Mathe Helfer

Simon Roske

18. Januar 2024

Inhaltsverzeichnis

Abbildungsverzeichnis

1 Zahlen

1.1 Die natürlichen Zahlen N

Der Zahlenbereich der natürlichen Zahlen bildet das Zählen als natürlichen Prozess ab. Ob die 0 teil von N ist, ist Definitionssache und nicht überall gleich. Angenommen, das sei nicht der Fall, dann gilt:

$$\mathsf{N} = \{1, 2, 3, 4, \dots, n, n+1, \dots\}.$$

Wie kann man mit natürlichen Zahlen rechnen? Man darf uneingeschränkt addieren und multiplizieren. Man sagt N ist bezüglich der Addition und Multiplikation abgeschlossen. Alle anderen Rechenoperationen sind nicht uneingeschränkt durchführbar, da negative Zahlen nicht unter die Natürlichen fallen. Eine Untermenge von N ist die Menge der **Primzahlen**, definiert als:

$$\mathsf{P} = \{1, 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, ...\}.$$

Primzahlen sind nicht durch 1 und sich selbst teilbar!

1.2 Die ganzen Zahlen Z

Erweitert man den Zahlenbereich der natürlichen Zahlen mit negativen Zahlen, hat man die ganzen Zahlen:

$$\mathsf{Z} = \{\dots -3, -2, -1, 0, 1, 2, 3, \dots\}$$

Nun kann man auch uneingeschränkt subtrahieren.

1.3 Rationale und irrationale Zahlen Q, $\mathsf{R}\backslash\mathsf{Q}$

Will man uneingeschränkt dividieren, braucht man Bruchzahlen:

- $\mathbb{Q}_+$ enthält alle positiven Brüche: $\mathbb{Q}_+ = \{\frac{a}{b} |$ a,b sei eine natürliche Zahl und $b \neq 0\}$.

Nimmt man negative Brüche dazu, hat man die rationalen Zahlen.

- $\mathbb{Q} = \{\frac{a}{b} |$ a,b sei eine ganze Zahl und $b \neq 0\}$.
- In $\mathbb{Q}$ sind alle Grundrechenarten uneingeschränkt ausführbar.
- $\mathbb{Q}$ enthält alle positiven und negativen Brüche, sowie alle abbrechenden Dezimalbrüche $z.B. -3,75 - 3,75$ und periodischen Dezimalbrüche (z.B. 0,66666...0,66666...).

Bei den rationalen Zahlen ist eines nicht vollständig erlaubt: Das Wurzelziehen, da es zu unendlichen Zahlen führt, die sich nicht als Bruch darstellen lassen, den **irrationalen Zahlen**: $\sqrt{2} = 1,41421356...$

1.4 Die reellen Zahlen $\mathbb{R}$

Vereint man die rationalen und die irrationalen Zahlen, erhält man die reellen Zahlen $\mathbb{R}$. Jedoch: Wurzelziehen aus negativen Zahlen ist nicht definiert. Zum Beispiel ist $\sqrt{-4}$ nicht definiert. Solche Zahlen sind nicht in $\mathbb{R}$ enthalten.

1.5 Komplexe Zahlen $\mathbb{C}$

Eine komplexe Zahl - z - ist beschreibar als reelles Zahlenpaar:

$$x + iy | x, y \in \mathbb{R}, i = \sqrt{-1}$$

Wichtig ist das mit der imaginären Zahl 1.6.3 das ziehen von negativen Wurzeln möglich ist. Man kann für c $\in \mathbb{C}$ auch (x,y) schreiben, wobei x der Realteil und y der Imaginärteil ist. So kann die Menge der komplexen Zahlen ($\mathbb{C}$) auch in Form der reellen Zahlenpaare ($\mathbb{R}^2$) geometrisch auf der komplexen Ebene (auch Gaußsche Zahlenebene) dargestellt werden, siehe Bild.
Analog definiert man Addition wiefolgt:

$$z_1 + z_2 = (x_1 + x_2) + (y_1 + y_2) \cdot i$$
$$z_1 - z_2 = (x_1 - x_2) + (y_1 - y_2) \cdot i$$

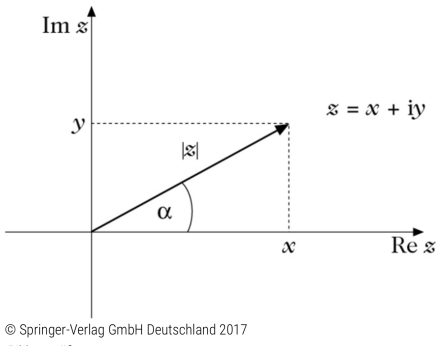

Abbildung 1: Gaußsche Zahlenebene; $c \in \mathsf{C}$ als reelles Zahlenpaar (x, y)

und Multiplikation so:

$$z_1 * z_2 = (x_1 + x_2) * (y_1 - y_2) \cdot i = x_1 y_1 + x_1 y_2 \cdot i + y_1 x_2 \cdot i + x_2 y_2 \cdot i^2$$
$$z_1 * z_2 = (x_y y_1 - x_2 y_2) + (x_1 y_2 + y_1 - x_2) \cdot i$$

Ist z eine komplexe Zahl, so ist z^* die komplexe Konjugation. In der Darstellung unten würde der Realteil(z) hierbei gespiegelt. Als Spezialfall gilt: $i^* = -i$, sodass $z^* = x - y \cdot i$. Multipliziert mit dem komplexen Konjugat ergibt sich der Betrag von z.

Den Betrag einer komplexen Zahl ergibt sich durch:

$$|z| = \sqrt{x^2 + y^2}$$

Die Division ist definiert als: $\frac{z_1}{z_2} = \frac{z_1}{z_2} \cdot \frac{z_2*}{z_2*}$. Das ist nicht sehr angenehm und kann vermieden werden, wenn es nicht unbedingt sein muss. Sollte kein Weg daran vorbeiführen, so multipliere die Faktorden des Nenners und des Zählers je für sich und vereinfache so gut es gut mit der Definition von i.

Eine andere Darstellung ist mithilfe von Polarkoordinaten (3.2) möglich:

$$z = a + i * b \ |a, b, r \in \mathsf{R}, \theta \in [0, 2\pi] \Leftrightarrow z = r * (\cos(\theta) + i * \sin(\theta)) \Leftrightarrow r * e^{i}\theta.$$

1.6 Besondere Zahlen

1.6.1 π - Die Kreiszahl

3.1415926535.....damit beginnt die irrationale Zahl π[1].Pi beschreibt das Verhältnis von Umfang zu Durchmesser eines Kreises. Es gibt viele Formeln, in denen π die tragende Rolle spielt:

Kreisumfang
$U = \pi \cdot d = 2 \cdot \pi \cdot r$
Kreisfläche
$F = \pi \cdot r^2$
Volumen
$V = \frac{4}{3} \cdot \pi \cdot r^3$

1.6.2 e - Euler's Zahl

Die Eulersche Zahl ist die Basis des natürlichen Logarithmus 2.7, also $ln(e) = 1$. Die Eulersche Zahl kann beschrieben werden durch $e = 2,71828...$, aber ähnlich wie für π gibt es keine exakte Lösung. Die Eulersche Zahl wurde nach dem Schweizer Mathematiker und Physiker Leonhard Euler $1707 - 1783$ benannt.

[1] 8979323846 2643383279 5028841971 6939937510 5820974944 5923078164 0628620899 8628034825 3421170679 8214808651 3282306647 0938446095 5058223172 5359408128 4811174502 8410270193 8521105559 6446229489 5493038196 4428810975 6659334461 2847564823 3786783165 2712019091 4564856692 3460348610 4543266482 1339360726 0249141273 7245870066 0631558817 4881520920 9628292540 9171536436 7892590360 0113305305 4882046652 1384146951 9415116094 3305727036 5759591953 0921861173 8193261179 3105118548 0744623799 6274956735 1885752724 8912279381 8301194912

Unter anderem beinhaltet die natürliche Exponentialfunktion $f(x) = e^x$ die Eulersche Zahl. Für uns ist e vor allem wichtig, weil man anhand von der Eulerschen Formel $e^{j\phi} = \cos(\phi) + j * \sin(\phi)$ die komplexen Zahlen in der Exponentialform darstellen kann, was die Berechnungen erheblich erleichtert.

Außerdem gilt $\partial e^t / \partial t = e^t$ was bedeutet, dass die Position von e gleich der Ableitung, also der Geschwindigkeitsänderung ist. Außerdem ist dieser Zusammenhang besonders schön:

$$\lim_{n\to\infty} \left(1 + \tfrac{1}{n}\right)^n = 2,718281828459...$$

Damit is e der Grenzwert (7.5) der (exponentiell) steigenden Zahlenfolge.
Weitere wichtige Exponentialfunktionen, die in der Natur auftreten sind:

$$y = e - t/\tau \quad \text{fallend}$$

$$y = (1 - e - t/\tau) \quad \text{steigend}$$

$$y = (1 - 2e - t/\tau) \quad \text{steigend, mit negativem Anfangswert.}$$

1.6.3 i - die imaginäre Zahl

i ist definiet als die Lösung der Gleichung $\sqrt{-1}$, daher $i \cdot i = -1$. Das kann man nicht ändern, also akzeptiere es besser und merke die folgenden Regeln:

$$i^2 = -1$$

$$i^3 = -i$$

$$i^4 = 1$$

$$i^{-1} = -i.$$

2 Grundlagen & Rechengesetze

Eine Verknüpfung lässt sich definieren als die Art und Weise wie 2 Objekte ein Drittes definieren. Die Verknüpfung wird abstrakt mit '$\circ$' ausgedrückt.

2.1 Kommutativgesetz

Die Reihenfolge der Verknüpfung spielt keine Rolle, e.g. $2 \circ 3 = 3 \circ 2$.

2.2 Assoziativgesetz

Die Verknüpfung dreier Zahlen hängt nicht davon ab, wie wir sie gruppieren, e.g. $(2 \circ 3) \circ 4 = 2 \circ (3 \circ 4)$.

2.3 Distributivgesetz

Der Umgang mit Klammern hängt vom Zahlenraum und der Art der Verknüpfung ab. Für Addition sowie Multiplikation für die Zahlenmenge R seien beide Verknüpfungen distributiv, sodass gilt: $2 \odot (3 \oplus 4) = 2 \odot 3 \oplus 2 \odot 4$.

2.4 Die binomischen Formeln

(a) : $(a+b)^2 = a^2 + 2ab + b^2$

(b) : $(a-b)^2 = a^2 - 2ab + b^2$

(c) : $(a+b) \cdot (a-b) = a^2 - b^2$

2.5 Potenzen

Eine Potenz ist eine abgekürzte Schreibweise der Multiplikation mit sich selber.

- $a^0 = 1$
- $a^1 = a$

- $a^{-1} = \frac{1}{a}$
- $a^{-n} = \frac{1}{a^n}$
- $a^n = \frac{1}{a^{-n}}$
- $a^p * a^q = a^{p+q}$
- $a^p : a^q = a^{p-q}$
- $a^q * b^q = (a * b)^q$
- $a^q : b^q = (a : b)^q$
- $(a^p)^q = a^{p*q}$
- $\frac{a^m}{a^n} = a^{m-n}$
- $\frac{a^n}{b^n} = (\frac{a}{b})^n$.
- $(\frac{a}{b})^{-n} = (\frac{b}{a})^n$

2.6 Wurzeln

Die Wurzel einer Zahl ergibt mit sich selbst multipliziert die Wurzel. Im Normalfall handelt es sich um eine qudratische Multiplikation, aber auch höhere Wurzlen - e.g. die Kubische $\sqrt[3]{x}$ - ist nicht undenkbar. Als Potenz geschrieben ist die qudratische Wurzel $\sqrt{x} = x^{\frac{1}{2}}$ und $\sqrt[n]{x} = x^{\frac{1}{n}}$, z.B. $\sqrt[3]{125} = 125^{\frac{1}{3}}$. Haben wir eine Potenz mit einer Lösung, dann gilt

$$x^n = a \Leftrightarrow x = \sqrt[n]{a}$$

wie in $3^4 = 81 \equiv \sqrt[4]{81} = 3$. Randnotiz: Damit ist die n-te Wurzel die inverse Funktion (7.1.3) zu der Funktion: x^n.

2.7 Logarithmus

Frage: Mit welcher Zahl muss ich a potenzieren um y zu bekommen? Geschrieben als $\log_a(x) = y$. Beachte, dass das Logarithmieren von Null und negativen Zahlen nicht definiert ist!

Außerdem ist der Logarithmus die Umkehrfunktion zur Exponentialfunktion:

$$f(x) = a^x = y,\ f^{-1}(y) = \log_a(y) = x$$

Also liefert der log. den Exponenten der Exponentialfunktion zur Basis a. Haben wir die Exponentialfunktion $f(x) = e^x$ mit e als Basis, so ist der Logarithmus Naturalis (ln) = f^{-1}. Der Logarithmus hat ein paar eigene Rechenregeln:

- $\log_a(1) = 0$
- $\log_a(a) = 1$
- $\log_a(p * q) = \log_a(p) + \log_a(q)$
- $\log_a(\frac{p}{q}) = \log_a(p) - \log_a(q)$
- $\log_a(p^q) = q * \log_a(p)$
- $\log_a(\sqrt[n]{p}) = \frac{\log_a(p)}{n}$
- $\log_a(p) = \frac{\log_b(p)}{\log_b(a)}$

Logarithmische Skalierungen ist hilfreich, wenn die Daten stark variieren oder wenn die Verhältnisunterschiede zwischen Werten von Bedeutung sind, denn logarithmische Skalierung ist dann angenehmer um Strukturen zu erkennen.

3 Geometrie

Geometrie ist der Bereich der Mathematik, der sich mit Formen, Größen, Eigenschaften des Raumes und den Beziehungen zwischen verschiedenen Formen beschäftigt. Sie umfasst Konzepte wie Punkte, Linien, Flächen und Volumen. Aus der Schule ist allgemein die *euklidische Geometrie* bekannt, es gibt aber auch noch andere! Im Folgenden soll es um diese, Topologie und Trignonometrie gehen.

3.1 Trignonometrie

Trignonometrie ist das Studium von Dreiecken und den Beziehungen zwischen ihren Seiten und Winkeln.

Ein Einheitskreis ist definiert als ein Kreis mit radius 1. Das Interessante hierbei ist, dass sich die x- und y-Koordinate mithilfe von sinus und cosinus bestimmen lassen, abhängig vom Winkel. Man könnte auch sagen:

$$\cos(\theta) = \frac{x}{1}$$

und

$$\sin(\theta) = \frac{y}{1}$$

Ist der *radius* $\neq 1$ spielt der Wert r eine Rolle da die Division mit einem anderen Wert x und y verschiebt. Schließlich gilt auch für den radius (oder die Hypothenuse):

$$\tan(\theta) = \frac{y}{x}.$$

Die gewöhnliche Definition von sinus, cosinus und tangens, die sich auf Dreiecke bezieht, ist sehr ähnlich und hat demenstprechend Gemeinsamkeiten, die sich darin finden lassen, dass wir diese Dreicke einfach in den Einheitskreis einsetzen (siehe Abbildung 3.1).

$$\begin{aligned}
\sin(\theta) &= \frac{y}{z} = \frac{\text{Gegenkathete}}{\text{Hypothenuse}} \\
\cos(\theta) &= \frac{x}{z} = \frac{\text{Ankathete}}{\text{Hypothenuse}} \\
\tan(\theta) &= \frac{y}{x} = \frac{\text{Gegenkathete}}{\text{Ankathete}}
\end{aligned}$$

Fassen wir zusammen: Die expliziten Formeln zeigen, dass **Sinus**, **Cosinus** und

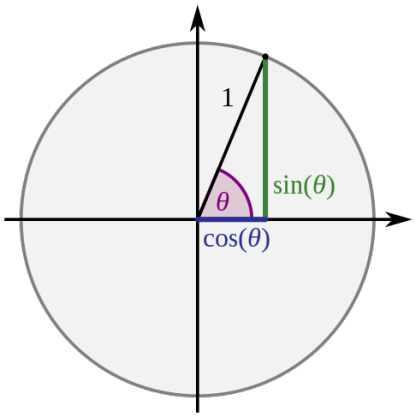

Abbildung 2: Einheitskreis von der Wikipedia Seite. Vom 13. Oktober, 2023.

Tangens einen Winkel nutzen um das Verhältnis zweier zueinander liegender Seite zu bestimmen. Da für den Einheitskreis der Divident = 1 ist, bilden sinus und cosinus so den x- und y-Wert ab, gegeben die X- und Y-Achse ist so orientiert. Dieses Bild reflektiert sehr schön, warum folgendes gilt:

$$\begin{aligned}
\sin(0) &= 0 \\
\cos(0) &= 1 \\
\sin(90) &= 1 \\
\cos(90) &= 0 \\
\sin(-\theta) &= -\sin(\theta) \\
\cos(-\theta) &= \cos(\theta)
\end{aligned}$$

Das ist auch der Grund, warum diese Funktionen (cos und sin) periodisch sind. Für Winkel > 360 kann man genauso gut Modulo rechnen und kommt wieder bei 0, bzw. 1 an.

3.2 Polarkoodinaten

In Polarkoordinaten wird ein Punkt durch Angabe des Abstands r zum Koordinatenursprung und durch einen Winkel θ (auch Azimut genannt) bezüglich einer vorgegebenen Achse (z.B. der X-Achse) beschrieben. Das Zahlenpaar (r, θ) wird als Polarkoordinaten des Punktes bezeichnet. Umrechnen lassen sich die Koordinaten in das andere System wiefolgt:

Von Polarkoodinaten in kartesische Koordinaten:

$$x = r \cdot \cos(\theta) \text{ and } y = r \cdot \sin(\theta)$$

Von kartesischen Koordinaten in Polarkoordinaten:

Hier müssen wir zwei Dinge bestimmen. Der radius ist mittels Phytagoras gegeben, i.e. $r = \sqrt{x^2 + y^2}$. Ist der radius = 0, sind alle möglichen θ mögliche, daher gilt als (praktische) Lösung $\theta = 0$.

Außerdem sollte der der Wertebereichs des Winkels mit einem haloffenen Intervall gegeben sein, damit die Transformation eindeutig ist, i.e. $[0, 2\pi)$. Der Winkel θ muss nach unterschiedlichen Fällen behandelt werden:

$$x > 0, y \geq 0 : \theta = \arctan(\frac{y}{x})$$

$$x > 0, y < 0 : \theta = \arctan(\frac{y}{x}) + 2\pi$$

$$x < 0 : \theta = \arctan(\frac{y}{x}) + \pi$$

$$x = 0, y > 0 : \theta = 90\,\text{deg}$$

$$x = 0, y < 0 : \theta = 270$$

An dieser Stelle sei eine Besonderheit genannt. Wir wissen, dass für die Um-

rechnung $f_{P\to CK} = r * sin(\theta) = y, r * cos(\theta) = x$. Da sich r aus der Formel für den Tangens herauskürzen lässt, erhalten wir $tan(\theta) = \frac{\sin(\theta)}{\cos(\theta)}$. Anders formuliert beudetet das folgendes: Über den Winkel θ lässt sich mit $\tan(\theta)$ das Verhältnis $y : x$ berechnen. Daher können wir den Winkel θ mit der inversen zu berechenen. $f^{-1} \equiv \tan^{-1}(\frac{y}{x})$ übersetzt sich zu $\arctan(\frac{y}{x})$, woraus folgt: $\arctan(Gegenkahete/Ankathete) = \theta$.[2]

Werden Kreiskoordinaten mit einer dritten Koordinate ergänzt, ergeben sich räumliche Polarkoordinaten, wie z. B. Kugelkoordinaten oder Zylinderkoordinaten.

3.2.1 Komplexe Zahlen mittels Polarkoodinaten

$z \in \mathsf{C} = a + b \cdot i \equiv z = |z| \cdot e^{i\theta} = |z|(\cos(\theta) + i \cdot \sin(\theta)) = |z| \cdot \cos(\theta) + |z| \cdot i \cdot \sin(\theta)$

3.3 Bogenmaß und Radian

Meistens werden Winkel in Grad angegeben. Aber ein Winkel von bsp. 45 Grad kann auch in Bogenmaß ($\frac{1}{4}\pi \approx 0,79$) angegeben werden. Es gilt: $360\,\text{deg} = 2\pi$, $1\pi = 180\,\text{deg}$ und $1,5\pi = 270\,\text{deg}$. Zu jedem Mittelpunktswinkel am Einheitskreis gehört ein Kreisbogen auf dem Einheitskreis. Die Länge des Kreisbogens ist ein Maß für die Größe des Winkels. Dieses wird als Bogenmaß bezeichnet und trägt die Einheit *Radiant*, abgekürzt *rad*.

$$\boxed{rad = \frac{\text{deg}}{180\,\text{deg}} * \pi \text{ und } \text{deg} = \frac{rad}{\pi} * 180\,\text{deg}}$$

3.4 Phytagoras

$a^2 + b^2 = c^2$

[2] θ is in radian.

4 Statistik

4.1 Die Termini der Statistik

Was man kennen sollte ...

4.1.1 Durchschnitt

Gibt es eine endliche Datenmenge D, die aus Stichproben der Form $x \sim P_X$ besteht, ist der Durchschnitt der arithmetische Mittelwert:

$$\bar{x} := \frac{1}{|D|} \sum_{x \in D} x$$

Der Durchschnitt ist also ein Schätzwert für den unbekannten Erwartungswert einer Verteilung von dem die Stichproben kommen.

4.1.2 Modus

Das am häufigsten vorkommende Element einer Datenmege:

$$Modus(D) = 1, 2, 3, 6, 7, 8, 3, 2, 2 = 2$$

4.1.3 Median

Jenes Element einer Datenmenge, zu dem jeweils 50% größer und 50% kleiner sind:

$$Median(D) = 1, 2, 3, 4, 5, 6, 7 = 4$$

4.1.4 Erwartungswert

Der Erwartungswert ist der förmliche Begriff für den Durschnitt einer Variablen, mit dem Unterschied, das der Erwartungswert prinzipiell ins Unendliche geht. Ausgedrückt wird das mit:

$$E[X] := \mu_X := \int_{x \in \Omega} x dP_X(x)$$

4.1.5 Varianz und Kovarianz

Die Varianz misst die erwartete Abweichung einer Variablen vom Erwartungswert. Mit der euklidischen Distanzt als Metrik (hoch 2) ergibt sich folgende Formel:

$$Var(X) = E[(X - E[X])^2]$$

Ein positiver Kovarianzwert bedeutet, dass hohe Werte für die erste Koordinate des Paares mit hohen Werten für die zweite Koordinate des Paares korrespondieren. Ein negativer Kovarianzwert drückt eine Korrespondenz von hohen Werten in der ersten Koordinate mit niedrigen Werten in der zweiten Koordinate aus. Ein Wert von 0 bedeutet, dass die Werte der beiden Koordinaten nicht miteinander korrespondieren. Gegeben eine Menge von n Datenpunkten in einem d-dimensionalen Datenspace als eine $n \times d$-Matrix D, wird die Kovarianzmatrix berechnet als

$$C = \frac{1}{n-1}(D - \mu) \cdot (D - \mu)^T$$

wobei μ den Mittelwertvektor der Datensatzes bezeichnet.

4.1.6 Standardabweichung

Die Standardabweichung ist die Wurzel der Varianz. Sie ist stets angegeben in der selber Größe wie das zu messende Merkmal.

$$\sigma = \sqrt[2]{Var(X)}$$

4.1.7 Korrelation

Die Korrelation ist eine statistische Technik, die verwendet wird, um die Beziehung zwischen zwei Variablen zu bestimmen. Sie misst, wie stark zwei Variablen miteinander zusammenhängen. Der Korrelationskoeffizient variiert zwischen -1 und +1. Ein Wert nahe +1 zeigt eine starke positive Korrelation, während ein Wert nahe -1 eine starke negative Korrelation anzeigt. Ein Wert nahe Null bedeutet, dass zwischen den Variablen keine Korrelation besteht. Die Formel zur Berechnung der Korrelation zwischen zwei Variablen X und Y ist:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2(y_i - \bar{y})^2}}$$

wobei x_i und y_i die einzelnen Datenpunkte, $\bar{x}$ und $\bar{y}$ die Durchschnittswerte und n die Anzahl der Datenpunkte sind.

4.2 Statistische Verfahren

4.2.1 MLE

4.2.2 Pearson's Chi-Quadrat-Test

Der Pearson's Chi-Quadrat-Test ist ein statistisches Verfahren, das verwendet wird, um festzustellen, ob ein beobachteter Datensatz mit einer erwarteten Verteilung übereinstimmt. Es basiert auf der Berechnung der Chi-Quadrat-Statistik, die die Abweichung zwischen den beobachteten und erwarteten Häufigkeiten misst. Der Test verwendet die Nullhypothese, dass es keine signifikante Abweichung gibt, und die Alternative, dass es eine signifikante Abweichung gibt. Der Chi-Quadrat-Test berechnet die Chi-Quadrat-Statistik als Summe der quadrierten Abweichungen zwischen den beobachteten und erwarteten Häufigkeiten, dividiert durch die erwarteten Häufigkeiten. Die resultierende Statistik folgt einer Chi-Quadrat-Verteilung. Die Freiheitsgrade des Chi-Quadrat-Tests werden durch die Anzahl der Kategorien im Datensatz minus 1 bestimmt. Sie geben an, wie viele Freiheitsgrade zur Verfügung stehen, um die Daten zu variieren, während die Nullhypothese aufrechterhalten wird. Der Pearson's Chi-Quadrat-Test ist ein weit verbreitetes Verfahren in der statistischen Analyse, insbesondere in der Kontingenztafelanalyse und der Anpassungstestung.

4.2.3 T-Test

4.2.4 Sampling

4.2.5 Varianzanalyse

Anova

5 Wahrscheinlichkeit

5.1 Abhängige und unabhängige Variable

Die unabhängige Variable ist der Faktor, den ein Forscher manipuliert, während die abhängige Variable das Ergebnis ist, das gemessen wird, um den Einfluss der unabhängigen Variable zu sehen.

5.2 Kolmogorow' Axiome

Kolmogorow's drei Axiome - so heißen die Grundsätze einer Theorie - sind die bekannteste Beschreibung der grundlegenden Eigenschaften der Wahrscheinlichkeitsrechnung. Seien $\Omega = \omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6$ die Ergebnismenge eines Zufallsexperiments, A und B Teilmengen von Ω und P eine Funktion, die jedem A eine reelle Zahl zwischen 0 und 1 zuordnet. P(A) wird Wahrscheinlichkeit genannt, falls folgende drei Bedingungen (Axiome) erfüllt werden:

1. $P(A) \geq 0$ Diese Bedingung besagt, dass jede Wahrscheinlichkeit für das Eintreffen einer Teilmenge von Ω (Ereignis) nicht negativ ist. Man nennt diese Eigenschaft daher auch: Nichtnegativität

2. $P(\Omega) = 1$ Das zweite Axiom bringt eine weitere Eingrenzung des Wertebereichs von der Funktion P. Mit Axiom 1 und 2 darf P(A) mit beliebigem A minimal Wert 0 und maximal Wert 1 annehmen.

3. $P(A \cup B) = P(A) + P(B)$ Dies bedeutet also, dass für kein Ergebnis beide Ereignisse erfüllt werden. A und B nennt man in diesem Fall auch disjunkt.

5.3 Wahrscheinlichkeitskonzepte

5.3.1 Wahrscheinlichkeitsformeln

In der Wahrscheinlichkeitstheorie gibt es verschiedene Konzepte, die uns helfen, die Beziehungen zwischen verschiedenen Ereignissen oder Variablen zu verstehen. Dazu gehören die gemeinsame Wahrscheinlichkeit, die Grenzwahrscheinlichkeit und die bedingte Wahrscheinlichkeit. Die gemeinsame Wahrscheinlichkeit $P(X, Y)$ gibt die Wahrscheinlichkeit an, dass sowohl Ereignis X als auch

Ereignis Y eintreten. Sie kann berechnet werden, indem man die bedingte Wahrscheinlichkeit eines Ereignisses, gegeben das andere, mit der Wahrscheinlichkeit des anderen Ereignisses multipliziert:

$$P(X,Y) = P(X|Y) \cdot P(Y) = P(Y|X) \cdot P(X) \tag{1}$$

Die Grenzwahrscheinlichkeit (im Englischen **marginal** oder **total probability**) $P(X)$ ist die Wahrscheinlichkeit, dass Ereignis X eintritt, unabhängig von allen anderen Ereignissen. Sie kann berechnet werden, indem man über alle möglichen Werte von Y die gemeinsamen Wahrscheinlichkeiten summiert:

$$P(X) = \sum_y P(X, Y = y) = \sum_y P(X|Y = y) \cdot P(Y = y) \tag{2}$$

Die bedingte Wahrscheinlichkeit $P(X|Y)$ gibt die Wahrscheinlichkeit an, dass Ereignis X eintritt, gegeben, dass Ereignis Y eingetreten ist. Sie kann berechnet werden, indem man die gemeinsame Wahrscheinlichkeit von X und Y durch die Wahrscheinlichkeit von Y teilt:

$$P(X|Y) = \frac{P(X,Y)}{P(Y)} \tag{3}$$

Wenn eine dritte Variable Z vorhanden ist, können diese Formeln erweitert werden, um die bedingten und gemeinsamen Wahrscheinlichkeiten unter Berücksichtigung von Z zu berechnen. Zum Beispiel ist die gemeinsame Wahrscheinlichkeit von X und Y, gegeben Z, definiert als:

$$P(X,Y|Z) = P(X|Y,Z) \cdot P(Y|Z) = P(Y|X,Z) \cdot P(X|Z) \tag{4}$$

Und die bedingte Wahrscheinlichkeit von X, gegeben Y und Z, ist definiert als:

$$P(X|Y,Z) = \frac{P(X,Y|Z)}{P(Y|Z)} \tag{5}$$

5.3.2 Bedingte Unabhängigkeit

Bedingte Unabhängigkeit tritt auf, wenn das Auftreten eines Ereignisses A keine Auswirkungen auf die Wahrscheinlichkeit des Auftretens eines Ereignisses B hat, gegeben das Ereignis C. In anderen Worten, A und B sind bedingt unabhängig, wenn die bedingte Wahrscheinlichkeit von B gegeben A und C gleich

der bedingten Wahrscheinlichkeit von B gegeben C ist. Dies wird mathematisch durch die Gleichung

$$P(B|A,C) = P(B|C) \tag{6}$$

ausgedrückt. Wenn A und B unabhängig sind gegeben C, so schreibt man $P(A,B|C) = P(A|C) \cdot P(B|C)$

5.3.3 Unabhängigkeit von Variablen

Zwei Variablen X und Y sind unabhängig, wenn das Auftreten einer Variable keinen Einfluss auf die Wahrscheinlichkeit des Auftretens der anderen Variable hat. Mathematisch ausgedrückt bedeutet dies, dass die gemeinsame Wahrscheinlichkeit von X und Y gleich dem Produkt der Einzelwahrscheinlichkeiten von X und Y ist: $P(X,Y) = P(X) \cdot P(Y)$. Wenn X und Y unabhängig sind, gilt auch $P(X|Y) = P(X)$ und $P(Y|X) = P(Y)$.

5.4 Wahrscheinlichkeitsverteilungen

Die Wahrscheinlichkeitsverteilung wird in zwei Arten unterteilt, die diskrete und die stetige Zufallsvariable. Diese sind dann jeweils noch mehrmals in verschiedene Kategorien unterteilt. Da es sich bei den Wahrscheinlichkeitsverteilungen um Funktionen handelt, gibt es immer einen Funktionswert und einen x-Wert. Die **Diskrete Zufallsvariable** zeichnet sich dadurch aus, dass sie eine begrenzte, abzählbare Anzahl an möglichen Ausprägungen hat. Beispiele dafür sind der Münz- oder Würfelwurf. Beide haben nur eine begrenzte Anzahl an möglichen Ausprägungen, der Münzwurf hat zum Beispiel zwei und der Würfelwurf hat dafür 6 Ausprägungen. Die **kontinuierliche oder stetige Zufallsvariable** dagegen hat eine unbegrenzte Anzahl an möglichen Ausprägungen. Als Beispiel kann man dafür die Haarlänge nehmen. Theoretisch könnte man sagen, dass es von keinen Haaren, bis zu den weltweit längsten Haaren eine begrenzte Anzahl an Zentimetern gibt. Jedoch, wenn man die Länge in immer genaueren Einheiten angeben würdest, hätte man unendlich viele verschiedene Haarlängen auf der Welt, zumal es keine festgelegte Grenze für das Haarwachstum gibt. Die **Wahrscheinlichkeits Dichtefunktion** ist eine Wahrscheinlichkeitsfunktion, die kontinuierlich ist. Die Summe der Wahrscheinlichkeiten aller Ergebnisse ist in einem Zufallsexperiment immer gleich 1. In einer stetigen Zufallsverteilung muss die 1 auf unendlich viele Ausprägungen verteilt werden. Das führt dazu,

dass die Wahrscheinlichkeit für eine einzelne Ausprägung praktisch gegen 0 geht. Ebendarum lässt sich in der Dichtefunktion nicht die Wahrscheinlichkeit einer einzelnen Ausprägung ableiten. Um aber trotzdem an ein Ergebnis zu gelangen, kannst du über mehrere Ausprägungen hinweg *integrieren* und erhältst so die Wahrscheinlichkeit für diese Menge an Ausprägungen (Verteilungsfunktion). Genannt werden hier der Erwartungswert E, die Varianz V (Standardabweichung ist $\sqrt{V}$), die Funktion der Wahrscheinlichkeitsverteilung, aber nicht die Verteilungsfunktion.

5.4.1 Normalverteilung

Die Normalverteilung wird auch Gaußsche Glockenkurve genannt. Die beiden Parameter (μ und σ) geben Mittelwert sowie Standardabweichung der Normalverteilung an. Der zentrale Grenzwertsatz besagt, dass unter bestimmten allgemeinen Voraussetzungen die Summe aus n unabhängigen, identisch verteilten Zufallsvariablen wiederum normalverteilt ist. Als Beispiel se nehmen wir den Wurf von n fairen Würfeln: Wenn man man nur einen Würfel wirft, so ist jede Augenzahl gleich wahrscheinlich. Wirft man hingegen n-viele Würfel, so wird die mittlere Augenzahl durch die Normalverteilung beschrieben. Daher ist die Normalverteilung die Wichtigste, da natürliche Phänomene mit ausreichend großem n sich ihr annähern. Die Formel zur Berechnung der Verteilung lautet $f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$. Es gilt: $E = \mu$ und $V = \sigma^2$, bzw. σ (Standardabweichung). Die gesamte Fläche, die von der Kurve der Normalverteilung eingeschlossen wird ist stets 1. Ist $\mu = 0$ und $\sigma = 1$ spricht man von der Standardnormalverteilung, die durch eine vereinfachte Gleichung (da $\mu = 0$ wegfällt und $\sigma = 1$) beschrieben wird: $\frac{1}{2\pi}e^{-\frac{1}{2}x^2}$. Der Vorfaktor stellt sicher, dass die gesamte Fläche unter der Kurve (und damit auch das Integral von -inf bis inf) eine Fläche von genau 1 hat. Die $\frac{1}{2}$ im Exponenten der e-Funktion gibt der Normalverteilung eine Einheitsvarianz. Jede Normalverteilung ist eine Variante der Standardnormalverteilung mit gestreckter Standardabweichung ($\frac{1}{\sigma}$) und *z-transfomiertem* $\frac{x-\mu_x}{\sigma}$. Geschrieben wird die Normalverteilung für gewöhnlich so: $\mathcal{N}(\mu, \sigma^2)$.

5.4.2 Gemischte Verteilung

Eine Mischverteilung besteht aus mehreren Untermengen die gemeinsame eine große Verteilung bilden. Ein solcher Ansatz kann verwendet werden, um ei-

ne große Population mit verschiedenen Unterpopulationen zu modellieren, von denen jede individuelle Eigenschaften hat. Formal bietet man für jede Unterpopulation z eine spezifische Verteilung $P(X \mid Z = z)$ an. Diese werden gemischt entsprechend der Wahrscheinlichkeit $P(Z = z)$, ein Individuum aus dieser Unterpopulation auszuwählen, d.h.

$$P(X = x) = \sum_z P(Z = z) \cdot P(X = x \mid Z = z)$$

5.4.3 Exponentialverteilung

Die Exponentialverteilung ist eine kontinuierliche Verteilung, die zur Modellierung der Dauer zufälliger Zeitintervalle genutzt wird. Der Parameter λ steht für die Zahl der erwarteten *Ereignisse* pro Zeitintervall. Als Beispiele nehmen wir hier die Länge eines Telefongesprächs oder der radioaktive Zerfall. Die Verteilung lässt keine negativen Werte zu, da negative Zeiten sinnlos sind. Sie wird in der Statistik häufig mit $\exp(\lambda)$ abgekürzt. Die Dichtefunktion ist folgendermaßen definiert:

$$f_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}.$$

Der Erwartungswert E ist defniert als $\frac{1}{\lambda}$, die Varianz V als $\frac{1}{\lambda^2}$. Der Modus (der Wert, bei dem die Wahrscheinlichkeit am größten ist) liegt bei dieser Dichtefunktion bei $x_{mod} = 0$. Möchte man die Wahrscheinlichkeit des Eintretens eines Ereignisses berechnen, so nutzt man dafür idealerweise die Verteilungsfunktion $F(x)$, die das Integral bis zu einem Wert x bildet. So entsteht eine akkumulierte Wahrscheinlichkeit $P(X \leq x)$.Oft ist die tatsächliche Verteilung keine Exponentialverteilung, jedoch ist die Exponentialverteilung einfach zu handhaben und wird zur Vereinfachung angewandt. Sie ist anwendbar, wenn ein Poisson-Prozess vorliegt, also die poissonschen Annahmen erfüllt sind. Die Exponentialverteilung ist ein Teil der viel größeren und allgemeineren Exponentialfamilie, einer Klasse von Wahrscheinlichkeitsmaßen, die sich durch eine leichte Handhabbarkeit auszeichnen.

5.4.4 Gleichverteilung

Der französische Mathematiker Pierre Simon de Laplace (1749 bis 1827) untersuchte als einer der Ersten intensiv Zufallsexperimente, bei denen angenommen werden kann, dass jedes seiner Ergebnisse mit der gleichen Wahrscheinlichkeit eintritt. Zufallsexperimente mit Gleichverteilung heißen Laplace-Experimente. Die Gleichverteilung ist ein Sonderfall unter den Wahrscheinlichkeitsverteilungen, denn sie existiert sowohl als *stetige* als auch als *diskrete* Verteilung. Hier seien einmal kurz die Formeln für deren Berechnung genannt. Zuerst für den Fall einer diskreten Verteilung: $f(x) = \frac{1}{n}, E(x) = \frac{n+1}{2}$ und $V(x) = \frac{1}{n}\sum_0^n (x_i - \mu)^2$. Und für eine stetige Verteilung: $f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$. a und b seien die Grenzen eines Intervalls, die x beinhalten. Da für alle x die gleiche Wahrscheinlichkeit gilt, ist diese von den Grenzen des Intervalls abhängig. $E(x) = \frac{a+b}{2} und V(x) = \frac{1}{12} \cdot (b-a)^2$.

5.4.5 Poissonverteilung

Die Poisson-Verteilung ist eine diskrete Wahrscheinlichkeitsverteilung, welche die Verteilung von Zählgrößen beschreibt. Oder mit anderen Worten: Wie oft tritt ein bestimmtes, zählbares Ereignis ein, wenn man es sehr oft wiederholt? Der Parameter gibt hierbei die mittlere Ereignisrate an.Die Wahrscheinlichkeit für die Zufallsvariable X der Poisson-Verteilung wird durch folgende Formel berechnet: $\frac{\lambda^x}{x!} \cdot e^{-\lambda}, x \in N_0$. Es besteht ein Zusammenhang zwischen Exponentialverteilung und Poissonverteilung. Beide betrachten denselben Sachverhalt aus verschiedenen Perspektiven. Die Exponentialverteilung gibt an, wie die Wahrscheinlichkeit der Dauer verschiedener Vorgänge verteilt ist. Die Poissonverteilung zählt, wie oft die gezählten Ereignisse in einem festgelegten Intervall auftreten. Ausgehend von der Exponentialverteilung soll ermittelt werden, wie die Wahrscheinlichkeit ist, dass genau n Ereignisse in einem Zeitintervall von t auftreten. Wie sich zeigen wird, ist das Ergebnis die Poissonverteilung. Da der Binomialkoeffiziert bei größeren Werten nur unter erhöhtem Rechenaufwand zu berechnen ist, kann man die Poisson-Verteilung benutzen, um die Binomialverteilung anzunähern. Man benutzt die Poisson-Verteilung im allgemeinen zu Annäherung der Binomialverteilung, wenn n groß ist und p klein. Als Erwartungswert $E = \mu$ der Poisson-Verteilung verwenden wir $\mu = \lambda = n \cdot p$, der mit der Varianz identisch istd. Allgemein approximiert die Poisson-Verteilung die

Binomialverteilung sehr gut für Werte von $n \geq 100$ und $\lambda \leq 10$. Neben den Geschwindigkeitsvorteilen bei der Berechnung, hat die Poission-Verteilung noch den Vorteil, dass sie unendlich abzählbar ist, sich also ins positiv Unendliche inf fortsetzt.

5.4.6 t-Verteilung oder Student-t-Verteilung

Die Normalverteilung wird bei vielen statistischen Verfahren eingesetzt. Allerdings unterschätzt die Normalverteilung bei kleinen Stichprobenumfängen die Standardabweichung. Dieser Effekt kann aber ausgeglichen werden, indem man bei manchen statischen Verfahren statt der Normalverteilung die t-Verteilung einsetzt. Die t-Verteilung ist eine der Normalverteilung verwandte Verteilung. Die t-Verteilung erhält man, wenn man den Mittelwert einer normalverteilten Population in Situationen schätzt, in denen der Stichprobenumfang klein[3] ist und die Standardabweichung der Population unbekannt ist. Diese Verteilung zeichnet sich dadurch aus, dass Sie breitere Enden als die Normalverteilung hat. Für steigende Stichprobenumfänge nähern sich die beiden Verteilungen an und sind schließlich identisch. Die t-Verteilung ist folgendermaßen definiert: $T = \frac{Z}{\sqrt{\frac{1}{n} * \chi^2}}$, $E = 0$ und $V = \frac{n}{n-2}$, wobei Z normalverteilt ist und χ^2 unabhängig und Chi^2 verteilt ist.

5.4.7 Chi^2 Verteilung

Die Chi^2 Verteilung ist eine kontinuierliche Verteilung, die häufig zur Testung statistischer Unabhängigkeit oder zur Gültigkeit einer Hypothese (goodness of fit) genommen wird, genannt sei hier der *Pearson's chi-square test*, **??**. Nur wenige weltliche Dinge sind mit der Chi^2 Verteilung gut beschrieben. Es gibt einen Parameter, der die Freiheitsgrade n festlegt[4]. Diese Freiheitsgraden bestimmen die Verteilung insofern, als dass $\chi^2_n = Z^2_1 + ... + Z^2_n$ gilt, also dass n unabhängige, quadrierte und standardnormalverteilte Zufallsvariablen ihr ungefähr äquivalent sind, d. h. $Z_k \sim \mathcal{N}(0,1) \forall k = 1, ...n$ und $\chi^2 \sim \chi^2_n$. Außerdem gilt: $E_{Chi^2} = n$ und $V_{Chi^2} = 2 \cdot k$.

[3] Eine gute Faustregel lautet, dass Sie bei einer Stichprobengröße von mindestens 30 die z-Verteilung anstelle einer t-Verteilung nutzen können.

[4] Würde man zufällige unabhängige Stichproben von n normalverteilten Größen nehmen, summierten sich diese Stichproben zu einer einer $Chi^2 - V$ mit n Freiheitsgraden.

5.4.8 Binominalverteilung

Prozesse bei denen nur 2 mögliche Ausgänge denkbar sind (e.g. ein Münzwurf) lassen sich mit der **Bi**nominalverteilung beschreiben. Voraussetzung dafür ist, dass das Experiment aus gleichen und von einander unabhängigen Versuchen besteht. Die Parameter n und k deuten schon darauf hin, dass es sich um eine diskrete Wahrscheinlichkeitsverteilung handelt, die Fragen nach k Erfolgen bei n Ereignissen beantwortet. Es gilt:

Variabel	Formel
$P(X = k)$	$\binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$
E	$n \cdot p$
V	$n \cdot p \cdot q$
σ	$\sqrt{n \cdot p \cdot q}$
$\binom{n}{k}$	$\frac{n!}{k!(n-k)!}$

(7)

Der Binominalkoeffizient beschreibt die Anzahl der Möglichkeiten, wie k Objekte in einer Gruppe aus n ohne Wiederholung angeordnet werden können. Die Binomialverteilung ist linksschief, wenn $p \geq 0,5$[5], rechtsschief wenn $p \leq 0,5$ und bei $p = 0,5$ symmetrisch. Wenn n hinreichend groß ist, kann die Normalverteilung als Annäherung zur Binomialverteilung verwendet werden, da die Schiefe mit zunehmenden n kleiner wird.

5.4.9 Bernoulliverteilung

Die Bernoulli Verteilung ist eine diskrete Verteilung, deren Zufallsvariable X nur zwei Werte annimmt: 0 = Misserfolg / Niete bzw. 1 = Erfolg / Treffer. Sie entsteht, wenn man ein Bernoulli Experiment (welches nur 2 mögliche Ausgänge hat) genau 1 Mal ausführt. Die Bernoulli Verteilung ist daher ein Spezialfall der Binomialverteilung für n=1.

Es gilt: $E_{bernoulli} = p, V_{bernoulli} = p \cdot (1-p)$ und $f_{bernoulli}(x) = \begin{cases} 1-p & if x = 0 \\ p & if x = 1 \\ 0 & sonst. \end{cases}$

[5] Greather than aber nicht gleich! Symbol fehlt.

5.5 Bayes

Der Satz von Bayes trifft Aussagen über Wahrscheinlichkeiten gegeben dass man Daten oder Observationen kennt, also von Ereignissen ($Data$) zu deren Ursachen (Parameter=Θ). Je mehr desto zuverlässiger wird die Verteilungswahrscheinlichkeit einer Variablen. Bayes Satz lautet:

$$P(\Theta|Data) = \frac{P(Data|\Theta) \cdot P(\Theta)}{P(Data)}$$

$P(\Theta|Data)$ für die **posteriore** Wahrscheinlichkeit der Parameter gegeben die Daten. $P(Data|\Theta)$ ist die **Likelihood** die Daten gegeben der Parameter zu beobachten. Für $P(\Theta)$ legt man in der Regel einen **Prior** fest, also dessen Wahrscheinlichkeiten ohne vorherige Kenntnis.

Nehmen wir das sogenannte "Monty Hall Problem". In diesem Spiel gibt es drei Türen, hinter einer davon ist ein Gewinn und hinter den anderen beiden sind Ziegen. Der Spieler wählt eine Tür aus, der Moderator öffnet eine andere hinter der sich eine Niete befindet. Der Spieler hat dann die Möglichkeit, bei seiner ursprünglichen Wahl zu bleiben oder zu wechseln. Die Frage ist nun: Sollte der Spieler bei seiner ursprünglichen Wahl bleiben? Mit Bayes-Theorems lässt sich diese Frage klar beantworten.

Sei A die gwählte Tür und C die Tür, die der Moderator öffnet. Wie hoch ist die Wahrscheinlichkeit, dass der Preis hinter Tür A ist, gegeben dass C geöffnet wurde, und wie hoch ist sie für B? ***Wir müssen also*** $P(A|C)$ ***und*** $P(B|C)$ ***berechnen.*** Wenn $P(B|C) > P(A|C)$ ist, sollte gewechselt werden! Wir wissen, dass $P(A) = P(B) = P(C) = \frac{1}{3}$ und dass nur eine Tür geöffnet wird hinter der nicht der Preis ist. Es gilt:

$$P(A|C) = \frac{P(C|A) \cdot P(A)}{P(C)} P(A) = \frac{P(C|A) \cdot P(A)}{P(C|A) \cdot P(A) + P(C|B) \cdot P(B) + P(C|C) \cdot P(C)}$$

wobei $P(C)$ nach dem Prinzip der Gesamtwahrscheinlichkeit die Wahrscheinlichkeiten normiert (die Summe aller Wahrscheinlichkeiten unter dene C geöffnet wird in unserem Fall!). Mit einsetzen erhalten wir

$$P(A|C) = \frac{0.5 \cdot \frac{1}{3}}{0.5 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = \frac{1}{3}$$

Um die Logik nachzuvollziehen, stelle man die Frage: *Wie wahrscheinlich ist es*

Tür X zu öffnen, gegeben dass der Preis hinter Tür Y ist? Andererseits ist

$$P(B|C) = \frac{P(C|B) \cdot P(B)}{P(C)} P(B) = \frac{P(C|B) \cdot P(B)}{P(C|B) \cdot P(B) + P(C|A) \cdot P(A) + P(C|C) \cdot P(C)}$$

$$= \frac{1 \cdot \frac{1}{3}}{1 \cdot \frac{1}{3} + 0.5 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = \frac{2}{3}$$

und damit größer für B, man sollte wechseln.

6 Lineare Algebra

Wir definieren eine lineare Transformation - nichts anderes als eine Matrix - als eine Abbildung, welche den Nullvektor unverändert und die Linien aller Dimensionen im Vektorraum in den gleichen Abständen zueinander lässt[1].

Im folgenden sind die 8 Axiome der Linearen Algebra:

1. $u + (v + w) = (u + v) + w$
2. $v + w = w + v$
3. Es gibt den Nullvektor, sodass $v + 0 = v$
4. Für jeden vektor v gibt es ein inverses, sodass gilt $v + (-v) = 0$
5. $a * (b * v) = (a * b) * v$
6. $1 * v = v$
7. $a * (v + w) = a * v + a * w$
8. $(a + b) * v = a * v + b * v$

[1] Der Begriff der Linearität liegt hier vor.

6.1 Inverse einer matrix

Die Umkehrtransformation einer beliebigen Transformation M, geschrieben als M^{-1} heißt Inverse zu M. Nicht jede Matrix ist invertierbar. Als generelle Voraussetzung gilt, dass

1. die Matrix quadratisch ist
2. die Determinante ungleich Null ist[6]

Ist das nicht der Fall, lässt sich trotzdem eine Pseudoinverse berechnen, siehe 6.1.1. Eine bekannte Methode zur Berechnung der (symmetrischen) Inversen ist das Gauss-Jordan-Eliminations-Verfahren:
Zu invertieren sei die Matrix

$$M = \begin{bmatrix} 2 & 1 & -3 \\ 4 & -1 & 2 \\ 1 & 3 & -1 \end{bmatrix}$$

Der Gedanke ist stufenweise jede Variable zu eliminieren indem man Reihenweise Verknüpfungen ausführt.

$$\left[\begin{array}{ccc|c} 2 & 1 & -3 & 0 \\ 4 & -1 & 2 & 0 \\ 1 & 3 & -1 & 0 \end{array}\right]$$

$$\left[\begin{array}{ccc|c} 1 & \frac{1}{2} & -\frac{3}{2} & 0 \\ 4 & -1 & 2 & 0 \\ 1 & 3 & -1 & 0 \end{array}\right]$$

$$\left[\begin{array}{ccc|c} 1 & \frac{1}{2} & -\frac{3}{2} & 0 \\ 0 & -3 & 8 & 0 \\ 0 & \frac{5}{2} & -\frac{1}{2} & 0 \end{array}\right]$$

$$\left[\begin{array}{ccc|c} 1 & \frac{1}{2} & -\frac{3}{2} & 0 \\ 0 & 1 & -\frac{8}{3} & 0 \\ 0 & \frac{5}{2} & -\frac{1}{2} & 0 \end{array}\right]$$

[6] Wenn die Determinante Null ist, hat die Matrix entweder keine Inverse oder unendlich viele Inversen.

$$\left(\begin{array}{ccc|c} 1 & 0 & \frac{1}{3} & 0 \\ 0 & 1 & -\frac{8}{3} & 0 \\ 0 & \frac{5}{2} & -\frac{1}{2} & 0 \end{array}\right)$$

$$\left(\begin{array}{ccc|c} 1 & 0 & \frac{1}{3} & 0 \\ 0 & 1 & -\frac{8}{3} & 0 \\ 0 & 0 & \frac{11}{6} & 0 \end{array}\right)$$

$$\left(\begin{array}{ccc|c} 1 & 0 & \frac{1}{3} & 0 \\ 0 & 1 & -\frac{8}{3} & 0 \\ 0 & 0 & 1 & 0 \end{array}\right)$$

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right)$$

6.1.1 Moore-Penrose Pseudoinverse

$$A^{\dagger} = (A^T A)^{-1} A^T$$

wobei A^T die Transponierte der Matrix A bedeutet und A^{-1} ihre Inverse darstellt. Die Moore-Penrose-Pseudoinverse hat die Eigenschaft, dass sie die "besteÄpproximation der Inversen ist, selbst wenn die Matrix A nicht quadratisch oder invertierbar ist.

6.2 Die Basisvectoren

6.3 Determinante

Zur Veranschulichung empfiehlt es sich die det als Faktor vorzustellen, der Flächeninhalt/Volumen des Raumes, der von den Basisvektoren aufgespannt wird, skaliert. Ein Wert von Null bedeutet, dass der Raum in seinen Dimensionen reduziert wird und damit ist die Matrix weder invertierbar noch hat sie linear unabhängige Spalten. Gilt $\det(A) \neq 0$, dann ist A invertierbar. Für **Gl.2** bedeutet das, dass A^{-1} eine eindeutige Lösung besitzt. Ist $\det(A) \neq 0$, so besitzt A entweder keine oder unendliche viele Lösungen.

7 Analysis

7.1 Funktionen/Abbildungen

7.1.1 Definitionsbereich und Wertebereich

Die Definitionsmenge ist die Menge aller x-Werte, die in die Funktion eingesetzt werden dürfen. Die Wertemenge ist die Menge aller y-Werte, die die Funktion unter Beachtung ihrer Definitionsmenge annehmen kann. Wird jedes Element des Wertebereichs 'getroffen', so heißt die Funktion **surjektiv**. Bilden keine zwei Elemente des Definitionsbereichs auf dasselbe im Wertebereich ab, so heißt die Funktion **injektiv**. Ist eine Abbildung sowohl surjektiv als auch injekttiv, so heißt sie **bijektiv**.

7.1.2 Spezielle Funktionen

Eine **lineare Funktion** ist einer Funktion der allgemeinen Form $f(x) = m \cdot x + b$.

Quadratische Funktionen haben die allgemeine Form $f(x) = a \cdot x^2 + b \cdot x + c$. Dabei ist $a \neq 0$ und b und c Konstanten. Da x^2 der höchste Term ist (die höchste Ordnung) ist er der entscheidende Faktor für das Aussehen der Funktion[7].

Eine Funktion f mit der Funktionsgleichung $f(x) = x^n | n \in Z/\{0\}$ heißt **Potenzfunktion**. Wichtig ist hier zwischen den Exponenten zu unterscheinden, da hier grade/ungrade und positve/negative Exponenten die Potenzfunktion verschienden aussehen lassen. So nennt man Potenzfunktionen mit positivem Exponentem größer als 1 **Parabel**, und jene mit negativem Exponetem **Hyperbel**.

Die Sigmoid Funktion: Es gibt mehrere Sigmoidsche Funktionen. Zum Beispiel die **logistische Sigmoid**, **die Arctangent** und **die hyperbolische tangent Funktion**. Charakteristisch ist eine s-förmige Kurve, die du auch als logistische Kurve kennst. Sigmoidsche Funktionen mappen alle x-Werte zwischen 0 und 1 ($S_{log} \rightarrow$ Wahrscheinlichkeit), zwischen -1 und 1 ($S_{arctangent}$) oder noch größere Intervalle. Damit kann sie großartig als Aktivierungsfunktion in einem NN genutzt werden, um die letzte Schicht (output) zu erstellen. Außerdem ist bemerkenswert, dass die logistische Sigmoid für plus und minus unendlich gegen

[7] https://www.schlauerlernen.de/quadratische-funktion/

ihre spezifischen Limits covergiert. Daher hat sie stets Gradienten ungleich 0. Die **logistische Sigmoid** wird häufig nur **Sigmoid** genannt und wird so geschrieben:

$S(x) = \frac{1}{1+\exp(-x)} \equiv \frac{\exp(x)}{\exp(x)+1}$

Die **hyperbolische tangent Funktion**, die alle Werte in den Wertebereich = [-1,1] mapped, hat folgende Form:

$S_{tan}(x) = \frac{\exp(x)-\exp(-x)}{\exp(x)+\exp(-x)}$

Zuletzt genannt sei die **Arctangent Funktion**:

$S_{arc}(x) = arctan(x)$

Bekannt ist sie aus der Trignonometrie(3.1), da sie die inverse Funktion (7.1.3) zur tangent Funktion ist. Ihr Wertebereich ist $[-2\pi, 2\pi]$. Die beiden letzteren Funktionen nähern sich viel schneller ihrem Grenzwert als für die klassische Sigmoid Funktion. Die **rectified linear Unit: ReLU** ist ebenfalls sehr bekannt aus KI. Sie wir geschrieben als:

$ReLU(x) = max(0, x)$

Das coole an ihr, dass ihre Ableitung stets 1 ist für $x > 0$ und dass sie nicht vom *vanishing gradient* betroffen ist, der für große NN eine schnelle Konvergenz gegen 0 in der Backpropagation hat. In diesem Aspekt ist ReLU der Sigmoid klar überlegen, vorausgesetzt man kompensiert für negative x (deren Gradient = 0 ist), indem ein kleiner Term dazugerechnet wird.

7.1.3 Inversibilität

Eine Funktion heißt umkehrbar eindeutige (**eineindeutige**) Funktion, wenn nicht nur jedem Argument eindeutig ein Funktionswert zugeordnet ist, sondern auch umgekehrt zu jedem Funktionswert genau ein Argument gehört. Um die Umkehrfunktion einer Abbildung zu ehalten, löst man die Funktionsgleichung nach x um und vertauscht dann die Variablen. Zum Beispiel: $f(x) = x^2 + 1 = y \rightarrow f^{-1} = x = \sqrt{y-1}$, vertauschen führt zu $f^{-1} = \sqrt{x-1} = y \mid \forall x \neq 1$.

Allerdings, ist diese Funktion teilweise nicht definiert und daher nicht invertierbar. Das gilt nur solange wir die negativen Zahlen zulassen. Nehmen wir R^+, ist $f(x)$ invertierbar.

7.1.4 Nullstellen

Nullstellen sind diejenigen Punkte einer Funktion, an denen der y-Wert den Wert 0 hat. Um sie auszurechnen setzen wir die Funktionsgleichung gleich 0 und lösen sie. Für Ableitungen ist das Verfahren dasselbe, doch hat es dann eine etwas andere Bedeutung, nämlich ist die (erste) Ableitung als Steigungsgraph der Funktion anzusehen, und diejenigen Punkte die im Ableitungsgraphen Nullstellen sind, sind in der **Stammfunktion** maxima oder minima. Die zweite Ableitung mit ihren Nullstellen gibt die **Wendepunkte** der Stammfunktion wieder.

7.2 Ableitungen

Was beschreibt eine Ableitung? Die Ableitung beschreibt das Änderungsverhalten von Funktionen. Am interessantesten sind die erste und zweite Ableitung, da man hier graphisch leicht Nullstellen mit Maxima und Minima gleichsetzen kann. Die Nullstellen in der zweiten Ableitung geben Nullstellen der ersten Ableitung und damit Wende- oder Sattelpunkte der Abbildung wieder. Ableitungen werden mit $f'(x) = \frac{df(x)}{dx} = \frac{d}{dx}f(x) = \frac{dy}{dx} = y'$ und mit ∂ für d (für partielle Ableitungen) kenntlichgemacht.

7.2.1 Spezielle Ableitungen

- $f(x) = c \Rightarrow f'(x) = 0$
- $f(x) = e^x \Rightarrow f'(x) = e^x$
- $f(x) = ln(x) \Rightarrow f'(x) = \frac{1}{x}$
- $f(x) = \sin(x) \Rightarrow f'(x) = \cos(x)$
- $f(x) = cos(x) \Rightarrow f'(x) = -\sin(x)$
- $f(x) = tan(x) \Rightarrow f'(x) = \frac{1}{cos^2(x)}$

7.2.2 Ableitungsregeln

- Potenzregel: $f(x) = x^n \Rightarrow f'(x) = n * x^{n-1}$
- Faktorregel: $f(x) = c * g(x) \Rightarrow f(x) = c * g'(x)$
- Summenregel/Differenzregel: $f(x) = h(x) + g(x) \Rightarrow f'(x) = h'(x) + g'(x)$
- Produktregel: $f(x) = h(x) * g(x) \Rightarrow f'(x) = h'(x) * g(x) + h(x) * g'(x)$
- Kettenregel: $f(x) = h(g(x)) \Rightarrow f'(x) = h'(g(x)) * g'(x)$
- Quotientenregel: $f(x) = \frac{g(x)}{h(x)} \Rightarrow f'(x) = \frac{g'(x)*h(x)-g(x)*h'(x)}{h(x)^2}$
- Linearität: $f'(a * x) = a * f'(x)$ und $f'(x + z) = f'(x) * f'(z)$

7.2.3 Partielle Ableitungen ∂

Die Ableitung einer Funktion mit mehreren Argumenten nach einem dieser Argumente heißt partielle Ableitung. Das Argument, nach dem nicht abgeleitet wird, verhält sich wie eine Konstante.

Sei f eine Funktion $f : \mathbb{R}^n \to \mathbb{R}, \vec{x} = (x_1, ..., x_n) \to f(x_1, ..., x_n)$.
Eine *partielle Ableitung* ist dann eine Ableitung von f nach nur einem x_j :

$$\frac{\partial f}{\partial x_j}$$

Das bedeutet, dass du alle anderen x_i als Konstanten annimmst, während du f ableitest.
Der *Gradient* ist nun einfach ein Vektor mit allen partiellen Ableitungen:

$$\nabla f = (\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n})$$

Jetzt kann man die partielle Ableitung schreiben als

$$\frac{\partial f}{\partial x_j} = \vec{e_j} \cdot \nabla f$$

wobei $e_j = (0, .., 1, .., 0)$ der j-te Vektor der Standardbasis ist. Die Richtungsableitung ist nun die Verallgemeinerung der obigen Gleichung, die es uns ermöglicht, die Ableitung in jeder Richtung zu berechnen (nicht nur in den Richtungen der

Vektoren der Standardbasis):

$$\nabla_v f = v \cdot \nabla f$$

7.2.4 Jacobimatrix

Die Jacobimatrix verallgemeinert das Konzept des Gradienten auf den Fall einer Funktion

$$f : \mathbb{R}^n \to \mathbb{R}^m, \vec{x} = (x_1, ..., x_n) \to \Big(f_1(x_1, .., x_n) \ : \ f_m(x_1, .., x_n) \Big)$$

Sie ist definiert als

$$J = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & ... & \frac{\partial f_1}{\partial x_n} \\ : & & : \\ \frac{\partial f_m}{\partial x_1} & ... & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

7.2.5 Hessische Matrix

Für die Hessische Matrix sind wir nun wieder bei einer Funktion $f : \mathbb{R}^n \to \mathbb{R}, \vec{x} = (x_1, ..., x_n) \to f(x_1, ..., x_n)$. Die Hessische Matrix enthält alle partiellen Ableitungen zweiter Ordnung. Das (i,j)-te Element der Hessischen Matrix ist

$$H_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j},$$

was bedeutet, dass man zuerst nach x_j ableitet und dann nach x_i.

7.3 Polynome

Polynome sind mathematische Ausdrücke, die aus einer Summe von Potenzfunktionen bestehen. Sie haben die allgemeine Form

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 \tag{8}$$

wobei n die Ordnung des Polynoms ist und die Koeffizienten $a_n, a_{n-1}, \ldots, a_1, a_0$ reale Zahlen sind. Die Variable x repräsentiert den unabhängigen Wert, während die Koeffizienten die Gewichtungen der einzelnen Potenzen darstellen.

7.4 Integrale

7.5 Der Grenzwert

7.6 Differentialgleichungen

Eine Differentialgleichung ist eine Gleichung, in der eine Funktion und auch Ableitungen von dieser Funktion auftauchen können. Die Lösung dieser Art von Gleichung ist eine **Funktion, keine Zahl**! Die explizite Darstellung einer DGL erhalten wir, wenn wir die DGL auf die höchste vorkommende Ableitung umstellen können. Falls das nicht möglich ist, kann die DGL nur in impliziter Darstellung geschrieben werden. Also explizit: y≐ Ausdruck, implizit: Audruck + y≐ Ausdruck. Wir unterscheiden zwischen gewöhnlichen DGL, bei denen die gesuchte Funktion von einer Variable abhängt, bzw. es tauchen nur Ableitungen nach einer Variablen auf, und die partielle DGL. Hier hängt die gesuchte Funktion von mehreren Variablen ab und es tauchen partielle Ableitungen der Funktion auf. Die Ordnung einer DGL ist stets n, die höchste enthaltene Ableitung. Alle anderen Ableitungen niedrigerer Ordnung sind Funktionsargumente. Außerdem unterscheidet man zwischen linearen und nicht-linearen DGL, wobei erstere als Linearkombination explizit geschrieben werden kann, $y^n + a_{n-1}(x)y^{n-1} + ... + a_2(x)y^2 + a_1(x)y = b(x)$. Ist das nicht der Fall, ist die DGL nicht-linear. In vielen Büchern und Skripten taucht die Typisierung autonome DGL auf. Eine DGL heißt autonom, wenn die Variable x nicht explizit in der DGL auftaucht (also lediglich versteckt als Funktionsargument in der Funktion y und deren Ableitungen).

1.

7.7 Divergenz

7.7.1 Kullback-Leibler

$D_{KL} = \mathbf{E} = [log(\frac{p(x)}{q(x)})] = \int_a^b \log(p(x)/q(x))\,dx$

7.7.2 Jensen-Shannon

7.8 Regression

Der Sinn von Regression besteht darin, die Beziehung zwischen einer abhängigen Variablen und einer oder mehreren unabhängigen Variablen zu modellieren.

Durch Regression können wir Vorhersagen über den Wert der abhängigen Variablen basierend auf den Werten der unabhängigen Variablen treffen. Es ermöglicht es uns, Muster und Trends in den Daten zu identifizieren, Zusammenhänge zu verstehen und zukünftige Werte vorherzusagen. Regression wird häufig in der Statistik, Ökonometrie und maschinellen Lernens eingesetzt, um Prognosen zu erstellen, Hypothesen zu testen und Entscheidungsgrundlagen zu schaffen.

7.8.1 Lineare Regression

In der linearen Regression werden die Parameter anhand der Loss Function **??** angepasst, um eine bestmögliche Anpassung an die Daten zu erreichen. Die Loss Function misst den Fehler zwischen den vorhergesagten Werten und den tatsächlichen Werten. In der linearen Regression wird oft die *Mean Squared Error* (MSE) Loss Function verwendet, definiert als

$$L(\theta) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

wobei (n) die Anzahl der Datenpunkte ist, (y_i) der tatsächliche Wert des (i)-ten Datenpunkts, ($\hat{y}_i$) der vorhergesagte Wert und (θ) die Modellparameter sind. Um die Parameter anzupassen, wird ein Optimierungsalgorithmus wie der Gradientenabstieg (Gradiend descent) verwendet. Der Gradientenabstieg berechnet den Gradienten der Loss Function nach den Parametern und aktualisiert die Parameter in Richtung des negativen Gradienten, um den Fehler zu minimieren. Die Aktualisierung der Parameter erfolgt iterativ mit der folgenden Formel:

$$\theta_{t+1} = \theta_t - \alpha\nabla L(\theta_t)$$

wobei (θ_t) die Parameter in der (t)-ten Iteration sind, (α) die Lernrate ist und ($\nabla L(\theta_t)$) der Gradient der Loss Function nach den Parametern ist. Der Gradient der MSE Loss Function nach den Parametern (θ) ist:

$$\nabla L(\theta) = \frac{2}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)\mathbf{x}_i$$

wobei ($\mathbf{x}_i$) der Vektor der unabhängigen Variablen des (i)-ten Datenpunkts ist. Durch wiederholte Anwendung des Gradientenabstiegs wird die Loss Function minimiert und die Parameter (θ) werden an die Daten angepasst, um eine

bestmögliche lineare Regression zu erreichen.

7.8.2 Polynomische Regression

Die Idee der Funktionsannäherung basiert hier auf einem Polynom 7.3 von Grad m:

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \cdots + \alpha_m x^m + \epsilon$$

Die einfachste Version der Polynomregression verwendet monomiale Basisfunktionen $1, x, x^2, x^3, \ldots$ in der Design-Matrix. Eine solche Matrix wird in der linearen Algebra als Vandermonde-Matrix bezeichnet.

$$\Phi = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^m \\ 1 & x_2 & x_2^2 & \ldots & x_2^m \\ 1 & x_3 & x_3^2 & \ldots & x_3^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^m \end{bmatrix}$$

Um die richtigen Parameter a_i zu lernen, verwenden wir den Vektor $\vec{a}$, der alle a_i enthält, und lösen die folgende Formel:

$$y = \Phi \vec{a}$$

$$\vec{a} = \Phi^\dagger y$$

Hierbei ist Φ die Design-Matrix und $\Phi^\dagger$ ihre Pseudoinverse 6.1.1, und y ist der Vektor der abhängigen Variablen.

7.8.3 Logistische Regression

Die logistische Regression ist eine Methode zur Modellierung von binären abhängigen Variablen (0 oder 1). Sie verwendet die logistische Funktion (auch als Sigmoid-Funktion 7.1.2 bekannt) zur Modellierung des Auftretens einer Klasse. Die logistische Funktion hat die Form:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{9}$$

Mit einer linearen Kombination der unabhängigen Variablen wird versucht, das Auftreten der abhängigen Variablen zu modellieren, was ausgeschrieben un-

gefähr so aussieht:

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_n x_n)}} \tag{10}$$

Hierbei ist $P(y = 1|x)$ die Wahrscheinlichkeit, dass die abhängige Variable y den Wert 1 annimmt, gegeben die Werte der unabhängigen Variablen $x_1, x_2, \ldots, x_n$.

8 Entropie

Entropie misst die *Unordnung* in einer Menge an Elementen (die minimale Menge an Bits nötig um ein Element zu kodieren).

$$E(S) = -p_+ \log_2 p_+ - p_- \log_2 p_- \tag{11}$$

Hierbei bezeichnet p_+ den Anteil der positiven und p_- den Anteil der negativen Beispiele im Datensatz. Ein Satz S mit nur positiven (oder nur negativen) Beispielen hätte keine Entropie (d.h. $E(S) = 0$), während ein Satz mit der gleichen Anzahl von positiven und negativen Beispielen maximale Entropie hätte ($E(S) = 1$).