0) \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

dann gibt es einen eindeutig bestimmten Lagrange-Multiplikator $\lambda_0 \in \mathbb{R}$, so dass (x_0, y_0, λ_0) das Gleichungssystem

$$\text{grad } L(x_0, y_0, \lambda_0) = \mathbf{0}$$

löst.

Untersuchung der stationären Punkte der Lagrange-Funktion:

Es sei

$$H(x, y | \lambda_0) = \begin{pmatrix} \frac{\partial^2}{\partial x^2} L(x, y, \lambda_0) & \frac{\partial^2}{\partial x \partial y} L(x, y, \lambda_0) \\ \frac{\partial^2}{\partial x \partial y} L(x, y, \lambda_0) & \frac{\partial^2}{\partial y^2} L(x, y, \lambda_0) \end{pmatrix}$$

die Hesse-Matrix der Funktion

$$(x, y) \mapsto L(x, y, \lambda_0)$$

(d.h. bei festem λ_0).

Dann gilt:

- 1) Ist $H(x_0, y_0 | \lambda_0)$ positiv (negativ) definit, so liegt in (x_0, y_0) eine lokale Minimalstelle (Maximalstelle) von $f(x, y)$ unter der gegebenen Restriktion vor.

2) Ist $H(x, y|\lambda_0)$ positiv (negativ) definit für alle $(x, y) \in \mathbb{R}^2$, so liegt in (x_0, y_0) eine globale Minimalstelle (Maximalstelle) von $f(x, y)$ unter der Restriktion vor.

Die generelle Vorgehensweise bleibt gleich, wenn man Funktionen in n Variablen unter einer Restriktion optimiert.

Optimierung unter mehreren Nebenbedingungen

Der Lagrange-Ansatz kann auch bei Vorliegen von mehreren Nebenbedingungen angewendet werden.

Min./Max. $f(x_1, \dots, x_n)$, $(x_1, \dots, x_n) \in D \subset \mathbb{R}^n$, unter den Nebenbedingungen

$$\begin{aligned} g_1(x_1, \dots, x_n) &= 0 \\ &\vdots \\ g_m(x_1, \dots, x_n) &= 0 \end{aligned}$$

Satz: Die Zielfunktion und die m Funktionen g_1, \dots, g_m seien auf $D \subset \mathbb{R}^n$ definiert und dort stetig differenzierbar. \mathbf{x}_0 sei lokale Extremalstelle von $f(\mathbf{x}_0)$ unter den Nebenbedingungen

$$g_1(\mathbf{x}) = 0, \dots, g_m(\mathbf{x}) = 0$$

Falls die $(m \times n)$ -Jakobi-Matrix (der partiellen Ableitungen aller g_i nach allen Variablen x_j),

$$Dg(\mathbf{x}_0) = \frac{\partial f(\mathbf{x}_0)}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial g_1(\mathbf{x}_0)}{\partial x_1} & \dots & \frac{\partial g_1(\mathbf{x}_0)}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial g_m(\mathbf{x}_0)}{\partial x_1} & \dots & \frac{\partial g_m(\mathbf{x}_0)}{\partial x_n} \end{pmatrix},$$

(In der i -ten Zeile dieser Matrix steht der Gradient von $g_i(\mathbf{x})$ an der Stelle \mathbf{x}_0).

vollen Rang m besitzt, dann gibt es eindeutig bestimmte Zahlen $\lambda_1, \dots, \lambda_m \in \mathbb{R}$, die Lagrange-Multiplikatoren, so dass gilt:

$$(+) \quad \text{grad } f(\mathbf{x}_0) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x}_0) = 0.$$

Vorgehensweise:

1. Lagrange-Funktion aufstellen:

$$L(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m) = f(x_1, \dots, x_n) + \sum_{i=1}^m \lambda_i g_i(x_1, \dots, x_n)$$

2. Stationäre Punkte berechnen, d.h. das Gleichungssystem (+) lösen.

3. Hesse-Matrix $H(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m)$ nach x_1, \dots, x_n bei festen Lagrange-Multiplikatoren berechnen.

4. Stationäre Punkte einsetzen und Hesse-Matrix auf positive bzw. negative Definitheit prüfen.

1.13. Mehrdimensionale Integration

Sei $f : D \rightarrow \mathbb{R}$ eine (stückweise) stetige Funktion und $(\mathbf{a}, \mathbf{b}] = (a_1, b_1] \times \dots \times (a_n, b_n]$, $\mathbf{a} = (a_1, \dots, a_n)$, $\mathbf{b} = (b_1, \dots, b_n) \in \mathbb{R}^n$ ein Intervall.

Das n -dimensionale Integral von $f(x_1, \dots, x_n)$ über $(\mathbf{a}, \mathbf{b}]$ ist:

$$I = \int_{(\mathbf{a}, \mathbf{b}]} f(x_1, \dots, x_n) dx_1 \dots dx_n$$

Berechnung: Durch sukzessives (schrittweises) Integrieren 'von innen nach außen':

$$I = \int_{a_1}^{b_1} \dots \left(\int_{a_n}^{b_n} f(x_1, \dots, x_n) dx_n \right) \dots dx_1.$$

Eigenschaft: Die Reihenfolge der Integrationen ist vertauschbar.

$$\int_a^b \left(\int_c^d f(x, y) dy \right) dx = \int_c^d \left(\int_a^b f(x, y) dx \right) dy.$$

2. Wahrscheinlichkeitsrechnung

2.1. Grundbegriffe

Ein **Zufallsexperiment** ist ein zufallsbehafteter Vorgang, dessen Ausgang nicht deterministisch festgelegt ist. Die Menge aller möglichen Ausgänge eines Zufallsexperiments heißt **Ergebnismenge (Grundmenge)** und wird mit Ω bezeichnet. Ein Element $\omega \in \Omega$ wird als **Ausgang (Ergebnis, Versuchsausgang)** bezeichnet.

Ist Ω eine höchstens abzählbar unendliche Grundmenge, dann heißt jede Teilmenge $A \subset \Omega$ **Ereignis**. Die **Potenzmenge**

$$\text{Pot}(\Omega) = \{A | A \subset \Omega\}$$

aller Teilmengen von Ω bildet die Menge aller Ereignisse und heißt in diesem Zusammenhang auch **Ereignisalgebra**. Das Ereignis A tritt ein, wenn $\omega \in A$. Ein Ereignis A mit $A = \{\omega\}$ für ein $\omega \in \Omega$ wird **Elementarereignis** genannt.

Die Schnittmenge

$$A \cap B = \{x | x \in A \text{ und } x \in B\}$$

zweier Ereignisse $A, B \subset \Omega$ heißt **UND-Ereignis** und die Menge

$$A \cup B = \{x | x \in A \text{ oder } x \in B\}$$

ODER-Ereignis. Das **komplementäre Ereignis** ist durch

$$\overline{A} = A^c = \{x | x \in \Omega \text{ und } x \notin A\} = \Omega \setminus A$$

gegeben und entspricht der logischen Negation.

Für Ereignisse A_1, A_2, \dots ist

$$\bigcup_{k=1}^{\infty} A_k = A_1 \cup A_2 \cup \dots = \{\omega \in \Omega : \omega \in A_k \text{ für mind. ein } k\}$$

das Ereignis, dass mindestens eines der Ereignisse A_k eintritt und

$$\bigcap_{k=1}^{\infty} A_k = A_1 \cap A_2 \cap \dots = \{\omega \in \Omega : \omega \in A_k \text{ für alle } k = 1, 2, \dots\}$$

das Ereignis, dass alle Ereignisse A_k eintreten.

Rechenregeln

1. Distributivgesetze:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad \text{und} \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

2. Regeln von DeMorgan:

$$\overline{(A \cup B)} = \overline{A} \cap \overline{B} \quad \text{und} \quad \overline{(A \cap B)} = \overline{A} \cup \overline{B}.$$

Wahrscheinlichkeitsmaß

Eine Abbildung P , die jedem Ereignis $A \subset \Omega$ eine Zahl $P(A)$ zuordnet, ist ein **Wahrscheinlichkeitsmaß** oder eine **Wahrscheinlichkeitsverteilung**, wenn die Kolmogorov-Axiome erfüllt sind:

1. $0 \leq P(A) \leq 1$ für alle Ereignisse A ,
2. Normierung: $P(\Omega) = 1$,
3. Für disjunkte Mengen A_1, A_2, \dots gilt

$$P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots = \sum_{k=1}^{\infty} P(A_k).$$

Rechenregeln

Seien $A, B \subset \Omega$ Ereignisse. Dann gilt:

1. $P(\overline{A}) = 1 - P(A)$.
2. Für $A \subset B$ gilt: $P(B \setminus A) = P(B) - P(A)$.
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
4. $P(A \cap B) = P(A) + P(B) - P(A \cup B)$.

Laplace-Raum

In einem **Laplace-Raum** (Ω, P) ist die Ergebnismenge $\Omega = \{\omega_1, \dots, \omega_K\}$ endlich und jedes Elementarereignis besitzt dieselbe

Wahrscheinlichkeit

$$P(\omega) = P(\{\omega\}) = \frac{1}{K}, \quad \omega \in \Omega.$$

P heißt auch (**diskrete**) **Gleichverteilung** auf Ω .

Regel

In einem Laplace-Raum (Ω, P) gilt für jedes Ereignis A :

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der für } A \text{ günstigen Fälle}}{\text{Anzahl aller Fälle}},$$

wobei $|A|$ die Kardinalität, d.h. die Anzahl der Elemente von A bezeichnet.

Chancen

Die **Chance** $o = o(A)$ eines Ereignissen A ist durch den Quotienten der Wahrscheinlichkeit $p = P(A)$ von A und der komplementären Wahrscheinlichkeit $P(\bar{A}) = 1 - p$ definiert:

$$o = o(A) = \frac{p}{1-p}.$$

Die **logarithmischen Chancen** sind gegeben durch

$$\log(o) = \log\left(\frac{p}{1-p}\right) = \log(p) - \log(1-p).$$

Es gilt:

$$\log o(\bar{A}) = -\log o(A).$$

Sind A und \bar{A} gleichwahrscheinlich, dann gilt $o = 1$ und $\log(o) = 0$.

Die Chancen $o(A)$ und $o(B)$ von zwei Ereignissen A und B können durch das **Chancenverhältnis**

$$r = \frac{o(A)}{o(B)} = \frac{P(A)/(1-P(A))}{P(B)/(1-P(B))}$$

verglichen werden.

Siebformel

Für Ereignisse $A_1, \dots, A_n \subset \Omega$ gilt:

$$\begin{aligned} P(A_1 \cup \dots \cup A_n) &= \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) \\ &\quad \mp \dots + (-1)^n P(A_1 \cap \dots \cap A_n). \end{aligned}$$

Drei Ereignisse

Es gilt:

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C). \end{aligned}$$

σ -Algebra

Eine **σ -Algebra (Ereignisalgebra)** ist ein Mengensystem $\mathcal{A} \subset \text{Pot}(\Omega)$ von Teilmengen von Ω , für das gilt:

1. $\Omega \in \mathcal{A}, \emptyset \in \mathcal{A}$.
2. Ist A in \mathcal{A} , dann auch \bar{A} .
3. Sind A_1, A_2, \dots Mengen aus \mathcal{A} , dann ist auch $\bigcup_{i=1}^{\infty} A_i$ in \mathcal{A} .

Die Elemente von \mathcal{A} heißen **Ereignisse**.

Ist $\varepsilon \subset \text{Pot}(\Omega)$ eine Menge von Teilmengen von Ω , dann existiert eine kleinste σ -Algebra $\sigma(\varepsilon)$, die ε umfasst. Dies ist gerade der Schnitt über alle σ -Algebren, die ε umfassen. ε heißt **Erzeuger** dieser σ -Algebra.

Wichtige Fälle

1. $\Omega = \mathbb{R}$: Die **Borel- σ -Algebra (Borelsche Ereignisalgebra)** \mathcal{B} wird durch alle endlichen Intervalle der Form $(a, b], a \leq b, a, b \in \mathbb{R}$, erzeugt.
Die Elemente von \mathcal{B} heißen **Borelsche Mengen**.
2. $\Omega = \mathcal{X} \subset \mathbb{R}$: Hier wählt man die Ereignisalgebra $\mathcal{B}(\mathcal{X}) = \{B \cap \mathcal{X} : b \in \mathcal{B}\}$, die auch **Spur- σ -Algebra** heißt.

3. $\Omega = \mathbb{R}^n$: Als Erzeuger wählt man die Menge aller Rechtecke

$$(\mathbf{a}, \mathbf{b}] = (a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_n, b_n],$$

wobei $\mathbf{a} = (a_1, \dots, a_n)$, $\mathbf{b} = (b_1, \dots, b_n) \in \mathbb{R}^n$. Die erzeugte σ -Algebra heißt **Borelsche σ -Algebra**.

4. $\mathcal{X} \subset \mathbb{R}^n$: Als σ -Algebra wählt man alle Mengen der Form $B \cap \mathcal{X}$, wobei B eine Borelsche Menge des \mathbb{R}^n ist.

Bedingte Wahrscheinlichkeit

Sind $A, B \in \Omega$ Ereignisse mit $P(B) > 0$, dann heißt

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

bedingte Wahrscheinlichkeit von A gegeben B .

In einem Laplace-Raum gilt

$$P(A|B) = \frac{|A \cap B|}{|B|}.$$

Sind $A, B \in \Omega$ Ereignisse mit $P(B) > 0$, dann gilt

$$P(A \cap B) = P(A|B)P(B).$$

Sind A_1, \dots, A_n Ereignisse mit $P(A_1 \cap \dots \cap A_n) > 0$, dann gilt

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2 \cap A_1) \dots P(A_n|A_{n-1} \cap \dots \cap A_1).$$

Totale Wahrscheinlichkeit

Ist A_1, \dots, A_K eine disjunkte Zerlegung von Ω , d.h.

$$\Omega = A_1 \cup \dots \cup A_K, \quad A_i \cap A_j = \emptyset, i \neq j,$$

dann gilt

$$P(B) = \sum_{i=1}^K P(B|A_i)P(A_i).$$

Sinngemäß ist auch $K = \infty$ möglich.

Satz von Bayes

Ist A_1, \dots, A_K eine disjunkte Zerlegung von Ω mit $P(A_i) > 0$ für alle $i = 1, \dots, K$, dann gilt für jedes Ereignis $P(B)$ mit $P(B) > 0$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{k=1}^K P(B|A_k)P(A_k)}.$$

Sinngemäß ist auch $K = \infty$ möglich.

Mehrstufige Wahrscheinlichkeitsmodelle

Betrachte ein Zufallsexperiment Ω , das aus n Teilexperimenten $\Omega_1, \dots, \Omega_n$ besteht. Dann definiert die **Startverteilung** auf Ω_1 ,

$$p(\omega_1), \quad \omega_1 \in \Omega_1,$$

die Wahrscheinlichkeiten von Ereignissen des ersten Teilexperiments. Allgemein sei

$$p(\omega_j|\omega_1, \dots, \omega_{j-1})$$

die bedingte Wahrscheinlichkeit, dass ω_j eintritt, wenn zuvor die Ausgänge $\omega_1, \dots, \omega_{j-1}$ eingetreten sind. Dann gilt für $\omega \in \Omega$

$$p(\omega) = p(\omega_1)p(\omega_2|\omega_1) \dots p(\omega_n|\omega_1, \dots, \omega_{n-1}).$$

Stochastische Unabhängigkeit

Zwei Ereignisse A und B heißen (**stochastisch**) **unabhängig**, wenn

$$P(A \cap B) = P(A)P(B).$$

Diese Identität wird **Produktsatz** genannt. Allgemein ist der Produktsatz für k Ereignisse $A_1, \dots, A_k \subset \Omega$ erfüllt, wenn

$$P(A_1 \cap \dots \cap A_k) = P(A_1) \dots P(A_k).$$

n Ereignisse $A_1, \dots, A_n \subset \Omega$ sind (**total**) **stochastisch unabhängig**, wenn der Produktsatz für jede Teilauswahl A_{i_1}, \dots, A_{i_k} von k Ereignissen erfüllt ist. A_1, \dots, A_n heißen **paarweise stochastisch unabhängig**, wenn alle Paare $A_i, A_j, i \neq j$, stochastisch unabhängig sind.

Eigenschaften

Sind $A_1, \dots, A_n \subset \Omega$ unabhängig und $B_1, \dots, B_k, k \leq n$, Ereignisse, von denen jedes B_i entweder A_i oder $\overline{A_i}$ ist, dann sind auch B_1, \dots, B_k unabhängig.

2.2. Zufallsvariablen und ihre Verteilung

Eine **Zufallsvariable** ist eine Abbildung

$$X : \Omega \rightarrow \mathcal{X} \subset \mathbb{R}, \quad \omega \mapsto X(\omega),$$

einer abzählbaren Ergebnismenge in die reellen Zahlen. Tritt $\omega \in \Omega$ ein, dann heißt $x = X(\omega)$ **Realisation**. Ist Ω überabzählbar und mit einer Ergebnisalgebra \mathcal{A} versehen, dann müssen zusätzlich alle Teilmengen der Form $\{\omega \in \Omega : X(\omega) \in B\}$, wobei B eine Borelsche Menge von \mathcal{X} ist, Ereignisse von Ω sein, d.h.

$$\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A} \quad \text{für alle Ereignisse } B \text{ von } \mathcal{X}.$$

Ist die Menge $\mathcal{X} = \{X(\omega) : \omega \in \Omega\}$ diskret, dann heißt X **diskrete Zufallsvariable**.

Eigenschaften

Wenn die Ergebnismenge Ω diskret ist, dann sind alle Zufallsvariablen $X : \Omega \rightarrow \mathcal{X}$ diskret.

Verteilung

Die **Verteilung von X** ordnet jedem Ereignis A von \mathcal{X} die Wahrscheinlichkeit $P(X \in A)$ zu und wird mit $P_X(A)$ bezeichnet.

Die **Verteilungsfunktion von X** ist durch die Funktion $F_X : \mathbb{R} \rightarrow [0, 1]$,

$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R},$$

gegeben.

Eigenschaften

1. $F_X(x)$ ist monoton wachend und rechtsstetig.

2. Es gilt

$$F(-\infty) := \lim_{x \rightarrow -\infty} F_X(x) = 0, \quad F(\infty) := \lim_{x \rightarrow \infty} F_X(x) = 1.$$

3. Es gilt:

$$P(X < x) = F(x-) = \lim_{z \uparrow x} F(z).$$

4. Es gilt:

$$P(X = x) = F(x) - F(x-).$$

Jede monoton wachsende und rechtsstetige Funktion $F : \mathbb{R} \rightarrow [0, 1]$ mit $F(-\infty) = 0$ und $F(\infty) = 1$ heißt Verteilungsfunktion (auf \mathbb{R}) und besitzt obige Eigenschaften.

Quantifunktion

Ist $F(x)$ eine Verteilungsfunktion, dann heißt die Funktion $F^{-1} : [0, 1] \rightarrow \mathbb{R}$,

$$F^{-1}(p) = \min\{x \in \mathbb{R} : F(x) \geq p\}, \quad p \in (0, 1),$$

Quantifunktion von F . Für ein festes p heißt $F^{-1}(p)$ (**theoretisches**) p -**Quantil**.

Ist $F(x)$ stetig und streng monoton steigend, dann ist $F^{-1}(p)$ die Umkehrfunktion von $F(x)$.

Dichte

Diskrete Zufallsvariablen

Ist X eine diskrete Zufallsvariable mit Werten in $\mathcal{X} = \{x_1, x_2, \dots\} \subset \mathbb{R}$, dann heißt die Funktion

$$p_X(x) = P(X = x), \quad x \in \mathbb{R},$$

Wahrscheinlichkeitsfunktion oder **Zähldichte** von X .

Es gilt:

$$\sum_{x \in \mathcal{X}} p_X(x) = \sum_{i=1}^{\infty} p_X(x_i) = 1.$$

Die Verteilung von X wird eindeutig durch die Zähldichte bestimmt. Die Zähldichte wird durch die Punktwahrscheinlichkeiten

$$p_i = P(X = x_i), \quad i = 1, 2, \dots,$$

festgelegt: Es gilt $p_X(x_i) = p_i$ und $p_X(x) = 0$, wenn $x \notin \mathcal{X}$. Wenn X nur endlich viele Werte x_1, \dots, x_k annehmen kann, wird (p_1, \dots, p_k) **Wahrscheinlichkeitsvektor** genannt.

Stetige Zufallsvariablen

Eine Zufallsvariable X ist **stetig (verteilt)**, wenn es eine nicht-negative, integrierbare Funktion $f_X(x)$ gibt, sodass für alle Intervalle $(a, b] \subset \mathbb{R}$ gilt:

$$P_X((a, b]) = P(a < X \leq b) = \int_a^b f(x) dx.$$

$f_X(x)$ heißt **Dichtefunktion** (kurz: **Dichte**).

Allgemein heißt jede Funktion $f : \mathbb{R} \rightarrow [0, \infty]$ mit $f(x) \geq 0, x \in \mathbb{R}$, und $\int_{-\infty}^{\infty} f(x) dx = 1$ Dichtefunktion.

Dichtetransformation

Sei $y = g(x)$ eine stetig differenzierbare Funktion, d.h. $g : (a, b) \rightarrow (c, d)$ mit Umkehrfunktion $x = g^{-1}(y)$ und $(g^{-1})'(y) \neq 0$ für alle $y \in (c, d)$. Dann ist die Dichtefunktion der Zufallsvariablen $Y = g(X)$ gegeben durch

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|, \quad y \in (c, d).$$

Stochastische Unabhängigkeit

1. Zwei Zufallsvariablen X und Y mit Werten in \mathcal{X} bzw. \mathcal{Y} sind **(stochastisch) unabhängig**, wenn für alle Ereignisse $A \subset \mathcal{X}$ und für alle Ereignisse $B \subset \mathcal{Y}$ gilt:

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B).$$

2. Allgemein heißen n Zufallsvariablen X_1, \dots, X_n mit Werten in $\mathcal{X}_1, \dots, \mathcal{X}_n$ **(stochastisch) unabhängig**, wenn für alle Ereignisse $A_1 \subset \mathcal{X}_1, \dots, A_n \subset \mathcal{X}_n$ die Ereignisse $\{X_1 \in A_1\}, \dots, \{X_n \in A_n\}$ (total) unabhängig sind. D. h. für alle $i_1, \dots, i_k \in \{1, \dots, n\}, 1 \leq k \leq n$, gilt:

$$P(X_{i_1} \in A_{i_1}, \dots, X_{i_k} \in A_{i_k}) = P(X_{i_1} \in A_{i_1}) \cdots P(X_{i_k} \in A_{i_k}).$$

Kriterium für diskrete Zufallsvariablen

Zwei diskrete Zufallsvariablen X und Y sind stochastisch unabhängig, wenn die Ereignisse $\{X = x_i\}$ und $\{Y = y_i\}$ für alle Realisationen x_i von X und y_i von Y stochastisch unabhängig sind, d.h.

$$P(X = x_i, Y = y_i) = P(X = x_i)P(Y = y_i).$$

Für stochastisch unabhängige diskrete Zufallsvariablen gilt

$$P(X = x_i | Y = y_i) = P(X = x_i) \quad \text{und} \quad P(Y = y_i | X = x_i) = P(Y = y_i).$$

Kriterium für stetige Zufallsvariablen

Zwei stetige Zufallsvariablen X und Y sind stochastisch unabhängig, wenn die Ereignisse $\{a < X \leq b\}$ und $\{c < Y \leq d\}$ für alle Intervalle $(a, b]$ und $(c, d]$ unabhängig sind, d.h.

$$P(a < X \leq b, c < Y \leq d) = \int_a^b f_X(x) dx \int_c^d f_Y(y) dy = \int_a^b \int_c^d f_X(x) f_Y(y) dy dx.$$

Zufallsstichprobe

Eine (**einfache**) **Zufallsstichprobe** besteht aus n unabhängig und identisch verteilten Zufallsvariablen X_1, \dots, X_n , d.h.

- X_1, \dots, X_n sind stochastisch unabhängig und
- X_1, \dots, X_n sind identisch verteilt:

$$P(X_i \in A) = P(X_1 \in A), \quad i = 1, \dots, n.$$

Sind die X_i nach der Verteilungsfunktion $F(x) = F_X(x)$ verteilt, dann schreibt man

$$X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} F(x).$$

Hier steht i.i.d. für **unabhängig und identisch verteilt** (engl.: *independent and identically distributed*).

Faltung

Diskrete Faltung

Sind $X \sim p_X(x)$ und $Y \sim p_Y(y)$ unabhängige Zufallsvariablen, dann ist die Verteilung der Summenvariable $Z = X + Y$ gegeben durch die **diskrete Faltung**

$$P(Z = z) = \sum_{y \in \mathcal{Y}} p_X(z - y)p_Y(y) = \sum_{x \in \mathcal{X}} p_Y(z - x)p_X(x)$$

für $z \in \mathcal{Z} = \{x + y : x \in \mathcal{X}, y \in \mathcal{Y}\}$.

Stetige Faltung

Sind $X \sim f_X(x)$ und $Y \sim f_Y(y)$ unabhängige stetige Zufallsvariablen, dann ist die Dichte der Summenvariablen $Z = X + Y$ gegeben durch die **stetige Faltung**

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z - y)f_Y(y)dy = \int_{-\infty}^{\infty} f_Y(z - x)f_X(x)dx.$$

2.3. Erwartungswert, Varianz, Momente und Entropie

Erwartungswert einer diskreten Zufallsvariablen

Ist X eine diskrete Zufallsvariable mit Werten in \mathcal{X} und Zähldichte $p_X(x)$, $x \in \mathcal{X}$, dann ist der **Erwartungswert von X** definiert durch

$$E(X) = \sum_{x \in \mathcal{X}} x \cdot p_X(x),$$

falls $\sum_{x \in \mathcal{X}} |x| p_X(x) < \infty$.

Erwartungswert einer stetigen Zufallsvariablen

Ist $X \sim f_X(x)$ eine stetige Zufallsvariable, dann heißt

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Erwartungswert von X , sofern $\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty$.

Rechenregeln

Sind X und Y Zufallsvariablen und $a, b \in \mathbb{R}$, dann gilt:

1. $E(X + Y) = E(X) + E(Y)$,
2. $E(aX + b) = aE(X) + b$,
3. $E|X + Y| \leq E|X| + E|Y|$,
4. **Jensen-Ungleichung:** Ist $g(x)$ konvex, dann gilt $E(g(X)) \geq g(E(X))$ und $E(g(X)) > g(E(X))$, wenn $g(x)$ strikt konvex ist. Ist $g(x)$ konkav, dann gelten die Ungleichungen mit umgekehrten Ungleichheitszeichen.

Produkteigenschaft

Sind X und Y stochastisch unabhängige Zufallsvariablen, dann gilt für alle Funktionen $f(x)$ und $g(x)$

$$E(f(X)g(Y)) = E(f(X)) \cdot E(g(Y)),$$

sofern $E|f(X)| < \infty$ und $E|g(Y)| < \infty$. Insbesondere gilt dann

$$E(XY) = E(X) \cdot E(Y).$$

Transformationsformel

Ist X eine Zufallsvariable und $g : \mathcal{X} \rightarrow \mathcal{Y}$ eine Funktion mit $E|g(X)| < \infty$, dann gilt für den Erwartungswert der Zufallsvariablen $Y = g(X)$:

1. Sind $X \sim p_X(x)$ und $Y \sim p_Y(y)$ diskrete Zufallsvariablen, dann gilt

$$E(Y) = \sum_{x \in \mathcal{X}} g(x)p_X(x) = \sum_{y \in \mathcal{Y}} y p_Y(y).$$

2. Sind $X \sim f_X(x)$ und $Y \sim f_Y(y)$ stetige Zufallsvariablen, dann gilt

$$E(Y) = \int_{-\infty}^{\infty} g(x)f_X(x)dx = \int_{-\infty}^{\infty} y f_Y(y) dy.$$

Varianz

Ist X eine Zufallsvariable mit $E(X^2) < \infty$, dann heißt

$$\sigma_X^2 = \text{Var}(X) = E((X - E(X))^2)$$

Varianz von X . Der Ausdruck

$$\sigma_X = \sqrt{\text{Var}(X)}$$

heißt **Standardabweichung von X** .

Varianz und Stichprobenvarianz

Ist X eine diskrete Zufallsvariable mit Werten in $\mathcal{X} = \{x_1, \dots, x_n\}$ und $P(X = x_i) = \frac{1}{n}$ für alle $i = 1, \dots, n$, dann gilt

$$E(X) = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{und} \quad \text{Var}(X) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Verschiebungssatz

Es gilt $\text{Var}(X) = E(X^2) - (E(X))^2$.

Rechenregeln

Sind X und Y Zufallsvariablen, deren Varianzen existieren, und $a \in \mathbb{R}$, dann gilt:

1. $\text{Var}(aX) = a^2 \text{Var}(X)$.
2. $\text{Var}(X) = E(X^2)$, wenn $E(X) = 0$.
3. Sind X und Y stochastisch unabhängig, dann gilt

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y).$$

Kovarianz

Für Zufallsvariablen X und Y , deren Varianzen existieren, heißt

$$\text{Cov}(X, Y) = E(X - \mu_X)(Y - \mu_Y)$$

Kovarianz von X und Y .

Rechenregeln

Sind X , Y und Z Zufallsvariablen mit existierenden Varianzen, dann gilt:

1. $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$ für alle $a, b \in \mathbb{R}$.
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
3. $\text{Cov}(X, Y) = 0$, falls X und Y stochastisch unabhängig sind.
4. $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$.

Unkorreliertheit

Zwei Zufallsvariablen X und Y sind **unkorreliert**, wenn $\text{Cov}(X, Y) = 0$.

Aus stochastischer Unabhängigkeit folgt Unkorreliertheit, die Umkehrung gilt i. A. nicht.

Korrelationskoeffizient

Sind X und Y Zufallsvariablen mit Varianzen $\sigma_X^2 \in (0, \infty)$ und $\sigma_Y^2 \in (0, \infty)$, dann wird

$$\rho = \rho(X, Y) = \text{Cor}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

Korrelationskoeffizient von X und Y genannt.

Eigenschaften

Für Zufallsvariablen X und Y gilt:

1. $\text{Cor}(X, Y) = \text{Cor}(Y, X)$.
2. $-1 \leq \text{Cor}(X, Y) \leq 1$.
3. $|\text{Cor}(X, Y)| = 1$ gilt genau dann, wenn X und Y linear abhängig sind.
4. $\text{Cov}(X, Y) = 1$ genau dann, wenn $Y = a + bX$ mit $b > 0, a \in \mathbb{R}$.
5. $\text{Cov}(X, Y) = -1$ genau dann, wenn $Y = a + bX$ mit $b < 0, a \in \mathbb{R}$.

Cauchy-Schwarz-Ungleichung

Für Zufallsvariablen X und Y mit Varianzen $\sigma_X^2 \in (0, \infty)$ und $\sigma_Y^2 \in (0, \infty)$ gilt

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)} = \sigma_X \sigma_Y.$$

Momente

Ist X eine Zufallsvariable mit $E|X|^k < \infty$ und $a \in \mathbb{R}$, dann heißt

$$m_k(a) = E(X - a)^k, \quad m_k = m_k(0),$$

Moment k -ter Ordnung von X bzgl. a . Der Ausdruck

$$m_k^*(a) = E|X - a|^k, \quad m_k^* = m_k^*(0),$$

heißt **zentriertes Moment k -ter Ordnung von X bzgl. a** . Die Ausdrücke

$$\mu_k = m_k(E(X)) \quad \text{und} \quad \mu_k^* = m_k^*(E(X))$$

heißen **zentrales Moment** und **zentrales absolutes Moment**.

Einige Momente

- $m_1 = E(X)$,
- $m_2 = E(X^2)$ und $\mu_2 = \text{Var}(X)$,
- Das vierte Moment von $X^* = \frac{X - E(X)}{\sqrt{\text{Var}(X)}}$ ist $\beta_2 = E(X^*)^4 = \frac{m_4(X)}{\sigma_X^4}$ und heißt **Kurtosis**.
- Ist $X \sim N(\mu, \sigma^2)$, dann ist $\beta_2 = 3$ und $\gamma_2 = \beta_2 - 3$ heißt **Exzess**.

Entropie

Ist X eine diskrete Zufallsvariable mit Werten in $\mathcal{X} = \{x_1, x_2, \dots\}$ und zugehörigen Wahrscheinlichkeiten $p_i = P(X = x_i)$, dann heißt

$$H(X) = - \sum_{i=1}^{\infty} p_i \log_2(p_i)$$

Entropie von X .

Man setzt $0 \cdot \log_2(0) = 0$.

2.4. Diskrete Verteilungsmodelle

Bernoulli-Verteilung

Ist

$$X = 1_A = \begin{cases} 1, & A \text{ tritt ein} \\ 0, & A \text{ tritt nicht ein.} \end{cases}$$

und $p = P(X = 1)$ bzw. $q = 1 - p = P(X = 0)$, dann heißt X **Bernoulli-verteilt** mit Parameter $p \in [0, 1]$. Notation: $X \sim \text{Ber}(p)$.

- Erwartungswert: $E(X) = p$,
- Varianz: $\text{Var}(X) = p(1 - p)$,
- Zähldichte: $p(k) = p^k(1 - p)^{1-k}, k \in \{0, 1\}$.

Binomialverteilung

Für $n \in \mathbb{N}$ und $k \in \{0, \dots, n\}$ ist die Anzahl der Möglichkeiten, aus n Objekten k Objekte ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge auszuwählen durch den **Binomialkoeffizienten**

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n \cdot (n-1) \cdot (n-2) \dots (n-k+1)}{k \cdot (k-1) \dots 2 \cdot 1}$$

gegeben.

Ist

$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, \dots, n,$$

dann heißt Y **binomialverteilt** mit Parametern $n \in \mathbb{N}$ und $p \in [0, 1]$. Notation: $Y \sim \text{Bin}(n, p)$.

- Erwartungswert: $E(X) = np$,
- Varianz: $\text{Var}(X) = np(1 - p)$,
- Zähldichte: $p(k) = \binom{n}{k} p^k (1-p)^{n-k}, k \in \{0, 1\}$.

Für zwei unabhängige Zufallsvariablen $X \sim \text{Bin}(n_1, p)$ und $Y \sim \text{Bin}(n_2, p)$ gilt $X + Y \sim \text{Bin}(n_1 + n_2, p)$.

Hypergeometrische Verteilung

Eine Zufallsvariable X mit Zähldichte

$$P(X = r) = p_r = \frac{\binom{R}{r} \binom{B}{n-r}}{\binom{R+B}{n}}, \quad \max(0, n-B) \leq r \leq \min(R, n), \quad n = r+b,$$

heißt **hypergeometrisch verteilt**.

Geometrische Verteilung

Eine Zufallsvariable T ist geometrisch verteilt mit Parameter $p \in (0, 1]$, wenn

$$P(T = n) = p(1-p)^{n-1}, \quad n = 1, 2, \dots$$

$W = T - 1$ heißt **Wartezeit**. Notation: $T \sim \text{Geo}(p)$.

- Erwartungswert: $E(T) = \frac{1}{p}$, $E(W) = \frac{1}{p} - 1$,
- Varianz: $\text{Var}(T) = \text{Var}(W) = \frac{1-p}{p^2}$.

Negative Binomialverteilung

Ist $S_k = T_1 + \dots + T_k$ die Summe von k unabhängig und identisch $\text{Geo}(p)$ -verteilten Zufallsvariablen, dann ist S_k **negativ binomialverteilt** mit

$$P(S_k = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k}, \quad n = k, k+1, \dots$$

- Erwartungswert: $E(S_n) = \frac{k}{p}$,
- Varianz: $\text{Var}(S_n) = \frac{k(1-p)}{p^2}$.

Poisson-Verteilung

Poisson-Grenzwertsatz

Sind $Y_n \sim \text{Bin}(n, p_n)$, $n = 1, 2, \dots$, binomialverteilte Zufallsvariablen mit $np_n \rightarrow \lambda$, $n \rightarrow \infty$, dann gilt für festes k :

$$\lim_{n \rightarrow \infty} P(Y_n = k) = p_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Eine Zufallsvariable Y mit $P(Y = k) = p_\lambda(k)$ heißt **poissonverteilt** mit Parameter λ . Notation: $Y \sim \text{Poi}(\lambda)$.

- Erwartungswert: $E(Y) = \lambda$,
- Varianz: $\text{Var}(Y) = \lambda$,
- Zähldichte: $p(k) = e^{-\lambda} \frac{\lambda^k}{k!}$.

Rechenregeln

1. Sind $X \sim \text{Poi}(\lambda_1)$ und $Y \sim \text{Poi}(\lambda_2)$ unabhängig, dann ist $X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$.
2. Ist $X \sim \text{Poi}(\lambda)$ die Anzahl der Ereignisse in $[0, T]$ und Y die Anzahl der Ereignisse in dem Teilintervall $[0, r \cdot T]$, dann ist $Y \sim \text{Poi}(r \cdot \lambda)$.

2.5. Stetige Verteilungsmodelle

Stetige Gleichverteilung

Eine Zufallsvariable X mit Dichtefunktion

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & x \notin [a, b] \end{cases}$$

heißt **(stetig) gleichverteilt auf dem Intervall** $[a, b]$. Notation: $X \sim U[a, b]$.

- Erwartungswert: $E(X) = \frac{b+a}{2}$,
- Varianz: $\text{Var}(X) = \frac{(b-a)^2}{12}$,
- Verteilungsfunktion:

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & x \in [a, b] \\ 1, & x > b \end{cases}$$

Exponentialverteilung

Eine Zufallsvariable Y mit Dichtefunktion

$$f(y) = \lambda e^{-\lambda y}, \quad y > 0,$$

heißt **exponentialverteilt** mit Parameter λ . Notation: $Y \sim \text{Exp}(\lambda)$.

- Erwartungswert: $E(Y) = \frac{1}{\lambda}$,
- Varianz: $\text{Var}(Y) = \frac{1}{\lambda^2}$,
- Verteilungsfunktion: $F(y) = 1 - e^{-\lambda y}, \quad y > 0$.

Normalverteilung

Ist eine Zufallsvariable **normalverteilt** mit Parametern $\mu \in \mathbb{R}$ und $\sigma^2 \in (0, \infty)$, dann ist die Dichte gegeben durch

$$\varphi_{(\mu, \sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$$

Notation: $X \sim N(\mu, \sigma^2)$. Die **Standardnormalverteilung** erhält man mit $\mu = 0$ und $\sigma^2 = 1$, man schreibt die Dichte dann meist verkürzt als $\varphi(x)$.

- Erwartungswert: $E(X) = \mu$,
- Varianz: $\text{Var}(X) = \sigma^2$.

Für die Verteilungsfunktion

$$\Phi_{(\mu, \sigma^2)}(x) = \int_{-\infty}^x \varphi_{(\mu, \sigma^2)}(t) dt, \quad x \in \mathbb{R},$$

und die p -Qantile $z_p = \Phi_{(\mu, \sigma^2)}^{-1}(p)$, $p \in (0, 1)$, der Normalverteilung gibt es keine explizite Formel.

Eigenschaften

$\varphi(x)$ und $\Phi(x)$ seien die Dichte und die Verteilungsfunktion der Standardnormalverteilung. Dann gilt:

1.

$$\Phi_{(\mu, \sigma^2)}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R},$$

und

$$\varphi_{(\mu, \sigma^2)}(x) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right),$$

sowie

$$\Phi_{(\mu, \sigma^2)}^{-1}(p) = \mu + \sigma \Phi^{-1}(p), \quad p \in (0, 1).$$

2. Sind $X \sim N(\mu_1, \sigma_1^2)$ und $Y \sim N(\mu_2, \sigma_2^2)$ unabhängig, dann gilt für $a, b \in \mathbb{R}$

$$aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

3. Ist $X \sim N(\mu, \sigma^2)$, dann ist

$$X^* = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

4. Sind $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ unabhängig, dann gilt

$$\bar{X} \sim N(\mu, \sigma^2/n) \quad \text{und} \quad \bar{X}^* = \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \sim N(0, 1).$$

5. Ist $X^* \sim N(0, 1)$, dann gilt $\mu + \sigma X^* \sim N(\mu, \sigma^2)$, wenn $\mu \in \mathbb{R}$ und $\sigma > 0$.

Betaverteilung

Eine Zufallsvariable X ist **betaverteilt**, wenn sie die Dichte

$$f_{(p,q)} = \frac{x^{p-1}(1-x)^{q-1}}{B(p,q)}, \quad x \in [0, 1],$$

besitzt, wobei

$$B(p, q) = \int_0^1 x^{p-1}(1-x)^{q-1}, \quad p, q \in [0, 1],$$

die **Betafunktion** ist. Notation: $X \sim \text{Beta}(p, q)$.

- Erwartungswert: $E(X) = \frac{p}{p+q}$,
- Varianz: $\text{Var}(X) = \frac{pq}{(p+q+1)(p+q)^2}$.

Gammaverteilung

Eine Zufallsvariable X mit Dichte

$$f(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}, \quad x > 0,$$

heißt **gammaverteilt** mit Parametern $a > 0$ und $\lambda > 0$. Hierbei ist $\Gamma(x)$ die **Gammafunktion**. Notation: $X \sim \Gamma(a, \lambda)$.

- Erwartungswert: $E(X) = \frac{a}{\lambda}$,
- Varianz: $\text{Var}(X) = \frac{a}{\lambda^2}$.

2.6. Zufallszahlen

Gemischter linearer Kongruenzoperator

Der gemischte lineare Kongruenzoperator erzeugt Zufallszahlen mit maximaler Periodenlänge m , die annähernd $U[0, 1]$ -verteilt sind:

1. Wähle einen Startwert $y_1 \in \{0, \dots, m - 1\}$!
2. Berechne die Folge

$$y_i = (ay_{i-1} + b) \mod m$$

mit $a, b \in \{1, \dots, m - 1\}$!

3. Gebe y_i/m aus!

- Gute Resultate erhält man für $m = 2^{35}$, $a = 2^7 + 1$ und $b = 0$.
- Für kryptografische Zwecke ist diese Methode nicht sicher genug.

Inversionsmethode

Ist $U \sim U[0, 1]$, dann hat die Zufallsvariable $X = F^{-1}(U)$ die Verteilungsfunktion $F(x)$.

Poissonverteilte Zufallszahlen

Sind $Y_1, \dots, Y_n \sim \text{Exp}(1)$ unabhängig, dann ist die Zahl $X \geq 0$ mit

$$\sum_{i=1}^X Y_i < \lambda \leq \sum_{i=1}^{X+1} Y_i$$

poissonverteilt mit Erwartungswert λ .

Box-Muller-Methode

Sind die beiden Zufallsvariablen U_1 und U_2 unabhängig und identisch $U[0, 1]$ -verteilt, dann sind $Z_1 = \sqrt{-2 \ln U_1} \cos(2\pi U_2)$ und $Z_2 = \sqrt{-2 \ln U_1} \sin(2\pi U_2)$ unabhängig und identisch $N(0, 1)$ -verteilt.

2.7. Zufallsvektoren

Ist Ω abzählbar, dann wird eine Abbildung

$$\mathbf{X} : \Omega \rightarrow \mathbb{R}^n, \quad \omega \mapsto \mathbf{X}(\omega) = (X_1(\omega), \dots, X_n(\omega)),$$

Zufallsvektor genannt.

Ist Ω überabzählbar und mit einer Ergebnisalgebra \mathcal{A} versehen, dann müssen alle $X_i, i = 1, \dots, n$, die folgende Bedingung erfüllen: Alle Teilmengen der Form $\{\omega \in \Omega : X_i(\omega) \in B\}$, wobei B eine Borelsche Menge ist, müssen Ereignisse von Ω sein.

Verteilung

Für einen Zufallsvektor $\mathbf{X} = (X_1, \dots, X_n)$ mit Werten in $\mathcal{X} \subset \mathbb{R}^n$ wird durch

$$P_{\mathbf{X}}(A) = P(\mathbf{X} \in A) = P((X_1, \dots, X_n) \in A)$$

eine Wahrscheinlichkeitsverteilung auf \mathcal{X} definiert, wobei jedem Ereignis $A \in \mathcal{X}$ die Wahrscheinlichkeit zugeordnet wird, dass \mathbf{X} in der Menge A realisiert wird. $P_{\mathbf{X}}$ heißt **Verteilung** von \mathbf{X} .

Sei $\mathbf{X} = (X_1, \dots, X_n)$ ein Zufallsvektor mit Werten in $\mathcal{X} \subset \mathbb{R}^n$. Dann heißt die Funktion $F : \mathbb{R}^n \rightarrow [0, 1]$,

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n), \quad x_1, \dots, x_n \in \mathbb{R},$$

Verteilungsfunktion von \mathbf{X} .

Eigenschaften

- $F(x_1, \dots, x_n)$ ist in jedem Argument monoton wachsend.

•

$$\lim_{x_i \rightarrow -\infty} F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_{i-1} \leq x_{i-1}, X_{i+1} \leq x_{i+1}, \dots, X_n \leq x_n).$$

•

$$\lim_{x_i \rightarrow \infty} F(x_1, \dots, x_n) = 0, \quad \lim_{x_1, \dots, x_n \rightarrow \infty} F(x_1, \dots, x_n) = 1.$$

Produktverteilung

Wenn $F_1(x), \dots, F_n(x)$ Verteilungsfunktionen auf \mathbb{R} sind, dann ist

$$F(x_1, \dots, x_n) = F_1(x_1) \cdot F_2(x_2) \cdot \dots \cdot F_n(x_n)$$

eine Verteilungsfunktion auf \mathbb{R}^n , deren zugehörige Verteilung **Produktverteilung** heißt.

Für einen Zufallsvektor $\mathbf{X} = (X_1, \dots, X_n) \sim F_1(x_1) \cdot F_2(x_2) \cdot \dots \cdot F_n(x_n)$ gilt:

1. $X_i \sim F_i(x), x \in \mathbb{R}$, für alle $i = 1, \dots, n$,
2. X_1, \dots, X_n sind stochastisch unabhängig.

Diskrete Zufallsvektoren

Nimmt ein Zufallsvektor nur Werte in einer diskreten Menge an, dann wird er **diskreter Zufallsvektor** genannt.

Zähldichte

Die Funktion $p_{\mathbf{X}} : \mathbb{R}^n \rightarrow [0, 1]$,

$$p_{\mathbf{X}}(\mathbf{x}) = P(\mathbf{X} = \mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^n,$$

wird **(multivariate) Zähldichte** (oder **Wahrscheinlichkeitsfunktion**) von \mathbf{X} genannt.

Sind $\mathcal{X} = \{x_1, x_2, \dots\} \subset \mathbb{R}^n$ eine diskrete Punktmenge und p_1, p_2, \dots Zahlen aus dem Intervall $[0, 1]$ mit $\sum_{i=1}^n p_i = 1$, dann erhält man durch

$$p(\mathbf{x}) = \begin{cases} p_i, & \mathbf{x} = \mathbf{x}_i \\ 0, & \mathbf{x} \notin \mathcal{X} \end{cases}$$

eine Zähldichte p .

Produkt-Zähldichte

Sind $p_1(x), \dots, p_n(x)$ Zähldichten auf den Mengen $\mathcal{X}_1, \dots, \mathcal{X}_n$, dann ist

$$p(x_1, \dots, x_n) = p(x_1) \cdot \dots \cdot p(x_n)$$

eine Zähldichte auf $\mathcal{X}_1 \times \dots \times \mathcal{X}_n$ und heißt **Produkt-Zähldichte**.

Ist $(X_1, \dots, X_n) \sim p(x_1) \cdot \dots \cdot p(x_n)$, dann sind die Koordinaten unabhängig mit $X_i \sim p_i(x), i = 1, \dots, n$.

Rand-Zähldichte

Ist $(X, Y) \sim p_{(X,Y)}(x, y)$, dann heißen

$$p_X(x) = \sum_{y \in \mathcal{Y}} p_{(X,Y)}(x, y) \quad \text{und} \quad p_Y(y) = \sum_{x \in \mathcal{X}} p_{(X,Y)}(x, y)$$

Rand-Zähldichten.

Allgemein erhält man die Rand-Zähldichte eines Teilvektors durch Summation über alle Komponenten, die nicht zum Teilvektor gehören.

Stetige Zufallsvektoren

Ein Zufallsvektor $\mathbf{X} = (X_1, \dots, X_n)$ ist **stetig (verteilt)**, wenn eine nichtnegative Funktion $f_{\mathbf{X}}(x_1, \dots, x_n)$ existiert, sodass

$$\begin{aligned} P(\mathbf{X} \in (\mathbf{a}, \mathbf{b})) &= P(a_1 < X_1 \leq b_1, \dots, a_n < X_n \leq b_n) \\ &= \int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n \end{aligned}$$

für alle Intervalle $(\mathbf{a}, \mathbf{b}) \subset \mathbb{R}^n$, $\mathbf{a} = (a_1, \dots, a_n)$, $\mathbf{b} = (b_1, \dots, b_n) \in \mathbb{R}^n$, gilt. Notation: $\mathbf{X} \sim f_{\mathbf{X}}$.

Dichtefunktion

Eine nicht-negative Funktion $f(x_1, \dots, x_n)$ heißt **(multivariate) Dichtefunktion**, wenn

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \dots dx_n = 1.$$

Durch sie ist eine eindeutige Wahrscheinlichkeitsverteilung auf \mathbb{R}^n definiert.

Randdichte

Ist $(X_1, \dots, X_n) \sim f(x_1, \dots, x_n)$, dann ist die i -te **Randdichte** gegeben durch

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n.$$

Allgemein erhält man die Randdichte eines Teilvektors, indem man die gemeinsame Dichte über alle anderen Variablen integriert.

Produktdichte

Sind $f_1(x), \dots, f_n(x)$ Dichtefunktionen auf \mathbb{R} , dann ist die **Produktdichte** definiert durch

$$f(x_1, \dots, x_n) = f_1(x_1) \cdot \dots \cdot f_n(x_n).$$

Ist $(X_1, \dots, X_n) \sim f_1(x_1) \cdot \dots \cdot f_n(x_n)$, dann sind die Koordinaten unabhängig mit $X_i \sim f_i(x_i), i = 1, \dots, n$.

Bedingte Verteilung

Bedingte Zähldichte für diskrete Zufallsvektoren

Ist (X, Y) ein diskret verteilter Zufallsvektor mit Zähldichte $p(x, y)$, dann definiert die **bedingte Zähldichte**

$$p_{X|Y}(x|y) = P(X = x|Y = y) = \begin{cases} \frac{p(x, y)}{p_Y(y)}, & y \in \{y_1, y_2, \dots\} \\ p_X(x), & y \notin \{y_1, y_2, \dots\} \end{cases}$$

die bedingte Verteilung von X gegeben Y . Notation: $X|Y = y \sim p_{X|Y}(x|y)$.

Bedingte Dichtefunktion für stetige Zufallsvektoren

Sind X und Y stetige Zufallsvektoren mit gemeinsamer Dichtefunktion $f(x, y)$, dann wird

$$f_{X|Y}(x|y) = \begin{cases} \frac{f(x, y)}{f_Y(y)}, & f_Y(y) > 0 \\ f_X(x), & f_Y(y) = 0 \end{cases}$$

bedingte Dichtefunktion von X gegeben $Y = y$ genannt. Notation: $X|Y = y \sim f_{X|Y}(x|y)$.

Faktorisierung

Ist $X|Y = y \sim f(x|y)$, dann gilt für die gemeinsame Dichtefunktion

$$f(x, y) = f(x|y)f(y) = f(y|x)f(x).$$

Bedingte Erwartung

Ist (X, Y) diskret verteilt mit Zähldichte $p(x, y)$, dann ist der **bedingte Erwartungswert von X gegeben $Y = y$** definiert durch

$$E(X|Y = y) = \sum_{x \in \mathcal{X}} xp_{X|Y}(x|y).$$

Ist $(X, Y) \sim f_{(X,Y)}(x, y)$ stetig, dann gilt

$$E(X|Y = y) = \int xf_{X|Y}(x|y)dx.$$

Kriterien für stochastische Unabhängigkeit

Sind X und Y diskret verteilt mit gemeinsamer Zähldichte $p_{X,Y}(x, y)$, dann sind X und Y genau dann stochastisch unabhängig, wenn

$$p_{X|Y}(x|y) = p_X(x) \quad \text{bzw.} \quad p_{Y|X}(y|x) = p_Y(y) \quad \text{für alle } x \text{ und } y.$$

Sind X und Y stetig verteilt mit gemeinsamer Dichte $f(x, y)$, dann sind X und Y genau dann stochastisch unabhängig, wenn

$$f_{X|Y}(x) = f_X(x) \quad \text{bzw.} \quad f_{Y|X}(y) = f_Y(y) \quad \text{für alle } x \text{ und } y.$$

Produktkriterium

Ein Zufallsvektor (X, Y) ist genau dann stochastisch unabhängig, wenn

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y) \quad \text{für alle } x, y \in \mathbb{R}.$$

Erwartungswertvektor und Kovarianzmatrix

Erwartungswertvektor

Ist $\mathbf{X} = (X_1, \dots, X_n)'$ ein Zufallsvektor, und existieren die n Erwartungswerte $\mu_i = E(X_i), i = 1, \dots, n$, dann heißt $\mu = (E(X_1), \dots, E(X_n))'$ **Erwartungswertvektor** von \mathbf{X} .

Die Rechenregeln für den Erwartungswert übertragen sich auf Erwartungswertvektoren.

Kovarianzmatrix

Ist $\mathbf{X} = (X_1, \dots, X_n)$ ein Zufallsvektor, dann heißt die $(n \times n)$ -Matrix $\text{Var}(\mathbf{X}) = (\text{Cov}(X_i, X_j))_{i,j}$ **Kovarianzmatrix** von \mathbf{X} .

2.8. Grenzwertsätze

Das Gesetz der großen Zahlen

Tschebyschow-Ungleichung

Sind X_1, \dots, X_n unabhängig und identisch verteilte Zufallsvariablen mit Erwartungswert μ und Varianz $\sigma^2 \in (0, \infty)$, dann gilt für das arithmetische Mittel \bar{X}_n :

$$P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Schwaches Gesetz der großen Zahlen

Sind X_1, \dots, X_n unabhängig und identisch verteilte Zufallsvariablen mit Erwartungswert μ und Varianz $\sigma^2 \in (0, \infty)$, dann gilt für jedes $\varepsilon > 0$

$$P(|\bar{X}_n - \mu| > \varepsilon) \rightarrow 0, \quad n \rightarrow \infty.$$

Starkes Gesetz der großen Zahlen

Sind X_1, \dots, X_n unabhängig und identisch verteilte Zufallsvariablen mit $E|X_1| < \infty$ und Erwartungswert μ , dann gilt

$$P(\bar{X}_n \rightarrow \mu) = P(\{\omega | \bar{X}_n(\omega) \text{ konvergiert gegen } \mu\}) = 1.$$

Der Hauptsatz der Statistik

Sind $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} F(x)$, dann gilt

$$P\left(\lim_{n \rightarrow \infty} \max_{x \in \mathbb{R}} |F_n(x) - F(x)| = 0\right) = 1.$$

Der zentrale Grenzwertsatz (ZGWS)

Sind X_1, \dots, X_n unabhängig und identisch verteilte Zufallsvariablen mit Erwartungswert $\mu = E(X_1)$ und Varianz $\sigma^2 \in (0, \infty)$, dann gilt

$$\bar{X}_n \sim_{approx} N(\mu, \sigma^2/n)$$

in dem Sinne, dass die Verteilungsfunktion der standardisierten Version gegen die Verteilungsfunktion der Standardnormalverteilung konvergiert:

$$P\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \leq x\right) \rightarrow \Phi(x), \quad n \rightarrow \infty.$$

Ist s_n eine Zufallsvariable mit $\lim_{n \rightarrow \infty} P(|s_n/\sigma - 1| > \varepsilon) = 0$ für alle $\varepsilon > 0$, dann kann man beim ZGWS σ durch s_n ersetzen.

Konvergenzbegriffe

Stochastische Konvergenz

Ist X_1, X_2, \dots eine Folge von Zufallsvariablen und $a \in \mathbb{R}$ konstant, dann sagt man, $(X_n)_{n \in \mathbb{N}}$ **konvergiert stochastisch gegen** a , wenn für alle $\varepsilon > 0$ gilt:

$$\lim_{n \rightarrow \infty} P(|X_n - a| > \varepsilon) = 0.$$

Notation: $X_n \xrightarrow{P} a, n \rightarrow \infty$.

Die Konstante a kann auch durch eine Zufallsvariable X ersetzt werden.

Fast sichere Konvergenz

Ist X_1, X_2, \dots eine Folge von Zufallsvariablen und $a \in \mathbb{R}$ eine Konstante, dann sagt man, $(X_n)_{n \in \mathbb{N}}$ **konvergiert fast sicher gegen** a , wenn

$$P(X_n \rightarrow a) = P\left(\lim_{n \rightarrow \infty} X_n = a\right) = 1.$$

Notation: $X_n \xrightarrow{f.s.} a, n \rightarrow \infty$.

Die Konstante a kann auch durch eine Zufallsvariable X ersetzt werden.

Konvergenz in Verteilung

Ist X_1, X_2, \dots eine Folge von Zufallsvariablen mit $X_i \sim F_i(x)$, $i = 1, 2, \dots$, dann sagt man, X_n konvergiert in Verteilung gegen $X \sim F(x)$, wenn

$$F_n(X) \rightarrow F(x), \quad n \rightarrow \infty,$$

in allen Stetigkeitsstellen x von $F(x)$. Notation: $X_n \xrightarrow{d} X$, $X_n \xrightarrow{P} F$ oder $F_n \xrightarrow{d} F$.

Eigenschaften

1. Es gilt

$$X_n \xrightarrow{f.s.} X \quad \Rightarrow \quad X_n \xrightarrow{P} X \quad \Rightarrow \quad X_n \xrightarrow{d} X.$$

2. Aus $E(X_n - X)^2 \rightarrow 0$ für $n \rightarrow \infty$ folgt $X_n \xrightarrow{P} X$ für $n \rightarrow \infty$.

3. Die Umkehrungen gelten i. A. nicht.

2.9. Mehrdimensionale Verteilungen

Multinomialverteilung

Der Ausdruck

$$\binom{n}{x_1 \dots x_k} = \frac{n!}{x_1! \cdot x_2! \cdots x_k!}$$

wird **Multinomialkoeffizient** genannt und gibt die Anzahl der Möglichkeiten an, eine n -elementige Menge in k Teilmengen mit x_1, \dots, x_k Elementen zu zerlegen.

Auf der Menge

$$\mathcal{X} = \{0, \dots, n\} \times \cdots \times \{0, \dots, n\}$$

wird durch

$$\begin{aligned} p_H(x_1, \dots, x_k) &= P((H_1, \dots, H_k) = (x_1, \dots, x_k)) \\ &= \begin{cases} \binom{n}{x_1 \dots x_k} p_1^{x_1} \cdots p_k^{x_k}, & \text{falls } x_1 + \dots + x_k = n, x_1, \dots, x_k \geq 0 \\ 0, & \text{sonst} \end{cases} \end{aligned}$$

die **Multinomialverteilung** mit Parametern n und $\mathbf{p} = (p_1, \dots, p_k)$ definiert. Notation: $(H_1, \dots, H_k) \sim M(n; p_1, \dots, p_k)$.

Eigenschaften

- $E(H_j) = n \cdot p_j$,
- $\text{Var}(H_j) = n \cdot p_j(1 - p_j)$,
- $\text{Cov}(H_I, H_j) = -n \cdot H_i \cdot H_j$.

Multivariate Normalverteilung

Multivariate Standardnormalverteilung

Seien X_1, \dots, X_n unabhängig und identisch standardnormalverteilte Zufallsvariablen. Dann hat der Zufallsvektor $\mathbf{X} = (X_1, \dots, X_n)'$ die gemeinsame Dichtefunktion

$$\varphi(x_1, \dots, x_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2} \sum_{i=1}^n x_i^2\right), \quad x_1, \dots, x_n \in \mathbb{R}.$$

\mathbf{X} heißt **multivariat** oder **n -dimensional standardnormalverteilt**. Notation: $\mathbf{X} \sim N_n(\mathbf{0}, \mathbf{I})$.

Ist $\mathbf{X} \sim N_n(\mathbf{0}, \mathbf{I})$ und $\mu \in \mathbb{R}^n$ ein Vektor, dann gilt

$$\mathbf{Y} = \mathbf{X} + \mu \sim N_n(\mu, \mathbf{I}).$$

Notation: $\mathbf{Y} \sim N_n(\mu, \mathbf{I})$.

Ist $\mathbf{X} = (X_1, \dots, X_n)' \sim N_n(\mu, \mathbf{I})$ und $\mathbf{a} = (a_1, \dots, a_n)' \in \mathbb{R}^n$ ein Spaltenvektor, dann gilt

$$\mathbf{a}' \mathbf{X} \sim N_n(\mathbf{a}' \mu, \mathbf{a}' \mathbf{a}).$$

Zwei Zufallsvariablen $U = \mathbf{a}' \mathbf{X}$ und $V = \mathbf{b}' \mathbf{X}$ sind genau dann unkorreliert und somit unabhängig, wenn $\mathbf{a}' \mathbf{b} = 0$.

Multivariate Normalverteilung

Seien $\mathbf{X} = (X_1, \dots, X_n)'$ ein multivariat standardnormalverteilter Zufallsvektor, $\mathbf{a}_1, \dots, \mathbf{a}_m$ linear unabhängige Spaltenvektoren und \mathbf{A} die $m \times n$ -Matrix mit Zeilenvektoren $\mathbf{a}'_1, \dots, \mathbf{a}'_m$. Dann ist der Spaltenvektor

$$\mathbf{Y} = (Y_1, \dots, Y_m)' = (\mathbf{a}'_1 \mathbf{X}, \dots, \mathbf{a}'_m \mathbf{X})' = \mathbf{A} \mathbf{X}$$

multivariat normalverteilt mit Erwartungswertvektor $\mathbf{0} \in \mathbb{R}^m$ und $(m \times m)$ -Kovarianzmatrix

$$\Sigma = (\text{Cov}(Y_i, Y_j))_{i,j} = (\mathbf{a}'_i \mathbf{a}'_j)_{i,j} = \mathbf{A} \mathbf{A}'.$$

Notation: $\mathbf{Y} \sim N_m(\mathbf{0}, \Sigma)$.

Der Zufallsvektor $\mathbf{Y} = \mathbf{A} \mathbf{X} + \mathbf{b}$, $\mathbf{b} \in \mathbb{R}^n$, ist dann multivariat normalverteilt mit Erwartungswertvektor \mathbf{b} und Kovarianzmatrix $\Sigma = \mathbf{A} \mathbf{A}'$. Notation: $\mathbf{Y} \sim N_m(\mathbf{b}, \Sigma)$.

Die Matrix Σ hat maximalen Rang m .

Zweidimensionale Normalverteilung

Die Dichtefunktion eines zweidimensional normalverteilten Zufallsvektors (X, Y) ist gegeben durch

$$f(x, y) = \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{x-\mu_X}{\sigma_X}\frac{y-\mu_Y}{\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}},$$

für $(x, y) \in \mathbb{R}^2$. Man schreibt dann

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N(\mu, \Sigma) \quad \text{mit} \quad \mu = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_X^2 & \rho \\ \rho & \sigma_Y^2 \end{pmatrix}.$$

Bei der zweidimensionalen Normalverteilung ist Unabhängigkeit äquivalent zu Unkorreliertheit: Im Fall $\rho = 0$ gilt

$$f(x, y) = \frac{\exp\left(-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right)}{\sqrt{2\pi}} \cdot \frac{\exp\left(-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right)}{\sqrt{2\pi}}.$$

Jede Linearkombination $aX + bY$ mit Koeffizienten $a, b \in \mathbb{R}$ ist wieder normalverteilt.

Schätzer

Für die Schätzer

$$\begin{aligned} \hat{\mu}_X &= \frac{1}{n} \sum_{i=1}^n X_i, & \hat{\mu}_Y &= \frac{1}{n} \sum_{i=1}^n Y_i, \\ \hat{\sigma}_X^2 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, & \hat{\sigma}_Y^2 &= \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2, \\ \hat{\rho}_{XY}^2 &= \frac{\sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y}}{\sqrt{\hat{\sigma}_X^2 \hat{\sigma}_Y^2}} \end{aligned}$$

gilt:

- Diese Schätzer sind Maximum-Likelihood-Schätzer.
- $\hat{\mu}_X$ ist erwartungstreu und stark konsistent für μ_X .
- $\hat{\mu}_Y$ ist erwartungstreu und stark konsistent für μ_Y .
- $\hat{\sigma}_X^2$ ist asymptotisch erwartungstreu und stark konsistent für σ_X^2 .
- $\hat{\sigma}_Y^2$ ist asymptotisch erwartungstreu und stark konsistent für σ_Y^2 .
- $\hat{\rho}_{XY}^2$ ist asymptotisch erwartungstreu und stark konsistent für ρ_{XY}^2 .

Eigenschaften

Sei (X, Y) bivariat normalverteilt mit Parametern $\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho$. Dann gilt:

1. $X \sim N(\mu_X, \sigma_X^2)$.
2. $Y \sim N(\mu_Y, \sigma_Y^2)$.
3. X und Y sind genau dann unabhängig, wenn $\rho = 0$.
4. Die bedingte Verteilung von Y gegeben $X = x$ ist eine Normalverteilung mit bedingtem Erwartungswert

$$\mu_Y(x) = E(Y|X = x) = \mu_Y + \rho\sigma_Y \frac{x - \mu_X}{\sigma_X}$$

und bedingter Varianz

$$\sigma_Y^2(x) = \text{Var}(Y|X = x) = \sigma_Y^2(1 - \rho^2).$$

Notation: $Y|X = x \sim N(\mu_Y(x), \sigma_Y^2(x))$.

5. Die bedingte Verteilung von X gegeben $Y = y$ ist eine Normalverteilung mit bedingtem Erwartungswert

$$\mu_X(y) = E(X|Y = y) = \mu_X + \rho\sigma_X \frac{y - \mu_Y}{\sigma_Y}$$

und bedingter Varianz

$$\sigma_X^2(y) = \text{Var}(X|Y = y) = \sigma_X^2(1 - \rho^2).$$

Notation: $X|Y = y \sim N(\mu_X(y), \sigma_X^2(y))$.

2.10. (Erzeugende Funktionen)

Erzeugende Funktion

Sei X eine Zufallsvariable mit diskreter Wahrscheinlichkeitsfunktion $p(k) = P(X = k)$, $k \in \mathbb{N}_0$. Dann heißt

$$g_X(t) = Et^X = \sum_{k=0}^{\infty} p_X(k)t^k$$

erzeugende Funktion von X . Für $|t| \leq 1$ konvergiert $g_X(t)$ sicher.

Eigenschaften

- Die Verteilung einer Zufallsvariablen wird durch die erzeugende Funktion eindeutig charakterisiert.
- Es gilt $g_X(0) = P(X = 0)$ und $g_X(1) = 1$.
- Es gilt

$$g_X^{(k)}(0) = k!p_X(k) \quad \text{und somit} \quad p_X(k) = \frac{g_X^{(k)}}{k!}.$$

- Es gilt $g_X^{(k)}(1) = E(X(X - 1) \cdot \dots \cdot (X - k + 1))$.

Sind X_1, X_2, \dots unabhängig und identisch verteilte Zufallsvariablen mit erzeugender Funktion $g_X(t)$ und N eine von X_1, X_2, \dots unabhängige Zufallsvariable mit erzeugender Funktion $g_N(t)$, dann hat $S_N = X_1 + \dots + X_N$ die erzeugende Funktion $g_{S_N}(t) = g_N(g_X(t))$.

Faltungseigenschaft

Seien X und Y zwei unabhängige Zufallsvariablen mit erzeugenden Funktionen $g_X(t)$ und $g_Y(t)$, dann ist die erzeugende Funktion von $X + Y$ gegeben durch $g_{X+Y}(t) = g_X(t)g_Y(t)$.

Momenterzeugende Funktion

Sei X eine Zufallsvariable, dann heißt

$$m_X(t) = E(e^{tX})$$

für alle $t \geq 0$, für die der Ausdruck (in \mathbb{R}) existiert, **momenterzeugende Funktion**. Wenn X stetig verteilt mit Dichte $f(x)$ ist, dann heißt

$$L_f(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

Laplace-Transformierte.

L_f ist in dieser Form nicht nur für Dichtefunktionen definiert.

$m_X(t)$ ist stets für $t = 0$ definiert. Wenn $m_X(t)$ für ein $t > 0$ existiert, dann auch für alle Werte im Intervall $(-t, t)$.

2.11. (Markov-Ketten)

Eine endliche Folge von Zufallsvariablen X_0, \dots, X_T heißt **Markov-Kette** mit **Zustandsraum** S und **Übergangsmatrix** $\mathbf{P} = (p(x_i, x_j))_{i,j \in S}$, falls

$$P(X_n = x_n | X_0 = x_0, \dots, X_{n-1} = x_{n-1}) = P(X_n = x_n | X_{n-1} = x_{n-1}) = p(x_{n-1}, x_n)$$

für alle $x_0, \dots, x_n \in S$ und $n = 1, \dots, T$ mit

$$P(X_0 = x_0, \dots, X_{n-1} = x_{n-1}) > 0.$$

Der Zeilenvektor $\mathbf{p}_0 = (p_0, \dots, p_m)$, $p_i = P(X_0 = x_i)$, heißt **Startverteilung**.

Eine $m \times m$ -Matrix mit Einträgen zwischen 0 und 1, die sich zeilenweise zu 1 addieren, nennt man **stochastische Matrix**.

Für den Zeilenvektor $\mathbf{p}^{(n)} = (p_1^{(n)}, \dots, p_m^{(n)})$ der Wahrscheinlichkeiten $p_i^{(n)} = P(X_n = i)$, dass sich das System nach n Schritten in Zustand i befindet, gilt:

$$\mathbf{p}^{(n)} = \mathbf{p}_0 \mathbf{P}^n.$$

\mathbf{P}^n heißt n -Schritt-Übergangsmatrix.

$H_i = \min\{j | X_{i+j} \neq X_i\}$ heißt **Verweilzeit im i -ten Zustand**. Es gilt $H_i | H_0 \sim \text{Geo}(p_{ii})$.

Chapman-Kolmogorov-Gleichung

Für alle $x, y \in S$ und $n, m \in \mathbb{N}$ gilt:

$$p^{(m+n)}(x, y) = \sum_{z \in S} p^{(m)}(x, z) p^{(n)}(z, y).$$

Stationäre Verteilung

Eine Verteilung π auf S mit $\pi = \lim_{n \rightarrow \infty} \mathbf{p}^{(n)}$ und $\pi = \pi \mathbf{P}$ heißt **stationäre Verteilung**.

Ist π stationäre Verteilung, dann ist π' normierter Eigenvektor zum Eigenwert 1 von \mathbf{P}' .

Irreduzibilität

Eine stochastische Matrix \mathbf{P} heißt **irreduzibel**, wenn es für beliebige Zustände $x, y \in S$ ein $n \in \mathbb{N}_0$ gibt, sodass man ausgehend vom Zustand x den Zustand y nach n Schritten erreichen kann, d.h. $p^{(n)}(x, y) > 0$.

Periodizität

Eine stochastische Matrix \mathbf{P} heißt **periodisch**, wenn das System alle $k \geq 2$ Zustände wieder in einen Zustand x zurückkehren kann, d.h. wenn $p^{(n)}(x, x) > 0$ für $n = kr$ mit $r \in \mathbb{N}$ gilt. Dann ist der ggT der Menge

$$\mathcal{N}(x) = \{n \in \mathbb{N} : p^{(n)}(x, x) > 0\}$$

größer als 1.

\mathbf{P} heißt **aperiodisch**, wenn für jeden Zustand $x \in S$ der ggT der Menge $\mathcal{N}(x)$ genau 1 ist.

\mathbf{P} heißt **ergodisch**, wenn es ein $k \in \mathbb{N}$ gibt, sodass alle Einträge \mathbf{P}^k positiv sind.

Eine stochastische Matrix \mathbf{P} ist genau dann ergodisch, wenn sie irreduzibel und aperiodisch ist.

Ergodensatz

Eine ergodische stochastische Matrix \mathbf{P} besitzt genau eine stationäre Verteilung $\pi = (\pi_1, \dots, \pi_m)$. Alle π_j sind positiv und die n -Schritt-Übergangswahrscheinlichkeiten konvergieren unabhängig vom Startzustand gegen die stationäre Verteilung, d.h. für alle $j = 1, \dots, m$ gilt:

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j, \quad \text{für alle } i = 1, \dots, n.$$

3. Inferenzstatistik

3.1. Statistische Schätzer

Grundbegriffe der schließenden Statistik

Stichprobe

X_1, \dots, X_n wird **Stichprobe** vom **Stichprobenumfang** n genannt, wenn X_1, \dots, X_n reellwertige Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) sind. Der Zufallsvektor $\mathbf{X} = (X_1, \dots, X_n)$ nimmt dann Werte im Stichprobenraum $\mathcal{X} = \{\mathbf{X}(\omega) : \omega \in \Omega\} \subset \mathbb{R}^n$ an, dessen Elemente (x_1, \dots, x_n) **Realisierungen** genannt werden.

Verteilungsmodell

Ein **Verteilungsmodell** ist eine Menge \mathcal{P} von (möglichen) Verteilungen auf \mathbb{R}^n (für die Stichprobe (X_1, \dots, X_n)). Bei einem **parametrischen Verteilungsmodell** wird jede Verteilung $P \in \mathcal{P}$ durch Angabe eines Parametervektors ϑ aus einer Menge $\Theta \in \mathbb{R}^n$ möglicher Vektoren definiert, wobei Θ dann **Parameterraum** heißt. Kann die Menge \mathcal{P} nicht durch einen Parameter parametrisiert werden, spricht man von einem **nichtparametrisierten Verteilungsmodell**.

Schätzer

Ist X_1, \dots, X_n eine Stichprobe und $T : \mathbb{R}^n \rightarrow \mathbb{R}^d$, $d \in \mathbb{N}$ (oft $d = 1$), eine Abbildung, dann heißt $T(X_1, \dots, X_n)$ **Statistik**. Die Statistik heißt **Schätzfunktion** oder **Schätzer** für den Parameter ϑ , wenn sie in den Parameterraum abbildet, d.h. $T : \mathbb{R}^n \rightarrow \Theta$, und den Parameter ϑ schätzen soll.

Um Funktionen $g(\vartheta)$ eines Parameters ϑ zu schätzen, verwendet man Statistiken $T : \mathbb{R}^n \rightarrow \Gamma$ mit $\Gamma = g(\Theta) = \{g(\vartheta) | \vartheta \in \Theta\}$. $T(X_1, \dots, X_n)$ heißt dann Schätzer für $g(\vartheta)$.

Nichtparametrische Schätzung

Empirische Verteilungsfunktion

Ein nichtparametrischer Schätzer für die Verteilungsfunktion $F(x) = P(X_i \leq x)$, $x \in \mathbb{R}$, ist die empirische Verteilungsfunktion

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, x]}(X_i), \quad x \in \mathbb{R}.$$

Es gilt

$$E(\hat{F}_n(x)) = P(X_i \leq x) = F(x), \quad \text{Var}(\hat{F}_n(x)) = \frac{F(x)(1 - F(x))}{n}.$$

$\hat{F}_n(x)$ konvergiert mit Wahrscheinlichkeit 1 gegen $F(x)$.

Dichteschätzung

Durch den Histogramm-Schätzer wird eine Vergrößerung $g(x)$ der Dichtefunktion $f(x)$ geschätzt, für die gilt:

$$g(x) = \int_{g_j}^{g_{j+1}} f(x) dx = P(X_1 \in (g_j, g_{j+1}]),$$

wenn $x \in (g_j, g_{j+1}]$.

Für festes $x \in (g_j, g_{j+1}]$ ist $n\hat{f}(x)$ binomialverteilt mit Parametern n und $p = p(x) = P(X_1 \in (g_j, g_{j+1}])$.

Das Likelihood-Prinzip

Likelihood-Prinzip

Ein Verteilungsmodell ist bei gegebenen Daten plausibel, wenn es die Daten mit hoher Wahrscheinlichkeit erzeugt. Entscheide dich für das plausibelste Verteilungsmodell!

Likelihood-Funktion und Maximum-Likelihood-Schätzer

Ist $p_\vartheta(x)$ eine Zähldichte (in $x \in \mathcal{X}$) und $\vartheta \in \Theta \subset \mathbb{R}^k$, $k \in \mathbb{N}$, ein Parameter, dann heißt die Funktion

$$L(\vartheta|x) = p_\vartheta(x), \quad \vartheta \in \Theta,$$

für eine gegebene (feste) Beobachtung $x \in \mathcal{X}$ **Likelihood-Funktion**. $\hat{\vartheta} = \hat{\vartheta}(x) \in \Theta$ heißt **Maximum-Likelihood-Schätzer (ML-Schätzer)**, wenn für festes x gilt:

$$p_{\hat{\vartheta}}(x) \geq p_\vartheta(x) \quad \text{für alle } \vartheta \in \Theta.$$

Ist $f_\vartheta(x)$ eine (stetige) Dichtefunktion (in x) und $\vartheta \in \Theta \subset \mathbb{R}^k$, $k \in \mathbb{N}$, dann heißt die Funktion

$$L(\vartheta|x) = f_\vartheta(x), \quad \vartheta \in \Theta,$$

für festes x **Likelihood-Funktion**. $\hat{\vartheta} \in \Theta$ heißt **Maximum-Likelihood-Schätzer**, wenn für festes x gilt:

$$f_{\hat{\vartheta}}(x) \geq f_\vartheta(x) \quad \text{für alle } \vartheta \in \Theta.$$

Dadurch wird eine Funktion $\hat{\vartheta} : \mathcal{X} \rightarrow \Theta$ definiert.

Likelihood einer Stichprobe

Ist X_1, \dots, X_n eine Stichprobe von unabhängig und identisch verteilten Zufallsvariablen, und wurde $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ beobachtet, dann ist die Likelihood gegeben durch

$$L(\vartheta|x) = L(\vartheta|x_1) \cdot \dots \cdot L(\vartheta|x_n).$$

Log-Likelihood

Die **Log-Likelihood** ist gegeben durch

$$l(\vartheta|\mathbf{x}) = \ln L(\vartheta|x) = \sum_{i=1}^n l(\vartheta|x_i).$$

Hierbei ist $l(\vartheta|x_i) = \ln f_\vartheta(x_i)$ der **Likelihood-Beitrag der i -ten Beobachtung**.

3.2. Gütekriterien für statistische Schätzer

Erwartungstreue

Ein Schätzer $\hat{\theta}_n$ für einen Parameter θ heißt **erwartungstreu, unverfälscht** oder **unverzerrt**, wenn

$$E_\theta(\hat{\theta}_n), \quad \text{für alle } \theta \in \Theta.$$

Der Schätzer heißt **asymptotisch erwartungstreu**, wenn

$$E_\theta(\hat{\theta}_n) \rightarrow \theta, \quad \text{für alle } \theta \in \Theta.$$

Verzerrung

Die **Verzerrung** (engl.: Bias) wird durch

$$\text{Bias}(\hat{\theta}_n; \theta) = E_\theta(\hat{\theta}) - \theta, \quad \theta \in \Theta,$$

gemessen. Im Allgemeinen ist $\text{Bias}(\hat{\theta}_n; \theta)$ eine Funktion von θ .

Konsistenz

Ein Schätzer $\hat{\theta}_n = T(X_1, \dots, X_n)$ basierend auf einer Stichprobe vom Umfang n heißt **(schwach) konsistent für θ** , wenn er ein schwaches Gesetz großer Zahlen erfüllt, d.h.

$$\hat{\theta}_n \xrightarrow{P} \theta, \quad n \rightarrow \infty.$$

$\hat{\theta}_n$ heißt **stark konsistent für θ** , wenn fast sichere Konvergenz gilt.

Effizienz

Sind T_1 und T_2 zwei erwartungstreue Schätzer für θ mit $\text{Var}(T_1) < \text{Var}(T_2)$, dann heißt T_1 **effizienter** als T_2 . Ist T_1 effizienter als jeder andere erwartungstreue Schätzer, dann heißt T_1 **effizient**.

Mittlerer quadratischer Fehler

Der **mittlere quadratische Fehler** (engl.: **mean square error, MSE**) misst die erwartete quadratische Abweichung eines Schätzers $\hat{\theta}_n$ vom wahren Parameter θ :

$$\text{MSE}(\hat{\theta}_n; \theta) = E_\theta(\hat{\theta}_n - \theta)^2.$$

Ist $\hat{\theta}$ ein Schätzer mit $\text{Var}_\theta(\hat{\theta}) < \infty$, dann gilt

$$\text{MSE}(\hat{\theta}_n; \theta) = \text{Var}_\theta(\hat{\theta}) + [\text{Bias}(\hat{\theta}_n; \theta)]^2.$$

3.3. Testverteilungen

t-Verteilung

Sind X_1, \dots, X_n unabhängig und identisch $N(\mu, \sigma^2)$ -verteilt, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ das arithmetische Mittel und $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$, dann heißt die Verteilung von

$$T = \sqrt{n} \frac{\bar{X} - \mu}{S}$$

t-Verteilung mit $n - 1$ Freiheitsgraden und wird mit $t(n - 1)$ bezeichnet.

Das p -Quantil wird mit $t(n - 1)_p$ notiert. Für $T \sim t(k)$ gilt $E(T) = 0$ und für $k \geq 3$ $\text{Var}(T) = \frac{k}{k-2}$.

χ^2 -Verteilung

Sind U_1, \dots, U_k unabhängig und identisch $N(0, 1)$ -verteilt, dann heißt die Verteilung der Statistik

$$Q = \sum_{i=1}^n U_i^2$$

χ^2 -Verteilung mit k Freiheitsgraden. Ist T eine Zufallsvariable und $c \in \mathbb{R}$, so dass $T/c \sim \chi^2(k)$, dann heißt T **gestreckt χ^2 -verteilt mit k Freiheitsgraden**.

Es gilt $E(Q) = k$ und $\text{Var}(Q) = 2k$.

F-Verteilung

Sind $Q_1 \sim \chi^2(n_1)$ und $Q_2 \sim \chi^2(n_2)$ unabhängig χ^2 -verteilt, dann heißt die Verteilung der Statistik

$$F = \frac{Q_1/n_1}{Q_2/n_2}$$

F-Verteilung mit n_1 und n_2 Freiheitsgraden und wird mit $F(n_1, n_2)$ bezeichnet.

Das p -Quantil wird mit $F(n_1, n_2)_p$ bezeichnet.

Es gilt $E(F) = \frac{n_2}{n_2-2}$ und $\text{Var}(F) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-1)^2(n_2-4)}$.

3.4. Konfidenzintervalle

Ein Intervall $[L, U]$ mit datenabhängigen Intervallgrenzen $L = L(X_1, \dots, X_n)$ und $U = U(X_1, \dots, X_n)$ heißt **Konfidenzintervall** oder **Vertrauensbereich zum Konfidenzniveau $1 - \alpha$** , wenn

$$P(\vartheta \in [L, U]) \geq 1 - \alpha.$$

Konfidenzintervall für den Erwartungswert einer Normalverteilung

Sind $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$, dann ist

$$\left[\bar{X} - t(n-1)_{1-\alpha/2} \frac{S}{\sqrt{n}}, \bar{X} + t(n-1)_{1-\alpha/2} \frac{S}{\sqrt{n}} \right]$$

ein Konfidenzintervall zum Konfidenzniveau $1 - \alpha$ für den Erwartungswert μ , wobei $T = \sqrt{n}(\bar{X} - \mu)/S$ eine $t(n-1)$ -verteilte Statistik ist.

Wenn σ bekannt ist, muss S durch σ und das $t(n-1)_{1-\alpha/2}$ -Quantil durch das Normalverteilungsquantil $z_{1-\alpha/2}$ ersetzt werden.

Konfidenzintervall für die Varianz einer Normalverteilung

Sind $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$, dann ist

$$\left[\frac{n-1}{\chi^2(n-1)_{1-\alpha/2}} \hat{\sigma}^2, \frac{n-1}{\chi^2(n-1)_{\alpha/2}} \hat{\sigma}^2 \right]$$

ein Konfidenzintervall zum Konfidenzniveau $1 - \alpha$ für die Varianz σ^2 , wobei $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

Konfidenzintervall für die Erfolgswahrscheinlichkeit einer Binomialverteilung

Sei $Y \sim \text{Bin}(n, p)$, dann ist

$$\left[\hat{p} - z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

ein (approximativer) $(1 - \alpha)$ -Konfidenzintervall für p , wobei \hat{p} eine Schätzung für p ist.

Konfidenzintervall für den Parameter λ einer Poisson-Verteilung

Sind $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \text{Poi}(\lambda)$, dann ist

$$\left[\bar{X} - \sqrt{\frac{\bar{X}}{n}} z_{1-\alpha/2}, \bar{X} + \sqrt{\frac{\bar{X}}{n}} z_{1-\alpha/2} \right]$$

ein (approximativer) $(1 - \alpha)$ -Konfidenzintervall.

3.5. Einführung in die statistische Testtheorie

Testproblem

Seien f_0 und f_1 zwei mögliche Verteilungen für eine Zufallsvariable X , dann wird das **Testproblem**, zwischen $X \sim f_0$ und $X \sim f_1$ zu entscheiden, in der Form

$$H_0 : f = f_0 \quad \text{gegen} \quad H_1 : f = f_1$$

notiert, wobei f die wahre Verteilung von X bezeichnet. H_0 heißt **Nullhypothese** und H_1 **Alternative**.

Meist kann das Datenmaterial X_1, \dots, X_n durch eine Zahl $T = T(X_1, \dots, X_n)$, **Statistik** genannt, verdichtet werden.

Statistischer Test

Ein (**statistischer**) **Test** ist eine Entscheidungsregel, die basierend auf T entweder zugunsten von H_0 (Notation: " H_0 ") oder zugunsten von H_1 (" H_1 ") entscheidet.

Kritischer Wert

H_0 wird angenommen, wenn $T \in (-\infty, c_{\text{krit}}]$ und abgelehnt, wenn $T \in (c_{\text{krit}}, \infty)$. $A = (-\infty, c_{\text{krit}}]$ heißt **Annahmebereich**, $A^c = (c_{\text{krit}}, \infty)$ heißt **Ablehnbereich** oder **Verwerfungsbereich** und c_{krit} **kritischer Wert**.

Fehler 1. und 2. Art

Eine Entscheidung für H_1 , obwohl H_0 richtig ist, heißt **Fehler 1. Art**. Eine Entscheidung für H_0 , obwohl H_1 richtig ist, heißt **Fehler 2. Art**.

Die **Fehlerwahrscheinlichkeit 1. Art** ist die unter H_0 berechnete Wahrscheinlichkeit, fälschlicherweise H_0 abzulehnen und heißt **Signifikanzniveau**. Die **Fehlerwahrscheinlichkeit 2. Art** ist die unter H_1 berechnete Wahrscheinlichkeit, fälschlicherweise H_0 zu akzeptieren.

Test zum Niveau α

Ein **statistischer Test zum Signifikanzniveau α** ist gegeben, wenn

$$P_{H_0}(\text{" } H_1 \text{ "}) \leq \alpha.$$

Schärfe

Sei β die Wahrscheinlichkeit eines Fehlers 2. Art, dann heißt

$$1 - \beta = P_{H_1}(\text{" } H_1 \text{ "}) = E_{H_1}(1 - \phi)$$

Schärfe oder **Power** des Testverfahrens. Die Funktion

$$G(\vartheta) = P_\vartheta(\text{" } H_1 \text{ "}) = E_\vartheta(1 - \phi), \quad \vartheta \in \Theta,$$

heißt **Gütfunktion** des Tests.

3.6. 1-Stichproben-Tests

Seien X_1, \dots, X_n unabhängig und normalverteilt.

Einseitiger Gaußtest

Sei

$$T = \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma}$$

und $z_{1-\alpha}$ das $(1 - \alpha)$ -Quantil der Standardnormalverteilung. Der einseitige Gaußtest verwirft die Nullhypothese $H_0 : \mu \leq \mu_0$ auf dem Signifikanzniveau α zugunsten von $H_1 : \mu > \mu_0$, wenn $T > z_{1-\alpha}$.

Der einseitige Gaußtest verwirft $H_0 : \mu \geq \mu_0$ auf dem Signifikanzniveau α zugunsten von $H_1 : \mu < \mu_0$, wenn $T < z_\alpha$.

Gütfunktion

Die Gütfunktion des einseitigen Gaußtests ist gegeben durch

$$G(\mu) = \Phi \left(-z_{1-\alpha} + \frac{\mu - \mu_0}{\sigma / \sqrt{n}} \right).$$

Die Gütfunktion ist differenzierbar in μ , monoton wachsend im Stichprobenumfang n , monoton wachsend in $\mu - \mu_0$ sowie monoton fallend in σ^2 .

Fallzahlplanung

Sei d_0 die Mindestabweichung der Lageänderung $d = \mu - \mu_0$, die mit einer Wahrscheinlichkeit von mindestens $1 - \beta$ aufgedeckt werden soll. Dann muss gelten:

$$n \geq \frac{\sigma^2}{|d_0|^2} (z_{1-\alpha} + z_{1-\beta})^2.$$

Zweiseitiger Gaußtest

Der zweiseitige Gaußtest verwirft die Nullhypothese $H_0 : \mu = \mu_0$ zugunsten der Alternative $H_1 : \mu \neq \mu_0$ (Abweichung vom Sollwert μ_0), wenn $|T| > z_{1-\alpha/2}$.

Gütfunktion

Die Gütfunktion des zweiseitigen Gaußtests ist gegeben durch

$$G(\mu) = 2\Phi \left(-z_{1-\alpha/2} + \frac{\mu - \mu_0}{\sigma / \sqrt{n}} \right).$$

Die Gütfunktion ist differenzierbar in μ , monoton wachsend im Stichprobenumfang n , monoton wachsend in $|\mu - \mu_0|$ sowie monoton fallend in σ^2 .

Fallzahlplanung

Sei Δ die Mindestabweichung von $|\mu - \mu_0|$, die mit einer Wahrscheinlichkeit von mindestens $1 - \beta$ aufgedeckt werden soll. Dann muss gelten:

$$n \geq \frac{\sigma^2}{|\Delta|^2} (z_{1-\alpha/2} + z_{1-\beta})^2.$$

Der t -Test

Der zweiseitige t -Test verwirft $H_0 : \mu = \mu_0$ zugunsten von $H_1 : \mu \neq \mu_0$ auf dem Signifikanzniveau α , wenn $|T| > t(n-1)_{1-\alpha/2}$. Der einseitige t -Test für das Testproblem $H_0 : \mu \leq \mu_0$ gegen $H_1 : \mu > \mu_0$ verwirft H_0 , wenn $T > t(n-1)_{1-\alpha}$. Die Nullhypothese $H_0 : \mu \geq \mu_0$ wird zugunsten von $H_1 : \mu < \mu_0$ verworfen, wenn $T < -t(n-1)_{1-\alpha}$.

p -Wert

Sei $t_{\text{obs}} = T(x_1, \dots, x_n)$ der realisierte Wert der Teststatistik T und T^* die Teststatistik bei Wiederholung des Experiments. Dann ist der p -Wert für das Testproblem

$$H_0 : \mu \leq \mu_0 \quad \text{gegen} \quad H_1 : \mu > \mu_0$$

definiert durch

$$p = P_{H_0}(T^* > t_{\text{obs}})$$

Binomialtest

Sei $Y \sim \text{Bin}(n, p)$. Der exakte Binomialtest verwirft $H_0 : p \leq p_0$ zugunsten von $H_1 : p > p_0$, wenn $Y > c_{\text{krit}}$, wobei c_{krit} die kleinste ganze Zahl ist, sodass

$$\sum_{k=c_{\text{krit}}+1}^n \binom{n}{k} p_0^k (1-p_0)^{n-k} \leq \alpha.$$

Asymptotischer Binomialtest

Der asymptotische Binomialtest verwirft $H_0 : p \leq p_0$ auf dem Niveau α zugunsten von $H_1 : p > p_0$, wenn

$$T = \frac{Y - np_0}{\sqrt{np_0(1-p_0)}} > z_{1-\alpha}.$$

Äquivalent dazu ist $Y > np_0 + z_{1-\alpha} \sqrt{np_0(1-p_0)}$.

3.7. 2-Stichproben-Tests

Verbundene Stichproben

Gegeben sei eine Zufallsstichprobe

$$(X_1, Y_1), \dots, (X_n, Y_n)$$

von bivariat normalverteilten Zufallsvariablen. Es soll untersucht werden, ob sich die Erwartungswerte

$$\mu_X = E(X_i) \quad \text{und} \quad \mu_Y = E(Y_i)$$

unterscheiden.

Sei

$$D_i = Y_i - X_i, \quad i = 1, \dots, n, \quad \text{und} \quad \bar{D} = \frac{1}{n} \sum_{i=1}^n D_i,$$

sowie $\delta = E(D_i) = \mu_Y - \mu_X$ und $S_D^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2$. Definiere das Testproblem

$$H_0 : \delta = 0 \Leftrightarrow \mu_X = \mu_Y \quad \text{gegen} \quad H_1 : \delta \neq 0 \Leftrightarrow \mu_X \neq \mu_Y.$$

Dann wird H_0 auf dem Signifikanzniveau α verworfen, wenn für

$$T = \frac{\bar{D}}{S_D / \sqrt{n}}$$

gilt, dass $|T| > t(n-1)_{1-\alpha/2}$.

Beim einseitigen Testproblem $H_0 : \delta \leq 0$ gegen $H_1 : \delta > 0$ wird H_0 verworfen, wenn $T > t(n-1)_{1-\alpha}$. Analog wird $H_0 : \delta \geq 0$ zugunsten von $H_1 : \delta < 0$ verworfen, wenn $T < t(n-1)_\alpha$.

Unverbundene Stichproben

Gegeben seien zwei unabhängige Stichproben

$$\begin{aligned} X_{11}, \dots, X_{1n_1} &\stackrel{i.i.d.}{\sim} N(\mu_1, \sigma_1^2) \\ X_{21}, \dots, X_{2n_2} &\stackrel{i.i.d.}{\sim} N(\mu_2, \sigma_2^2). \end{aligned}$$

F-Test auf Varianzhomogenität

Gilt $\sigma_1 \neq \sigma_2$, spricht man von **Varianzhomogenität** oder **Heteroskedastizität** und **heteroskedastischen Daten**. Ansonsten liegt **Varianzhomogenität** vor.

Seien

$$S_1^2 = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (X_{1j} - \bar{X}_1)^2 \quad \text{und} \quad S_2^2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)^2,$$

dann ist die F -Teststatistik definiert durch

$$F = \frac{S_1^2}{S_2^2}.$$

Der F -Test auf Varianzgleichheit verwirft $H_0 : \sigma_1 = \sigma_2$, wenn $F < F(n_1 - 1, n_2 - 1)_{\alpha/2}$ oder $F > F(n_1 - 1, n_2 - 1)_{1-\alpha/2}$.

t-Test auf Lageunterschied

Der 2-Stichproben t -Test verwirft $H_0 : \mu_1 = \mu_2$ zugunsten von $H_1 : \mu_1 \neq \mu_2$, wenn $|T| > t(n-2)_{1-\alpha/2}$. Analog verwirft der einseitige Test $H_0 : \mu_1 \leq \mu_2$ zugunsten von $H_1 : \mu_1 > \mu_2$, wenn $T < t(n-2)_\alpha$, und $H_0 : \mu_1 \geq \mu_2$ zugunsten von $H_1 : \mu_1 < \mu_2$, falls $T > t(n-2)_{1-\alpha}$.

Welchs Test

Sei $\sigma_1 \neq \sigma_2$. Weiter sei

$$T = \frac{\bar{X}_2 - \bar{X}_1}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

und

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\left(\frac{S_1^2}{n_1}\right)^2 \frac{1}{n_1-1} + \left(\frac{S_2^2}{n_2}\right)^2 \frac{1}{n_2-1}}.$$

Der Welch-Test verwirft $H_0 : \mu_1 \leq \mu_2$ zugunsten von $H_1 : \mu_1 > \mu_2$, wenn $T > t(df)_\alpha$. $H_0 : \mu_1 \geq \mu_2$ wird zugunsten von $H_1 : \mu_1 < \mu_2$ verworfen, wenn $T > t(df)_{1-\alpha}$.

Fallzahlplanung

Sei $n_1 = n_2 = n$.

Einseitiger Test: Wähle

$$n \geq \frac{\sigma^2}{\Delta^2} (z_{1-\alpha} + z_{1-\beta})^2,$$

um eine Schärfe von $1 - \beta$ bei einer Abweichung von $\Delta = |\mu_A - \mu_B|$ näherungsweise zu erzielen.

Zweiseitiger Test: Wähle

$$n \geq \frac{\sigma^2}{\Delta^2} (z_{1-\alpha/2} + z_{1-\beta})^2,$$

um eine Schärfe von $1 - \beta$ bei einer Abweichung von $\Delta = |\mu_A - \mu_B|$ näherungsweise zu erzielen.

Wilcoxon-Test

Seien

$$X_{i1}, \dots, X_{in_i} \sim F_i(x), \quad i = 1, 2.$$

Weiter sei $R_{ij} = k$, wenn $X_{ij} = W_{(k)}$ der k -te Wert in der Ordnungstatistik der Gesamtstichprobe ist. Definiere

$$W = \sum_{j=1}^{n_2} R_{2j}$$

und

$$T = \frac{W - n_1 n_2 / 2}{\sqrt{n_1 n_2 (n+1)/12}} \sim_n N(0, 1).$$

Betrachte das Testproblem

$$H_0 : \Delta = 0 \Leftrightarrow F_1 = F_2 \quad \text{gegen} \quad H_1 : \Delta \neq 0 \Leftrightarrow F_1 \neq F_2.$$

Der Wilcoxon-Test verwirft H_0 auf dem Niveau α , wenn $|T| > z_{1-\alpha/2}$ bzw. wenn

$$W > \frac{n_1 n_2}{2} + z_{1-\alpha/2} \sqrt{n_1 n_2 (n+1)/12}$$

oder

$$W < \frac{n_1 n_2}{2} - z_{1-\alpha/2} \sqrt{n_1 n_2 (n+1)/12}.$$

2-Stichproben Binomialtest

Seien

$$Y_1 \sim \text{Bin}(n_1, p_1), \quad Y_2 \sim \text{Bin}(n_2, p_2)$$

unabhängig.

Betrachte das Testproblem

$$H_0 : p_1 = p_2 \quad \text{gegen} \quad H_1 : p_1 \neq p_2.$$

Seien

$$\hat{p}_1 = \frac{Y_1}{n_1} \quad \text{und} \quad \hat{p}_2 = \frac{Y_2}{n_2}$$

sowie

$$T = \frac{\hat{p}_2 - \hat{p}_1}{\sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n_2} + \frac{\hat{p}_1(1-\hat{p}_1)}{n_1}}}.$$

Der 2-Stichproben Binomialtest verwirft H_0 zugunsten von H_1 auf dem Niveau α , wenn $|T| > z_{1-\alpha/2}$. Analog wird $H_0 : p_1 \geq p_2$ zugunsten von $H_1 : p_1 < p_2$ verworfen, wenn $T > z_{1-\alpha}$, und $H_0 : p_1 \leq p_2$ zugunsten von $H_1 : p_1 > p_2$, wenn $T < z_\alpha$.

3.8. Korrelationstests

Test auf Korrelation

Sei $(X_1, Y_1), \dots, (X_n, Y_n)$ eine Stichprobe von unabhängig und identisch bivariat normalverteilten Zufallsvariablen X, Y mit Korrelationskoeffizient $\rho = \rho(X, Y) = \text{Cor}(X, Y)$.

Betrachte das Testproblem

$$H_0 : \rho = 0 \quad \text{gegen} \quad H_1 : \rho \neq 0.$$

Sei

$$T = \frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}}.$$

Der Korrelationstest für normalverteilte bivariate Stichproben verwirft H_0 auf dem Signifikanzniveau α , wenn $|T| > t(n-2)_{1-\alpha/2}$.

Rangkorrelationstest

Betrachte das Testproblem H_0 : "Es besteht kein monotoner Trend zwischen X und Y " gegen H_1 : "Es besteht ein monotoner Trend zwischen X und Y ". Definiere als Teststatistik

$$T = \frac{R_{Sp}\sqrt{n-2}}{\sqrt{1-R_{Sp}^2}}.$$

Dann wird H_0 auf dem Niveau α abgebildet, wenn $|T| > t(n-2)_{1-\alpha/2}$.

3.9. Analyse von Kontingenztafeln

Vergleich diskreter Verteilungen

Definitionen

- Eine Kontingenztafel habe r Zeilen und s Spalten mit insgesamt N Beobachtungen.
- Zeilenweise liegen diskrete Verteilungen einer Zielgröße mit s Ausprägungen vor, der Stichprobenumfang ist fest.
- Sei N_{ij} die Anzahl der Beobachtungen in Zeile i und Spalte j , dann ist (N_{i1}, \dots, N_{is}) die Häufigkeitsverteilung der i -ten Zeile vom Stichprobenumfang $N_{i*} = \sum_{j=1}^s N_{ij}$.

Testproblem

- H_0 : Alle Zeilenverteilungen gleich, d.h. nur eine Verteilung p_1, \dots, p_s .
- Die Daten können dann spaltenweise zusammengefasst werden zur Randverteilung (N_{*1}, \dots, N_{*s}) , wobei $N_{*j} = \sum_{i=1}^r N_{ij}$ die j -te Spaltensumme ist.
- Die p_j werden durch

$$\hat{p}_j = \frac{N_{*j}}{N}, \quad j = 1, \dots, s,$$

geschätzt.

- Unter H_0 ist der Erwartungswert von N_{ij}

$$E_{ij} = E_{H_0}(N_{ij}) = N_{i*}p_j,$$

da $N_{ij} \sim \text{Bin}(N_{i*}, p_j)$. E_{ij} wird geschätzt durch

$$\hat{E}_{ij} = N_{i*}\hat{p}_j = \frac{N_{i*}N_{*j}}{N}.$$

Chiquadrat-Test

Sei

$$Q = \sum_{i=1}^r \sum_{j=1}^s \frac{(N_{ij} - N_{i*}N_{*j}/N)^2}{N_{i*}N_{*j}/N}.$$

Der Chiquadrat-Test zum Vergleich diskreter Verteilungen verwirft H_0 , wenn $Q > \chi^2((r-1)(s-1))_{1-\alpha}$.

Bei einer 2×2 -Kontingenztafel mit Einträgen a, b, c, d ist

$$Q = \frac{n(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}.$$

Chiquadrat-Unabhängigkeitstest

Definitionen

- Die Kontingenztafel habe r Zeilen und s Spalten.
- Sie bestehe aus durch Kreuzklassifikation von N zufällig ausgewählten statistischen Einheiten nach zwei nominal skalierten Merkmalen X und Y .
- X habe r Ausprägungen a_1, \dots, a_r , Y habe s Ausprägungen b_1, \dots, b_s .
- N_{ij} ist dann die Häufigkeit der Kombination (a_i, b_j) .

Testproblem

- H_0 : X und Y sind stochastisch unabhängig.
- Sind (p_1, \dots, p_r) bzw. (q_1, \dots, q_s) die Verteilungen von X bzw. Y , dann ist $E_{ij} = E_{H_0}(N_{ij}) = Np_iq_j$, da $N_{ij} \sim \text{Bin}(N, p_iq_j)$ mit $p_iq_j = p_iq_j$, und wird durch

$$\hat{E}_{ij} = \frac{N_{i*}N_{*j}}{N}$$

geschätzt.

Chiquadrat-Test

Sei

$$Q = \sum_{i=1}^r \sum_{j=1}^s \frac{(N_{ij} - N_{i*}N_{*j}/N)^2}{N_{i*}N_{*j}/N}.$$

Der Chiquadrat-Test zum Test auf Unabhängigkeit verwirft H_0 , wenn $Q > \chi^2((r - 1)(s - 1))_{1-\alpha}$.

Unter H_0 ist Q in großen Stichproben $\chi^2((r - 1)(s - 1))$ -verteilt.

3.10. Lineares Regressionsmodell

Modell

Seien $(Y_1, x_1), \dots, (Y_n, x_n)$ unabhängige Beobachtungen. Sei

$$Y_i = a + bx_i + \varepsilon_i, \quad i = 1, \dots, n.$$

Annahmen

- $\varepsilon_1, \dots, \varepsilon_n$ sind unabhängig und identisch normalverteilte Zufallsvariablen mit $E(\varepsilon_i) = 0$, $\text{Var}(\varepsilon_i) = \sigma^2 > 0$, $i = 1, \dots, n$. σ^2 heißt auch **Modellfehler**.
- x_1, \dots, x_n sind vorgegeben.
- a und b sind unbekannte Parameter, genannt **Regressionskoeffizienten**.

Dann heißt

$$f(x) = a + bx$$

wahre Regressionsfunktion.

Schätzung

Eine (**geschätzte**) **Regressionsgerade (Ausgleichsgerade)** erhält man durch

$$\hat{f}(x) = \hat{a} + \hat{b}x, \quad x \in [x_{\min}, x_{\max}],$$

mit

$$\hat{b} = \frac{s_{xy}}{s_x^2}, \quad \hat{a} = \bar{Y} - \hat{b}\bar{x},$$

wobei

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n x_i Y_i - \bar{x} \bar{Y}, \quad s_x^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2.$$

Die (**geschätzten**) **Residuen** sind gegeben durch $\hat{\varepsilon}_i = Y_i - \hat{Y}_i, i = 1, \dots, n$. Eine erwartungstreue Schätzung des Modellfehlers σ^2 erhält man durch

$$\hat{\sigma}^2 = s_n^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{\varepsilon}_i^2.$$

Statistische Eigenschaften

- \hat{a} und \hat{b} sind erwartungstreu und konsistent für a und b .
- Ihre Varianzen können durch

$$\hat{\sigma}_b^2 = \frac{\hat{\sigma}^2}{ns_x^2}, \quad \hat{\sigma}_a^2 = \frac{\sum_{i=1}^n x_i^2}{n^2 s_x^2} \hat{\sigma}^2$$

geschätzt werden,

Konfidenzintervalle

- für b zum Niveau $1 - \alpha$:

$$\left[\hat{b} - t(n-2)_{1-\alpha/2} \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}, \hat{b} + t(n-2)_{1-\alpha/2} \frac{\hat{\sigma}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} \right]$$

- für σ^2 zum Niveau $1 - \alpha$:

$$\left[\frac{(n-2)\hat{\sigma}^2}{\chi^2(n-2)_{1-\alpha/2}}, \frac{(n-2)\hat{\sigma}^2}{\chi^2(n-2)_{\alpha/2}} \right]$$

- untere Schranke für die Regressionsfunktion zum Niveau $1 - \alpha$:

$$l(x) = \hat{a} + \hat{b}x - \hat{\sigma} \sqrt{2F(2, n-2)_{1-\alpha} \left(\frac{1}{n} + \frac{(\bar{x} - x)^2}{ns_{xx}} \right)}$$

- obere Schranke für die Regressionsfunktion zum Niveau $1 - \alpha$:

$$u(x) = \hat{a} + \hat{b}x + \hat{\sigma}\sqrt{2F(2, n-2)_{1-\alpha} \left(\frac{1}{n} + \frac{(\bar{x}-x)^2}{ns_{xx}} \right)}$$

Hypothesentests

Es gilt:

- $T_b = \frac{\hat{b}-b}{\hat{\sigma}_b} \sim t(n-2)$,
- $T_a = \frac{\hat{a}-a}{\hat{\sigma}_a} \sim t(n-2)$,
- $Q = \frac{(n-2)\hat{\sigma}^2}{\sigma_0^2} \sim \chi^2(n-2)$.

Test der Regressionskoeffizienten

1. $H_0 : b = b_0$ gegen $H_1 : b \neq b_0$. H_0 ablehnen, wenn $|T_b| > t(n-2)_{1-\alpha/2}$.
2. $H_0 : b \leq b_0$ gegen $H_1 : b > b_0$. H_0 ablehnen, wenn $T_b > t(n-2)_{1-\alpha}$.
3. $H_0 : b \geq b_0$ gegen $H_1 : b < b_0$. H_0 ablehnen, wenn $T_b < -t(n-2)_{1-\alpha} = t(n-2)_\alpha$.

Analog für a .

Test des Modellfehlers

1. $H_0 : \sigma^2 = \sigma_0^2$ gegen $H_1 : \sigma^2 \neq \sigma_0^2$. H_0 ablehnen, wenn $Q < \chi^2(n-2)_{\alpha/2}$.
2. $H_0 : \sigma^2 \leq \sigma_0^2$ gegen $H_1 : \sigma^2 > \sigma_0^2$. H_0 ablehnen, wenn $Q > \chi^2(n-2)_\alpha$.
3. $H_0 : \sigma^2 \geq \sigma_0^2$ gegen $H_1 : \sigma^2 < \sigma_0^2$. H_0 ablehnen, wenn $Q < \chi^2(n-2)_\alpha$.

Heteroskedastizität

Der Schätzer

$$\tilde{\sigma}_b^2 = \frac{1}{ns_x^2(n-2)} \sum_{i=1}^n (X_i - \bar{X})^2 \hat{\varepsilon}_i^2$$

ist konsistent, wenn die Varianzen der Fehlerterme $\varepsilon_1, \dots, \varepsilon_n$ nicht identisch sind.

3.11. (Multiple Lineare Regression)

Modell

Es wird angenommen, dass

$$Y = f(x_1, \dots, x_p) + \varepsilon$$

mit $E(\varepsilon) = 0$. f heißt dann (**wahre**) **Regressionfunktion**. Hat f die Form

$$F(x_1, \dots, x_p) = b_0 + b_1 x_1 + \dots + b_p x_p,$$

kann f in der Form $f(x_1, \dots, x_p) = \mathbf{b}' \mathbf{x}$ mit $\mathbf{b} = (b_0, \dots, b_p)' \in \mathbb{R}^{p+1}$ und $\mathbf{x} = (1, x_1, \dots, x_p) \in \mathbb{R}^{p+1}$ geschrieben werden, und heißt **linearer Prädiktor**.

Sei

$$Y_i = f(x_{i1}, \dots, x_{ip}) + \varepsilon_i, \quad i = 1, \dots, n,$$

wobei $\varepsilon_1, \dots, \varepsilon_n$ unabhängig und identisch verteilte Zufallsvariablen mit $E(\varepsilon_i) = 0$, $\text{Var}(\varepsilon_i) = \sigma^2 \in (0, \infty)$, $i = 1, \dots, n$, sind. Dann gilt mit

$$\mathbf{Y} = (Y_1, \dots, Y_n)' \in \mathbb{R}^n, \quad \varepsilon = (\varepsilon_1, \dots, \varepsilon_n)' \in \mathbb{R}^n, \quad \mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nk} \end{pmatrix},$$

dass

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \varepsilon.$$

Die $(n \times k)$ -Matrix \mathbf{X} heißt dann **Designmatrix**.

KQ-Schätzung

Zu minimieren ist die Zielfunktion

$$Q(\mathbf{b}) = \sum_{i=1}^n (Y_i - \mathbf{x}_i' \mathbf{b})^2, \quad \mathbf{b} \in \mathbb{R}^k.$$

Jedes Minimum $\hat{\mathbf{b}} = (\hat{b}_0, \hat{b}_1, \dots, \hat{b}_p)'$ von $Q(\mathbf{b})$ heißt KQ-Schätzer für \mathbf{b} .

Schätzung der Residuen

$\hat{\varepsilon} = \mathbf{Y} - \mathbf{X}\hat{\mathbf{b}}$ mit

$$\hat{\varepsilon}_i = Y_i - \mathbf{x}_i' \hat{\mathbf{b}}.$$

Schätzung des Modellfehlers

$$\hat{\sigma}^2 = \frac{1}{n-k} \sum_{i=1}^n \hat{\varepsilon}_i^2.$$

Normalgleichungen

Ist $\hat{\mathbf{b}}$ der KQ-Schätzer für \mathbf{b} , dann gelten die Normalgleichungen

$$\mathbf{X}' \mathbf{X} \hat{\mathbf{b}} = \mathbf{X}' \mathbf{Y}.$$

Hat \mathbf{X} (vollen) Rang k , dann ist

$$\hat{\mathbf{b}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}, \quad \hat{\varepsilon} = (\mathbf{I} - \mathbf{X}(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}') \mathbf{Y}.$$

Verteilungseigenschaften

Sind $\varepsilon_1, \dots, \varepsilon_n$ unabhängig und identisch $N(0, \sigma^2)$ -verteilt, dann gilt

$$\varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I}) \quad \text{und} \quad \mathbf{Y} \sim N(\mathbf{X}\mathbf{b}, \sigma^2 \mathbf{I}).$$

Hat \mathbf{X} vollen Spaltenrang, dann gilt:

1. $\hat{\mathbf{b}} \sim N(\mathbf{b}, \sigma^2 (\mathbf{X}' \mathbf{X})^{-1})$
2. $\hat{\varepsilon} \sim N(\mathbf{0}, (\mathbf{I} - \mathbf{X}(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'))$
3. $\sum_{i=1}^n \hat{\varepsilon}_i^2 \sim \chi^2(n-k)$
4. $\hat{\sigma}^2$ ist erwartungstreue für σ^2 .

5. \hat{b} und $\hat{\sigma}^2$ sind unabhängig.

Test der Regressionskoeffizienten

Die Statistik

$$T_j = \frac{\hat{b}_j - b_j}{\hat{\sigma} h_i}$$

ist $t(n - k)$ -verteilt, wobei h_i das i -te Diagonalelement der Matrix $(\mathbf{X}' \mathbf{X})^{-1}$ ist.

Hypothesentest:

$$H_0 : b_j = 0 \quad \text{gegen} \quad H_1 : b_j \neq 0.$$

H_0 ablehnen, wenn $|T_j| > t(n - k)_{1-\alpha/2}$.

3.12. (Elemente der Bayes-Statistik)

Grundbegriffe

X_1, \dots, X_n seien unabhängig und identisch verteilt mit

$$X_i \sim f_\vartheta(x),$$

wobei f_ϑ eine (Zähl-)Dichte aus einer parametrischen Verteilungsfamilie $\mathcal{F} = \{f_\vartheta : \vartheta \in \Theta\}$ ist, und $\Theta \subset \mathbb{R}^k$ den Parameterraum bezeichne.

Entscheidungsfunktion

Eine **Entscheidungsfunktion** δ ist eine Statistik $\delta : \mathbb{R}^n \rightarrow \mathcal{A}$, und \mathcal{A} bezeichnet dann den **Aktionsraum**. Beobachtet man $\mathbf{x} = (x_1, \dots, x_n)$, wird die Entscheidung $\delta(x_1, \dots, x_n)$ getroffen. \mathcal{D} bezeichne die Menge der möglichen Entscheidungsfunktionen.

Verlustfunktion

Eine nicht-negative Funktion $L : \Theta \times \mathcal{A}$ heißt **Verlust** oder **Verlustfunktion**.

Im Fall $\mathcal{A} = \Theta$ heißt $L(\vartheta, a) = (\vartheta - a)^2$ **quadratische Verlustfunktion**.

Risikofunktion

Der erwartete Verlust der Entscheidungsfunktion $\delta(X)$ im Punkt ϑ definiert die Risikofunktion $R : \Theta \times \mathcal{D} \rightarrow \mathbb{R}$,

$$R(\vartheta, \delta) = E_\vartheta L(\vartheta, \delta(X)).$$

Minimax-Prinzip

δ^* heißt Minimax-Regel, wenn

$$\max_{\vartheta \in \Theta} R(\vartheta, \delta^*) \leq \max_{\vartheta \in \Theta} R(\vartheta, \delta) \quad \text{für alle } \delta \in \mathcal{D}.$$

Bayes-Prinzip

Es wird angenommen, dass $\vartheta \sim \pi(\vartheta)$. $\pi(\vartheta)$ heißt **a-priori-Verteilung** oder **Prior**.

Statt $f_\vartheta(x)$ schreibt man nun $f(x|\vartheta)$. Für die gemeinsame Dichte von X und ϑ gilt

$$f(x, \vartheta) = f(x|\vartheta)\pi(\vartheta).$$

sowie

$$f(x) = \int f(x, \vartheta)d\vartheta \quad \text{bzw.} \quad f(x) = \sum_{\vartheta} f(x, \vartheta)$$

und

$$f(\vartheta|x) = \frac{f(x, \vartheta)}{f(x)}.$$

$f(\vartheta|x)$ heißt **a posteriori-Verteilung (Posterior-Verteilung)** von ϑ .

Die Risikofunktion wird geschrieben als $R(\vartheta, \delta) = E(L(\vartheta, \delta(x))|\vartheta)$. Ist X stetig, dann gilt

$$R(\vartheta, \delta) \int L(\vartheta, \delta(x))f(x|\vartheta)dx$$

und wenn X diskret ist

$$R(\vartheta, \delta) = \sum_x L(\vartheta, \delta(x))f(x|\vartheta).$$

Bayes-Risiko

Der Mittelwert des bedingten Risikos $R(\vartheta, \delta)$ über ϑ ,

$$R(\pi, \delta) = E_\pi R(\vartheta, \delta),$$

heißt **Bayes-Risiko** von δ .

Ist $\pi(\vartheta)$ eine Dichte, dann ist

$$R(\pi, \delta) = \int R(\vartheta, \delta)\pi(\vartheta)d\vartheta$$

und bei diskretem Prior ist

$$R(\pi, \delta) = \sum_{\vartheta} R(\vartheta, \delta) \pi(\vartheta).$$

Bayes-Regel

Eine Entscheidungsfunktion $\delta^* \in \mathcal{D}$ heißt **Bayes-Regel**, wenn

$$R(\pi, \delta^*) = \min_{\delta} R(\pi, \delta).$$

Ist X stetig, ist der Bayes-Schätzer der Erwartungswert der Posterior-Verteilung.

Konjugierte Prior-Familie

$\pi(\vartheta), \vartheta \in \Theta$, heißt **konjugierte Prior-Familie** (oder **konjugierter Prior**) zu einem bedingten Verteilungsmodell $f(x|\vartheta)$, wenn die a posteriori-Verteilung ein Element der Prior-Familie ist.

$f(x \vartheta)$	$\pi(\vartheta)$	$f(\vartheta x)$
$N(\vartheta, \sigma^2)$	$N(\mu, \tau^2)$	$N\left(\frac{\sigma^2\mu+x\tau^2}{\sigma^2+\tau^2}, \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}\right)$
$\Gamma(\nu, \beta)$	$\Gamma(\alpha, \beta)$	$\Gamma(\alpha + \nu, \beta + x)$
Bin(n,p)	Beta(α, β)	Beta($\alpha + x, \beta + n - x$)

4. Deskriptive Statistik

4.1. Grundbegriffe der deskriptive Statistik

- erhobene Daten: **statistische Einheiten, Untersuchungseinheiten** oder **Merkmalsträger, Versuchseinheiten** (durch Experimente gewonnen), **Beobachtungseinheiten** (Beobachtungsstudien)
- Menge der statistischen Einheiten: **Grundgesamtheit** oder **Population**
- Teilauswahl einer Grundgesamtheit: **Stichprobe**; Eine Stichprobe ist **repräsentativ**, wenn sie bezüglich wichtiger Charakteristika strukturgleich zur Grundgesamtheit ist
- vorgegebene Quoten, die Stichprobe einhalten muss: **quotierte Teilauswahl**
- (**einfache**) **Zufallsstichprobe**: jede Teilmenge der Grundgesamtheit hat gleiche Wahrscheinlichkeit, gezogen zu werden
- interessierende Größen der statistischen Einheiten: **Merkmale** oder **Variablen**
- Mögliche Werte eines Merkmals: **Merkmalsausprägungen** oder (**mögliche**) **Ausprägungen**
- nur endlich viele oder abzählbar unendlich viele Ausprägungen möglich: **diskretes Merkmal**, sonst **stetiges Merkmal**

Skalen

- **Nominalskala**: Ausprägungen stehen in keiner Beziehung zueinander; nur zwei Ausprägungen: **dichotome** oder **binäre Variable**
- **Ordinalskala**: Ausprägungen sind vergleichbar
- **Metrische Skala (Kardinalskala)**: Ausprägungen sind Vielfache einer Grundeinheit; **Intervallskala**: Nullpunkt willkürlich, sonst **Verhältnis-, Quotienten- oder Ratioskala**

Studiendesigns

- **Querschnittsstudie** (engl.: cross-sectional study): fester Zeitpunkt, interessierende Merkmale werden an statistischen Einheiten erhoben
- **Longitudinalstudie**: Merkmale werden zu verschiedenen Zeitpunkten an festem Kollektiv (Panel) erhoben

Aufbereitung von Daten

- **Rohdaten (Pimärdaten, Urliste)** mit p Merkmalen (**Dimension der Daten**) an n statistischen Einheiten (**Stichprobenumfang**) in $n \times p$ -Matrix (**Datenmatrix**) darstellbar
- $p=1$: **univariate Daten**, sonst **multivariate Daten**
- Der Vektor

$$\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$$

der n Ausprägungen eines Merkmals heißt **Datenvektor**.

4.2. Aufbereitung von univariaten Daten

Absolute und relative Häufigkeiten

Die **absoluten Häufigkeiten** (engl.: frequencies, counts) h_1, \dots, h_k sind durch

$$h_j = \text{"Anzahl der } x_i \text{ mit } x_i = a_j\text{"} = \sum_{i=1}^n 1(x_i = a_j),$$

$j = 1, \dots, k$, gegeben. Die (tabellarische) Zusammenstellung der absoluten Häufigkeiten h_1, \dots, h_k heißt **absolute Häufigkeitsverteilung**.

Die **relativen Häufigkeiten** f_1, \dots, f_k sind gegeben durch

$$f_j = \frac{h_j}{n}.$$

f_j ist der Anteil aller Beobachtungen mit Wert a_j . Die (tabellarische) Zusammenstellung der f_1, \dots, f_k heißt **relative Häufigkeitsverteilung**.

Eigenschaften

- $h_1 + \dots + h_k = n$.
- $f_1 + \dots + f_k = 1$.

Ordnungsstatistiken

Sortiere die n Beobachtungen aufsteigend der Größe nach, sodass

$$x_{\min} = x_{(1)} \leq \dots \leq x_{(n)} = x_{\max}.$$

Dann heißt $x_{(i)}$ **i-te Ordnungsstatistik** und $(x_{(1)}, \dots, x_{(n)})$ **Ordnungsstatistik** der Stichprobe x_1, \dots, x_n .

Kumulierte Häufigkeitsverteilung

Für Rohdaten x_1, \dots, x_n ist die **kumulierte Häufigkeitsfunktion** definiert durch

$$H(x) = \sum_{i=1}^n 1(x_i \leq x) = \sum_{j: a_j \leq x} h_j.$$

Die **empirische Verteilungsfunktion (relative kumulierte Häufigkeitsverteilung)** ist für $x \in \mathbb{R}$ definiert durch

$$\hat{F}(x) = \frac{H(x)}{n} = \text{Anteil der } x_i \text{ mit } x_i \leq x = \sum_{j: a_j \leq x} f_j.$$

Eigenschaften

- $H(x)$ und $\hat{F}(x)$ sind monoton wachsende Treppenfunktionen, die an den Ordnungsstatistiken $x_{(i)}$ Sprungstellen besitzen.
- Bei $H(x)$ ist die Sprunghöhe genau die Anzahl der Beobachtungen, die gleich $x_{(i)}$ sind.
- Bei $\hat{F}(x)$ ist die Sprunghöhe genau der Anteil der Beobachtungen, die gleich $x_{(i)}$ sind.

Gruppierung

Das Intervall $[x_{\min}, x_{\max}]$ heißt **Messbereich**.

Seien

$$I_1 = [g_1, g_2], I_2 = (g_2, g_3], \dots, I_k = (g_k, g_{k+1}]$$

k Intervalle, die den gesamten Messbereich überdecken. Dann heißt I_j **j-te Gruppe** oder **Klasse** und ist für $j = 2, \dots, k$ gegeben durch $I_j = (g_j, g_{j+1}]$. Die Zahlen g_1, \dots, g_{k+1} sind die **Gruppengrenzen**,

$$b_j = g_{j+1} - g_j, \quad j = 1, \dots, k,$$

die **Gruppenbreiten** und

$$m_j = \frac{g_{j+1} - g_j}{2}, \quad j = 1, \dots, k,$$

die **Gruppenmitten**.

Stamm-Blatt-Diagramm

Angenommen, die Zahlen des Datensatzes bestehen aus d Ziffern. Wähle die Zahl mit den ersten $d - 1$ Ziffern von x_{\min} und zähle gleichmäßig bis zur Zahl mit den ersten $d - 1$ Ziffern von x_{\max} hoch, sodass sich die letzte Ziffer immer um 1 erhöht. Diese Zahlen definieren die

Gruppengrenzen und bilden den **Stamm** des Diagramms, die untereinander aufgeschrieben werden. Schreibe nun für jede Zahl des Datensatzes die d -te Ziffer rechts neben den jeweiligen Stamm.

Histogramm

Gruppiere den Datensatz in k Klassen mit relativen Häufigkeiten f_j und Gruppenbreiten $b_j, j = 1, \dots, k$. Zeichne nun über der j -ten Klasse ein Rechteck der Höhe

$$l_j = \frac{f_j}{b_j}, \quad j = 1, \dots, k.$$

Dann erhält man das **Histogramm**, bei dem die Rechtecke die relativen Häufigkeiten repräsentieren.

Der obere Rand des Histogramms definiert eine Treppenfunktion

$$\hat{f}(x) = \begin{cases} 0, & x < g_1 \\ l_1, & x \in [g_1, g_2] \\ l_j, & x \in (g_j, g_{j+1}], j = 2, \dots, k, \\ 0, & x > g_{k+1}. \end{cases}$$

$\hat{f}(x)$ heißt **Häufigkeitsdichte** oder auch **Dichteschätzer**.

Eigenschaften

- Es gilt:

$$f_j = \int_{g_j}^{g_{j+1}} \hat{f}(x) dx$$

und

$$\int_{-\infty}^{\infty} \hat{f}(x) dx = 1.$$

- Die Höhe repräsentiert die Dichte der Daten.

Gleitendes Histogramm

Sei $x \in \mathbb{R}$ und $\tilde{f}(x)$ der Anteil der Beobachtungen x_i mit $|x - x_i| \leq h$, dividiert durch $2h$. Dann heißt $\tilde{f}(x)$ gleitendes Histogramm und h Bandbreite, und es gilt

$$\tilde{f}(x) = \frac{1}{2nh} \sum_{i=1}^n \mathbf{1}(|x - x_i| \leq h).$$

Kerndichteschätzer

Gegeben seien Daten x_1, \dots, x_n . Die Funktion

$$\tilde{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right), \quad x \in \mathbb{R},$$

heißt Kerndichteschätzer (nach Parzen-Rosenblatt) zur Bandbreite h , wenn $K(z)$ eine stetige Funktion mit

$$K(z) \geq 0, \quad \int_{-\infty}^{\infty} K(z) dz = 1$$

ist, die symmetrisch um 0 ist. $K(z)$ heißt dann Kernfunktion.

Gängige Kernfunktionen sind:

der **Gauß-Kern**

$$K(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad z \in \mathbb{R},$$

der **Epanechnikov-Kern**

$$K(z) = \begin{cases} \frac{3}{4}(1 - z^2), & |z| < 1, \\ 0, & \text{sonst,} \end{cases}$$

der **Gleichverteilungskern**

$$K(z) = \frac{1}{2} \mathbf{1}(|z| \leq 1) = \begin{cases} \frac{1}{2}, & |z| \leq 1, \\ 0, & \text{sonst.} \end{cases}$$

4.3. Lagemaße

Median

Der **Median** wird in der Regel definiert durch

$$x_{\text{med}} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & n \text{ ungerade}, \\ \frac{1}{2}(x_{n/2} + x_{n/2+1}), & n \text{ gerade}. \end{cases}$$

Eigenschaften

- Der Median vollzieht affin-lineare Transformationen nach:

$$y_i = a + b \cdot x_i \quad \Rightarrow \quad y_{\text{med}} = a + b \cdot x_{\text{med}}.$$

- Der Median minimiert den Ausdruck

$$Q(m) = \sum_{i=1}^n |x_i - m|.$$

- Der Median ist sehr robust gegenüber Ausreißern.

Arithmetisches Mittel

Das **arithmetische Mittel** ist definiert durch

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Eigenschaften

- Das arithmetische Mittel vollzieht affin-lineare Transformationen nach:

$$y_i = a + b \cdot x_i \quad \Rightarrow \quad \bar{y} = a + b \cdot \bar{x}.$$

- Das arithmetische Mittel minimiert den Ausdruck

$$Q(m) = \sum_{i=1}^n (x_i - m)^2.$$

- Das arithmetische Mittel ist sensitiv bzgl. Ausreißern.

Geometrisches Mittel

Das **geometrische Mittel** von n nichtnegativen Zahlen x_1, \dots, x_n ist definiert durch

$$\bar{x}_{\text{geo}} = (x_1 \cdot \dots \cdot x_n)^{(1/n)}.$$

Es gilt: $\bar{x}_{\text{geo}} \leq \bar{x}$.

Wachstumsfaktor und Wachstumsrate

Seien x_1, \dots, x_n Bestandsgrößen. Dann nennt man

$$w_1 = 1, \quad w_i = x_i / x_{i-1}, \quad i = 2, \dots, n,$$

i -ten **Wachstumsfaktor** und

$$r_i = w_i - 1 \quad \Leftrightarrow \quad x_i = (1 + r_i)x_{i-1}$$

i -te **Wachstumsrate (Zinssatz)**.

Der **mittlere Wachstumsfaktor** wird als derjenige Wachstumsfaktor w definiert, der bei Anwendung in allen n Perioden zum Wert x_n führt. Die

mittlere Wachstumsrate (effektiver Zinssatz) ist durch $r = w - 1$ gegeben.

Weitere Mittelwerte

Harmonisches Mittel

Seien x_1, \dots, x_n Zahlen ungleich Null mit $\sum_{i=1}^n \frac{1}{x_i} \neq 0$. Dann ist das **harmonische Mittel** definiert durch

$$\bar{x}_{\text{harm}} = \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right)^{-1}.$$

Getrimmtes Mittel

Sei $k = \lfloor na \rfloor$, wobei vermutet werde, dass es im Datensatz nicht mehr als $2a \cdot 100\%$ Ausreißer gibt. Dann ist das **getrimmte Mittel** definiert durch

$$\bar{x}_a = \frac{x_{(\lfloor k+1 \rfloor)} + \dots + x_{(\lfloor n-k \rfloor)}}{n - 2k}.$$

Winorisiertes Mittel

k sei wie beim getrimmten Mittel definiert. Ersetze die k größten und die k kleinsten Beobachtungen durch den jeweils nächstgelegenen der zentralen $n - 2k$ Beobachtungen. Der Mittelwert dieses modifizierten Datensatzes heißt **winorisiertes Mittel**.

4.4. Streuung

Entropie

Die Maßzahl

$$H = - \sum_{j=1}^n f_j \cdot \log_2(f_j)$$

wird **Shannon-Wiener-Index** oder **(Shannon-) Entropie** genannt.

Eigenschaften

- Je ähnlicher die Häufigkeitsverteilung der diskreten Gleichverteilung ist, desto größer ist der Wert von H .
- Nachteil: der Wert hängt vom verwendeten Logarithmus ab.

Normierte Entropie

Die **relative** oder **normierte Entropie** ist gegeben durch

$$J = \frac{H}{\log(k)}.$$

Sie hängt nicht von der Wahl des Logarithmus ab.

Stichprobenvarianz und Standardabweichung

Die **Stichprobenvarianz** oder **empirische Varianz** der Beobachtungen x_1, \dots, x_n wird definiert durch

$$\text{var}(\mathbf{x}) = s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Der Wert

$$s = \sqrt{s^2} = \sqrt{\text{var}(\mathbf{x})}$$

heißt **Standardabweichung**.

Ist die Häufigkeitsverteilung f_1, \dots, f_k mit Gruppenmitteln m_1, \dots, m_k gegeben, verwendet man

$$s_g^2 = \sum_{j=1}^n f_j (m_j - \bar{x}_g)^2.$$

Analog definiert man für Ausprägungen a_1, \dots, a_k eines metrisch skalierten Merkmals

$$s_a^2 = \sum_{j=1}^n f_j (a_j - \bar{x})^2.$$

Eigenschaften

Für alle $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ und $a, b \in \mathbb{R}$ gilt:

- Invarianz unter Lageänderungen:

$$\text{var}(a + \mathbf{x}) = \text{var}(\mathbf{x}).$$

- Quadratische Reaktion auf Maßstabsänderungen:

$$\text{var}(b\mathbf{x}) = b^2 \text{var}(\mathbf{x}).$$

- Verschiebungssatz:

$$s^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2.$$

- Für gruppierte Daten:

$$s_g^2 = \sum_{i=1}^n f_j m_j^2 - \bar{x}_g^2.$$

Mittlere absolute Abweichung (MAD)

Die **mittlere absolute Abweichung** (Mean Absolute Deviation, MAD) ist definiert durch

$$\text{MAD} = \frac{1}{n} \sum_{i=1}^n |x_i - \tilde{x}_{med}|.$$

4.5. Schiefe und Symmetrie

Eine Funktion $f(x)$ heißt **symmetrisch mit Symmetriezentrum m** , wenn für alle $x \in \mathbb{R}$ gilt:

$$f(m+x) = f(m-x).$$

Eine empirische Verteilung ist **linksschief**, wenn für alle $a > 0$ der Anteil der Beobachtungen mit $x_i > m + a$ größer ist als der Anteil der Beobachtungen mit $x_i < m - a$. Im umgekehrten Fall heißt die empirische Verteilung **rechtsschief**, bei Gleichheit ist die Verteilung **symmetrisch**, dann ist m der Median.

Bestimmung der Schiefe

Das **dritte standardisierte Moment** ist definiert durch

$$m_3^* = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3.$$

- $m_3^* > 0$: Rechtsschiefe.
- $m_3^* < 0$: Linksschiefe.
- $m_3^* = 0$: Symmetrie.

4.6. Quantile

Ein (**empirisches**) p -**Quantil**, $p \in (0, 1)$, eines Datensatzes x_1, \dots, x_n ist ein Wert $\tilde{x}_p \in \{x_1, \dots, x_n\}$, sodass

- mindestens $p \cdot 100\%$ der Beobachtungen kleiner oder gleich \tilde{x}_p sind, und zugleich
- mindestens $100 \cdot (1 - p)\%$ der Beobachtungen größer oder gleich \tilde{x}_p sind.

Fälle

- $np \in \mathbb{N}$: $x_{(np)}$ und $x_{(np+1)}$ sind p -Quantile. Metrische Skalierung: jede Zahl dazwischen ist ebenfalls p -Quantil.
- $np \notin \mathbb{N}$: $\tilde{x}_p = x_{(\lfloor np \rfloor + 1)}$ ist das eindeutige p -Quantil.

Quartile

Das 0.25-Quantil heißt **erstes** oder **unteres Quartil** Q_1 , das **dritte Quartil** Q_3 ist das 0.75-Quantil. Der Median ist das zweite Quartil Q_2 .

Zusammen mit dem Minimum und dem Maximum der Beobachtungen wird der Datensatz durch diese drei Quantile in vier Bereiche mit gleichen Anteilen aufgeteilt.

Quartilsabstand

Der **Quartilsabstand** (engl.: interquartile range) ist gegeben durch

$$\text{IQR} = Q_3 - Q_1.$$

Fünf-Punkte-Zusammenfassung

Die **Fünf-Punkte-Zusammenfassung** eines Datensatzes besteht aus dem Minimum x_{\min} , dem ersten Quartil $Q_1 = \tilde{x}_{0.25}$, dem Median x_{med} , dem dritten Quartil $Q_3 = \tilde{x}_{0.75}$ und dem Maximum x_{\max} .

Boxplot

Ein **Boxplot** ist eine grafische Darstellung der Fünf-Punkte-Zusammenfassung. Hierbei wird eine Box von Q_1 bis Q_3 gezeichnet, die beim Median einen vertikalen Strich enthält. An die Box werden Striche (**Whiskers**) angesetzt, die bis zum Minimum x_{\min} bzw. Maximum x_{\max} reichen.

Beobachtungen, die unterhalb der Grenze

$$Q_1 - 1.5 \cdot (Q_3 - Q_1)$$

bzw. oberhalb der Grenze

$$Q_3 + 1.5 \cdot (Q_3 - Q_1)$$

liegen, heißen **äußere Beobachtungen**. Die Grenzen nennt man **innere Zäune**, wählt man statt dem Faktor 1.5 den Faktor 3, erhält man die **äußersten Zäune**.

QQ-Plot

Gegeben seien zwei Datensätze

$$x_1, \dots, x_n \quad \text{und} \quad y_1, \dots, y_m.$$

Die Datensätze sollen verglichen werden, indem verschiedene p -Quantile gegeneinander aufgetragen werden. Für $n = m$ werden die p_i -Quantile mit

$$p_i = i/n, \quad i = 1, \dots, n,$$

benutzt, welche gerade die Ordnungsstatistiken $x_{(i)}$ und $y_{(i)}$ sind. Ist $n \neq m$, werden die p_i -Werte des kleineren Datensatzes verwendet. Der QQ-Plot kann wie folgt interpretiert werden:

- In Bereichen, in denen die Punkte unterhalb der Winkelhalbierenden liegen, sind die x -Quantile größer als die y -Quantile, d.h., dass die y -Verteilung mehr Masse bei den kleinen Werten hat als die x -Verteilung.
- Wenn die Punkte (nahezu) auf einer Geraden liegen, dann können die Datensätze durch eine lineare Transformation $y_i = a + b \cdot x_i$ ineinander überführt werden (Lage- und Skalenänderung).

4.7. Konzentrationsmessung

Ausgangspunkt ist die Modellierung eines Marktes durch n Merkmalsträger $1, \dots, n$, für die n kardinalisierte Merkmalsausprägungen $x_1, \dots, x_n \geq 0$ gegeben sind. Wir gehen davon aus, dass die Merkmalsausprägungen sortiert sind:

$$x_1 \leq x_2 \leq \dots \leq x_n$$

Die j kleinsten Marktteilnehmer vereinen die Merkmalssumme $x_1 + \dots + x_j$ auf sich. In Anteilen ausgedrückt bedeutet dies: Die $j/n \cdot 100\%$ kleinsten Marktteilnehmer vereinen den (Markt-)Anteil

$$a_j = \frac{x_1 + \dots + x_j}{x_1 + \dots + x_n}$$

auf sich.

Lorenzkurve

Die **Lorenzkurve** $L(t)$, $t \in [0, 1]$, ist die grafische Darstellung der $n + 1$ Punktpaare

$$(0, 0), \left(\frac{1}{n}, a_1 \right), \dots, (1, a_n)$$

durch einen Streckenzug. Man verbindet also diese Punktpaare durch Linien.

Gini-Koeffizient

Der **Gini-Koeffizient** G ist gegeben durch

$$G = 2 \cdot \text{Fläche zwischen Lorenzkurve und Diagonale.}$$

Es gilt:

$$G = \frac{n+1 - 2 \sum_{j=1}^n a_j}{n}.$$

Normierter Gini-Koeffizient

Der **normierte Gini-Koeffizient** G^* berechnet sich zu

$$G^* = \frac{n}{n-1} G.$$

Herfindahl-Index

Der **Herfindahl-Index** ist gegeben durch

$$H = \sum_{i=1}^n p_i^2,$$

wobei

$$p_i = \frac{x_i}{x_1 + \dots + x_n}$$

den Marktanteil des i -ten Merkmalsträgers notiert.

4.8. Kontingenztafeln, Randverteilungen und Kontingenzkoeffizienten

Gegeben seien n Punktpaare $(x_1, y_1), \dots, (x_n, y_n)$, generiert durch simultane Erhebung der Merkmale X und Y an n statistischen Einheiten. Wir sprechen auch von einer **zweidimensionalen** oder **bivariaten Stichprobe**.

Nomiale Merkmale

Für nominal skalierte Merkmale X und Y , die simultan an statistischen Einheiten beobachtet wurden, geht man wie folgt vor:

Die Merkmalsausprägungen von X seien a_1, \dots, a_r , diejenigen von Y notieren wir mit b_1, \dots, b_s . Liegt nun eine bivariate Stichprobe $(x_1, y_1), \dots, (x_n, y_n)$ vom Umfang n vor, so stimmt jedes Beobachtungspaar mit einer Ausprägung (a_j, b_j) überein.

Absolute Häufigkeit

Zählt man aus, wie oft die Kombination (a_i, b_j) in der Stichprobe vorkommt, so erhält man die zugehörige **absolute Häufigkeit** h_{ij} .

Kontingenztafel

Die $r \cdot s$ absoluten Häufigkeiten werden übersichtliche in einer Tabelle mit $r \cdot s$ Zellen zusammengestellt. Diese Tabelle heißt Kontingenztafel.

Randverteilungen

Die Zeilensummen der Kontingenztafel bilden die absolute Häufigkeitsverteilung des Merkmals X , die Spaltensummen der Kontingenztafel bilden die absolute Häufigkeitsverteilung des Merkmals Y . Diese Verteilungen heißen **Randverteilungen**.

Notation:

$$h_{i\bullet} = h_{i1} + \dots + h_{is} = \sum_{j=1}^s h_{ij}$$

$$h_{\bullet j} = h_{1j} + \dots + h_{rj} = \sum_{i=1}^r h_{ij}$$

Division durch n gibt die relativen Häufigkeitsverteilungen der Merkmale.

Bedingte Häufigkeitsverteilung

Die **bedingte Häufigkeitsverteilung von Y unter der Bedingung $X = a_i$** ist gegeben durch

$$f_Y(b_j|a_i) = \frac{h_{ij}}{h_{i\bullet}} = \frac{f_{ij}}{f_{i\bullet}}, \quad j = 1, \dots, s,$$

sofern $h_{i\bullet} > 0$. Entsprechend heißt

$$f_X(a_i|b_j) = \frac{h_{ij}}{h_{\bullet j}} = \frac{f_{ij}}{f_{\bullet j}}, \quad i = 1, \dots, r,$$

bedingte Häufigkeitsverteilung von X unter der Bedingung $Y = b_j$.

Empirische Unabhängigkeit

Die Merkmale einer Kontingenztafel heißen **empirisch unabhängig**, falls

$$h_{ij} = \frac{h_{i\bullet} \cdot h_{\bullet j}}{n} \Leftrightarrow f_{ij} = f_{i\bullet} \cdot f_{\bullet j}$$

für alle $i = 1, \dots, r$ und $j = 1, \dots, s$ gilt.

Chiquadrat-Statistik, χ^2 -Koeffizient

Die Maßzahl

$$Q = \sum_{i=1}^r \sum_{j=1}^s \frac{(h_{ij} - e_{ij}^2)}{e_{ij}}, \quad e_{ij} = \frac{h_{i\bullet} \cdot h_{\bullet j}}{n},$$

heißt **Chiquadrat-Statistik (χ^2 -Koeffizient)** und wird auch mit dem Symbol χ^2 bezeichnet. Es gilt:

$$Q = n \sum_{i=1}^r \sum_{j=1}^s \frac{(f_{ij} - f_{i\bullet} \cdot f_{\bullet j})^2}{f_{i\bullet} \cdot f_{\bullet j}}.$$

Für eine (2×2) - Kontingenztafel gilt die einfache Formel:

$$Q = n \frac{(h_{11}h_{22} - h_{12}h_{21})^2}{h_{1\bullet}h_{2\bullet}h_{\bullet 1}h_{\bullet 2}}.$$

Kontingenzkoeffizient, normierter Kontingenzkoeffizient

Der **Kontingenzkoeffizient nach Pearson** ist gegeben durch

$$K = \sqrt{\frac{Q}{n + Q}}$$

und nimmt die Werte zwischen 0 und $K_{\max} = \sqrt{\frac{\min(r,s)-1}{\min(r,s)}}$ an. Der normierte Kontingenzkoeffizient ist definiert als

$$K^* = \frac{K}{K_{\max}}$$

und nimmt Werte zwischen 0 und 1 an.

4.9. Empirische Kovarianz, Korrelationskoeffizient und Rangkorrelationskoeffizient

Die **empirische Kovarianz** einer bivariaten Stichprobe $(x_1, y_1), \dots, (x_n, y_n)$ ist definiert als

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

Die empirische Kovarianz ist eine Funktion der beiden Datenvektoren $\mathbf{x} = (x_1, \dots, x_n)$ und $\mathbf{y} = (y_1, \dots, y_n)$. Mitunter verwenden wir daher auch die Notation $\text{cov}(\mathbf{x}, \mathbf{y})$:

$$s_{xy} = \text{cov}(\mathbf{x}, \mathbf{y}).$$

Eigenschaften: Für $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ und Zahlen $a, b \in \mathbb{R}$ gilt:

- $\text{cov}(\mathbf{x}, \mathbf{y}) = \text{cov}(\mathbf{y}, \mathbf{x})$
- $\text{cov}(a\mathbf{x}, b\mathbf{y}) = ab\text{cov}(\mathbf{x}, \mathbf{y})$
- $\text{cov}(\mathbf{x}, \mathbf{y}) + \mathbf{z} = \text{cov}(\mathbf{x}, \mathbf{y}) + \text{cov}(\mathbf{x}, \mathbf{z})$
- $\text{cov}(\mathbf{x}, \mathbf{x}) = s_x^2$
- $\text{var}(\mathbf{x} + \mathbf{y}) = \text{var}(\mathbf{x}) + \text{var}(\mathbf{y}) + 2\text{cov}(\mathbf{x}, \mathbf{y})$

Korrelationskoeffizient nach Bravais - Pearson

Für eine bivariate Stichprobe $(x_1, y_1), \dots, (x_n, y_n)$ ist der **Korrelationskoeffizient nach Bravais-Pearson** gegeben durch

$$r_{xy} = \hat{\rho} = \text{cor}(\mathbf{x}, \mathbf{y}) = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}},$$

wobei $s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ und $s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$.

Eigenschaften des Korrelationskoeffizienten: Für alle $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ und Zahlen $a, b, c, d \in \mathbb{R}$ gilt:

- $-1 \leq r_{xy} \leq 1$
- $\text{cor}(a\mathbf{x} + b, c\mathbf{y} + d) = \text{cor}(\mathbf{x}, \mathbf{y})$
- $|r_{xy}| = 1$ gilt genau dann, wenn \mathbf{x} und \mathbf{y} linear abhängig sind. Speziell:

1. $r_{xy} = 1$ genau dann, wenn $\mathbf{y} = a + b\mathbf{x}$ mit $b > 0$
2. $r_{xy} = -1$ genau dann, wenn $\mathbf{y} = a + b\mathbf{x}$ mit $b < 0$.

Rangkorrelationskoeffizient

Die der bivariaten Stichprobe $(x_1, y_1), \dots, (x_n, y_n)$ zugrunde liegenden Merkmale X und Y seien nun ordinal skaliert. Dann können wir den x - und y -Werten **Rangzahlen** zuordnen:

Die Beobachtung x_i erhält den Rang $r_{X,i} = k$, wenn x_i an der k -ten Stelle in der Ordnungsstatistik $x_{(1)}, \dots, x_{(n)}$ steht: $x_i = x_{(k)}$. Ist die Position k nicht eindeutig, da es mehrere Beobachtungen mit dem Wert x_i gibt, dann verwendet man das arithmetische Mittel dieser Positionen (Mittelränge).

Für $n \geq 4$ ist der **Rangkorrelationskoeffizient nach Spearman** gegeben durch

$$R_{Sp} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n+1)(n-1)}$$

mit $d_i = r_{Y,i} - r_{X,i}$, $i = 1, \dots, n$.

4.10. Einfache lineare Regression und KQ-Methode

Gesucht werden Koeffizienten a, b , so dass die Gerade

$$f(x) = a + bx, \quad x \in \mathbb{R},$$

den Datensatz $(x_1, y_1), \dots, (x_n, y_n)$ bestmöglichst approximiert.

Kleinste - Quadrate Methode (KQ-Methode)

Bei der **KQ-Methode** wird die Zielfunktion

$$Q(a, b) = \sum_{i=1}^n (y_i - (a + bx_i))^2, \quad (a, b) \in \mathbb{R}^2,$$

minimiert. Die Minimalstelle (\hat{a}, \hat{b}) ist gegeben durch

$$\begin{aligned}\hat{b} &= \frac{s_{xy}}{s_x^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \hat{a} &= \bar{y} - \hat{b}\bar{x}.\end{aligned}$$

Ausgleichsgerade, Regressionsgerade

Sind \hat{a}, \hat{b} die KQ-Schätzer für a, b , dann ist die **Ausgleichsgerade (geschätzte Regressionsgerade)** gegeben durch

$$\hat{f}(x) = \hat{a} + \hat{b} \cdot x, \quad x \in [x_{\min}, x_{\max}].$$

Das Intervall $[x_{\min}, x_{\max}]$ heißt **Stützbereich der Regression**.

Die Werte

$$\hat{y}_i = \hat{a} + \hat{b} \cdot x_i, \quad i = 1, \dots, n$$

heißen **Prognosewerte** oder auch **Vorhersagewerte**. Die Differenzen zu den Zielgrößen Y_i ,

$$\hat{\epsilon}_i = y_i - \hat{y}_i, \quad i = 1, \dots, n,$$

sind **geschätzte Residuen** (kurz: **Residuen**).

Weiterhin sind

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2, \quad SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad \text{und} \quad SSE = \sum_{i=1}^n \hat{\epsilon}_i^2.$$

Es gilt die folgende **Streuungszerlegung**:

$$SST = SSR + SSE$$

Der durch die Regression erklärte Anteil $R^2 = \frac{SSR}{SST}$ heißt **Bestimmtheitsmaß** und es gilt $R^2 = r_{xy}^2 = \text{cor}(\mathbf{x}, \mathbf{y})^2$