Definitionen und Eigenschaften mit ausgewählten Beispielen

Mathematik

Westra, Wenzgasse

1 Mengen

Eine Menge ist eine Sammlung mit Objekten, die wir **Elemente** nennen. Die Mengen können als Listen dargestellt werden:

Tage der Woche: {Sonntag, Montag, Dienstag, Mittwoch, Donnerstag, Freitag, Samstag}.

Positive Vielfache von 3: $\{3, 6, 9, 12, \ldots\}$.

Eine besondere Menge ist die leere Menge $\emptyset$, die kein einziges Element enthält.

Falls a ein Element von der Menge A ist, schreibt man $a \in A$, falls a nicht ein Element von A ist, dann $a \notin A$.

1.1 Zahlenmengen

Die Menge der **natürlichen Zahlen** $\mathbb{N}$ ist die Menge $\{0, 1, 2, 3, \ldots\}$.

Die Menge der **ganzen Zahlen** $\mathbb{Z}$ ist die Menge $\{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$.

Die Menge der **rationalen Zahlen** $\mathbb{Q}$ ist die Menge aller Zahlen $\frac{a}{b}$, wobei $a, b \in \mathbb{Z}$, aber $b \neq 0$.

Die Menge der **reellen Zahlen** $\mathbb{R}$ ist die Menge aller Zahlen, die man mit einer Dezimalentwicklung darstellen kann. Diese Dezimalentwicklung kann endlich sein, aber auch unendlich.

Die **reelle Ebene** ist die Menge $\mathbb{R}^2$ aller Zahlenpaare (x, y) mit $x, y \in \mathbb{R}$.

Eine Zahl q ist **irrational**, falls sie eine reelle Zahl ist, die keine rationale Zahl ist. Die Dezimalentwicklung einer rationalen Zahl ist endlich oder ab einer bestimmten Stelle periodisch. Eine irrationale Zahl hat eine nicht-periodische Dezimalentwicklung.

Beispiel: Die Zahl $3,4567878787\overline{878}\cdots$ ist letztendlich periodisch und somit eine Bruchzahl. Die Zahl $1,101001000100001000001\cdots$ ist irrational.

1.2 Mengen duch Bedingungen definiert

Eine **Bedingung** B ist eine Funktion auf einer Menge und kann nur die Werte 0 und 1 annehmen. Falls die Bedingung für x stimmt, dann $B(x) = 1$, falls die Bedingung für x nicht stimmt, dann ist $B(x) = 0$.

Beispiel: Sei B die Bedingung auf der natürlichen Zahlen definiert durch $B(x) = 1$ genau dann, wenn x eine Primzahl ist. Somit ist $B(0) = 0$, $B(1) = 0$, $B(2) = 1$, $B(3) = 1$, $B(4) = 0$ und so weiter.

Eine Menge M ist durch eine Bedingung B auf A definiert, falls sie genau die Elemente enthält, für welche $B(x) = 1$. Wir schreiben dann $M = \{x \in A \mid B(x) = 1\}$.

In der Praxis kann man die Bedingung auch umschreiben.

Beispiel: Die Menge $\{x \in \mathbb{Z} \mid 2|x\}$ enthält alle Vielfache von zwei, also die geraden Zahlen. Die Menge $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 4\}$ ist ein Kreis mit Radius 2 in der Ebene.

Ein **Intervall** ist eine Teilmenge von $\mathbb{R}$ von einer der folgenden Formen:

Ein **geschlossenes Intervall**: $[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$, wobei $a \leq b$.

Ein **offenes Intervall**: $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$, wobei $a < b$.

Ein **halb-offenes Intervall**: $(a, b] = \{x \in \mathbb{R} \mid a < x \leq b\}$, $[a, b) = \{x \in \mathbb{R} \mid a \leq x < b\}$, wobei $a < b$.

1.3 Operationen mit Mengen

Seien A und B Mengen, dann:

Die **Vereinigung** von A und B ist die Menge aller x, für welche $x \in A$ oder $x \in B$, wobei x auch in beiden enthalten sein darf. Wir schreiben $A \cup B$ für die Vereinigung von A und B. Klarerweise liegen sowohl A wie auch B in der Menge $A \cup B$, also $A \subset A \cup B$ und $B \in A \cup B$. Auch evident ist $A \cup B = B \cup A$.

Der **Durchschnitt** von A und B ist die Menge aller x, für welche $x \in A$ und $x \in B$. Wir schreiben $A \cap B$ für den Durchschnitt von A und B. Es gilt $A \cap B \subset A$ und $A \cap B \subset B$ und $A \cap B = B \cap A$ und $A \cap B \subset A \cup B$.

Die Menge $A \setminus B$ besteht aus allen Elementen $x \in A$, die nicht in B liegen. Es gilt $A \setminus B \subset A$, aber $A \setminus B \neq B \setminus A$. Es gilt sogar $(A \setminus B) \cap B = \emptyset$.

2 Algebra

Eine **Tautologie** ist eine Aussage, die immer wahr ist. Eine Gleichung $G(x) = F(x)$ ist eine Tautologie, falls sie für alle x wahr ist.

Ein **Widerspruch**, auch wohl **Contradictio** ist eine Aussage, die niemals wahr ist. Eine Gleichung $G(x) = F(x)$ ist ein Widerspruch, falls für keinen Wert für x die Aussage $G(x) = F(x)$ wahr ist.

Eine Gleichung $G(x) = F(x)$ ist **lösbar**, falls sie kein Widerspruch ist. Die Menge aller x, für welche die Gleichung erfüllt ist, heißt die Lösungsmenge der Gleichung.

Beispiel. Die Gleichung $(x+2)^2 = 4+4x+x^2$ ist eine Tautologie. Die Gleichung $x+1 = x$ ist ein Widerspruch. Die Gleichung $x^2 = x$ ist lösbar und die Lösungsmenge L finden wir durch

$$L = \{x \in \mathbb{R} \mid x^2 = x\} = \{0, 1\}\,.$$

Achtung: Bei der Lösungsmenge muss immer darauf geachtet werden, was die **Grundmenge der Gleichung** ist, also aus welcher Menge die Lösungen genommen werden dürfen. Betrachte zum Beispiel:

$$\{x \in \mathbb{Z} \mid 2x = 1\} = \emptyset\,, \quad \{x \in \mathbb{Q} \mid 2x = 1\} = \{\frac{1}{2}\}\,, \quad \{x \in \mathbb{R} \mid x^2 + 1 = 0\} = \emptyset\,.$$

Bekanntlich aber $\{x \in \mathbb{C} \mid x^2 + 1 = 0\} = \{-i, +i\}$.

Achtung: Eine Gleichung ist eine Tautologie, falls die Lösungsmenge mit der Grundmenge zusammenfällt: $\{x \in \mathbb{N} \mid 1 = 1\} = \mathbb{N}$.

2.1 Lineare Gleichungen

Eine **lineare Gleichung** ist eine Gleichung in einer Unbekannten x, die auf die Form $ax + b = c$ gebracht werden kann. Falls $a \neq 0$ ist sie immer lösbar.

2.2 Quadratische Gleichungen

Eine **quadratische Gleichung** ist eine Gleichung in einer Variablen x, die in die Form $ax^2 + bx + c = 0$ geschrieben werden kann. Hierbei sind a, b und c reelle Zahlen, wobei aber $a \neq 0$. Die Lösungen werden in der Menge der reellen Zahlen gesucht.

Die **Diskriminante** der quadratischen Gleichung $ax^2 + bx + c = 0$ ist der Term $D = b^2 - 4ac$.

Falls $D > 0$ gibt es zwei Lösungen: $x = \frac{-b\pm\sqrt{D}}{2a}$.

Falls $D = 0$ gibt es genau eine Lösung: $x = -\frac{b}{2a}$.

Falls $D < 0$ gibt es keine Lösungen.

2.3 Exponente

Für natürliche Zahlen n ist klar was mit a^n gemeinst ist; es ist das Produkt von n Exemplaren a. Für $n = 0$ und $a \neq 0$ setzt man $a^0 = 1$. Es gibt keine gute Definition von 0^0, also lassen wir das undefiniert.

Sei nun weiter $a \neq 0$ und $n \in .\mathbb{N}$, dann definiert man $a^{-n} = \frac{1}{a^n}$.

Falls $a > 0$ so existiert auch $\sqrt[n]{a}$, und dies ist definiert als die positive Zahl b mit der Eigenschaft, dass $b^n = a$. Es bedarf hier der Erläuterung, dass diese Zahl b eindeutig definiert ist, sodass $\sqrt[n]{a}$ eindeutig berechenbar ist. Wir schreiben auch wohl $\sqrt[n]{a} = a^{\frac{1}{n}}$ und es gilt $(a^{\frac{1}{n}})^n = a^{\frac{n}{n}} = a^1 = a$.

Achtung! In der Schulmathematik in der Oberstufe werden solche Wurzeln $\sqrt[n]{a}$ nur für positive a erlaubt. Obwohl man $\sqrt[3]{-1}$ eindeutig als -1 definieren kann, gibt es einfache Gründe, dies nicht zu erlauben! Da diese Gründe keine Universalität besitzen, lasse ich sie hier weg. Merkt es euch einfach; gebrochene Exponente nur bei positiver Basiszahl.

Falls r eine Bruchzahl ist und $a > 0$, so gibt es $n, m \in \mathbb{Z}$, sodass $r = \frac{n}{n}$ und somit kann man definieren

$$a^r = a^{\frac{n}{m}} = \sqrt[m]{a^n} = (\sqrt[m]{a})^n .$$

Falls r eine reelle Zahl ist, so gibt es eine Folge von Bruchzahlen $r_1, r_2, r_3, \ldots$, sodass $\lim_{n\to\infty} r_n = r$, m.a.W., die Folge der Bruchzahlen r_n konvergiert zu r. Dann definieren wir

$$a^r = \lim_{n\to\infty} a^{r_n} .$$

Beispiel. Wie berechnet man eigentlich $2^{\sqrt{3}}$? Nun, wir kennen die Dezimalentwicklung $\sqrt{3} = 1,732050\ldots$. Wir nehmen nun $r_1 = 1,7 = \frac{17}{10}$, $r_2 = 1,73\frac{173}{100}$, $r_3 = 1,732 = \frac{1732}{1000}$, $r_4 = 1,7320 = \frac{17320}{10000}$ (was gleich r_3 ist, aber das ist uns jetzt egal), $r_5 = 1,73205 = \frac{173205}{100000}$, und so weiter. So können wir dann berechnen

$$2^{r_1} = \sqrt[10]{2^{17}} = 3,249\ldots , \quad 2^{r_2} = 3,317\ldots , \quad 2^{r_3} = 3,321880\ldots$$

und so weiter. Diese Reihe konvergiert zu einer Zahl, und diese Zahl nennen wir $2^{\sqrt{3}}$. Diese Definition ist etwas umständig, aber, wir brauchen nur eine bestimmte Genauigkeit, somit könnten wir auch schon mit $2^{\sqrt{3}} \approx 3,3$ recht zufrieden sein.

2.4 Logarithmus

Gleichungen von der Form $a^x = b$ mit $a, b \in \mathbb{R}^{>0}$ sind nicht mit den vorigen Methoden immer lösbar, es sei denn, du hast einen Zufallstreffer wie $3^x = 9$. Wir definieren nun den a-Logarithmus wie folgt:

Der **a-Logarithmus von b** ist die Zahl x, zu welcher man a potenzieren muss, um b zu bekommen. Wir schreiben dann ${}^a\log(b)$ oder $\log_a(b)$ für den a-Logarithmus von b.

Die Identität $a^{\log_a(b)} = b$ ist eine Tautologie; schreiben wir $x = \log_a(b)$, dann heißt die Definition nichts anderes als: x ist die Zahl zu welcher man a potenzieren muss, um b zu bekommen. Somit, $a^x = b$.

Die Identität $\log_a(a^r) = r$ ist auch fast eine Tautologie; um a^r zu bekommen, muss man a mit r potenzieren.

Für Informationen zu e siehe unten bei Funktionen.

3 Geometrie

Eine **Gerade** in der Ebene ist die Lösungsmenge einer Gleichung von der Form $ax+by = c$, wobei a, b und c fixe Zahlen sind, die die Gerade charakterisieren. Das Verhältnis $\frac{-b}{a}$ heißt die Steigung der Geraden.

Das **Skalarprodukt** zweier dreidimensionalen Vektoren $\vec{a} = (a_1, a_2, a_3)$ und $\vec{b} = (b_1, b_2, b_3)$ ist die Zahl $\vec{a} \cdot \vec{b} = a_1b_1 + a_2b_2 + a_3b_3$. In zwei Dimensionen geht es genau so, nur gibt es dann keine dritten Komponenten.

Zwei Vektoren **stehen normal aufeinander**, wenn ihr Skalarprodukt Null ist.

Zwei zweidimensionale Vektoren (a_1, a_2) und (b_1, b_2) sind **parallel zu einander**, falls $a_1b_2 = a_2b_1$.

Zwei dreidimensionale Vektoren $\vec{a} = (a_1, a_2, a_3)$ und $\vec{b} = (b_1, b_2, b_3)$ sind **parallel zu einander**, falls $a_1b_2 = a_2b_1$, $a_2b_3 = a_3b_2$ und $a_1b_3 = a_3b_1$. M.a.W. $\vec{a} \times \vec{b} = 0$.

Ein **Normalvektor** zu einer Geraden g ist ein Vektor $\vec{n}$, der normal auf dem Richtungsvektor $\vec{v}_g$ von g steht, $\vec{n} \cdot \vec{v}_g = 0$. Zu einer gegebenen Gerade gibt es mehrere Normalvektoren, die aber alle parallel zu einander sind.

Parameterdarstellung einer Gerade: $L: \ \vec{X} = P_0 + t \cdot \vec{g}$. Hierbei ist $\vec{g}$ ein Richtungsvektor von L. P_0 ist ein Punkt auf L. P_0 ist frei auf L wählbar. $\vec{g}$ ist bis auf ein Vielfaches bestimmt. Der Parameter t nimmt Werte in $\mathbb{R}$ an.

Satz der Mehrdeutigkeit der Parameterdarstellung: Es seien zwei Parameterdarstellungen gegeben $L: \ \vec{X} = P_0 + t \cdot \vec{g}$ und $L': \ \vec{X} = Q_0 + s \cdot \vec{h}$. Dann sind L und L' dieselbe Gerade genau dann, wenn $\vec{g}$ und $\vec{h}$ und $\overrightarrow{P_0Q_0} = Q_0 - P_0$ alle drei parallel sind. L und L' sind parallel, genau dann wenn $\vec{g}$ und $\vec{h}$ parallel sind.

Normalform, Normalvektordarstellung einer Geraden L: $\vec{X} \cdot \vec{n} = c$, auch wohl $ax + by = c$. Hierbei ist $(a|b) = \vec{n}$ ein Normalvektor zu L. Somit wäre dann $(-b|a)$ ein Richtungsvektor zu L. Wenn $a \neq 0$, so ist $(\frac{c}{a}|0)$ ein Punkt auf L. Wenn $b \neq 0$, so ist $(0|\frac{c}{b})$ auch ein Punkt auf L.

4 Funktionen

Eine Funktion f ist **monoton steigend**, falls $a \leq b$ impliziert $f(a) \leq f(b)$.

Eine Funktion f ist **monoton fallend**, falls $a \leq b$ impliziert $f(a) \geq f(b)$.

Eine Funktion f ist **streng monoton steigend**, falls $a < b$ impliziert $f(a) < f(b)$.

Eine Funktion f ist **streng monoton fallend**, falls $a < b$ impliziert $f(a) > f(b)$.

Eine Funktion $f : A \to \mathbb{R}$ ist **stetig an der Stelle** $x \in A$, falls $\lim_{x\uparrow a} f(x) = f(a) = \lim_{x\downarrow a} f(x)$. Um sauber zu sein, muss A hierbei offen sein.

Eine Funktion f ist **stetig**, falls sie an jeder Stelle in der (offenen) Definitionsmenge stetig ist.

4.1 Besondere Punkte im Graphen

Eine Nullstelle einer Funktion f ist eine Stelle a, sodass $f(a) = 0$.

Ein **lokales Maximum** einer Funktion f ist eine Stelle a, sodass es ein offenes Intervall I um a gibt, und $f(x) < f(a)$ für alle $x \in I$ ungleich a.

Ein **lokales Minimum** einer Funktion f ist eine Stelle a, sodass es ein offenes Intervall I um a gibt, und $f(x) > f(a)$ für alle $x \in I$ ungleich a.

Wenn eine Funktion differenzierbar ist, haben wir aber auch folgende Charakterisierungen:

(a) f hat an der Stelle a ein Minimum, wenn $f'(a) = 0$ und f' wechselt von Vorzeichen bei $x = a$, und zwar so, dass f' von negativ auf positiv wechselt.

(a) f hat an der Stelle a ein Maximum, wenn $f'(a) = 0$ und f' wechselt von Vorzeichen bei $x = a$, und zwar so, dass f' von positiv auf negativ wechselt.

Beispiel: Die Funktion $f(x) = |\sin(x)|$ hat ein Minimum bei jeder Nullstelle, obwohl die Ableitung dort nicht existiert. Schau dir den Graphen mit GeoGebra an!

Eine **Wendestelle** einer Funktion f ist eine Stelle a, sodass $f''(a) = 0$ und $f'(a)$ wechselt das Vorzeichen bei a.

Eine **Terrassenstelle** einer Funktion f ist eine Stelle a, sodass $f'(a) = f''(a) = 0$ und $f'''(a) \neq 0$. Allgemeiner kann man sagen, dass eine Terrassenstelle eine Stelle ist, wo die ersten beiden Ableitungen verschwinden, aber f'' von Vorzeichen wechselt. So hat die Funktion $f(x) = x^5$ auch eine Terrassenstelle.

5 Analysis

Eine Funktion $f : D \to \mathbb{R}$ ist **differenzierbar an der Stelle** a, falls $\lim_{x \to a} \frac{f(x)-f(a)}{x-a}$ existiert. Eine Funktion ist **differenzierbar**, falls sie in der ganzen Definitionsmenge D differenzierbar ist.

Sei I ein geschlossenes Intervall, dann ist P eine **Partition von** I, falls P eine Vereinigungen von geschlossenen Intervallen $J \subset I$ ist, die einander nur in den Randpunkten schneiden dürfen. M.a.W. P teilt das Intervall I in kleineren, aber eventuell nicht gleich großen Intervallen auf.

Beispiel: Falls $I = [0, 1]$, dann ist $P = \{[0, \frac{1}{2}], [\frac{1}{2}, \frac{3}{4}], [\frac{3}{4}, \frac{4}{5}], [\frac{4}{5}, 1]\}$ eine Partition von I.

Sei $f : [a, b] \to \mathbb{R}$ eine (stetige) Funktion. Eine **Obersumme** von f für das Intervall $[a, b]$ ist eine Summe, die auf folgende Weise durch eine Partition P von $[a, b]$ bestimmt wird: für jedes Intervall J in P suchen wir das Maximum von f im Intervall und multiplizieren dieses Maximum mit der Breite von J. Die Summe all dieser Produkte für alle Intervalle in der Partition in P ist die Obersumme. Wir kunnen diese Obersumme mit O_P andeuten, wobei wir in der Notation f und $[a, b]$ nicht andeuten.

Sei $f : [a, b] \to \mathbb{R}$ eine stetige Funktion. Eine **Untersumme** von f für das Intervall $[a, b]$ ist eine Summe, die auf folgende Weise durch eine Partition P von $[a, b]$ bestimmt wird: für jedes Intervall J in P suchen wir das Minimum von f im Intervall und multiplizieren dieses Minimum mit der Breite von J. Die Summe all dieser Produkte für alle Intervalle in der Partition in P ist die Untersumme. Wir kunnen diese Untersumme mit U_P andeuten, wobei wir in der Notation f und $[a, b]$ nicht andeuten.

Die Obersummen und Untersummen haben eine wichtige Eigenschaft: Wenn wir die Partition P verfeinern, also alle Intervalle wieder partitionieren, sodass wir eine feinere Partition Q bekommen. Dann gilt $U_P \leq U_Q \leq O_Q \leq O_P$. Die Unter- und Obersummen werden sich dann annähern (können), wenn wir die Partitionen immer feiner machen.

Eine Funktion f ist **integrierbar** im Intervall $[a, b]$, falls die Obersummen O_P und Untersummen U_P für immer feiner werdende Partitionen gegen einander konvergieren. Dieser Grenzwert heißt dann das **Integral von** f **über das Intervall** $[a, b]$. Alle stetige Funktionen sind integrierbar, aber viele nicht stetige Funktionen sind auch integrierbar.

5.1 Funktionstypen

Eine Funktion f heißt **periodisch**, wenn es eine Zahl $p > 0$ gibt, sodass $f(x+p) = f(x)$ für alle x. Solche Zahlen nennt man dann eine **Periode**. Wenn man über **die Periode** einer Funktion spricht, meint man die kleinste positive Zahl mit dieser Eigenschaft.

Achtung: Es ist wesentlich, dass $p > 0$, weil für $p = 0$ findet man immer $f(x+0) = f(x)$. Konstante Funktionen sind zwar periodisch, haben aber keine Periode im strengen Sinne, da es keine kleinste positive Zahl gibt, die diese Bedingung erfüllt.

Die Funktionen $f(x) = a \cdot \sin(bx)$ und $g(x) = a \cdot \cos(bx)$, mit $a \in \mathbb{R}$ und $b > 0$, haben Periode $\frac{2\pi}{b}$. Es gilt $f'(x) = ab\cos(bx)$ und $g'(x) = -ab\sin(bx)$.

Die Funktion $f(x) = a\tan(bx)$ hat Periode $\frac{\pi}{b}$.

Die **linearen** Funktionen sind von der Form $f(x) = kx + d$ mit $k, d \in \mathbb{R}$. k heißt die **Steigung**, d heißt der Achsenabschnitt. Wenn $k = 0$, ist die Funktion **konstant**. Wenn $d = 0$, geht der Graph durch den Ursprung, und solche linearen Funktionen sind **direkte Proportionalitäten**. Direkte Proportionalitäten haben die Eigenschaft, dass das Verhältnnis $f(x) : x$ konstant ist, und der Steigung entspricht. Bei linearen Funktionen sind die mittleren Änderungsraten von x unabhängig und der Steigung gleich. So gilt zum Beispiel $f(x+1) - f(x) = k$.

Die **quadratischen Funktionen** sind von der Form $f(x) = ax^2+bx+c$, wobei $a, b, c \in \mathbb{R}$ und $a \neq 0$. Falls $a = 0$ wäre, dann wäre f eine lineare Funktion. Der Scheitel liegt bei $x = -\frac{b}{2a}$. In diesem Punkt ist die Tangente horizontal. Die Achse der Parabel geht durch den Scheitel und ist parallel zur zweiten Achse. Die Parabel ist spiegelsymmetrisch um diese Achse der Parabel.

Sekante-Tangente-Eigenschaft der Parabel: Die Sekante der Parabel durch die Punkte $(x_1|f(x_1))$ und $(x_2|f(x_2))$ ist zur Tangente an der Parabel an der Stelle $\frac{x_1+x_2}{3}$ parallel. Eine Sonderform dieser Eigenschaft: Der Scheitel liegt genau in der Mitte der beiden Nullpunkte, falls es sie gibt.

Die **Exponentialfunktionen** haben zwei bequeme Schreibweisen: $f(x) = c \cdot a^x$, wobei $c, a \in \mathbb{R}$ und $a > 0$ einerseits und $f(x) = be^{\lambda x}$ andererseits. Es gilt $f'(x) = \ln(a) \cdot c \cdot a^x$ und $f'(x) = \lambda b e^{\lambda x}$.

Potenzfunktionen haben die Form $f(x) = a \cdot x^r$ mit $r \in \mathbb{R}$.

Exponentialfunktionen: Sei $a > 0$ und $a \neq 1$, dann definiert $f(x) = b \cdot a^x$ auch eine Funktion. Eine elementare Berechnung ergibt

$$f'(x) = \lim_{h\to 0} \frac{ba^{x+h} - ba^x}{h} = \lim_{h\to 0} b \cdot a^x \cdot \frac{a^h - 1}{h} = b \cdot a^x \lim_{h\to 0} \frac{a^h-1}{h}.$$

Nun hängt $\lim_{h\to 0} \frac{a^h-1}{h}$ nicht von x ab! Somit gilt $f'(x) = C \cdot f(x)$, wobei C irgendeine Zahl ist, die durch a selbst bestimmt ist. Die Ableitung einer Exponentialfunktion ist direkt proportional zur Funktion selbst. Damit können wir folgendes definieren:

Die natürliche Zahl e ist die Zahl, für welche die Identität $\lim_{h\to 0} \frac{e^h-1}{h} = 1$. Somit gilt für die Funktion $f(x) = e^x$ dass $f'(x) = f(x)$.

Der natürliche Logarithmus ist der Logarithmus mit Basiszahl e. Man schreibt $\ln(x) = \log_e(x)$.

Satz zur Zahl e: Es gilt, dass $e \approx 2,71828$, dass e irrational ist, und dass

$$\lim_{h\to 0} \frac{a^h - 1}{h} = \ln(a)$$

und dass

$$e = \lim_{n\to\infty} (1 + \frac{1}{n})^n \,.$$

5.2 Differenzieren - Ergänzung

Sei $f : \mathbb{R} \to \mathbb{R}$ eine reelle Funktion. Der Quotient $\frac{f(x_1)-f(x_2)}{x_1-x_2}$ ist ein **Differenzenquotient** und gibt die Steigung der Sekante durch die Punkte $(x_1|f(x_1))$ und $(x_2|f(x_2))$. Sie ist die mittlere Steigung der Funktion f im Intervall $[x_1; x_2]$.

Falls der Grenzwert $\lim_{x_2\to x_1} \frac{f(x_1)-f(x_2)}{x_1-x_2} = \lim_{h\to 0} \frac{f(x_1+h)-f(x_1)}{h} = \lim_{h\to 0} \frac{f(x_1)-f(x_1-h)}{h}$ existiert (und alle drei Ausdrücke gleich sind), nennen wir f **differenzierbar** in x_1. Falls f überall in ihrem Definitionsbereich differenzierbar ist, so heißt f selbst differenzierbar. Wir nennen den Audruck

$$\lim_{z\to x} \frac{f(x)-f(z)}{x-z} = f'(x)$$

einen **Differentialquotienten**. $f'(x)$ gibt die Steigung der Tangente an der Stelle x an.

Example. Die Funktion $f(x) = |x|$ ist nicht überall differenzierbar; an der Stelle $x = 0$ geht es schief. Man sieht dies geometrisch am leichtesten; es gibt keine eindeutige Tangente in $(0|0)$.

Wichtige Ableitungen, die man nicht nachschauen sollte, sind in der folgenden Tabelle gegeben:

$f(x)$	$f'(x)$
x^n	nx^{n-1}
e^x	e^x
a^x	$\ln(a) \cdot a^x$
$\sin(x)$	$\cos(x)$
$\cos(x)$	$-\sin(x)$
$\ln(x)$	$\frac{1}{x}$
1	0

Es gelten die folgenden Regeln:

Summenregel: Falls $f(x) = f_1(x) + f_2(x)$, so gilt $f'(x) = f_1'(x) + f_2'(x)$. Also, die Ableitung der Summe ist die Summe der Ableitungen.

Produktregel: Falls $f(x) = g(x) \cdot h(x)$, so gilt $f'(x) = g'(x)h(x) + g(x)h'(x)$. Also, das Produkt der Ableitungen ist nicht die Ableitung des Produkts.

Kettenregel aka **Verknüpfungsregel**: Falls f zusammengesetzt ist, und zwar $f(x) = g(h(x))$, dann $f'(x) = g'(h(x)) \cdot h'(x)$.

Beispiel: Zur Produktregel: $f(x) = x^2 = x \cdot x$, also $f'(x) = (x)' \cdot x + x \cdot (x)' = 1 \cdot x + x \cdot 1 = x + x = 2x$. So auch $(x^3)' = (x \cdot x^2)' = 1 \cdot x^2 + x \cdot 2x = x^2 + 2x^2 = 3x^2$. Zur Kettenregel: Sei $f(x) = e^{x^2}$, dann $f(x) = e^{h(x)}$ mit $h(x) = x^2$. Also $f'(x) = e^{h(x)} \cdot h'(x)$. Noch ein Beispiel: $f(x) = \sin(x^2)$, dann $f'(x) = 2x \cdot \cos(x^2)$. Wenn aber $f(x) = (\sin(x))^2$ (also, Reihenfolge von g und h umgedreht), dann $f'(x) = 2\sin(x)\cos(x)$.

Achtung: Die Kettenregel sorgt oft für Probleme, da man die richtige Reihenfolge nicht findet. $f(x) = g(h(x))$ bedeutet, zuerst wird aus x das Bild $h(x)$ gemacht, darauf wird dann g angewandt: $x \mapsto h(x) \mapsto g(h(x))$. Aber, die letzte Aktion die angewandt wurde, ist hier eindeutig g, und so liest man es auch; von außen nach innen. Zum Beispiel $g(x) = x^2$ und $h(x) = e^x$, dann $g(h(x)) = h(x)^2 = e^{2x}$ und $h(g(x)) = e^{g(x)} = e^{x^2}$, und die beiden sind sicher nicht gleich!

Mehrfache Ableitungen entstehen, wenn man die Ableitung $f'(x)$ noch einmal differenziert: $f''(x) = (f'(x))'$. Und so weiter, also die n-te Ableitung ist die Ableitung von der $(n-1)$-ten Ableitung.

6 Wahrscheinlichkeit

Ein **Zufallsexperiment** ist ein Vorgang mit Anfang und Ende, bei dem das Ergebnis nicht deterministisch beschrieben wird. Die Menge der möglichen Ergebnisse ist die **Ereignismenge**, und ein Ergebnis wird auch wohl ein **Ereignis** genannt.

Zwei Ereignisse A und B sind disjunkt, falls A und B nicht beide gleichzeitig auftreten können.

Eine Wahrscheinlichkeit ist eine Zuordnung $P : E \to [0, 1]$, wobei E eine Ereignismenge ist, wobei P aber bestimmte Eigenschaften haben muss. Falls A und B disjunkt sind ist $P(A \cup B)$ die Summe von $P(A)$ und $P(B)$; die Wahrscheinlichkeit $P(E) = 1$, also, irgendein Ereignis tritt sowieso ein. Man nennt diese Wahrscheinlichkeitenzuordnung auch wohl eine Verteilung.

Achtung: Die Ereignismenge zu geben, ist ab und zu recht aufwändig. Grundsätzlich ist sie aber eine Menge, also eine Liste muss oft möglich sein. Die Ereignisse können einzelne Ereignisse sein, aber auch mehrere zusammen, also wieder Mengen.

Die Bedeutung einer Wahrscheinlichkeit liegt außerhalb der Mathematik. Daher nur zwei Bemerkungen:

Ansatz von Laplace: Die Wahrscheinlichkeit $P(A)$ eines Ereignisses A ist definiert als der Anteil an elementaren Ereignissen, die A induzieren.

Das Gesetz der großen Zahlen: Sei die Wahrscheinlichkeit eines Ereignisses $P(A)$. Falls wir ein Zufallsexperiment N mal wiederholen, und dabei mit $N(A)$, die Anzahl der

Experimente notieren, bei denen A eingetreten ist, dann wird sih der Quotient $N/N(A)$ dem Wert $P(A)$ annähern, in dem Übergang $N \to \infty$.

6.1 Stochastische Variablen

Bei diskreten Verteilungen gibt es nur eine diskrete Liste an Wahrscheinlichkeiten. Bei stetigen Verteilungen gibt es eine Dichtefunktion f, definiert auf einer Menge $E \subset \mathbb{R}$.

Eine Zufallsvariable (stochastische Variable) ordnet jedem Ereignis eines Zufallsexperiments eine Zahl zu. Somit ist eine Zufallsvariable eine Funktion $X : E \to \mathbb{R}$.

Für diskrete Zufallsvariablen wird der **Erwartungswert** wie folgt definiert: $EX = \sum_k k \cdot P(X = k)$, wobei die Summe über alle Werte von X geht.

Falls es sich um eine stetige Verteilung mit Dichtefunktion f handelt, so definiert man den **Erwartungswert** von X durch $EX = \int X(x)f(x)\mathrm{d}x$.

Beispiel. Sei $E = \{1, 2, 3, 4, 5, 6\}$ der Ereignisraum beim einmaligen Würfeln. Sei dann $X : E \to \mathbb{R}$ definiert durch $X(1) = 0$, $X(2) = 0$, $X(3) = 0,5$, $X(4) = 1$, $X(5) = 1,5$ und $X(6) = 3$, dann $EX = \frac{1}{6} \cdot 0 + \frac{1}{6} \cdot 0 + \frac{1}{6} \cdot 0,5 + \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 1,5 + \frac{1}{6} \cdot 3 = 1$.

Beispiel. Sei $X = [0, \infty)$ mit Dichtefunktion $f(x) = e^{-x}$. Sei $X(x) = x^2$, dann $EX = \int_0^\infty x^2 e^{-x}\mathrm{d}x = 2$, mit partieller Integration.

Die **Varianz** einer stochastischen Variable V_X oder $\mathrm{Var}(X)$ ist definiert durch $\mathrm{Var}(X) = E(X - EX)^2$.

Satz. Es gilt folgende Identität: $E(X - EX)^2 = E(X^2) - (EX)^2$.

Beispiel. Sei $E = \{1, 2, 3\}$ mit Verteilung $P(1) = P(2) = P(3) = \frac{1}{3}$ und sei X definiert durch $X(1) = -2$, $X(2) = 2$, $X(3) = 6$. Somit gilt $EX = 2$. Daher ist $Y = (X - EX)^2$ die Variable die die Werte $(-2-2)^2 = 16$, $(2-2)^2 = 0$ und $(6-2)^2 = 16$ annimmt. Die Varianz von X ist der Erwartungswert von X und somit $\mathrm{Var}(X) = \frac{2}{3} \cdot 16 + \frac{1}{3} \cdot 0 = 10\frac{2}{3}$. Andererseits hätten wir auch zuerst $E(X^2) = \frac{(-2)^2+2^2+6^2}{3} = \frac{44}{3}$ und $(EX)^2 = 4 = \frac{12}{3}$ bestimmen können. Und tatsächlich $\mathrm{Var}(X) = E(X^2) - (EX)^2 = \frac{44}{3} - \frac{12}{3}$.

Beispiel. Sei $X = [0, \infty)$ mit Dichtefunktion $f(x) = e^{-x}$. Sei $X(x) = x^2$, sodass $EX = 2$. Dann $E(X^2) = \int_0^\infty x^4 e^{-x}\mathrm{d}x = 24$. Somit $\mathrm{Var}(X) = 24 - 4 = 20$. Andererseits $E(X - EX)^2 = E(X-2)^2 = \int_0^\infty (x^2-2)^2 e^{-x}\mathrm{d}x = \int_0^\infty (x^4 - 4x^2 + 4)e^{-x}\mathrm{d}x = 24 - 8 + 4 = 20$.

Die **Standardabweichung** einer Zufallsvariable X ist die Wurzel der Varianz: $\sigma_X = \sqrt{\mathrm{Var}(X)}$.

Satz. Es gelten folgende Identitäten für Zufallsvariablen X_1 und X_2, $a \in \mathbb{R}$: $E(X_1+X_2) = EX_1 + EX_2$, $E(aX) = aEX$, $\mathrm{Var}(aX) = |a|\mathrm{Var}(X)$.

Die **Verteilungsfunktion** für eine Zufallsvariable X ist die Funktion F, die jeder Zahl x die positive Zahl $P(X \leq x)$ zuordnet. Somit $F(x) = P(X \leq x)$: Diese Funktion ist monoton steigend und positiv. Es gilt auch $F(-\infty) = 0$ und $F(\infty) = 1$.

6.2 Binomialverteilung

Die Binomialverteilung ist mittels eines Zufallsexperiments, oder eigentlich durch eine Reihe identischer und unabhängiger Zufallsexperimente, definiert: Bei einem einzelnen Experiment sei p die Wahrscheinlichkeit auf 1, und $q = 1 - p$ die Wahrscheinlichkeit auf 0. Jetzt wiederholen wir dieses Experiment n-mal und sei dabei X die Anzahl der Einzer. Es gilt die Formel

$$P(X = k) = \binom{n}{k} p^k q^{n-k} .$$

Man schreibt dann $X \sim \text{Bin}(n, p)$ und sagt, X ist binomialverteilt mit Parametern n und p. Es gilt $EX = np$ und $\text{Var}(X) = npq$. Merke auf, dass X nur die Werte $0, 1, 2, \ldots, n$ annehmen kann. X ist somit eine diskrete Zufallsvariable, die Binomialverteilung ist eine diskrete Verteiliung. Die Verteilungsfunktion ist nicht stetig.

Wichtige Ungleichung. Die Funktion $f(x) = x(1 - x)$ definiert eine Parabel mit einem Maximum bei $x = \frac{1}{2}$. Das Maximum ist dann $f(\frac{1}{2}) = \frac{1}{4}$. Somit haben wir bewiesen: $npq \leq \frac{n}{4}$, sodass $\text{Var}(X) = npq \leq \frac{n}{4}$ für $X \sim \text{Bin}(n, p)$.

6.3 Normalverteilung

Die Normalverteilung mit Mittelwert μ und Standardabweichung σ ist durch folgende Dichtefunktion definiert

$$f(x) = \tfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} .$$

Falls $\mu = 0$ und $\sigma = 1$ nennt man sie die Standard Normalverteilung. Falls X normalverteilt ist, gilt $EX = \mu$, $\text{Var}(X) = \sigma^2$. Die Dichtefunktion ist stetig, die Verteilungsfunktion ist auch stetig. Die Verteilungsfunktion der Standardnormalverteilung wird mit Φ angedeutet:

$$\Phi(z) = \int_{-\infty}^{z} \tfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, \mathrm{d}x .$$

Sei Z standardnormalverteilt: Es gilt $\Phi(-z) = P(X \geq z) = 1 - \Phi(z)$.

Satz. Falls $X \sim \mathcal{N}(\mu, \sigma)$, dann $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$.

7 Statistik

7.1 Beschreibende Statistik

Im Folgenden ist L eine Liste mit Daten. Wir nehmen mal an, dass die Liste L aus endlich vielen reellen Zahlen besteht. Wir werden auch annehmen, dass L geordnet ist, falls dies nicht der Fall ist, dann werden wir L zuerst ordnen. Wir schreiben dann $L = \{x_1, x_2, \ldots, x_n\}$, und nehmen also an $x_i \leq x_{i+1}$ ür alle i.

Die **Median** einer Liste L ist eine Zahl m, sodass mindestens 50% der Daten kleiner gleich und mindestens 50% der Daten größer gleich als m ist. Falls L eine ungerade Anzahl (sagen wir $n = 2N+1$) von Elementen hat, dann ist m die mittlere Zahl (Nummer $N+1$). Falls die Liste eine gerade Anzahl hat (sagen wir $n = 2N$), so nehmen wir den Mittelwert von den zwei Zahlen in der Mitte (von $N-1$ und $N+1$). Der Median L macht aus L zwei neue Listen L_1 und L_2, die beide gleich groß sind, und m liegt in keiner der beiden.

Das **arithmetische Mittel** von L ist die Summe aller Elemente von L dividiert durch die Anzahl der Elemente von L.

Das **geometrische Mittel** von L ist nur definiert, wenn alle Elemente von L positiv sind. In dem Fall ist das geometrische Mittel durch den Ausdruck $\sqrt[n]{x_1 \cdot x_2 \cdots x_n}$ gegeben.

Das **erste Quartil** einer Liste L ist eine Zahl q_1, sodass mindestens 25% der Daten kleiner gleich und mindestens 75% der Daten größer gleichen sind. Die Bildung geschieht genau wie beim Median; falls die Zahl mit der genannten Eigenschaften nicht in der Liste selbst ist, so liegt die Grenze zwischen den kleinsten 25% und den größten 75% zwischen zwei Zahlen in L, und dann bildet man den Mittelwert der beiden.

Das **dritte Quartil** einer Liste L ist eine Zahl q_3, sodass mindestens 75% der Daten kleiner gleich und mindestens 25% der Daten größer gleich sind. Die Bildung geschieht genau wie beim Median; falls die Zahl mit der genannten Eigenschaften nicht in der Liste selbst ist, so liegt die Grenze zwischen den kleinsten 75% und den größten 25% zwischen zwei Zahlen in L, und dann bildet man den Mittelwert der beiden.

Der Median ist das **zweite Quartil**.

Ein **gewichtetes Mittel** von den Zahlen $x_1, \ldots, x_n$ mit den Gewichten $w_1, \ldots, w_n$ ist die Summe $x_1 w_1 + x_2 w_2 + \ldots + x_n w_n$, wobei die Gewichte $w_1, \ldots, w_n$ die Bedingungen $w_1 + w_2 + \ldots + w_n = 1$ und $0 \leq w_i \leq 1$ für alle $i = 1, 2, \ldots, n$ erfüllen müssen. Die Wahl $w_i = \frac{1}{n}$ gibt uns das arithmetische Mittel zurück.

Die empirische Standardabweichung s ist definiert als die Wurzel von der empirischen Varianz. Letztere ist durch $s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$ definiert. Somit ist s ein Maß dafür, wie weit die Zahlen im Schnitt vom Mittelwert $\bar{x}$ liegen. Tatsächlich liegt x_i eine Distanz $|x_i - \bar{x}|$ vom Mittelwert. Der Grund warum man das Quadrat nimmt ist technisch. Eine Übung ist es, zu zeigen, dass es ohne Quadrat auf jeden Fall falsch geht: $\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}) = 0$.

7.2 Hypothesen und Konfidenzintervalle

8 Präfixe

Nano: Milliardstel 10^{-9}

Mikro-: Millionstel 10^{-6}

Milli-: Tausendstel 10^{-3}

Centi-: Hundertstel 10^{-2}

Dezi-: Zehntel 10^{-1}

Deka-: Zehn 10^{1}

Hekto-: Hundert 10^{2}

Kilo-: Tausend 10^{3}

Mega-: Million 10^{6}

Giga-: Milliard 10^{9}