Mathe Helfer

Simon Roske

27. September 2023

Inhaltsverzeichnis

1 Zahlen

1.1 Die natürlichen Zahlen N

Der Zahlenbereich der natürlichen Zahlen bildet das Zählen als natürlichen Prozess ab.

- Die kleinste natürliche Zahl ist die 0
- Die Menge der natürlichen Zahlen enthält alle Nachfolger der 0 bis unendlich: $\mathsf{N} = \{0, 1, 2, 3, 4, \ldots, n, n+1, \ldots\}$

Wie kannst du mit natürlichen Zahlen rechnen? Du darfst 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.

1.2 Die ganzen Zahlen Z

Erweiterst du den Zahlenbereich der natürlichen Zahlen mit den negativen Zahlen, hast du die ganzen Zahlen:

- In der Menge der negativen Zahlen sind alle positiven und negativen Zahlen ohne Komma: $\mathsf{Z} = \{\ldots -3, -2, -1, 0, 1, 2, 3, \ldots\}$

Nun kannst du auch uneingeschränkt subtrahieren.

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

Willst du uneingeschränkt dividieren, brauchst du die Bruchzahlen.

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

Nimmst du die negativen Brüche hinzu, hast du die rationalen Zahlen.

- $\mathsf{Q} = \{\frac{a}{b}|$ a,b sei eine ganze Zahl und $b \neq 0\}$.
- In Q darfst du alle Grundrechenarten uneingeschränkt ausführen.
- 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 R

Vereinst du die rationalen und die irrationalen Zahlen, erhältst du die reellen Zahlen R. In diesem Zahlenbereich sind alle positiven und negativen Bruchzahlen sowie alle Wurzeln. Aus negativen Zahlen kannst du keine Wurzel ziehen. Zum Beispiel ist $\sqrt{-4}$ nicht definiert. Solche Zahlen sind nicht in den reellen Zahhlen R enthalten.

1.5 Komplexe Zahlen C

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

Eine komplexe Zahl ist nichts anderes als ein reelles Zahlenpaar! Anstelle von $x + iy$ können wir genauso gut (x,y) schreiben. Nun kann die Menge aller reellen Zahlenpaare (die Menge R^2) geometrisch als Zeichenebene dargestellt werden. Indem wir komplexe Zahlen in der soeben beschriebenen Weise addieren und multiplizieren, haben wir diese Rechenoperationen auf den R^2 übertragen. Die Zeichenebene zusammen mit diesen Operationen ist die Menge der komplexen Zahlen und wird mit C bezeichnet. Ist z eine komplexe Zahl, so ist z^* die komplexe Konjugation. In der Darstellung unten würde der Realteil(z) hierbei gespiegelt. Als Spezialfall gilt hier: $i^* = -i$.

1.6 Besondere Zahlen π, e, i...

1.6.1 π - Die Kreiszahl

3.1415926535…..damit beginnt diese irrationale Zahl und wenn du mal Zeit hast, lerne weitere 490 Ziffern hier unten.[1]. Pi beschreibt das Verhältnis von Umfang zu Durchmesser eines Kreises. Es gibt viele Formeln, in denen π die tragende Rolle spielt. Sei es bei der Berechnung von Kreisumfang ($U = \pi d = 2\pi r$), Kreisfläche ($F = \pi r^2$), Volumen ($V = 4/3\pi r^3$) oder Oberfläche von Kugeln, Zylindern, Kegeln usw. Und damit ist klar, wenn es bei irgendeiner Zahl so richtig rund geht, dann bei der magischen Kreiszahl π.

1.6.2 e - Euler's Zahl

Die Eulersche Zahl ist die Basis des natürlichen Logarithmus, also $ln(e) = 1$. Die Eulersche Zahl kann beschrieben werden durch $e = 2,71828...$, aber ähnlich wie für π gibt es für e 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

Warum ist die Zahl e wichtig? Unter anderem beinhaltet die natürliche Exponentialfunktion $f(x) = e^x$ die Eulersche Zahl. Diese Funktion ist in vielen naturwissenschaftlichen Bereichen von großer Bedeutung. Für uns ist aber 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. Und diese Darstellung erleichtert die Berechnungen erheblich! Außerdem gilt $\partial e^t / \partial t = e^t$ was bedeutet, dass die Position von e gleich der Ableitung, also der Geschwindigkeitsänderung ist. Kennst du die schönste Formel der Welt? Die Eulersche Identität $e^{j\pi} + 1 = 0$ wurde von den Mathematikern zur schönsten Formel gewählt, weil sie die 5 wichtigsten Zahlen, also e, j, π, 1 und 0 miteinander verknüpft. Dieser Zusammenhang ist besonders schön: $\lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = 2,718281828459$...Damit is e der Grenzwert der (exponentiell) steigenden Zahlenfolge.

1.6.3 Die imaginäre Zahl i

$i^2 = -1 \to i^3 = -i \to i^4 = 1$, $i^{-1} = -i$.

2 Grundlagen & Rechengesetze

2.1 Verknüpfungsgesetze

2.1.1 Kommutativgesetz

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

2.1.2 Assoziativgesetz

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

2.1.3 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: $2 * (3 + 4) = 2 * 3 + 2 * 4$.

2.2 Potenzen

Beginnen wir mit den wichtigsten Regeln:

- $a^0 = 1$
- $a^1 = a$
- $a^{-1} = \frac{1}{a}$

- $(\frac{a}{b})^{-n} = (\frac{b}{a})^n$

Generell schreiben wir $a^{-n} = \frac{1}{a^n}$ *sowie* $a^n = \frac{1}{a^{-n}}$. Gibt es mehr als eine Potenz, muss man sie miteinander verrechnen, unzwar wie folgt:

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

2.3 Wurzeln

Als Wurzeln von etwas bezeichnet man die Zahl die mit sich selbst multipliziert die Wurzel ergibt. Bei der Wurzel muss es sich nicht um eine qudratische Handeln, sondern jegliche Wurzel (kubisch, ...) ist denkbar. Als Potenz geschrieben ist die qudratische Wurzel $\sqrt{x} = x^{\frac{1}{2}}$ und jede andere Wurzel $\sqrt[n]{x} = x^{\frac{1}{n}}$. Haben wir eine Potenz mit einer Lösung, dann gilt $x^n = a \Leftrightarrow x = \sqrt[n]{a}$. Viel mehr gibt es hier nicht zu wissen!

2.4 Logarithmus

Die Idee hinter dem Logarithmus ist: Mit welcher Zahl muss ich die Basis a potenzieren um y zu bekommen? Die Schreibweise hierbei ist $\log_a(x) = y$. Dabei muss beachtet werden, dass das Logarithmieren von Null und negativen Zahlen nicht definiert ist. Außerdem ist der Logarithmus die Umkehrfunktion zur Exponentialfunktion $f(x) = a^x = y$, denn $\log_a(y) = x$ (Beide haben die gleiche Basis). Haben wir die Exponentialfunktion $f(x) = e^x$ mit e als Basis, so ist der Logarithmus Naturalis (ln) f^{-1}. Hier sind die Rechengesetze für den Log:

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

Zum Abschluss noch eine Bemerkung zur Logskale: Logarithmische Skalierungen sind hilfreich, wenn die angezeigten Daten weitaus höhere oder niedrigere Werte aufweisen als die restlichen Daten oder wenn der Prozentsatz oder die Verhältnisunterschiede zwischen Werten von Bedeutung sind. Sie können angeben, ob eine logarithmische Skala verwendet werden soll, wenn die Werte im Diagramm einen sehr großen Bereich abdecken. Eine logarithmische Skala potenziert den Basiswert 10 mit einem Wert. Beispiel: 10 hat einen Logarithmus von 1, da 10 potenziert mit 1 gleich 10 ist. 100 hat einen Logarithmus von 2, da 10 potenziert mit 2 gleich 100 ist, und so weiter.

3 Geometrie

3.1 Phytagoras

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

3.2 Trignonometrie

(A) $sin(\theta) = y/z \Leftrightarrow sin(\theta) = Gegenkahete/Hypothenuse$
(B) $cos(\theta) = x/z \Leftrightarrow cos(\theta) = Ankathete/Hypothenuse$
(C) $tan(\theta) = y/x \Leftrightarrow tan(\theta) = Gegenkahete/Ankathete$

3.3 Polarkoodinaten

Mit $r = z$ folgt aus (C): $r * sin(\theta) = y$ und $r * cos(\theta) = x$ wobei r der radius in einem Polarkoordinatensystem ist. Da sich r aus dem Term herauskürzen lässt, erhalten wir $tan(\theta) = \frac{\sin(\theta)}{\cos(\theta)}$. Einfach formuliert beudetet dieser Zusammenhang folgendes: Über den Winkel θ kannst du mit $\tan(\theta)$ das Verhältnis $y : x$ berechnen. Außerdem können wir den Winkel θ mit der inversen zu tan, sin und cos berechenen. $f^-1 \equiv \tan^- 1(\theta)$ übersetzt sich zu $\arctan(\theta)$, woraus folgt: $\tan^- 1(Gegenkahete/Ankathete) = \theta$.[2]

3.4 Bogenmaß und Radian

Meistens werden Winkel in Grad angegeben. Aber ein Winkel von 45 deg kann auch im Bogenmaß, $\frac{1}{4}\pi \approx 0,79$, angegeben werden. Es gilt: $360\,\text{deg} = 2\pi$, $1\pi = 180\,\text{deg}$ und daher $1,5\pi = 270\,\text{deg}$. 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*.

Die Umrechnung geschieht folgendermaßen: $\boxed{rad = \frac{\text{deg}}{180\,\text{deg}} * \pi \text{ und } \text{deg} = \frac{rad}{\pi} * 180\,\text{deg}}$

[2] θ is in radian.

4 Wahrscheinlichkeit

In diesem Abschnitt geht es um Wahrscheinlichkeit. Es werden Verteilungen, Bayes' Ansatz und verschiedene Konzepte zur Berechnung der Wahrscheinlichkeit thematisiert. Am Ende wird ein wenig über statistische Verfahren zu deren Testung gehen.

4.1 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.

4.2 Wahrscheinlichkeitskonzepte

Was ist die joint probability, die marginal und die conditional probability[3]? Darum geht es hier. Schauen wir uns folgendes Diagramm dafür an: Die *Wahrscheinlichkeit gleichzeitiger Erreignisse (joint prob.)* bedeutet, dass für die zwei Variablen des Diagramms ein Wert gegeben ist, also dass $P(X = male, Y = rugby) = \frac{100}{500}$. Man kommt auch zu diesem Ergebnis, indem man die bedingte Wahrscheinlichkeit mit der marginalen Wahrscheinlichkeit multipliziert, also $P(X,Y) = P(x) \cdot P(y|x) = P(y) \cdot P(x|y) = \frac{125}{500} \cdot \frac{100}{125} = \frac{270}{500} \cdot \frac{100}{270}$. Die *marginale Wahrscheinlichkeit* ist die Wahrscheinlichkeit einer Kategorie anzugehören ohne Weiteres, also dass X=x oder Y=y für bestimmte x und y ist. Ein Beispiel: $P(X = female) = \frac{230}{500} = 0,46$ oder $P(Y = other) = \frac{180}{500} = 0,36$. Schließlich die bedingte Wahrscheinlichkeit, also der Fall dass für eine Variable schon ein Wert gesetzt ist der den Wert der Zielvariable einschränkt bzw. spezifiziert, wir schreiben ganz generell: $P(X|Y) = \frac{P(X,Y)}{P(Y)}$. Bayes 4.5 wird diesen Satz noch ausweiten.

[3] Was sind die deutschen Konzepte?

4.3 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 (probability density)** 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.

4.3.1 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: $\mathcal{N}(\mu, \sigma^2)$.

4.3.2 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.

4.3.3 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. $E(x) = \frac{a+b}{2} und V(x) = \frac{1}{12} \cdot (b-a)^2$.

4.3.4 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.

4.3.5 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[4] 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:

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

$T = \frac{Z}{\sqrt{\frac{1}{n} * \chi^2}}$, $E = 0$ und $V = \frac{n}{n-2}$, wobei Z normalverteilt ist und χ^2 unabhängig und Chi^2 verteilt ist.

4.3.6 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*, 4.4. Nur wenige weltliche Dinge sind mit der Chi^2 Verteilung gut beschrieben. Es gibt einen Parameter, der die Freiheitsgrade n festlegt[5]. Diese Freiheitsgraden bestimmen die Verteilung insofern, als dass $\chi_n^2 = Z_1^2 + ... + Z_n^2$ 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 $\chi^2 \sim \chi_n^2$. Außerdem gilt: $E_{Chi^2} = n$ und $V_{Chi^2} = 2 \cdot k$.

4.3.7 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	$\sqrt{n \cdot p \cdot q}$
$\binom{n}{k}$	$\frac{n!}{k!(n-k)!}$

(1)

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 $p \geq 0,5$[6], 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.

4.3.8 Bernoulliverteilung

Die Bernoulli Verteilung ist eine diskrete Verteilung, deren Zufallsvariable X nur zwei Werte annimmt: 0 = Misserfolg / Niete bzw. 1 = Erfolg / Treffer. Sie

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

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

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 $f_{bernoulli}(x) = \begin{cases} 1-p & ifx = 0 \\ p & ifx = 1 \\ 0 & sonst. \end{cases}$

4.4 Pearson's chi-square test

4.5 Bayes

5 Lineare Algebra

Ausgehend von dem Wissen was ein Vektor, ein Vektorraum und Basisvektoren sind, definieren wir eine lineare Transformation als eine Abbildung, welche den Nullvektor unverändert und die Linien aller Dimensionen im Vektorraum in den gleichen Abständen zueinander lässt[1]. Für die Basisvektoren bedeutet dass, dass ein Vektor v, für den gilt **Gl.1**: $v = 3 * \hat{i} + 4 * \hat{j}$ nach einer willkürlichen Transformation - wir können hier sagen nach Anwendung der Matrix A auf v - nach wie vor durch den Ausdruck **Gl.1** gegeben ist. Sei A eine quadratische Matrix in R^n, x und b Vektoren im selben Raum, dann ist A eine lineare Transformation auf x, geschrieben **Gl.2**: $A * x = b$. s

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. $(a + b) * v = a * v + b * v$

5.1 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

[1]Der Begriff der Linearität liegt hier vor.

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.

6 Analysis

6.1 Funktionen/Abbildungen

6.1.1 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.

6.1.2 Die wichtigsten 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].

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 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:

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

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

Zuletzt genannt sei die **Arctangent Funktion**:

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

Bekannt ist sie aus der Trignonometrie(3.2), da sie die inverse Funktion (6.1.1) 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.

6.2 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.

6.2.1 Spezielle Ableitungen

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

6.2.2 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 $f'(x + z) = f'(x) * f'(z)$

6.2.3 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.

$$\boxed{f_x = \frac{\partial f}{\partial x},\ f_y = \frac{\partial f}{\partial y},\ f_{xy} = \frac{\partial^2 f}{\partial xy},\ f_x = \frac{\partial^2 f}{\partial x},\ f_{yx} = \frac{\partial^2 f}{\partial yx}}$$

6.3 Polynome

6.4 Integrale

6.5 Der Grenzwert

6.6 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$\hat{=}$ Ausdruck, implizit: Audruck + y$\hat{=}$ 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.

6.7 Divergenz

6.7.1 Kullback-Leibler

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

6.7.2 Jensen-Shannon