



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

**„Das Problem des positiven Tests. Warum bedingte
Wahrscheinlichkeiten paradox erscheinen können“**

verfasst von / submitted by

Luca Garofalo, BEd

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Education (MEd)

Wien, 2021 / Vienna 2021

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 199 520 523 02

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Lehramt Sek (AB)
UF Mathematik UF Physik

Betreut von / Supervisor:

ao- Univ. Prof. Mag. Dr. Peter Raith

Abstract

Das Ziel der vorliegenden Masterarbeit war es, das Thema bedingte Wahrscheinlichkeit auf der einen Seite mathematisch und auf der anderen Seite fachdidaktisch aufzuarbeiten. Sie sollte einen Überblick über bedingte Wahrscheinlichkeiten bieten. Dabei wurden zunächst die wichtigsten mathematischen Begriffe zum Thema behandelt. Auf dieser Grundlage wurde dann ein Beispiel angeführt, welches als möglicher Einstieg zum Thema in der Schule dienen kann. Dabei wurde noch auf mögliche Lernschwierigkeiten Bezug genommen, auf welche die Lehrperson Rücksicht nehmen sollte. Abschließend wurden noch diverse stochastische Paradoxa vorgestellt, welche ebenfalls für die Schule geeignet sind.

Inhaltsverzeichnis

1	Einleitung	6
2	Stochastik in der Schule	7
3	Grundlagen	10
3.1	Zufallsexperiment	10
3.2	Zufallsvariable	11
3.3	Wahrscheinlichkeit	11
3.3.1	Wahrscheinlichkeit nach Laplace	13
3.3.2	Wahrscheinlichkeit als Grenzwert relativer Häufigkeiten	14
3.3.3	Die axiomatische Definition der Wahrscheinlichkeit	15
3.3.4	Zufallsvariable	17
3.3.5	Erwartungswert und Varianz	19
3.3.6	Das schwache Gesetz der großen Zahlen	24
3.4	Das Urnenmodell	29
3.5	Bedingte Wahrscheinlichkeit	33
3.6	Satz von Bayes	37
3.7	Baumdiagramm	38
3.7.1	Pfadregeln	39
4	Fachdidaktische Überlegung zur bedingten Wahrscheinlichkeit	41
4.1	Möglicher Einstieg in das Thema bedingte Wahrscheinlichkeit für die Schule	41
4.1.1	Beispiel zum Einstieg mit dem „combined test“	42
4.1.2	Fachdidaktische Überlegungen	49
4.1.3	Die bedingte Wahrscheinlichkeit und die Rolle der Intuition	51
5	Paradoxa	53

5.1	Simpson-Paradoxon	53
5.2	Monty-Hall Paradoxon	59
5.2.1	Historischer Verlauf	59
5.2.2	Das Paradoxa	60
5.3	Spielerparadoxa	65
5.3.1	Das Paradoxa	65
5.3.2	Bemerkung	67
5.4	Roulette bzw. Geburtstagsparadoxon	69
5.4.1	Geburtstagsparadoxon	69
5.4.2	Rouletteparadoxon	72
5.5	Der Fall Sally Clark	73
6	Resümee	76
7	Literaturverzeichnis	78

1 Einleitung

Der Begriff des Zufalls beschäftigt die Menschen schon seit der Antike. Aristoteles hat in seinem Buch „Physica“ schon eine Begriffserklärung für den Zufall (damals noch „tyche“ und „automaton“) formuliert [Vogt 108-116]. Besonders in der römischen Zeit wurde der Zufall, oder auch die „fortuna“ (ital. und lat. für Glück, Zufall) als etwas Mythisches angesehen. Aus diesem Grund hatten die Römer auch eine eigene Göttin für Schicksal und Glück, die Fortuna. Welchen Stellenwert die „fortuna“ in dieser Zeit hatte, lässt sich durch die große Anzahl an Tempeln erraten, die der Fortuna geweiht waren. Diese Vorstellung des Zufalls fand noch bis in das Mittelalter Gebrauch [Vogt 509].

In unserem heutigen Leben beeinflusst uns der Zufall in mehreren Bereichen unseres Lebens. Zum einen hat sich mit dem Glücksspiel ein ganzer Wirtschaftsbereich um den Zufall gebildet, aber auch bei normalen Gesellschaftsspielen kommt der Zufall häufig zum Tragen.

Diese Masterarbeit soll einen Einblick in die Welt der bedingten Wahrscheinlichkeit liefern. Dabei soll nicht mehr als das Wissen benötigt werden, welches in der Schule bei der Behandlung des Stoffes den Schüler*innen auch bekannt ist. Außerdem soll ein fachdidaktischer Exkurs die möglichen Probleme der Lernenden aufzeigen. Insbesondere soll auf die Anwendung des Satzes von Bayes weitgehend verzichtet werden. Dadurch soll ein intuitiver Zugang zu den bedingten Wahrscheinlichkeiten möglich sein.

Im Anschluss sollen noch Paradoxa vorgestellt werden, die nur mit dem Wissen von Schüler*innen erklärbar sind, die Schüler*innen aber vor eine Herausforderung stellen. Anhand dieser Beispiele soll ein gutes Verständnis für die (bedingte) Wahrscheinlichkeit erreicht werden. Abschließend soll noch gezeigt werden, was passiert, wenn kein ausreichendes Verständnis von der Wahrscheinlichkeitsrechnung zu einem groben Fehlurteil führt.

2 Stochastik in der Schule

Stochastik ist in unserer heutigen Zeit sehr präsent, auch wenn es uns in unserem Leben nicht immer bewusst ist. In unserem Alltag sind wir oft mit Situationen konfrontiert, bei denen wir mit der Hilfe von dem stochastischen Wissen, welches wir in der Schule erlernt haben, bessere Entscheidungen treffen können.

Nicht umsonst wird in vielen Gesellschaftsspielen der Zufall verwendet. Dieses Konzept ist leicht zu verstehen und ist bei vielen Spielen ein wichtiger und grundlegender Faktor, der oft für den Spielspaß und die strategische Tiefe, aber auch manchmal für Frustration sorgt.

Auch in der Wissenschaft oder Medizin finden Statistik und Wahrscheinlichkeiten Eingang. Bei den meisten Beipackzetteln eines Medikaments werden Nebenwirkungen in der Form „1 von 100“ oder „jeder tausendste Patient“ angegeben. Diese Angaben sind im Prinzip Wahrscheinlichkeiten für die einzelnen Nebenwirkungen. Um Stochastik in unserem Alltag zu erleben, müssen wir nicht erst ein Medikament aus der Apotheke holen. Oft reicht es, in der U-Bahn eine der Gratiszeitungen zu nehmen und aufzuschlagen, um ein Diagramm oder Ähnliches darin zu entdecken.

In weiterer Folge sind auch bedingte Wahrscheinlichkeiten wichtig für unseren Alltag. Eine der bekanntesten Anwendungen ist der HIV-Test. Viele Menschen sehen die Sicherheit, die ein solcher Test hat (Diese liegt oft bei über 99%) und schließen daraus, dass ein positiver Test automatisch auch bedeutet, dass sie HIV haben müssen. Dabei kann dieser Schluss auch falsch sein, und wie später gezeigt wird, ist er es oft auch. Das Menschen bedingten Wahrscheinlichkeiten und andere sogenannte Paradoxa oft falsch interpretieren ist bekannt. Daher ist es umso wichtiger, dass dieses Thema schon in der Schule behandelt wird, damit der Umgang mit diesen Situationen leichter fällt und es nicht zu oben beschriebenen Situationen kommt.

Bedingte Wahrscheinlichkeiten finden sich auch im Lehrplan wieder. Im aktuellen Lehrplan der AHS (Stand 27.02.2021) findet sich in der 6. Klasse, dem 4. Semester der Oberstufe im Fach Mathematik folgender Punkt [Lehrplan]:

- Beschreibende Statistik; Wahrscheinlichkeit
- Darstellungen und Kennzahlen der beschreibenden Statistik kennen und damit arbeiten können
- Die Begriffe Zufallsversuch, Ereignis und Wahrscheinlichkeit kennen; Methoden zur Ermittlung von Wahrscheinlichkeiten kennen: Bestimmung eines relativen Anteils, Ermittlung einer relativen Häufigkeit durch eine Versuchsserie, Angabe des subjektiven Vertrauens; wissen, dass diese Methoden nur näherungsweise bzw. unsichere Ergebnisse liefern
- Den Zusammenhang zwischen relativen Häufigkeiten und Wahrscheinlichkeiten kennen
- Mit Wahrscheinlichkeiten rechnen können (Baumdiagramme; Additions- und Multiplikationsregel)
- Bedingte Wahrscheinlichkeiten und (stochastische) Unabhängigkeit von Ereignissen kennen
- *Den Satz von Bayes kennen und anwenden können*

Hier ist klar erkennbar, dass bedingte Wahrscheinlichkeiten im Mathematikunterricht in der Oberstufe für alle Formen, welche zumindest drei Wochenstunden Mathematik besitzen, obligatorisch sind. Der Satz von Bayes wird im Lehrplan *kursiv* angegeben. Das bedeutet, dass der Satz von Bayes nur dann obligatorisch ist, wenn entweder mehr als drei Wochenstunden Mathematik vorgesehen sind oder es Möglichkeiten gibt, den Stoff aspektenreich und vertiefend zu behandeln.

Außerdem sollen Kompetenzen vermittelt werden, die außerhalb des Mathematikunterrichts zum Tragen kommen. Der Lehrplan hat dazu folgende Passage [Lehrplan]:

Der Mathematikunterricht soll beitragen, dass Schülerinnen und Schülern ihrer Verantwortung für lebensbegleitendes Lernen besser nachkommen können. Dies geschieht vor allem durch die Erziehung zu analytisch-folgerichtigem Denken und durch die Vermittlung von mathematischen Kompetenzen, die

von grundlegender Bedeutung für das Fach und relevant für viele Lebensbereiche sind. Beim Erwerben dieser Kompetenzen sollen die Schülerinnen und Schüler die vielfältigen Aspekte der Mathematik und die Beiträge des Gegenstandes zu verschiedenen Bildungsbereichen erkennen.

Außerdem finden sich verschiedene Aspekte im Lehrplan, die der Mathematikunterricht aufgreifen soll. Da wäre auf der einen Seite der *erkenntnistheoretische Aspekt*:

Mathematik ist eine spezielle Form der Erfassung unserer Erfahrungswelt. Sie ist eine spezifische Art, die Erscheinungen der Welt wahrzunehmen und durch Abstraktion zu verstehen. Mathematisierung eines realen Phänomens kann die Alltagserfahrung wesentlich vertiefen.

Auf der anderen Seite soll der *autonome Aspekt* gefördert werden, welcher dazu dient, einen Menschen als Individuum heranzuziehen:

Mathematische Gegenstände und Sachverhalte bilden als geistige Schöpfungen eine deduktiv geordnete Welt eigener Art, in der Aussagen – von festgelegten Prämissen ausgehend – stringent abgeleitet werden können. Mathematik befähigt damit, dem eigenen Denken mehr zu vertrauen als fremden Meinungsmachern, und fördert so den demokratischen Prozess.

Außerdem soll die Mathematik laut Lehrplan Beiträge zu den verschiedenen Bildungsbereichen erwirken. Dazu zählt zum Beispiel, dass Mathematik in vielen verschiedenen Bereichen unseres Lebens (Medizin, Psychologie, usw...) eine wichtige Rolle spielt. Dass dies der Fall ist, wurde bereits weiter oben gezeigt.

Im Zuge der didaktischen Grundsätze sollen im Mathematikunterricht unter anderem folgende Themen beim Lernen beachtet werden [Lehrplan]:

Lernen in anwendungsorientierten Kontexten: Anwendungsorientierte Kontexte verdeutlichen die Nützlichkeit der Mathematik in verschiedenen Lebensbereichen und motivieren so dazu, neues Wissen und neue Fähigkeiten

zu erwerben. Vernetzungen der Inhalte durch geeigneten fächerübergreifenden Unterricht ist anzustreben. Die minimale Realisierung besteht in der Thematisierung mathematischer Anwendungen bei ausgewählten Inhalten, die maximale Realisierung in der ständigen Einbeziehung anwendungsorientierter Aufgaben- und Problemstellungen zusammen mit einer Reflexion des jeweiligen Modellbildungsprozesses hinsichtlich seiner Vorteile und seiner Grenzen.

Aus diesem Grund werden angeführte Paradoxa und Beispiele zu bedingten Wahrscheinlichkeiten diesen Aspekt verstärkt berücksichtigen.

3 Grundlagen

In diesem Kapitel sollen zuerst die mathematischen Grundlagen für das Arbeiten mit Wahrscheinlichkeiten und dann in weiterer Folge mit bedingten Wahrscheinlichkeiten in der Schule geklärt werden. Außerdem wird noch kurz die historische Entwicklung des Satzes von Bayes thematisiert.

3.1 Zufallsexperiment

Bevor Wahrscheinlichkeiten überhaupt behandelt werden können, muss erst der Kontext definiert werden, in dem sie vorkommen. Der Ausgangspunkt dafür sind Zufallsexperimente. In diesen Experimenten gibt es verschiedene Ergebnisse (mindestens zwei), die vollständig vom Zufall abhängig sind. Diese müssen unter gleichen Bedingungen beliebig oft wiederholbar sein. Diese Definition ist für die Mathematik nicht sehr exakt und trifft in der Praxis höchstens annähernd zu, ist aber im Kontext dieser Arbeit ausreichend präzise.

Zu jedem Zufallsexperiment gehört auch die Ergebnismenge Ω , die Menge aller Ergebnisse, die im Experiment möglich sind. Die Ergebnismenge kann unendlich verschiedene Ergebnisse besitzen, in dieser Arbeit werden aber nur endliche Ergebnismengen betrachtet. [Fischer, S. 78]

1. Beispiel: Ein mögliches Zufallsexperiment wäre der Wurf mit einem sechsseitigen Würfel. In diesem Fall ist die Ergebnismenge $\Omega = \{1,2,3,4,5,6\}$.

2. Beispiel: Beim Roulette ist die Ergebnismenge $\Omega = \{0,1,2,3, \dots, 36\}$.

3.2 Zufallsvariable

Beim Roulette werden die Ergebnisse durch die Höhe ihres Gewinns unterschieden. So gibt die richtige Farbe beim Roulette einen weniger hohen Gewinn her als wenn man zum Beispiel die richtige Zahl trifft. Um diese potentiellen Gewinne abwägen zu können wird daher in der Mathematik die Zufallsvariable (oder Zufallsgröße) verwendet.

Die Zufallsvariable X ist eine Zuordnung, welche jedem Ergebnis ω ($\omega \in \Omega$) eine Zahl $X(\omega)$ zuordnet. Ein Beispiel dazu wäre, dass beim Werfen eines Würfelpaares (sechseitiger fairer Würfel) die Augensumme gebildet wird [Bewersdorff 2011, S. 49]. Eine genauere Definition der Zufallsvariable findet später in der Arbeit statt.

3.3 Wahrscheinlichkeit

Die moderne Definition der Wahrscheinlichkeit geht auf Andrei Kolmogoroff (*1903-1987, Sowjetunion) zurück. Sie ist nicht die aktuelle Definition der Wahrscheinlichkeit, welche in der Mathematik benutzt wird. Diese kommt später in der Arbeit. Kolmogoroff hat mit dieser Definition der Wahrscheinlichkeit jedoch einen Grundstein in der Stochastik mit seiner Arbeit gelegt. Diese sagte, dass die Chance für das Eintreten eines

Ereignisses durch eine Funktion P erfolgt. Diese Funktion P ordnet jedem Ereignis A eine Zahl $P(A)$ zu. Die Ereignisse beinhalten das unmögliche Ereignis $\emptyset$ und das sichere Ereignis Ω . Diese Funktion P muss folgenden Bedingungen genügen: [Mittag-Schüller, S. 162]

- (1) Die Zahl für jedes Ereignis A ist eine reelle Zahl zwischen 0 und 1: $0 \leq P(A) \leq 1$
- (2) Die Wahrscheinlichkeit für das sichere Ereignis $P(\Omega)$ muss 1 sein, die Wahrscheinlichkeit für das unmögliche Ereignis $P(\emptyset)$ muss 0 sein: $P(\Omega)=1$ und $P(\emptyset)=0$
- (3) Die Addition der Wahrscheinlichkeiten von zwei Ereignissen A und B muss gleich der Wahrscheinlichkeit des vereinigten Ereignisses $A \cup B$ sein. Diese Ereignisse müssen disjunkt sein: $P(A \cup B) = P(A) + P(B)$ wenn gilt $A \cap B = \emptyset$

Bedingung (3) gilt nur für ein endliches Ω .

Aus dieser Definition lassen sich verschiedene Rechenregeln ableiten. Die wichtigsten werden hier kurz vorgestellt:

$$P(\Omega \setminus A) = 1 - P(A) \tag{1}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{2}$$

$$P(A \setminus B) = P(A) - P(A \cap B) \tag{3}$$

Formel (1) ist die Gegenwahrscheinlichkeit, also die Wahrscheinlichkeit, dass Ereignis A nicht eintritt. Im Venn-Diagramm (Abbildung 1) können die wichtigsten Wahrscheinlichkeiten gesehen werden. Formel (2) wird durch das Venn-Diagramm auch verdeutlicht, da wenn $P(A)$ und $P(B)$ aufsummiert wird, die Schnittmenge $P(A \cap B)$ noch einmal abgezogen werden muss, da sie sonst doppelt gezählt werden würde. Formel (3) besagt, dass die Schnittmenge $A \cup B$ zwar zum Ereignis A dazugehört, diese jedoch abgezogen werden muss, damit das reine Ereignis A , ohne den Anteil der auch in B enthalten ist, übrig bleibt.

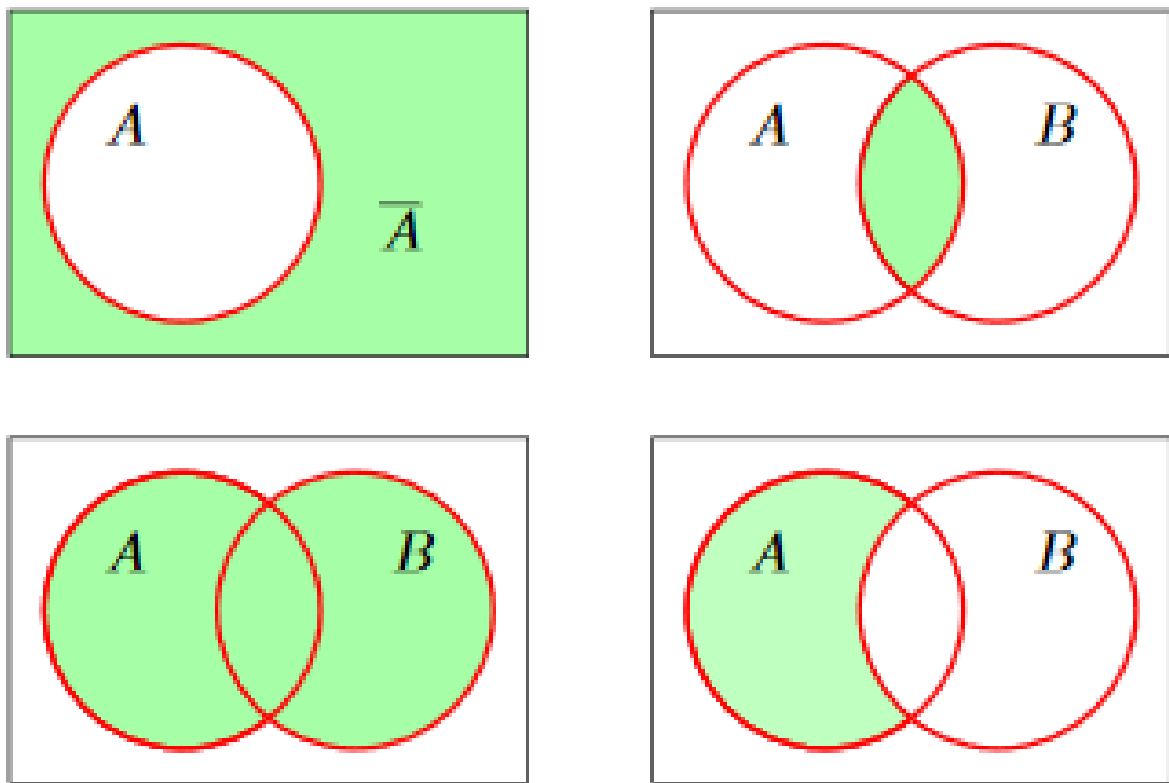


Abbildung 1: Venn-Diagramme zu $\bar{A}$, $P(A \cap B)$, $P(A \cup B)$, $P(A \setminus B)$

3.3.1 Wahrscheinlichkeit nach Laplace

Mit der Wahrscheinlichkeitsdefinition nach Laplace lassen sich endliche Experimente, wie zum Beispiel beim Würfeln oder beim Münzwurf, einfach darstellen. Diese Definition hat aber zwei wichtige Bedingungen:

1. Die Ergebnismenge muss endlich sein, also $\Omega = \{\omega_1, \omega_2, \omega_3, \dots, \omega_n\}$.
2. Die n Elementarereignisse $\{\omega_1, \omega_2, \omega_3, \dots, \omega_n\}$ müssen alle dieselbe Wahrscheinlichkeit besitzen, also: $P(\omega_1) = P(\omega_2) = P(\omega_3) = \dots = P(\omega_n)$.

Experimente, die diese beiden Bedingungen erfüllen, werden Laplace - Experimente genannt. Soll jetzt die Wahrscheinlichkeit eines Ereignisses A berechnet werden, wobei für A gelten muss, dass es eine Teilmenge von Ω ist, gilt:

$$P(A) = \frac{\text{Anzahl der für A günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} \quad (4)$$

Diese Definition der Wahrscheinlichkeit erfüllt immer die Bedingungen (1)-(3).

Beispiel:

Ein fairer sechsseitiger Würfel wird gewürfelt. Die Person, die diesen Würfel wirft will, dass entweder eine 5 oder 6 gewürfelt wird. Die Ergebnismenge ist $\Omega = \{1, 2, 3, 4, 5, 6\}$ und das günstige Ereignis ist $A = \{5, 6\}$. Da beim Würfeln mit einem fairen sechsseitigen Würfel alle Ergebnisse gleich wahrscheinlich sind, kann die Wahrscheinlichkeit mit Formel (4) berechnet werden.

$$P(A) = \frac{|A|}{|\Omega|} = \frac{2}{6} = \frac{1}{3} = 0,3$$

Der Ausdruck $|A|$ gibt die Anzahl der Elemente in der Ereignismenge A an.

3.3.2 Wahrscheinlichkeit als Grenzwert relativer Häufigkeiten

Um die Wahrscheinlichkeit für ein Ergebnis eines Experiments zu bestimmen, gibt es zwei Möglichkeiten.

Entweder wird theoretisch (oder mathematisch) versucht, eine Wahrscheinlichkeit zu bestimmen. Bei einem fairen sechsseitigen Würfel zum Beispiel kann mit ein paar Überlegungen schnell festgestellt werden, dass die Wahrscheinlichkeit für jede Zahl gleich $\frac{1}{6}$ sein muss. Da bei einem fairen sechsseitigen Würfel sechs Ergebnisse gewürfelt werden können und kein Ergebnis bevorzugt wird (sonst wäre es kein fairer Würfel), dann erhält man zwangsläufig mit (4) die Wahrscheinlichkeit $P(X=n)=\frac{1}{6}$ für $n \in \{1, 2, 3, 4, 5, 6\}$.

Die zweite Methode ist, ein Experiment n Mal durchzuführen und dabei darauf zu achten, wie sich die relative Häufigkeit r_n für das Ereignis A verhält. Diese lässt sich wie folgt berechnen [Zwerenz Statistik S. 61]:

$$r_n = \frac{h_n}{n} \quad (5)$$

Dabei ist h_n die absolute Häufigkeit des Ereignisses. Die Wahrscheinlichkeit kann jetzt ausgedrückt werden, indem die Folge der relativen Häufigkeiten mit dem Index n gegen unendlich laufen gelassen wird. Im Grenzwert wird die Wahrscheinlichkeit dann zu [Bewersdorff Statistik S. 90-94]:

$$P(A) = \lim_{n \rightarrow \infty} r_n \quad (6)$$

Grundlage für diese Definition ist, dass n unendlich groß wird. Für ein bestimmtes n muss die relative Häufigkeit und die Wahrscheinlichkeit nicht gleich sein. Außerdem muss sich mit größer werdenden n die relative Häufigkeit der Wahrscheinlichkeit nicht konstant annähern.

3.3.3 Die axiomatische Definition der Wahrscheinlichkeit

Die oben angeführten Definitionen sind eher für den Gebrauch in der Schule gedacht. In der universitären Mathematik wird aber eine axiomatische Definition für die Wahrscheinlichkeit verwendet. Für endliche Ereignisse ist diese Art der Definition nicht so weit von der Definition nach Kolmogoroff (Die Bedingungen auf Seite 9) entfernt.

In dieser Definition werden die günstigen Fälle zu einem Ereignis $A \subset \Omega$ zusammengefasst. Alle Ereignisse sind in der Potenzmenge $\mathcal{P}(\Omega)$ enthalten. Wie schon bei der Definition von Laplace sind $\omega_i \in \Omega$ mit $i \in \mathbb{N}$, $i \leq n$, die Elementarereignisse. Daraus folgt dann:

Ist Ω eine endliche Ereignismenge, so heißt die Abbildung

$$P : \mathcal{P}(\Omega) \rightarrow [0, 1], A \rightarrow P(A)$$

ein *Wahrscheinlichkeitsmaß* auf Ω , wenn folgende Bedingungen erfüllt sind:

$$(1) P(\Omega) = 1$$

$$(2) P(A \cup B) = P(A) + P(B) \text{ für } A, B \subseteq \Omega \text{ mit } A \cap B = \emptyset$$

Das Paar (Ω, P) heißt dann *endlicher Wahrscheinlichkeitsraum*

$P(A)$ heißt *Wahrscheinlichkeit des Ereignisses A*

[Fischer 2015, S. 81]

Diese Definition gilt nur dann, wenn Ω endlich ist. Wenn Ω unendlich ist, dann ist es nicht möglich, ein geeignetes Wahrscheinlichkeitsmaß dafür anzugeben. In diesem Fall wird eine Teilmenge $\mathcal{A} \subset \mathcal{P}$ ausgewählt, sodass zumindest für diese Teilmenge und das Ereignis $A \in \mathcal{A}$ ein geeignetes Wahrscheinlichkeitsmaß $P(A)$ existiert. Die für dieses $\mathcal{A}$ wichtigen Eigenschaften werden dann beschrieben durch:

Sei $\Omega \neq \emptyset$ eine beliebige Menge. Dann heißt eine Teilmenge $\mathcal{A} \subset \mathcal{P}(\Omega)$ eine *σ -Algebra* in Ω , wenn

$$(1) \Omega \in \mathcal{A}$$

$$(2) A \in \mathcal{A} \rightarrow \bar{A} \in \mathcal{A}$$

$$(3) \text{ Falls } A_1, A_2, A_3, \dots \in \mathcal{A} \text{ abzählbar viele Mengen sind, dann ist auch} \\ \rightarrow A_1 \cup A_2 \cup A_3 \cup \dots \in \mathcal{A}$$

[Fischer 2015, S. 218]

Bei (3) muss es sich jedoch um abzählbar viele Mengen handeln. In einer überabzählbaren Menge Ω ist die Wahrscheinlichkeit eines Elementarereignisses im Allgemeinen Null. Die Wahrscheinlichkeit eines Ereignisses A ist nur durch eine σ -Algebra erklärbar. Die allgemeine Definition dazu lautet:

Sei $\Omega \neq \emptyset$ eine beliebige Menge und $\mathcal{A} \subset \mathcal{P}(\Omega)$ eine σ -Algebra in Ω . Ein *Wahrscheinlichkeitsmaß* P auf $\mathcal{A}$ ist eine Abbildung

$$P : \mathcal{A} \rightarrow [0, 1], A \rightarrow P(A)$$

mit folgenden Eigenschaften

$$(1) \quad P(\Omega) = 1$$

(2) Für abzählbar viele, paarweise disjunkte Ereignisse $A_1, A_2, A_3, \dots \in \mathcal{A}$ gilt $P(A_1 \cup A_2 \cup A_3 \cup \dots) = P(A_1) + P(A_2) + P(A_3) + \dots$.

Die Eigenschaft (1) nennt man Normiertheit und die Eigenschaft (2) heißt σ -Additivität.

Ein *Wahrscheinlichkeitsraum* ist ein Tripel $(\Omega, \mathcal{A}, \mathcal{P})$ bestehend aus

- einer Menge $\Omega \neq \emptyset$
- einer σ -Algebra $\mathcal{A} \subset \mathcal{P}(\Omega)$
- einem Wahrscheinlichkeitsmaß P auf $\mathcal{A}$

Die Elemente $\omega \in \Omega$ nennt man Ergebnisse, die Mengen $A \in \mathcal{A}$ heißen Ereignisse

[Fischer 2015, S. 220]

3.3.4 Zufallsvariable

Wie schon weiter oben beleuchtet ist die Zufallsvariable eine Zuordnung, welche einem Ergebnis eine Zahl zuordnet. Mit der axiomatischen Definition der Wahrscheinlichkeit lässt sich jetzt die Zufallsvariable definieren als [Fischer 2015, S. 89]:

In einem endlichen Wahrscheinlichkeitsraum (Ω, P) wird die beliebige Abbildung

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

Zufallsvariable genannt. Außerdem ist die Menge $X(\Omega) \subset \mathbb{R}$ für ein endliches Ω endlich. Aus dem Wahrscheinlichkeitsraum und der Zufallsvariable erhält man einen neuen

Wahrscheinlichkeitsraum $(X(\Omega), P_X)$.

Mit dieser Definition der Zufallsvariable folgt dann für die Wahrscheinlichkeitsfunktion und das Wahrscheinlichkeitsmaß

Ist (Ω, P) ein endlicher Wahrscheinlichkeitsraum und $X: \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable, dann ist durch

$$P_X(a) := P(\{X = a\}) \text{ für } a \in X(\Omega)$$

eine Wahrscheinlichkeitsfunktion auf $x(\Omega)$ gegeben. Zu dieser Wahrscheinlichkeitsfunktion gehört das Wahrscheinlichkeitsmaß

$$P_X(A) = P(\{\omega \in \Omega : X(\omega) \in A\}) \text{ für } A \subset X(\Omega)$$

[Fischer 2015, S. 89]

Das gilt, da für $X(\Omega) = \{a_1, \dots, a_n\}$ gilt, dass

$$P_X(a_1) + \dots + P_X(a_m) = 1$$

ist, was wiederum aus der Tatsache folgt, dass

$$\{X = a_1\} \cup \dots \cup \{X = a_m\} = \Omega$$

eine disjunkte Zerlegung ist.

In einem allgemeinen Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$ gilt für die Zufallsvariable (mit $B(\mathbb{R})$ als die Borel σ -Algebra)

Sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine Abbildung $X: \Omega \rightarrow \mathbb{R}$ heißt Zufallsvariable, falls für $A \in B(\mathbb{R})$ stets $\{\omega \in \Omega : X(\omega) \in A\} \in \Sigma$ liegt.

Ist das gültig, dann heißt das Wahrscheinlichkeitsmaß $\mathbb{P}_X$ auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ die Verteilung von X und ist gegeben durch

$$P_X(A) := \mathbb{P}(X \in A)$$

[Kampf, Kunze 2016, S. 32]

Es ist schwer, die Verteilung konkret anzugeben, da sie nicht mehr durch Werte von einelementigen Mengen bestimmt werden kann. Für die Verteilungsfunktion gilt daher

Sei $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable, dann heißt $F_X : \mathbb{R} \rightarrow [0,1]$ Verteilungsfunktion von X und ist gegeben durch

$$F_X := \mathbb{P}(X \leq x)$$

Ein Beispiel für eine Verteilungsfunktion wäre, wenn X Bernoulliverteilt ist.

Dann gilt

$$F_X(x) = 0 \text{ für } x < 0; \quad 1 - p \text{ für } 0 \leq x < 1 \text{ und } 1 \text{ für } x \geq 1$$

[Kampf, Kunze 2016, S. 32]

3.3.5 Erwartungswert und Varianz

Eine sehr wichtige Kennzahl im Zusammenhang mit der Zufallsvariable ist der Erwartungswert $E(X)$. Der Erwartungswert gibt an, welches Ergebnis man in einer Versuchsreihe erwarten kann. Dieser muss nicht gleich sein mit einem möglichen Ergebnis. Der Erwartungswert $E(X)$ ist gegeben als

$$E(X) = P(\omega_1) \cdot \omega_1 + \dots + P(\omega_n) \cdot \omega_n$$

Ein Beispiel ist ein fairer, sechsseitiger Würfel. Dieser hat einen Erwartungswert von

$$E(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3,5$$

Für zwei Zufallsvariablen X und Y, die auf der Basis desselben Zufallsexperimentes beruhen, gelten folgende Sätze [Bewersdorff 2011, S. 55-56]:

- 1) $E(aX+b) = aE(X) + b$, $a, b \in \mathbb{R}$ und konstant
- 2) $E(X+Y) = E(X) + E(Y)$

Sind die zwei Zufallsvariablen unabhängig voneinander (Also sind alle Ergebnisse von X stochastisch unabhängig von Y und umgekehrt) dann gelten außerdem noch

- 1) $P(X = x \wedge Y = y) = P(X = x) \cdot P(Y = y)$
- 2) $E(XY) = E(X) \cdot E(Y)$ (Multiplikationssatz)

Auf die Beweise wird hier verzichtet, sie sind aber durch leichte Überlegungen beweisbar.

In der Mathematik wird mit der Varianz die Streuung einer Zufallsvariable um den Erwartungswert beschrieben. Diese ist definiert als

$$V(X) = P(X = x_1) \cdot (x_1 - E(X))^2 + \dots + P(X = x_n) \cdot (x_n - E(X))^2 \quad (7)$$

wobei X die Werte $x_1, \dots, x_n$ annehmen kann. Die Standardabweichung ist dann gegeben als

$$\sigma_x = \sqrt{V(X)}$$

Beim Beispiel mit dem Würfel würde nun für die Varianz

$$V(X) = \frac{1}{6} \cdot \left(\frac{5}{6}\right)^2 + \frac{1}{6} \cdot \left(\frac{3}{6}\right)^2 + \frac{1}{6} \cdot \left(\frac{1}{6}\right)^2 + \frac{1}{6} \cdot \left(\frac{1}{6}\right)^2 + \frac{1}{6} \cdot \left(\frac{3}{6}\right)^2 + \frac{1}{6} \cdot \left(\frac{5}{6}\right)^2 \approx 2,92$$

und für die Standardabweichung

$$\sigma_x = \sqrt{V(X)} \approx 1,71$$

Exkurs: Die Tschebyschev'sche Ungleichung

Wenn die Verteilungsfunktion von einer Zufallsvariable bekannt ist, dann kann die Wahrscheinlichkeit

$$P(|X - \mu| \geq c)$$

genau bestimmt werden. Die Tschebyschev'sche Ungleichung trifft eine Aussage über die Abweichung einer unbekannten Zufallsvariable von einem Erwartungswert mithilfe der dazugehörigen Varianz. In vielen Fällen ist zwar aus verschiedenen Erfahrungen ein Erwartungswert μ und eine Varianz V , jedoch keine dazugehörige Zufallsvariable X bekannt. In diesen Fällen kann man die Varianz hernehmen, um eine Aussagen über die Abweichung der Werte einer Zufallsvariable zu einem μ zu treffen.

Tschebyschev'sche Ungleichung

Sei X eine Zufallsvariable mit dem Erwartungswert μ und Varianz V . Dann gilt für jede Zahl $c > 0$ die Ungleichung

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

Beweis: Der Beweis wird für eine diskrete Zufallsvariable geführt. Für stetige und allgemeine Zufallsvariablen funktioniert der Beweis analog. Die Verteilung von X sei für jedes x_i existiert ein $P(X=x_i)$, $i = 1, 2, \dots$. Wird in (7) nur über die Werte x_i mit $|x_i - \mu| \geq c$ summiert, so folgt

$$V \geq \sum_{|x_k - \mu| \geq c} (x_k - \mu)^2 \cdot P(X = x_k)$$

Auf der rechten Seite der Ungleichung gilt wegen $(x_k - \mu)^2 \geq c^2$ für die einzelnen Summanden

$$(x_k - \mu)^2 \cdot P(X = x_k) \geq c^2 \cdot P(X = x_k)$$

Daraus folgt dann

$$V \geq c^2 \sum_{|x_k - \mu| \geq c} P(X = x_k) = c^2 \cdot P(|X - \mu| \geq c)$$

Dividiert man jetzt noch die ganze Ungleichung durch c^2 , dann erhält man die Aussage des Satzes

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

[Bosch 2010, S. 149-150]

Ist die Wurzel der Varianz größer oder gleich c , dann kann mit dieser Ungleichung keine Information gewonnen werden, da dann der Quotient auf der rechten Seite nicht kleiner als 1 ist.

Geht man jetzt zurück zur Definition der Wahrscheinlichkeit über die relative Häufigkeit, kann man eben diese relative Häufigkeit jetzt als Zufallsvariable betrachten. Um den Abstand zwischen der relativen Häufigkeit und der Wahrscheinlichkeit p zu ermitteln, kann mit der Ungleichung von Tschebyschow eine Ungleichung bilden. Diese trifft eine Aussage darüber, wie stark die relative Häufigkeit eines Ereignisses A von der Wahrscheinlichkeit p abweicht:

$$P(|R_n - p| \geq \sigma) \leq \frac{p(1-p)}{n\sigma^2} \quad (8)$$

Man beachte dabei, dass die Varianz gleich $p(1-p)$ ist. Aus dieser Ungleichung lässt sich herauslesen, dass der Bruch auf der rechten Seite bei größerem n immer kleiner wird, da p konstant ist (und in weiterer Folge auch $(1-p)$) und der Bruch durch größeres n auch immer kleiner wird. Dadurch wird die Wahrscheinlichkeit einer großen Abweichung zwischen r_n und p immer kleiner.

Beispiel:

Bei einem Spiel, wie *Mensch ärgere dich nicht*, wird über viele Spiele hinweg dokumentiert, wie oft ein Sechser gewürfelt wird. Dabei wird immer unter den selben Bedingungen gespielt, also auch mit dem selben Würfel.

Die Spieler haben in ihren letzten Partien insgesamt 12000 Mal den Würfel geworfen und dabei 1600 Mal die Sechs gewürfelt. Dieser Wert liegt um 400 Mal unter dem Erwartungswert von 2000, welcher sich mit der Formel

$$E(X) = n \cdot p \quad (9)$$

berechnen lässt.

Der Verdacht der Spieler ist, dass der Würfel gar kein fairer Würfel ist. Sie haben den Verdacht, dass dieser Würfel den Sechser nicht mit einer Wahrscheinlichkeit von $p = \frac{1}{6}$, sondern mit einer geringeren Wahrscheinlichkeit würfelt. Jetzt kann die Ungleichung (8) herangezogen werden.

$$P(|r_n - \frac{1}{6}| \geq \frac{1}{30}) \leq \frac{15}{36 \cdot 40} = \frac{1}{96}$$

Diese Rechnung zeigt, dass die Wahrscheinlichkeit bei einer so großen Abweichung der Messung vom Erwartungswert bei 12000 Würfeln bei $\frac{1}{96}$ oder ungefähr bei 1% liegt.

Das zeigt, dass ein solches Ereignis sehr unwahrscheinlich, jedoch nicht unmöglich ist.

Die Vermutung, dass dieser „faire“ Würfel in Wahrheit kein fairer Würfel sei, wird durch diese Rechnung zwar bestärkt, es lässt sich jedoch der Verdacht dadurch nicht bestätigen, da unwahrscheinliche Ereignisse in der Realität sehr häufig vorkommen, wie in Folge dieser Masterarbeit noch stärker beleuchtet wird.

3.3.6 Das schwache Gesetz der großen Zahlen

Bei einem Bernoulli-Experiment kann die Zufallsvariable X nur die Werte 0 oder 1 annehmen. Aus diesem Grund wird diese auch Null-Eins Verteilung genannt. Die zwei möglichen Ausgänge werden oft mit Erfolg oder 1 bzw. Misserfolg oder 0 bezeichnet. Die Wahrscheinlichkeit p ist dabei die Wahrscheinlichkeit für den Erfolg (auch Erfolgswahrscheinlichkeit) und $q = 1 - p$ die Gegenwahrscheinlichkeit. Der Erwartungswert $E(X)$ ist hier p und die Varianz ist pq . Die Bernoulli-Verteilung Beispiele für Zufallsexperiment, welches einer Bernoulli-Verteilung genügt ist ein Münzwurf oder der Wurf eines Würfels, bei dem man auf ein bestimmtes Ergebnis setzt [Mittag-Schüller, S. 183].

Ein Bernoulli-Prozess ist eine Aneinanderreihung von stochastisch unabhängigen Bernoulli-Experimenten. Hier werden n Durchführungen in einer zeitlichen Abfolge dargestellt. Das Ergebnis von der ersten Durchführung wird mit der Zufallsvariable X_1 beschrieben, die zweite Durchführung X_2 und so weiter. Die entstandene Folge ist entweder endlich oder abzählbar unendlich. Ein Beispiel für ein Bernoulli-Prozess mit $n=10$ und $p=0,4$ ist $(1,0,0,0,1,1,0,1,0,0)$.

Interessiert man sich in einem Bernoulli-Prozess mit n Durchführungen für die Anzahl der Erfolge k , dann wäre die dazugehörige Zufallsvariable binomialverteilt. Die Binomialverteilung ist eine diskrete Verteilung und verlangt nach einem Bernoulli-Prozess. Im einfachsten Fall der Binomialverteilung ($n=1$) hat man wieder eine Bernoulli-Verteilung. Eine zweite Verteilung, welche aus dem Bernoulli-Prozess resultiert ist die geometrische Verteilung. Die Zufallsvariable X gibt dabei an, wie viele Versuche es gebraucht hat, um den ersten Erfolg zu erzielen. Eine andere Möglichkeit die geome-

trische Verteilung zu interpretieren wäre, wenn die Zufallsvariable Y die Anzahl der Fehlversuche bis zum ersten Erfolg angibt. Für die beiden Zufallsvariablen gilt, dass $X = Y + 1$ ist.

Der Sachverhalt aus (5) lässt sich als das schwache Gesetz der großen Zahlen interpretieren. Der genaue Satz variiert von Quelle zu Quelle, die zugrunde liegende Idee ist aber immer sehr ähnlich. Die einfachste Form das schwache Gesetz der großen Zahl zu beschreiben, geht über ein Bernoulli-Experiment. Bernoulli hat anhand des Bernoulli-Experiments das schwache Gesetz der großen Zahlen postuliert. Dabei hat er wie in (6) mit relativen Häufigkeiten argumentiert.

Wichtig zu erwähnen ist dabei, dass sich die Annäherung der relativen Häufigkeit an die theoretische Wahrscheinlichkeit nicht gleichmäßig annähern muss, da es in der Realität immer Ausreißer geben kann. In Abb. 2 kann der Verlauf der relativen Häufigkeit für eine faire Münze mit $n=1000$ Wiederholungen betrachtet werden. In der Abbildung kann man erkennen, dass sich die relative Häufigkeit r_n für größere n an die Wahrscheinlichkeit p (der blaue Strich) zwar annähert, jedoch nicht gleichmäßig. Bei einem n von ca. 400 entfernt sich r_n tatsächlich wieder von p .

Allgemeine Definition und Beweis

In einem Bernoulli-Prozess mit der Länge n und der relative Häufigkeit R_n gilt:

Sei A ein Ereignis in einem Zufallsexperiment mit der dazugehörigen Wahrscheinlichkeit $P(A) = p$. Die relative Häufigkeit von A bei n Wiederholungen in einem Bernoulli-Prozess (n unabhängige Wiederholungen) ist R_n und eine Zufallsvariable. Dann gilt für jede positive Zahl a :

$$\lim_{n \rightarrow \infty} P(|R_n - p| < a) = 1$$

Beweis

Wir fassen die absolute Häufigkeit, die das Ereignis A in den n Wiederholungen auftritt, als Zufallsvariable auf und bezeichnen sie mit X. Für diese Zufallsvariable gilt für den Erwartungswert und der Varianz:

$$E(X) = n \cdot p \qquad V(X) = n \cdot p \cdot (1 - p)$$

Für die relative Häufigkeit R_n gilt dann

$$R_n = \frac{X}{n}$$

Daraus folgt dann für den Erwartungswert und der Varianz

$$E(R_n) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{n \cdot p}{n} = p$$

$$V(R_n) = V\left(\frac{X}{n}\right) = \frac{1}{n^2}V(X) = \frac{n \cdot p \cdot (1 - p)}{n^2} = \frac{p \cdot (1 - p)}{n}$$

Hier wird jetzt die Ungleichung von Tschebyschev hergenommen und für jede positive Zahl a erhält man dann

$$P(|R_n - E(R_n)| \geq a) \leq \frac{V(R_n)}{a^2}$$

setzt man für $E(R_n)$ und $V(R_n)$ die Terme von oben ein, dann erhält man

$$P(|R_n - p| \geq a) \leq \frac{p \cdot (1 - p)}{n \cdot a^2}$$

Und daraus folgt dann

$$\lim_{n \rightarrow \infty} P(|R_n - p| \geq a) = 0$$

Mit der Gegenwahrscheinlichkeit erhält man dann schlussendlich

$$\lim_{n \rightarrow \infty} P(|R_n - p| < a) = 1$$

[Kütting; Sauer 2013, S. 279-281]

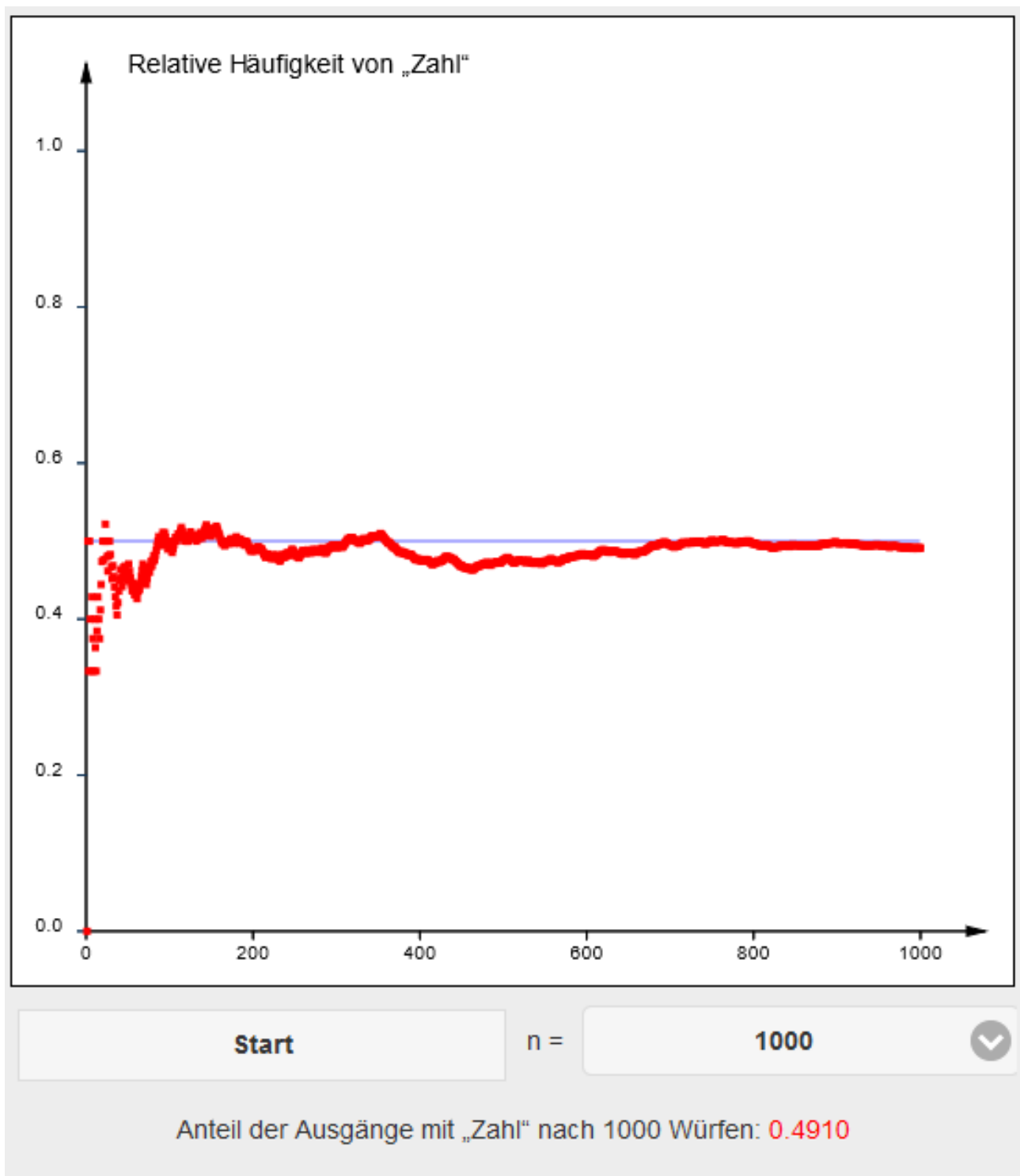


Abbildung 2: Entwicklung der relativen Häufigkeit des Ergebnisses Zahl für eine faire Münze bei $n=1000$ [<https://stat.iks-hagen.de/app/muenzwurf/>]

3.4 Das Urnenmodell

Bei der Laplace-Wahrscheinlichkeit muss zuerst ermittelt werden, wie groß die Anzahl der Elemente in der Menge der Elementarereignisse ist. Um die Anzahl der günstigen Ereignisse zu ermitteln, müssen kombinatorische Überlegungen gemacht werden. Die mathematische Disziplin ist dazu die Kombinatorik [Mittag-Schüller S. 166].

Ein Modell, welches die Realität gut abbilden kann, ist das Urnenmodell. Dieses ist besonders praktisch, da es im Fall einer endlichen Ereignismenge, welche in der Realität fast immer gegeben ist, sehr gut funktioniert und anschaulich ist. Beim Urnenmodell stellt man sich eine Urne mit einer endlichen Anzahl an Kugeln n vor. Diese werden dann durchnummeriert. Die günstigen Fälle des Ereignisses A haben dann einen Umfang von k , wobei $k \leq n$ sein muss. Jede Kugel muss mit gleicher Wahrscheinlichkeit aus der Urne gezogen werden. Also hat jede Kugel eine Wahrscheinlichkeit aus der Urne gezogen zu werden von $p = \frac{1}{n}$ [Mittag-Schüller S. 166].

Dieser Vorgang der Auswahl von k Kugeln aus der Grundgesamtheit n wird als Zufallsstichprobe bezeichnet. Jetzt wird noch unterschieden, ob nach dem Ziehen einer Kugel, die Kugel zurückgelegt wird oder nicht. Im Fall, dass die gezogene Kugel zurückgelegt wird, spricht man von einer Ziehung mit Zurücklegen, im Fall, dass sie nicht zurückgelegt wird spricht man von einer Ziehung ohne Zurücklegen.

Ein Beispiel für eine Ziehung mit Zurücklegen ist das Werfen eines Würfels. In diesem Beispiel sind sechs Kugeln in der Urne, durchnummeriert von eins bis sechs. Wenn eine Kugel gezogen wurde, wird diese anschließend wieder für den nächsten Wurf zurückgelegt. Wenn ein Würfel 10 Mal geworfen wird, ist in diesem Fall die Ausgangslage vor jedem Wurf gleich. In diesem Beispiel kann das Zufallsexperiment auch unendlich oft wiederholt werden.

Ein Beispiel für eine Ziehung ohne Zurücklegen ist die Lottoziehung. In Österreich gibt es 45 verschiedene Zahlen zur Auswahl, von denen sechs gezogen werden. Nach jeder Ziehung wird die gezogene Zahl nicht mehr zurück in die Urne geworfen. Dadurch ändert sich nach jeder Ziehung die Wahrscheinlichkeit für die verbliebenen Zahlen in der Urne. Außerdem können nur 45 Zahlen aus der Urne gezogen werden. Danach kann das Experiment nicht mehr fortgeführt werden, da sich keine Kugeln mehr in der Urne befinden.

Beim Ziehen von Kugeln aus einer Urne muss nicht nur beachtet werden, ob die Kugeln nach der Ziehung wieder zurückgelegt werden oder nicht, sondern auch ob die Reihenfolge der gezogenen Kugeln eine Rolle spielt oder nicht. Spielt die Reihenfolge der Ziehung eine Rolle, dann spricht man von einer geordneten Auswahl. Sollte die Reihenfolge der Ziehung keine Rolle spielen, dann spricht man von einer ungeordneten Auswahl.

Ein Beispiel für eine geordnete Auswahl wäre, wenn in einer Schulklasse der/die Klassensprecher*in und der/die Stellvertreter*in zufällig ermittelt werden. Die erste gezogene Person wird der/die neue Klassensprecher*in und die zweite Person der/die Stellvertreter*in.

Ein Beispiel für eine ungeordnete Auswahl ist wieder die Lottoziehung. In dieser spielt die Reihenfolge, in der die Kugeln gezogen werden, keine Rolle.

Um konkret die Wahrscheinlichkeit bestimmen zu können, muss zuerst die Art des Experiments bestimmt werden. Bei einer Lottoziehung liegt eine ungeordnete Auswahl ohne Zurücklegen vor. In der ersten Ziehung wird aus 45 Kugeln gezogen. Es gibt daher 45 verschiedene Zahlen, die gezogen werden können. Da die gezogene Kugel nicht zurückgelegt wird, sind bei der zweiten Ziehung nur noch 44 Kugeln in der Urne. In weiterer Folge sind es dann bei der dritten Ziehung 43, bei der vierten Ziehung 42, bei

der fünften Ziehung 41 und bei der sechsten Ziehung 40 Kugeln. In einer Lottoziehung müssen sechs verschiedene Kugeln gezogen werden. Daher ist die Anzahl der möglichen Fälle, das Produkt der Möglichkeiten der einzelnen Ziehungen von der Anzahl an Kugeln in der Urne, die es bei den sechs Ziehungen gibt.

In der ersten Ziehung gibt es sechs günstige Fälle (da beim Lotto aus 45 Zahlen sechs ausgesucht werden und noch keine gezogen wurde). In weiterer Folge verringert sich nach jeder Ziehung die Anzahl der günstigen Fälle um eins. Um die Wahrscheinlichkeit für einen Lotto-Sechser zu bestimmen, wird die Formel (4) für die Laplace Wahrscheinlichkeit verwendet:

$$P(\text{Sechs richtige}) = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{45 \cdot 44 \cdot 43 \cdot 42 \cdot 41 \cdot 40} = \frac{720}{586444} = 1,23 \cdot 10^{-7} \quad (10)$$

Alternativ kann bei einer Zufallsstichprobe ohne Zurücklegen, bei der die Reihenfolge keine Rolle spielt, der Binomialkoeffizient verwendet werden, wobei gilt $n, k \in \mathbb{N}$:

$$\binom{n}{k} = \frac{n!}{k! \cdot (n - k)!} \quad (11)$$

Hier gibt die Zahl n die Anzahl der Kugeln in der Urne an und die Zahl k die Anzahl der Kugeln, die aus der Urne gezogen werden sollen. Das wäre im Fall der Lottoziehung:

$$\binom{45}{6} = 8145060$$

Beim Lotto-Sechser sollen aus sechs Ziehungen sechs bestimmte Zahlen gezogen werden. In diesem Fall ist der Binomialkoeffizient $\binom{6}{6} = 1$. Daher gibt es genau einen günstigen Fall. Daraus folgt dann wieder aus Formel (4):

$$P(\text{Sechs richtige}) = \frac{1}{\binom{45}{6}} = \frac{1}{8145060} = 1,23 \cdot 10^{-7} \quad (12)$$

Sowohl (10) als auch (12) führen zum gleichen Ergebnis. Welcher dieser beiden Wege gewählt wird, spielt für das Ergebnis also keine Rolle.

Würden bei der Lottoziehung die Kugeln wieder zurückgelegt werden, dann müsste für die Anzahl der möglichen Fälle anders gerechnet werden. Die Anzahl der Kugeln bleibt für jede Ziehung konstant bei 45. Wird von n Kugeln in der Urne ausgegangen, dann ist die Anzahl der Möglichkeiten bei k Ziehungen [Bourier 2014, S. 121]:

$$\binom{n+k-1}{k} \quad (13)$$

Ein Beispiel für eine Ziehung ohne Zurücklegen, bei der die Reihenfolge wichtig ist, wäre, wenn in einer Urne 7 Kugeln enthalten sind und drei Mal daraus gezogen wird. Für die erste Ziehung gibt es 7 verschiedene Möglichkeiten, bei der zweiten dann nur noch 6 und bei der dritten Ziehung sind es dann noch 5 Kugeln, die gezogen werden können. Daraus ergibt sich dann:

$$\frac{7!}{4!} = 7 \cdot 6 \cdot 5 = 210$$

Allgemein lässt sich daraus die Formel für n Kugeln und k Ziehungen angeben als [Bourier 2014, S. 122]:

$$\frac{n!}{(n-k)!} = n \cdot n-1 \cdot \dots \cdot n-k+1 \quad (14)$$

Spielt bei einer Ziehung die Reihenfolge eine Rolle, wird jedoch nach jeder Ziehung die Kugel wieder zurück in die Urne gelegt, dann gibt es bei jeder Ziehung genau n verschiedene Möglichkeiten für das Ziehen einer Kugel. Bei k Ziehungen ergibt sich die Formel dann als [Bourier 2014, S. 120]:

$$n^k \quad (15)$$

3.5 Bedingte Wahrscheinlichkeit

Wahrscheinlichkeiten können sich durch verschiedene Zusatzbedingungen ändern. Man betrachtet als Beispiel eine Familie mit zwei Kindern. In diesem Beispiel wird davon ausgegangen, dass die Kinder keine Zwillinge sind (dass es also ein älteres und ein jüngeres Kind gibt) und dass die Wahrscheinlichkeit, dass das Kind bei der Geburt weiblich w oder männlich m ist, gleich groß ist. Mit diesen Bedingungen ergibt sich dann, dass die möglichen Fälle in der Form (Geschlecht älteres Kind, Geschlecht jüngeres Kind) gegeben sind. Daraus resultieren dann ohne Zusatzinformation die vier möglichen Fälle (m,m) , (m,w) , (w,m) und (w,w) .

Jetzt soll die Wahrscheinlichkeit berechnet werden, die den Fall „beide Kinder sind weiblich“ beschreibt. Der einzige Fall, der diesen Sachverhalt beschreibt, ist (w,w) . Mit (4) folgt dann die Anzahl der möglichen Fälle als 4 und die Anzahl der günstigen Fälle als 1 und daher:

$$P(\text{beide Kinder weiblich}) = \frac{1}{4}$$

Weiß man jetzt zusätzlich, dass das ältere Kind sicher weiblich ist, dann reduzieren sich die möglichen Fälle, da nur noch die Fälle (w,m) und (w,w) möglich sind. Daraus resultiert die Wahrscheinlichkeit:

$$P(\text{beide Kinder weiblich} \mid \text{das ältere Kind ist weiblich}) = \frac{1}{2}$$

Die Schreibweise $P(\text{Ereignis A} \mid \text{Bedingung B})$ wird bedingte Wahrscheinlichkeit genannt. Diese gibt die Wahrscheinlichkeit an, dass ein Ereignis A unter der Bedingung B eintritt. Theoretisch kann das Eingangsbeispiel auch als bedingte Wahrscheinlichkeit angeschrieben werden, und zwar als $P(\text{alle Kinder sind weiblich} \mid \text{es sind zwei Kinder})$. In diesem Fall ist diese Schreibweise aber unnötig kompliziert und sollte, vor allem in der Hinsicht auf den Schulunterricht, so nicht angeschrieben werden. Dieser Punkt wird

in der Masterarbeit später noch aufgegriffen.

Die Bedingung bei der bedingten Wahrscheinlichkeit kann als Ereignis betrachtet werden, welches mit einer Wahrscheinlichkeit von 1 eintritt. Wenn man es sich nun als Venn-Diagramm vorstellt, wie in Abb. 3 zu sehen, dann kann für die bedingte Wahrscheinlichkeit die Menge des Ereignisses B als alle möglichen Fälle interpretiert werden, eben mit der Begründung, dass Fälle, die nicht in B enthalten sind, durch die Bedingung schon im Vorhinein ausgeschlossen sind. Daher sind die Fälle, die nicht in B enthalten sind, nicht mehr möglich.

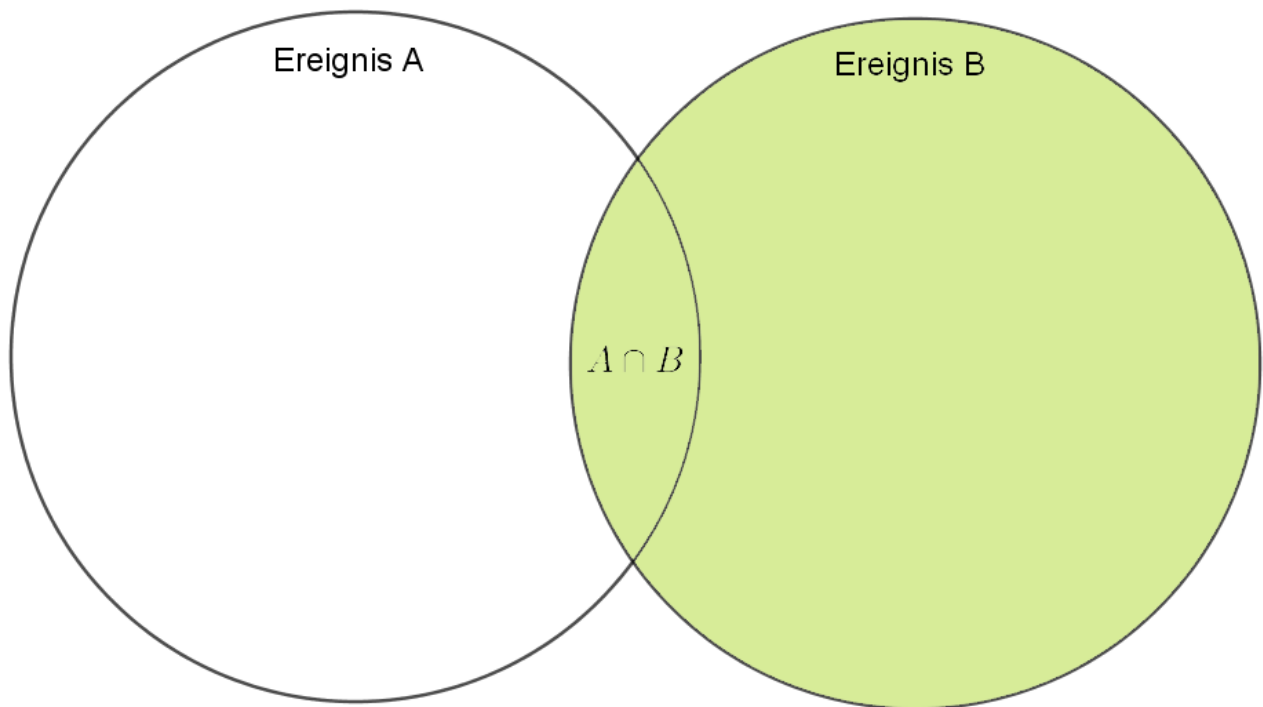


Abbildung 3: Venn-Diagramm mit der Ereignismenge B als neue Menge der möglichen Fälle

Betrachtet man jetzt das Ereignis A, dann muss der gesamte Teil der Menge, welcher zwar in A enthalten ist, in B jedoch nicht, ebenfalls ignoriert werden. Die günstigen Fälle beinhalten daher nur noch die Fälle, welche ebenfalls in B enthalten sind. Daher sind die günstigen Fälle nur noch Fälle aus der Menge $A \cap B$. Das dazugehörige Venn-Diagramm ist in Abb. 4 zu sehen.

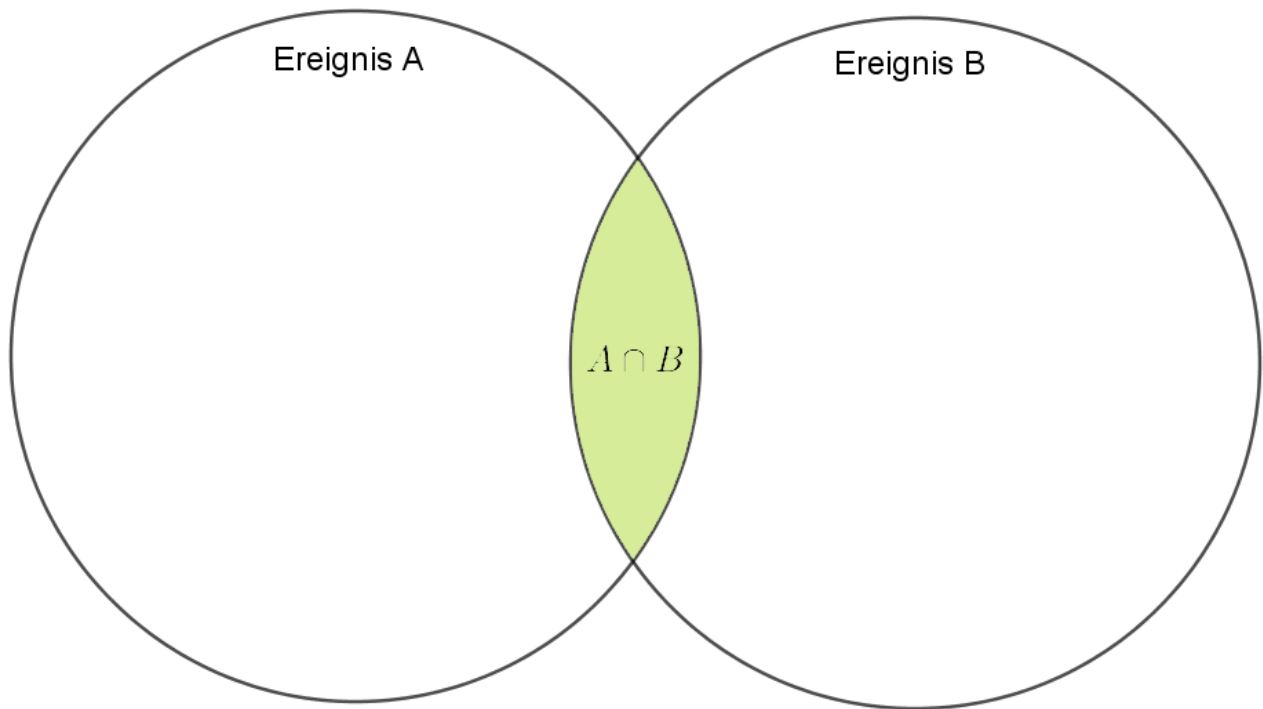


Abbildung 4: Venn-Diagramme mit der Ereignismenge $A \cap B$ als neue Menge der günstigen Fälle

Aus diesen Überlegungen folgt, dass die Anzahl der Ereignisse in der Menge $A \cap B$ die Anzahl der günstigen Ereignisse und die Anzahl der Ereignisse in B die Anzahl der möglichen Ereignisse ist. Wenn Ω eine endliche Grundmenge ist und die Einzelereignisse in Ω gleich wahrscheinlich sind, dann gilt für die Ereignisse A und B , $A, B \subset \Omega$ folgt aus (4) die bedingte Wahrscheinlichkeit:

$$P(A | B) = \frac{\text{Anzahl der Ereignisse in } A \cap B}{\text{Anzahl der Ereignisse in } B} \quad (16)$$

Diese naive Definition der bedingten Wahrscheinlichkeit ist für die Schule ausreichend. Auf axiomatischer Ebene betrachtet man einen Wahrscheinlichkeitsraum $(\Omega, \mathbb{A}, P)$ mit den Ereignissen $A, B \in \mathbb{A}$ und $P(B) > 0$. Dann ist die Wahrscheinlichkeit von A unter der Bedingung B als

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

Im Wahrscheinlichkeitsraum $(\Omega, \mathbb{A}, P)$ sind die Ereignisse $A, B \in \mathbb{A}$ dann unabhängig, wenn gilt

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

Mithilfe dieser Definition lässt sich dann noch zeigen, dass für denselben Wahrscheinlichkeitsraum $(\Omega, \mathbb{A}, P)$ die Ereignisse genau dann unabhängig sind, wenn $P(A|B)=P(A)$ ist ($A, B \in \mathbb{A}$, $P(B) > 0$).

Beweis:

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

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

$$(\Leftarrow): P(A \cap B) = P(A|B) \cdot P(B) = P(A) \cdot P(B) \quad (\text{wegen } P(A|B) = P(A))$$

3.6 Satz von Bayes

Aus (4) folgt dann analog zu $P(A)$ noch folgendes [Mittag-Schüller 2020, S. 171]:

$$P(B) = \frac{\text{Anzahl der für B günstige Ereignisse}}{\text{Anzahl der möglichen Ereignisse}} \quad (17)$$

$$P(A \cap B) = \frac{\text{Anzahl der für } A \cap B \text{ günstige Ereignisse}}{\text{Anzahl der möglichen Ereignisse}} \quad (18)$$

Wird jetzt (18) mit dem Kehrwert von (17) multipliziert, dann resultiert daraus

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad (19)$$

Das $P(A \cap B)$ ist die Wahrscheinlichkeit, dass beide Ereignisse A und B auftreten (Abb. 4). Wenn (Ω, P) ein endlicher Wahrscheinlichkeitsraum und $A, B \subset \Omega$ ist und außerdem noch $P(B) > 0$ gilt, dann ist (19) die bedingte Wahrscheinlichkeit von Ereignis A unter der Bedingung B .

Analog dazu resultiert aus (4) und (18)

$$P(B | A) = \frac{P(A \cap B)}{P(A)} \quad (20)$$

Aus (19) und (20) folgt dann für $P(A \cap B)$ wiederum der Multiplikationssatz

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

Daraus folgt dann der Satz von Bayes

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \quad (21)$$

Hier sei noch angemerkt, dass der Satz von Bayes zwar mathematisch seine Gültigkeit

hat, jedoch die Definition der bedingte Wahrscheinlichkeit auf einer schulischen Ebene viel wichtiger ist. Ein gutes Verständnis zu diesem Thema wird nicht durch das Lernen einer Formel erzielt sondern viel eher durch das Verstehen der Definition der bedingten Wahrscheinlichkeit.

3.7 Baumdiagramm

Eine Grundlage und wichtiges Werkzeug für das Verstehen von Wahrscheinlichkeiten im Allgemeinen und speziell bedingten Wahrscheinlichkeiten ist das Baumdiagramm. Dieses bietet eine anschauliche Möglichkeit komplexere Sachverhalte darzustellen. Nachteil ist, dass langwierige Sachverhalte mit einem Baumdiagramm sehr viel Platz benötigen und unübersichtlich werden.

Ein Baumdiagramm stellt ein Experiment da. In diesem werden Pfade und Knoten eingezeichnet. Knoten repräsentieren Ergebnisse, welche unter bestimmten Umständen eintreten können. Pfade führen von Knoten zu Knoten, für gewöhnlich von oben nach unten. Diese repräsentieren die möglichen Wege, die von Knoten zu Knoten genommen werden können. Diesen Pfaden wird für gewöhnlich eine Wahrscheinlichkeit zugeordnet. Ein Beispiel für ein Baumdiagramm ist in Abbildung 5 zu sehen. Die Ereignisse A, B und C sind die möglichen Ereignisse im ersten Schritt und die Ereignisse D und E die Möglichkeiten im zweiten Schritt. Die Ereignisse A, B und C im ersten Schritt repräsentieren die Knoten des Diagramms und die Pfeile von Knoten zu Knoten sind die Äste. Die Wahrscheinlichkeiten von D und E sind vom vorhergegangenen Ergebnis abhängig. Wenn zuerst Ereignis A und dann Ereignis D eintritt, dann spricht man hier in diesem Kontext vom Ereignis $A \cap D$.

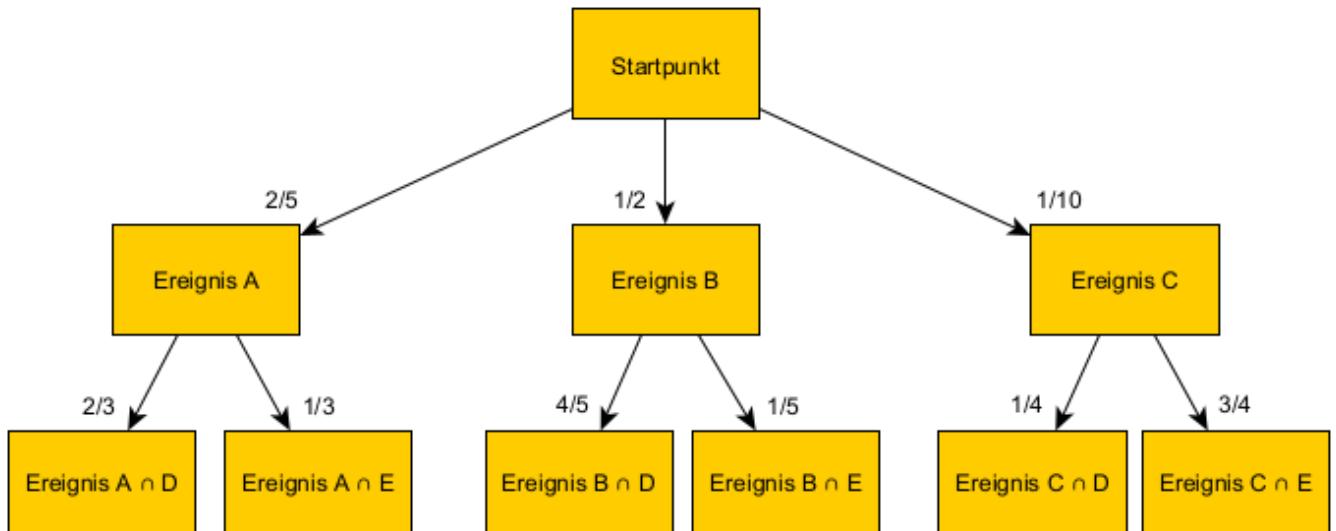


Abbildung 5: Zweistufiges Baumdiagramm für Ereignisse

3.7.1 Pfadregeln

Grundlage für das Benutzen von Baumdiagrammen ist die Kenntnis der drei Pfadregeln. [Mittelbach 1992, S. 70].

1. Pfadregel

Die erste Pfadregel besagt, dass die Summe der Wahrscheinlichkeiten aus allen aus einem Knoten ausgehenden Ästen immer Eins sein muss. Das bedeutet, dass in einem mehrstufigen Experiment von einem Knoten aus betrachtet die nächste Stufe von diesem Knoten aus garantiert erreicht werden muss. Es kann nicht sein, dass zwar Äste von diesem Knoten ausgehen, aber dieser Knoten nicht zu 100% verlassen werden kann. Die zweite Möglichkeit ist, wenn das Experiment nach diesem Schritt beendet ist. In diesem Fall führen keine Äste mehr aus diesem Knoten heraus.

In Abbildung 5 zum Beispiel muss die Summe der Ereignisse, die aus den Knoten „Ereignis A“ herausgehen genau eins sein. Diese Voraussetzung ist in diesem Fall auch erfüllt. Der Knoten „Ereignis A $\cap$ D“ ist ein Endknoten, daher gehen aus diesen Knoten auch keine Äste mehr.

2. Pfadregel

Die zweite Pfadregel besagt, dass in einem mehrstufigen Zufallsexperiment die Wahrscheinlichkeit eines Ereignisses das Produkt der Wahrscheinlichkeiten auf den Ästen ist, welche zu diesem Ereignis führen. Anders gesagt ist die Wahrscheinlichkeit für ein Ereignis in einem Knoten das Produkt der Wahrscheinlichkeiten, die auf dem Weg liegen, die zu diesem Knoten führen.

In Abbildung 5 ist zum Beispiel die Wahrscheinlichkeit vom Ereignis A gegeben als $P(A) = \frac{2}{5}$, da die Wahrscheinlichkeit vom Start zum Knoten mit $\frac{2}{5}$ gegeben ist und dies auch der einzige Schritt ist. Die Wahrscheinlichkeit vom Knoten mit „Ereignis $B \cap E$ “ ist dann gegeben als

$$P(B \cap E) = \frac{1}{2} \cdot \frac{1}{5} = \frac{1}{10}$$

3. Pfadregel

Die dritte Pfadregel besagt, dass die Wahrscheinlichkeiten von parallelen günstigen Ereignissen summiert werden müssen. Führen in einem Baumdiagramm zwei verschiedene Wege zu einem günstigen Ereignis, dann können diese zwei Pfade zusammenaddiert werden. Das funktioniert analog für mehr als zwei günstige Pfade. Werden alle Pfade miteinander aufsummiert, dann erhält man in Folge wieder eins.

Sind in Abbildung 5 die Ereignisse A und C günstige Ereignisse, dann ist die Wahrscheinlichkeit $P(A \vee C) = \frac{1}{2}$. Sind die Ereignisse „ $A \cap E$ “, „ $B \cap D$ “ und „ $C \cap E$ “ günstige Ereignisse, dann ist die Wahrscheinlichkeit gegeben als

$$P((A \cap E) \vee (B \cap D) \vee (C \cap E)) = \frac{2}{5} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{4}{5} + \frac{1}{10} \cdot \frac{1}{4} = \frac{67}{120}$$

4 Fachdidaktische Überlegung zur bedingten Wahrscheinlichkeit

In diesem Kapitel soll auf das benötigte Vorwissen der Schüler*innen eingegangen werden und anschließend ein mögliches Konzept zur Einführung der bedingten Wahrscheinlichkeit sowie mögliche Probleme beim Erlernen bzw. Fehlvorstellungen erläutert werden.

4.1 Möglicher Einstieg in das Thema bedingte Wahrscheinlichkeit für die Schule

Das Thema bedingte Wahrscheinlichkeiten in der Schule hat den Vorteil, dass es einen angeregten Unterricht lässt. Die Möglichkeit, spannende und anwendungsorientierte Beispiele zu bringen, welche im Alltag der Schüler*innen Anwendung finden, ist genauso ein Vorteil, wie die Möglichkeit, über Ergebnisse und Wahrscheinlichkeiten eine Diskussion zu führen. Weil Wahrscheinlichkeiten im Alltag der Schüler*innen Anwendung finden, ist eine gewisse Grundintuition da, welche den Austausch über das Thema fördert. [Hauer Typpelt 2007, S. 1]

Bei diesem Stoffgebiet bietet sich auch gut an, den Einstieg in die Unterrichtssequenz mit einem praktischen Beispiel zu begehen. In der Physik ist diese Art des Einstieges fachdidaktisch hoch im Kurs. Im 5E-Modell, welches ein Unterrichtsthema in fünf Phasen aufteilt, ist die erste Phase das Engage. In dieser Phase soll das Interesse der Schüler*innen durch eine Fragestellung, ein Problem oder ein Beispiel geweckt werden. [Hofer, Abels, Lembens 2016, S. 4]

Zu diesem Thema gibt es ein recht bekanntes Beispiel, in diesem geht es um die Erkennung von Down-Syndrom bei Ungeborenen mittels des „combined test“ [nach Hauer

Typfelt 2007, S. 1-5]. Ähnliche Beispiele, die sich als Einstieg anbieten, wie eines mit einem HIV-Test, gibt es auch. Es spielt dabei keine Rolle, welches dieser Beispiele genau genommen wird, solange es ungefähr den selben Zweck erfüllt.

4.1.1 Beispiel zum Einstieg mit dem „combined test“

Das Beispiel zum „combined test“ aus einem Vortrag von Petra Hauer-Typfelt wurde in dieser Masterarbeit leicht abgeändert [Hauer-Typfelt, 2007]. Der „combined test“ ist eine Methode, um Trisomie 21 (In folge Trisomie 21) bei einem ungeborenen Kind festzustellen. Dieser Test wird in Österreich für schwangere Frauen angeboten. Dabei liefert der Test mit einer Wahrscheinlichkeit von 0,92 ein positives Ergebnis, wenn das Ungeborene tatsächlich Trisomie hat. Sollte keine Trisomie vorliegen, liefert der Test mit einer Wahrscheinlichkeit von 0,9 ein negatives Ergebnis. In der Medizin ist die Spezifität eines Tests die Wahrscheinlichkeit, dass der Test ein negatives Ergebnis richtig anzeigt. Die Sensitivität ist die Wahrscheinlichkeit, dass der Test ein positives Ergebnis richtig anzeigt [Strick 2020, S. 15-19]. Außerdem weiß man, dass in Österreich 1 von 800 Ungeborenen Trisomie hat.

Nachdem die Fakten geklärt sind, widmet man sich jetzt den Fall, bei dem der Test durchgeführt wird und dieser eben durchgeführte Test als positiv bezüglich Trisomie zurückkommt. In dieser Situation bietet es sich an, die Schüler*innen raten zu lassen, mit welcher Wahrscheinlichkeit das ungeborene Kind jetzt tatsächlich Trisomie hat. Entweder lässt man die Kinder eine Prozentzahl zwischen 0 und 100 aufschreiben, oder die Lehrkraft gibt verschiedene Möglichkeiten in diesem Intervall vor und die Schüler*innen müssen ihrer Intuition nach einen Tipp abgeben. Wichtig dabei ist es, die Begründungen bzw. die Denkprozesse der Lernenden in den Mittelpunkt zu stellen und diese auch zu beleuchten. Diese Art von Frage hat außerdem noch den Effekt, dass die Schüler*innen jetzt natürlich auch wissen wollen, ob sie auch richtig liegen.

Im Anschluss sollte das erste Mal der Begriff der bedingten Wahrscheinlichkeit eingeführt werden. Zwei bedingte Wahrscheinlichkeiten lassen sich direkt aus der Angabe entnehmen, daher sollte anhand dieser die, für die bedingte Wahrscheinlichkeit typische, Schreibweise das erste Mal behandelt werden:

$$P(\text{pos}|T) = 0,92 \quad \text{und} \quad P(\text{neg}|\neg T) = 0,92$$

Der erste Ausdruck gibt die Wahrscheinlichkeit an, dass eine Mutter ein positives Testergebnis bekommt unter der Bedingung, dass das Ungeborene tatsächlich Trisomie hat. Der zweite Ausdruck gibt die Wahrscheinlichkeit an, dass eine Mutter ein negatives Testergebnis bekommt, wenn das Ungeborene keine Trisomie hat. Es ist wichtig, dass gleich am Anfang klar gemacht wird, was die bedingte Wahrscheinlichkeit aussagt und wie sie gelesen wird, da es dabei zu Problemen kommen kann. Folglich muss als Lehrender darauf geachtet werden, dass die Beschreibung der bedingten Wahrscheinlichkeit nicht zu komplex wird, da es die Schüler*innen schnell überfordern kann.

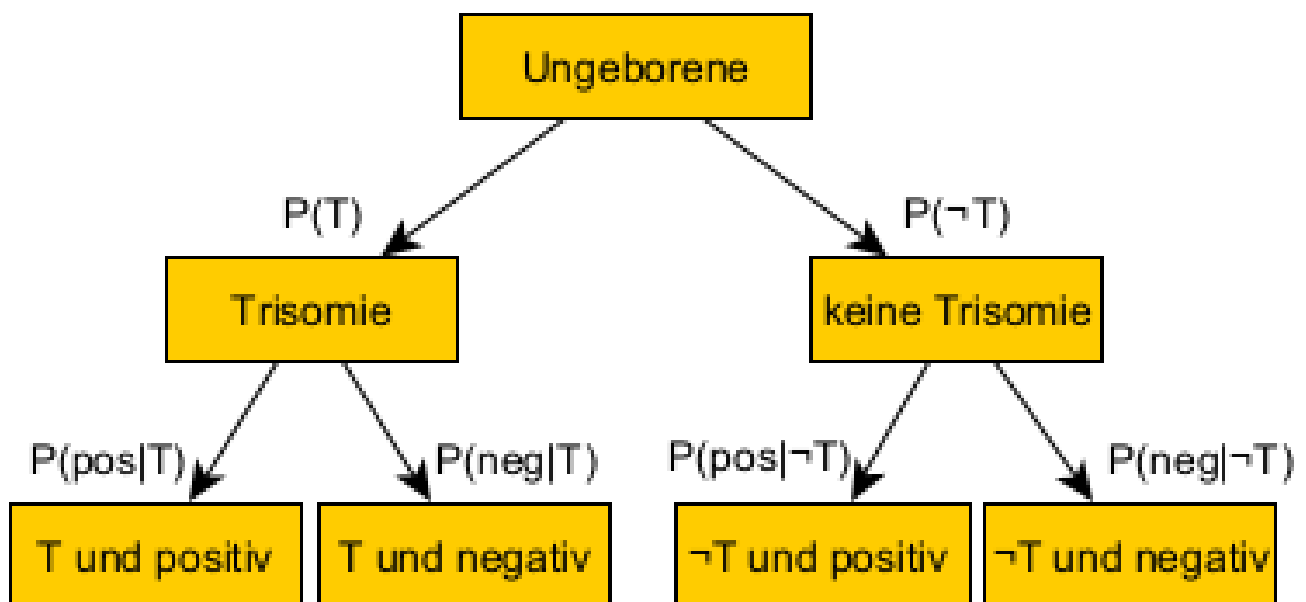


Abbildung 6: Allgemeines Baumdiagramm zum Beispiel „combined test“

Außerdem kann aus der Angabe noch die Wahrscheinlichkeit $P(T)=0,00125$ herausgelesen werden, mit welcher ein Ungeborenes Trisomie hat. Das allgemeine Baumdiagramm für diese Art von Aufgabe ist in Abbildung 6 zu sehen. Jetzt können die drei bekannten Wahrscheinlichkeiten in diesem Baumdiagramm ergänzt werden. Wenn das getan ist, merkt man, dass aus jedem Knoten, aus dem Äste ausgehen, eine Wahrscheinlichkeit bekannt ist. Daher können alle anderen Wahrscheinlichkeiten nach der ersten Pfadregel ebenfalls in das Baumdiagramm eingetragen werden. Diese Wahrscheinlichkeiten können natürlich nicht über die erste Pfadregel berechnet werden. Die Gegenwahrscheinlichkeit würde hier ebenfalls funktionieren. Das vollständige Baumdiagramm ist in Abbildung 7 zu sehen.

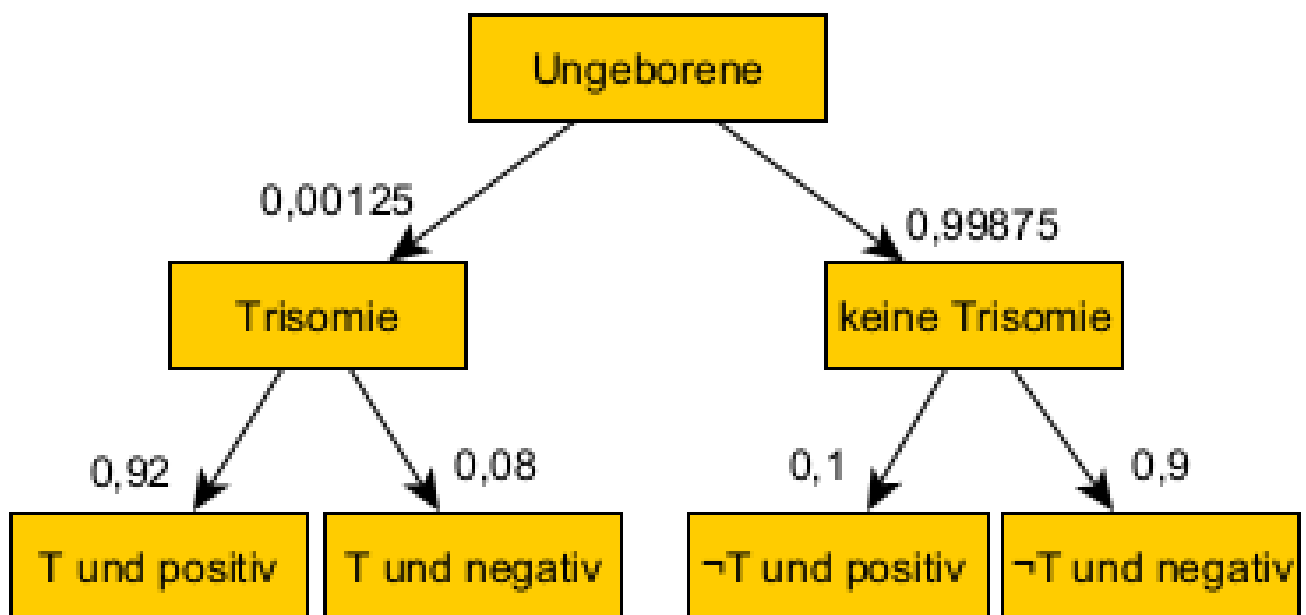


Abbildung 7: Vollständiges Baumdiagramm zum Beispiel „combined test“

Im nächste Schritt wird jetzt ein Baumdiagramm erstellt, welches im ersten Schritt das Ergebnis des Tests und im zweiten Schritt den Fakt, ob das Ungeborenen Trisomie hat oder nicht, darstellt. Das hat den Zweck, dass dadurch die Wahrscheinlichkeit auf einem Ast angegeben werden kann, ob das Ungeborene Trisomie unter der Bedingung, dass der Test positiv ist, also $P(T|\text{pos})$ hat.

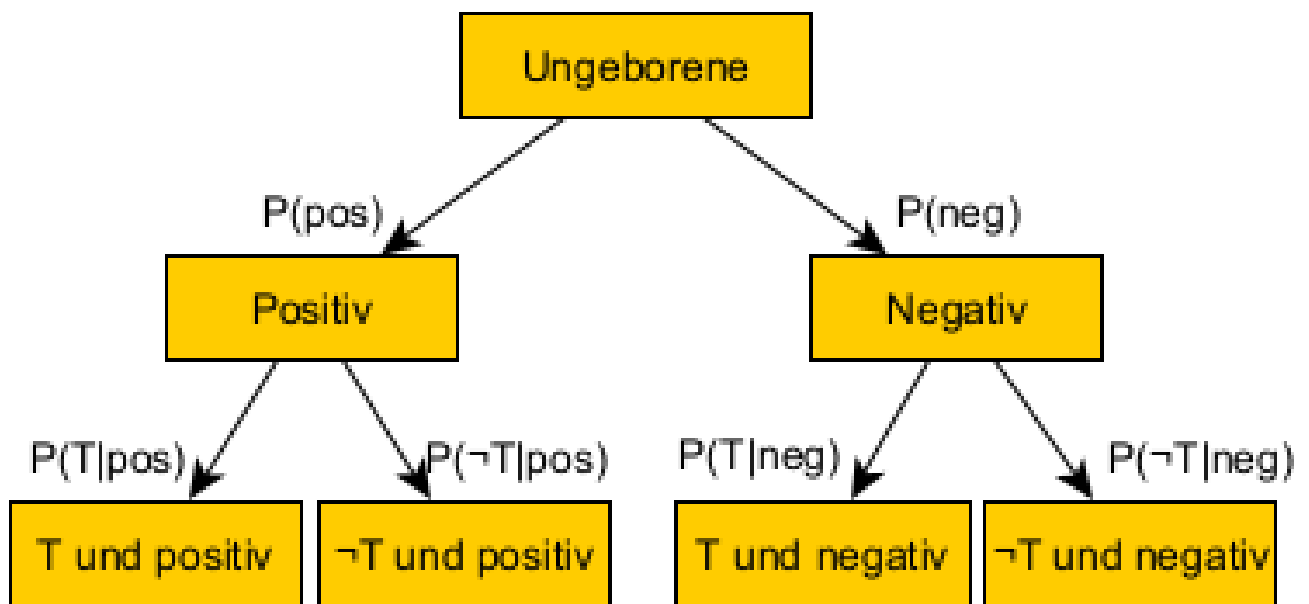


Abbildung 8: Umgekehrtes Baumdiagramm zum Beispiel „combined test“

Die Wahrscheinlichkeit die angibt, dass das Ungeborene Trisomie bei einem positiven Test hat, kann aus dem Baumdiagramm aus Abbildung 7 sowie der zweiten Pfadregel bestimmt werden.

$$P(\text{T und positiv}) = P(T) \cdot P(pos|T) = 0,00125 \cdot 0,92 = 0,00115$$

Außerdem kann aus dem Baumdiagramm aus Abbildung 8 und der zweiten Pfadregel dieselbe Wahrscheinlichkeit berechnet werden.

$$P(\text{T und positiv}) = P(pos) \cdot P(T|pos)$$

Aus dieser Formel lässt sich durch einen Umformungsschritt die Wahrscheinlichkeit $P(T|pos)$ ausdrücken als

$$P(T|pos) = \frac{P(\text{T und positiv})}{P(pos)} \quad (22)$$

Diese Herleitung lässt sich durch ein Baumdiagramm, welches die Diagramme aus Abbildung 7 und 8 zusammenfügt, noch deutlicher hervorheben. Dieses ist in Abbildung 9 dargestellt.

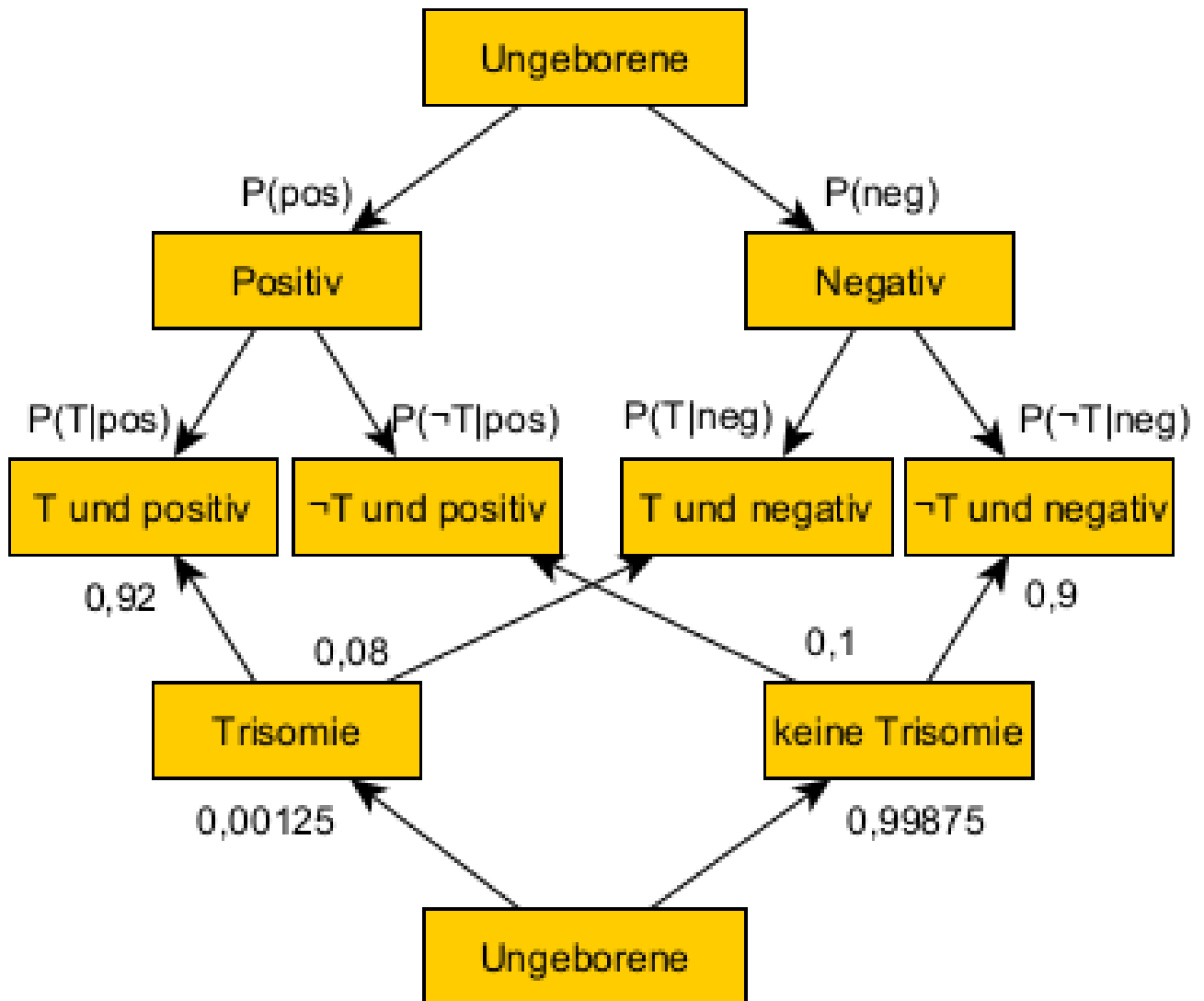


Abbildung 9: Zusammengeführtes Baumdiagramm zum Beispiel „combined test“

Anschließend will man die Wahrscheinlichkeit aus Gleichung (22) berechnen. Die Wahrscheinlichkeit für $P(T \text{ und positiv})$ wurde oben schon berechnet. Die Wahrscheinlichkeit $P(pos)$ lässt sich ebenfalls aus dem Baumdiagramm mit Anwendung der zweiten und dritten Pfadregel berechnen.

$$P(pos) = 0,00125 \cdot 0,92 + 0,99875 \cdot 0,1 = 0,101025$$

Damit ergibt sich dann eine Wahrscheinlichkeit für Trisomie, wenn der Test positiv ist als

$$P(T|pos) = \frac{0,00125}{0,101025} \approx 0,012$$

Spätestens jetzt werden die Schüler*innen stutzig, da die Wahrscheinlichkeit sehr gering ist. Die soeben berechnete Wahrscheinlichkeit gibt an, dass die Wahrscheinlichkeit, dass ein Ungeborenes bei einem positiven Test Trisomie hat, bei ungefähr 1,2% liegt. Dann stellt sich wahrscheinlich oft die Frage, wieso dieser Test überhaupt durchgeführt wird, wenn der Test so „unzuverlässig“ ist. Um dieses Problem bei der Wahrnehmung der Schüler*innen zu lösen, kann es hilfreich sein dieses Rechenbeispiel nochmal mit absoluten Zahlen durchzuführen. Dadurch kann gut gezeigt werden, dass die absolute Zahl der Ungeborenen, die keine Trisomie, jedoch aber ein positives Testergebnis haben, hoch ist, es dann aber in relativen Teilen verglichen mit der Anzahl negativer Testungen dann eben doch wieder hinkommt.

Man kann auch für ein Beispiel mit absoluten Zahlen ein Baumdiagramm verwenden. Im Fall von 100000 Ungeborenen würde das Baumdiagramm wie in Abbildung 10 aussehen. In diesem Diagramm sieht man, dass der größte Teil der Ungeborenen keine Trisomie hat und negativ getestet wird. Der Anteil der Ungeborenen, die keine Trisomie haben, welche aber fälschlicherweise positiv getestet wurde, ist in absoluten Zahlen groß. Daraus resultiert auch die Wahrscheinlichkeit von 1,2% von richtigen positiven Meldungen. Das wird außerdem dadurch verstärkt, dass der genetische Fehler nicht sehr oft bei Ungeborenen vorkommt. Ein weiterer Fakt, der noch erwähnt werden sollte ist, dass von 100000 Ungeborenen nur 10 fälschlicherweise negativ getestet werden, das sind nur 0,01%.

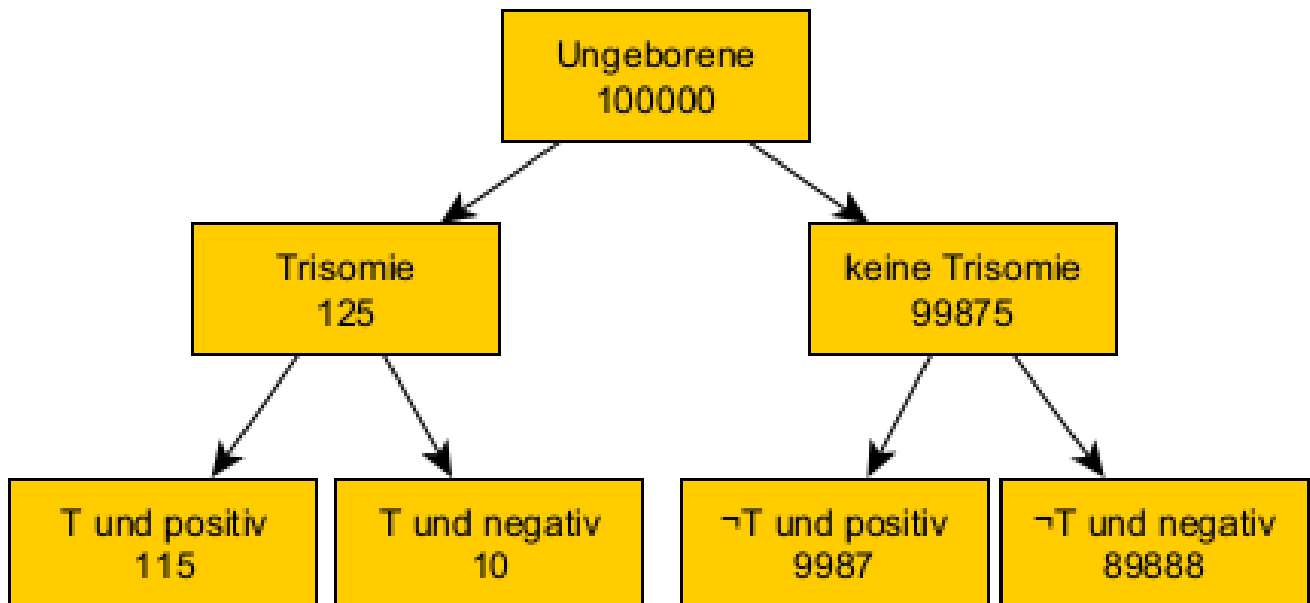


Abbildung 10: Baumdiagramm mit absoluten Zahlen zum Beispiel „combined test“

4.1.2 Fachdidaktische Überlegungen

Eine Überlegung, die prinzipiell in der Schule gemacht werden muss, ist das Beachten der möglichen Fehlvorstellungen. Anschließend werden kurz die gängigsten Fehlvorstellungen gezeigt [Hauer-Typpelt 2007, S. 7-10].

1.) Kausaler Zusammenhang zwischen den Ereignissen A und B

Nur weil eine bedingte Wahrscheinlichkeit gegeben ist, bedeutet es nicht, dass diese Ereignisse einen kausalen Zusammenhang haben.

2.) $P(A|B)$ und $P(B|A)$ werden verwechselt/gleichgesetzt

Eine sehr gängige Schwierigkeit beim Lernen der bedingten Wahrscheinlichkeit ist, dass diese neue Schreibweise die Schüler*innen vor Probleme stellen kann. Dies zeigt sich in der Verwechslung von $P(A|B)$ mit $P(B|A)$. Dies kann gleich am Anfang durch ein einfaches und übertriebenes Beispiel geklärt werden. Das Beispiel $P(\text{Mutter}|\text{Frau})$ und

$P(\text{Frau}|\text{Mutter})$ kann in diesem Zusammenhang die Verwirrung bei den Schüler*innen verringern.

Diese Problematik kann auch werden durch den Fakt, dass Aufgabenstellungen in ihrer Formulierung oft schwammig bzw. zweideutig sind, verstärkt. Hier sollte jeden Falls auf klare Formulierungen geachtet werden.

3.) Zeitgebundenes Denken bei Ereignissen

Als Beispiel für dieses Problem kann Folgendes erwähnt werden [Tietze, Klika, Wolpers 2002, S. 9]:

In einer Urne sind zwei rote und zwei blaue Kugeln. Es werden nacheinander zwei Kugeln ohne Zurücklegen gezogen.

a.) Wie groß ist die Wahrscheinlichkeit, dass die zweite Kugel rot ist, wenn die erste auch rot ist. b.) Angenommen die erste Kugel wird verdeckt gezogen und die zweite Kugel ist rot. Mit welcher Wahrscheinlichkeit war die erste Kugel auch rot?

Für Beispiel a.) ist die Antwort $p = \frac{1}{3}$. Das lässt sich leicht durch Anwenden von (4) berechnen. Bei Beispiel b.) wäre die intuitive Antwort $p = \frac{1}{2}$, da ja am Anfang des Experiments sich vier Kugeln in der Urne befinden, wovon zwei rot sind. Dies ist allerdings ein Fehlschluss! Nach der zweiten Ziehung kennt man nur die Farbe der zweiten Kugel und zwar rot. Die erste Kugel kann daher eine von drei möglichen Kugeln sein, wovon aber nur eine rot ist. Daher wäre hier die richtige Antwort, dass die Wahrscheinlichkeit in Beispiel b.) dann $p = \frac{1}{3}$ ist. Dies lässt sich auch mit der Definition von der bedingten Wahrscheinlichkeit (18) einfach zeigen.

4.) Verwechslung von bedingter und konjunktiver Wahrscheinlichkeit

Ein Fehler, der durch Baumdiagramme leicht gelöst werden kann. Hier wird $P(A \text{ und } B)$ und $P(A|B)$ gleichgesetzt.

5.) Wahrscheinlichkeiten werden falsch auf eins summiert

Diese Fehlvorstellung ist ähnlich wie Fehlvorstellung 2.). Ein Beispiel, welches weiter oben erwähnt wurde („combined test“), wäre, dass die Wahrscheinlichkeiten $P(\text{pos}|T) + P(\neg T|\text{pos}) = 1$ ist. Eine Deutung davon wäre zum Beispiel die Aussage: „Wenn ich krank bin, dann bekomme ich zu 90% ein positiven Test, also bekomme ich, wenn ich nicht krank bin, einen positiven Test zu 10%!“ [Hauer-Typelt 2007, S.8].

6.) Nichterkennen der bedingten Wahrscheinlichkeit

Lernende ignorieren manchmal eine bedingte Wahrscheinlichkeit. Die Bedingung wird dabei entweder übersehen oder nicht als solche erkannt.

4.1.3 Die bedingte Wahrscheinlichkeit und die Rolle der Intuition

Die bedingte Wahrscheinlichkeit ist mathematisch gesehen kein komplexes Thema. Trotzdem verursacht dieses Thema bei den Schüler*innen häufig Probleme. Das passiert weniger auf mathematischer Ebene, als auf einer verständlichen bzw. intuitiven Ebene. Dabei ist die Intuition, vor allem bei der bedingten Wahrscheinlichkeit, wichtig. Es wird prinzipiell nach zwei Arten von Intuition unterschieden [Wolpers, Götz 2002, S. 139].

Primäre Intuition

Die primäre Intuition ist die Vorstellung, die sich zu einem Sachverhalt gebildet hat, ohne dass man sich damit zuvor systematisch beschäftigt hätte. Man kann sagen, dass die primäre Intuition eine Art von Vorwissen ist, welches der/die Schüler*in mitbringt. Der Mathematikunterricht würde idealerweise auf diese Intuition der Lernenden aufbauen. Das Problem ist nur, dass diese Intuition nicht immer richtig sein muss (wie beim Beispiel „combined test“ zu sehen ist). Es ist möglich, mit 10 jährigen Kindern über Dreiecke zu sprechen, ohne dabei auf komplexere Sachverhalte einzugehen, da zu diesem Begriff schon eine primäre Intuition vorhanden ist.

Beim Thema Wahrscheinlichkeiten können primäre Intuitionen aber oft ein Problem darstellen. Im Schulunterricht müssen diese falschen Intuitionen sorgfältig beseitigt werden, damit sich neues Wissen zu diesem Thema festsetzen kann. Dieses neue Wissen wird im Anschluss in der Form von sekundärer Intuition vermittelt.

Sekundäre Intuition

Die sekundäre Intuition ist eine auf vorheriger Auseinandersetzung mit einem Wissensgebiet oder Thema basierende Vorstellung zu einem bestimmten Begriff bzw. Sachverhalt. Eine sekundäre Intuition kann sich erst bilden, nachdem es zu einer primären Intuition gekommen ist und sich die Person mit dem Thema anschließend auseinandergesetzt hat. Im Unterricht ist die Richtigkeit der Sekundärintuition ein maßgeblicher Indikator dafür, ob ein Thema bzw. Sachverhalt verstanden wurde. Die Verbindung zwischen Primärintuition und Sekundärintuition durch den Unterricht, ist einer der zentralen Aspekte des (Mathematik)Unterrichts.

5 Paradoxa

5.1 Simpson-Paradoxon

Schon in den 50er Jahren des 20. Jhd. hatte der Mathematiker Edward Simpson (Der Namensgeber dieses Paradoxons) darauf hingewiesen, dass verschiedene Datensätze, welche ungleich gewichtet sind, in ihrer Gesamtheit zu widersprüchlichen Aussagen zu verschiedenen Merkmalen führen können. Ein ziemlich bekanntes Beispiel zu dem ist die Berkeley-Studie von 1973 [Strick 2020, S. 35-38].

Berkeley-Studie von 1973 zum Simpson-Paradoxon

Im Zuge einer Studie wurde im Jahr 1973 an der Berkeley Universität in Kalifornien eine Studie durchgeführt, welche untersuchen sollte, ob weibliche Studierende bei der Studienzulassung benachteiligt werden oder nicht. In erster Instanz wurden die absoluten Zahlen an Studienbewerber*innen, nach Geschlecht differenziert. Dies ist in Abbildung 11 zu sehen.

	zugelassen	nicht zugelassen	Bewerbungen
Männer	3738	4704	8442
Frauen	1419	2827	4321
Gesamt	5232	7513	12763

Abbildung 11: Tabelle zur Studie von Berkeley in absoluten Zahlen

Aus diesen Zahlen lässt sich auf den ersten Blick nicht viel erkennen, daher würde sich hier eine Betrachtung der relativen Anteile anbieten. Dies ist in Abbildung 12 abgebildet.

	zugelassen	Nicht zugelassen	Bewerbungen
Männer	29,3 %	36,9 %	66,2 %
Frauen	11,7 %	22,1 %	33,8 %
Gesamt	41,0 %	59,0 %	100 %

Abbildung 12: Tabelle zur Studie von Berkeley in relativen Anteilen

Durch diese zwei Abbildungen ist gut zu erkennen, dass der Anteil der zugelassenen Männer höher ist als der der Frauen. Dieser Sachverhalt kann ähnlich wie beim „combined test“ Beispiel dargestellt werden. Mithilfe der Daten aus den Abbildungen kann man sich die bedingten Wahrscheinlichkeiten ausrechnen, mit der männliche bzw weibliche Bewerber an der Universität aufgenommen werden. Dafür wird (22) (W/M steht für Weiblich/Männlich und Z/¬Z steht für zugelassen/ nicht zugelassen) benötigt. Für männliche Bewerber wird

$$P(Z|M) = \frac{P(M \text{ und zugelassen})}{P(M)} = \frac{0,293}{0,293 + 0,369} \approx 0,4426$$

berechnet. Es werden somit ungefähr 44,26% der männlichen Bewerber für ein Studium in Berkeley zugelassen. Im Gegenzug werden daher ca. 55,74% der Bewerber abgelehnt. Analog wird für die Frauen folgendes berechnet.

$$P(Z|W) = \frac{P(W \text{ und zugelassen})}{P(W)} = \frac{0,117}{0,117 + 0,221} \approx 0,3462$$

Es werden daher ca. 34,62% der Bewerberinnen für ein Studium in Berkeley zugelassen. Daher werden im Gegenzug ca. 65,38% der Bewerberinnen abgelehnt. Diese Zahlen legen nahe, dass Bewerber den Bewerberinnen vorgezogen werden. Dieser Sachverhalt wird in Abbildung 13 dargestellt.

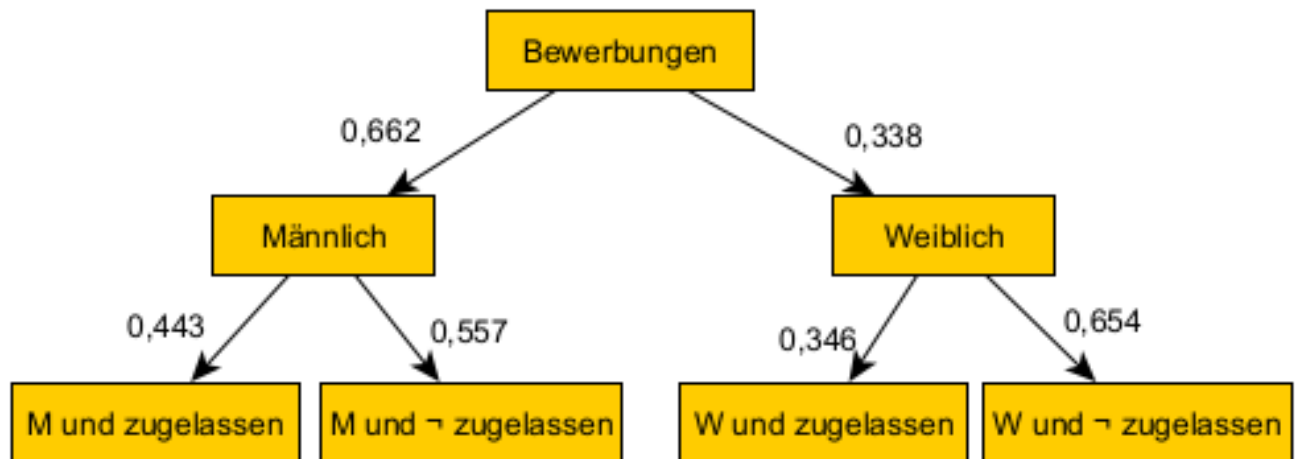


Abbildung 13: Baumdiagramm zur Zulassung an der Universität Berkeley unter der Bedingung des Geschlechts des Bewerbers

Um diesen Sachverhalt zu verstärken darzustellen, kann jetzt noch gezeigt werden, wie der Anteil der Geschlechter unter den zugelassenen Bewerber*innen ist. Der Anteil von Männern von allen zum Studium zugelassenen Bewerbern ist

$$P(M|Z) = \frac{P(M \text{ und zugelassen})}{P(Z)} = \frac{0,293}{0,293 + 0,117} \approx 0,7146$$

Der männliche Anteil, der zum Studium zugelassenen Bewerber, beträgt ca. 71,46 %. Daraus folgt, dass der Anteil der zum Studium zugelassenen, weiblichen Bewerber „nur“ 28,54% beträgt. Im Kontrast dazu beträgt der Anteil der männlichen Bewerber, welche nicht zugelassen wurden.

$$P(M|\neg Z) = \frac{P(M \text{ und } \neg \text{ zugelassen})}{P(\neg Z)} = \frac{0,369}{0,369 + 0,221} \approx 0,6254$$

Der Anteil der nicht zum Studium zugelassenen männlichen Bewerber beträgt 62,54%. Daher beträgt der Anteil der nicht zum Studium zugelassenen, weiblichen Bewerber 37,46%. Dieser Sachverhalt wird in Abbildung 14 dargestellt.

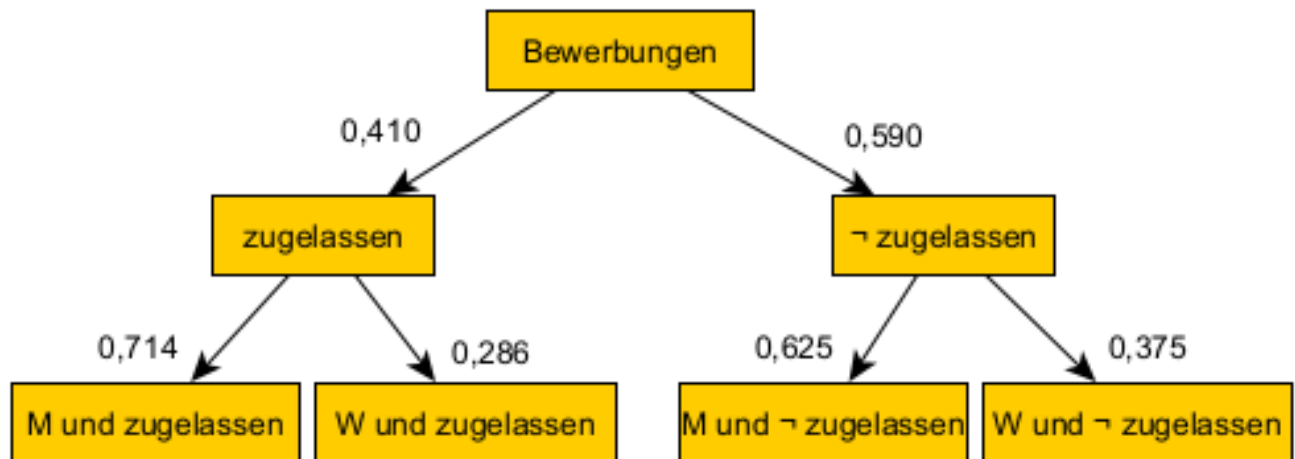


Abbildung 14: Baumdiagramm zum Geschlecht der Bewerber der Universität von Berkeley unter der Bedingung, dass sie zum Studium zugelassen wurden

Daraus folgt, dass der Anteil der zum Studium zugelassenen, männlichen Bewerber ca. 71,46% beträgt, während nur ca. 62,54% der männlichen Bewerber nicht zugelassen werden. Daher scheint es so, dass in Berkeley weibliche Bewerber benachteiligt würden.

Betrachtet man aber die Zahlen der einzelnen Institute der Universität, zeigt sich ein anderes Bild. Da es in Berkeley im Jahre 1973 85 verschiedene Institute gab, wurden nur die größten zwei beleuchtet. Die Ergebnisse der anderen Institute waren ähnlich. Für das erste Institut gab es 933 Bewerber*innen und für das zweite Institut 545 Bewerber*innen. Die relativen Anteile zu diesen zwei Instituten können aus Abbildung 15 entnommen werden.

Institut 1	zugelassen	nicht zugelassen	Bewerbungen
Männer	54,9 %	33,6 %	88,5 %
Frauen	9,5 %	2,0 %	11,5 %
Gesamt	64,4 %	35,6 %	100 %

Institut 2	zugelassen	Nicht zugelassen	Bewerbungen
Männer	57,4 %	38,0 %	95,4 %
Frauen	3,1 %	1,5 %	4,6 %
Gesamt	60,5 %	39,5 %	100 %

Abbildung 15: Tabelle zu den relativen Anteilen der Bewerber nach Geschlecht und Zulassung in den Instituten 1 und 2

Aus diesen Zahlen lassen sich wieder die bedingten Wahrscheinlichkeiten berechnen.

Für das Institut 1 würde man zu einer Wahrscheinlichkeit von

$$P(Z|M) = \frac{P(\text{M und zugelassen})}{P(M)} = \frac{0,549}{0,549 + 0,336} \approx 0,6203$$

für männliche Bewerber und zu

$$P(Z|W) = \frac{P(\text{W und zugelassen})}{P(W)} = \frac{0,095}{0,095 + 0,020} \approx 0,8261$$

für weibliche Bewerber kommen. Hier werden weibliche Bewerber zu ca. 82,61% genommen, während männliche Bewerber zu ca. 62,03% genommen werden. Im Institut 1 kann man also sagen, dass weibliche Bewerber den männlichen Bewerbern vorgezogen werden.

Im zweiten Institut erhält man ein ähnliches, jedoch bei weitem nicht so deutliches Bild. Hier ergibt sich für männliche Bewerber die Wahrscheinlichkeit

$$P(Z|M) = \frac{P(\text{M und zugelassen})}{P(M)} = \frac{0,574}{0,0,574 + 0,380} \approx 0,6017$$

und für weibliche Bewerber die Wahrscheinlichkeit

$$P(Z|W) = \frac{P(\text{W und zugelassen})}{P(W)} = \frac{0,031}{0,031 + 0,015} \approx 0,6739$$

Die Zahlen im zweiten Institut zeigen eine kleine Bevorzugung der weiblichen Bewerber. Diese Zahlen lassen jedoch keinen signifikanten Unterschied erkennen, sodass hier nicht von einer Männer-Benachteiligung gesprochen werden kann.

In den übrigen Instituten kann in ca. 10 der insgesamt 85 von einer signifikante Geschlechter-Benachteiligung die Rede sein. Jedoch sind beide Geschlechter davon betroffen, es zeigt sich keine Tendenz einer Benachteiligung eines bestimmten Geschlechts. Die angebliche Benachteiligung der weiblichen Bewerber ergab sich vor allem aus der Tatsache, dass Frauen sich überwiegend in Fächern beworben haben, welche nur eine geringe Kapazität hatten. Bei männlichen Bewerbern wiederum dominierten Studiengänge mit einer entweder erhöhten Zulassungsquote oder einer hohen Kapazität. Durch das Zusammenfügen der einzelnen Institute ohne entsprechender Gewichtung der einzelnen Daten, kam es zu einer verzerrten Darstellung der Daten. Insgesamt betrachtet waren männliche Bewerber scheinbar bevorzugt gegenüber weiblichen Bewerbern, berücksichtigt man jedoch die Zahlen im Detail merkt man, dass diese scheinbare Bevorzugung sich in Luft auflöst.

5.2 Monty-Hall Paradoxon

5.2.1 Historischer Verlauf

Im Jahr 1963 war die Premiere einer TV-Show in Amerika namens „*Let's make a deal*“. In dieser TV-Show hat der Moderator, Monty Hall (1921-2017), in jeder Folge die Kandidaten vor verschiedene Herausforderungen gestellt. Das Grundprinzip war, dass die Kandidaten vor der Frage standen, ob sie einen sicheren, jedoch kleinen Preis nehmen oder lieber ihr Glück herausfordern und versuchen wollen, einen größeren Preis zu erspielen.

Im Finale der Show hatte der Kandidat eine Wahl zwischen drei Türen. Hinter zwei der drei Türen war eine Ziege, hinter der dritten jedoch war der Hauptpreis, ein Auto. In der TV-Show wurde so verfahren, dass sobald der Kandidat eine Tür auswählte, hinter der sich eine Ziege verbarg, öffnete Monty Hall die Tür für gewöhnlich sofort und die Ziege kam zum Vorschein. In diesem Fall bekam der Kandidat die Ziege als Preis. Wenn der Kandidat das Auto wählte, und auch in manchen Fällen wenn die Ziege ausgewählt wurde, öffnete der Moderator eine andere Tür, hinter der sich eine Ziege befunden hat und gibt dem Kandidat die Möglichkeit, die Tür zu wechseln. Oft bietet der Moderator in dieser Situation dem Kandidaten Geld an, damit dieser auf den Preis hinter der Tür verzichtet und das Geld nimmt. In dieser Form gab es die TV-Show mit Monty Hall bis 1977. [Rosenhouse 2009, S. 20-22]

In einem Leserbrief wurde die Frage gestellt, ob es eine „beste“ Entscheidung in dieser TV-Show gibt. Im Jahr 1975 wurde ein Brief an Prof. Steve Selvin, damals Professor an der Universität von Kalifornien in Berkeley im „*Academic Journal of American Statistician*“ veröffentlicht. In diesem Brief wurde erstmals hinterfragt, ob die Wahrscheinlichkeit im Falle eines Wechsels der Tür eine 50%ige Chance für den Hauptpreis bringt oder doch eine größere sei. In der Antwort schrieb er, dass die Chance sich nicht ändere, also die Chance das Auto zu bekommen sich nicht ändern würde, unabhängig

davon, ob schon eine Türe mit Ziege gezeigt wurde. [Rosenhouse 2009, S. 22]

Im Jahr 1990 schrieb ein Leser einen Brief in der „Parade“ an Marilyn vos Savant:

Suppose you're on a game show, and you're given the choice of three doors. Behind one door is a car, behind the others, goats. You pick a door, say 1, and the host, who knows what's behind the doors, opens another door, say 3, which has a goat. He says to you, "Do you want to pick door 2?" Is it to your advantage to switch your choice of doors? [Craig F. Whitaker in der „Parade“, Ausgabe 9. September 1990]

Die Antwort von ihr:

Yes; you should switch. The first door has a $1/3$ chance of winning, but the second door has a $2/3$ chance. Here's a good way to visualize what happened. Suppose there are a million doors, and you pick door 1. Then the host, who knows what's behind the doors and will always avoid the one with the prize, opens them all except door 777,777. You'd switch to that door pretty fast, wouldn't you? [Marilyn vos Savant als Antwort auf den oben zitierten Leserbrief in der „Parade“, Ausgabe 9. September 1990]

Diese Antwort hat starke Reaktionen in der mathematischen Welt und darüber hinaus hervorgerufen. Viele Mathematiker haben sich zu Wort gemeldet und behauptet, dass diese Antwort falsch sei. Und auf den ersten Blick scheint dem auch so. Doch wenn man sich näher damit beschäftigt, kann man schnell sehen, dass diese Antwort sehr wohl richtig ist.

5.2.2 Das Paradoxa

Ein Kandidat in einer Spielfshow muss sich zwischen eine von drei Türen entscheiden. Hinter einer Türe ist der Hauptgewinn, ein neues Auto, während hinter den anderen beiden Türen jeweils eine Ziege ist. Hinter welcher Türe das Auto bzw. die Ziegen sind

ist vollkommen zufällig. Die Wahrscheinlichkeit auf Anhieb die Türe mit dem Auto dahinter zu treffen ist $p=\frac{1}{3}$, daher ist die Wahrscheinlichkeit $p=\frac{2}{3}$ eine Ziege zu treffen. Nach der Auswahl wählt der Moderator eine nicht gewählte Türe, hinter der sich aber keine Ziege befindet. Befindet sich hinter beiden Türen eine Ziege, dann wählt der Moderator eine der beiden Türen zufällig. Nachdem der Moderator eine Türe mit einer Ziege geöffnet hat, fragt er den Kandidaten, ob dieser die Türe wechseln möchte.

In dieser Frage liegt das gesamte Paradoxa. Ist es besser zu wechseln, oder doch besser bei der ersten Wahl zu bleiben? Oder spielt es im Endeffekt keine Rolle?

1. Fall: Der Kandidat wechselt die Türe nicht

Im ersten Fall wird davon ausgegangen, dass der Kandidat die Türe bei der Frage des Moderators nicht wechselt. In diesem Fall bleibt dieser bei der Wahl, die er am Anfang getroffen hat. Die gängige Meinung wäre, dass die Chance, dass die gewählte Türe das Auto birgt, nachdem der Moderator die Türe mit der Ziege geöffnet hat, bei $p=\frac{1}{2}$ sein muss. In Abbildung 16 ist dieser Sachverhalt dargestellt.

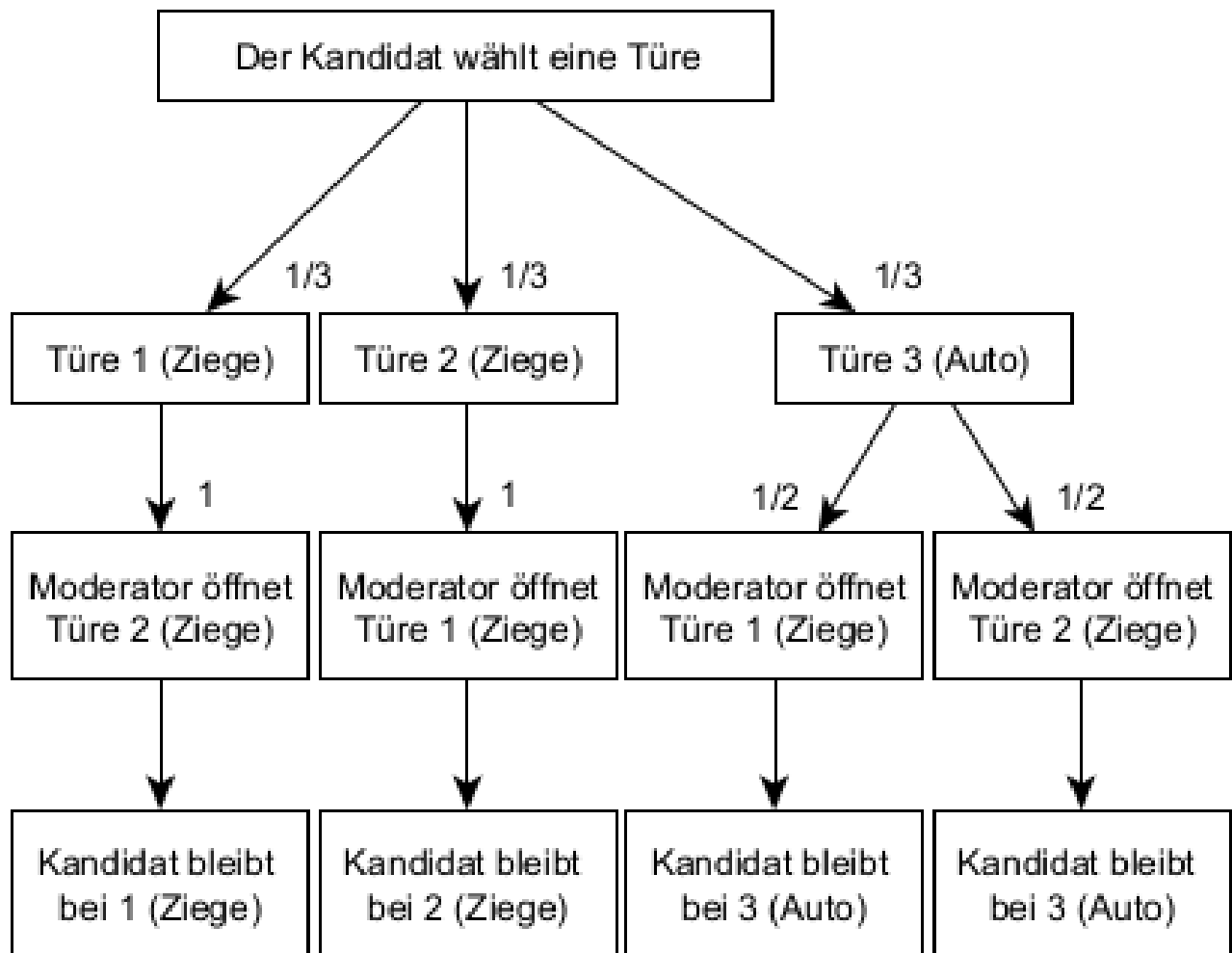


Abbildung 16: Baumdiagramm für den Fall, dass der Kandidat die Türe nicht wechselt

Im ersten Schritt wählt der Kandidat eine Türe. Wenn zwischen der Ziegen hinter den beiden Türen unterschieden werden kann, was ja hier der Fall ist, da die Ziege hinter der ersten Türe klar von der Ziege hinter der zweiten Türe unterschieden werden kann, liegt die Wahrscheinlichkeit für jede Türe bei $p=\frac{1}{3}$. Im Fall, dass der Kandidat eine Ziege auswählt, zeigt der Moderator die jeweils andere Türe her, hinter dieser sich eine Ziege befindet. Im Fall, dass der Kandidat die Türe mit dem Auto auswählt, sucht sich der Moderator zufällig mit einer Wahrscheinlichkeit von $p=\frac{1}{2}$ eine der beiden Türen aus, hinter der sich eine Ziege befindet. Das Ereignis A beschreibt den Fall, dass der Kandidat eine Ziege gewinnt, unter der Bedingung B, dass er die Türe nicht wechselt.

Nach der ersten Pfadregel im Baumdiagramm (Abbildung 16) ergibt der linke Ast eine Wahrscheinlichkeit von $p=\frac{1}{3}$. Da der Moderator garantiert die andere Türe mit der Ziege öffnet, wenn der Kandidat eine Türe mit einer Ziege wählt, ist die Wahrscheinlichkeit dafür 1. Nach der zweiten Pfadregel müssen die Äste zusammenaddiert werden, welche den Ausgang nehmen „Kandidat gewinnt Ziege“. Daher ergibt sich dann eine Wahrscheinlichkeit

$$P(A|B) = \frac{1}{3} \cdot 1 \cdot 1 + \frac{1}{3} \cdot 1 \cdot 1 = \frac{2}{3}$$

Die Wahrscheinlichkeit, dass der Kandidat eine Ziege gewinnt ist also, $p=\frac{2}{3}$. Im Gegenzug dazu ist die Wahrscheinlichkeit, dass der Kandidat ohne wechselnder Türe das Auto gewinnt bei $p=1 - \frac{2}{3} = \frac{1}{3}$. Im Fall, dass der Kandidat die Türe nicht wechselt, liegt seine Erfolgschance bei ca. 33,3%.

2. Fall: Der Kandidat wechselt die Türe

Jetzt wird davon ausgegangen, dass der Kandidat sicher die Türe wechselt. Die Schritte sind wie beim ersten Fall, mit dem Unterschied, dass der Kandidat die Türe wechselt, nachdem der Moderator die Ziege offenbart. Jetzt ist das Ereignis A der Fall, dass der Kandidat eine Ziege gewinnt unter der Bedingung C, dass er die Türe wechselt.

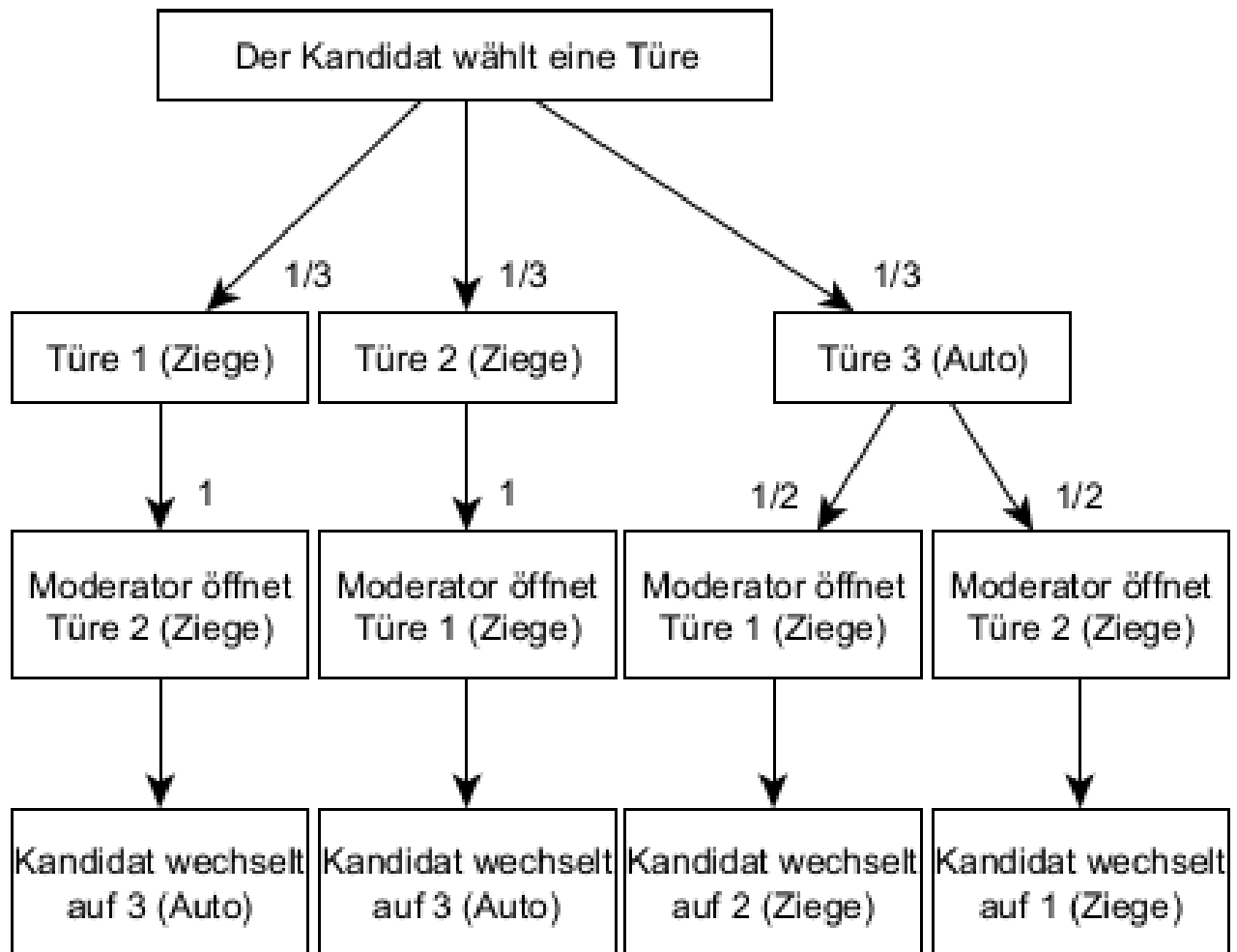


Abbildung 17: Baumdiagramm für den Fall, dass der Kandidat die Türe wechselt

Im Baumdiagramm von Abbildung 17 gilt wieder nach der ersten Pfadregel, dass die Wahrscheinlichkeiten an den beiden rechten Ästen multipliziert werden müssen. Das Ergebnis wird nach der zweiten Pfadregel zusammenaddiert. Daher ergibt sich die Wahrscheinlichkeit

$$P(A|C) = \frac{1}{3} \cdot \frac{1}{2} \cdot 1 + \frac{1}{3} \cdot \frac{1}{2} \cdot 1 = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

Hier liegt also die Wahrscheinlichkeit, dass der Kandidat eine Ziege gewinnt bei $p = \frac{1}{3}$. Daraus folgt im Umkehrschluss, dass die Wahrscheinlichkeit für den Gewinn des Autos bei $p = 1 - \frac{1}{3} = \frac{2}{3}$ liegt.

Daher sollte man im Falle des Ziegdilemmas das Angebot des Moderators immer annehmen. Die Wahrscheinlichkeit verdoppelt sich, wenn man den Entschluss zieht, die Türe zu wechseln. Diese TV-Show hat funktioniert, weil die Menschen das damals nicht wussten. Und damit ist nicht nur der normale, durchschnittliche Mensch mit einem durchschnittlichen mathematischen Verständnis gemeint, sondern auch die damaligen Professoren der Mathematik an den Universitäten.

Der zweite Grund ist außerdem, dass für viele Menschen in dieser Situation der psychische Faktor auch eine Rolle spielt. Der Fall, dass durch das Wechseln das Auto „verloren“ wird, spielt in den Köpfen der Kandidaten eine größere Rolle als der Fall, dass durch das nicht Wechseln das Auto nicht gewonnen wird. In der menschlichen Psyche werden Verluste prinzipiell als emotional viel stärker empfunden als Gewinne. Es ist für viele Menschen emotional schlimmer 100 zu verlieren als 200 zu gewinnen. [Glaser 2019, S. 205-208]

5.3 Spielerparadoxa

5.3.1 Das Paradoxa

Ein Satz der häufiger in Spielerunden zu hören ist, ist folgender: „Jetzt habe ich schon ewig keinen Sechser gewürfelt, langsam muss dieser ja kommen!“ Ein weiterer Satz ist: „Jetzt ist schon 5 Mal in Folge Kopf gekommen, daher kommt beim nächsten Mal ziemlich sicher Zahl!“

Diese Wahrnehmung von Wahrscheinlichkeiten passiert im Alltag relativ häufig. Wenn etwas schon lange nicht mehr passiert ist, dann ist die Wahrscheinlichkeit, dass es beim nächsten Mal eintrifft größer. So ist zumindest die Denkweise bei diesen Gedanken.

Im Beispiel mit der fairen Münze mit „Kopf oder Zahl“ gehen viele Spieler davon aus,

dass nach einer Serie von gleichen Ergebnissen, zum Beispiel „Kopf“, das Ergebnis „Zahl“ mit größerer Wahrscheinlichkeit kommen muss. Der zugrundeliegende Gedanke dazu liegt im Gesetz der großen Zahlen. Da das Gesetz der großen Zahlen besagt, dass bei vielen Wiederholungen die Anzahl der Ereignisse „Kopf“ und „Zahl“ annähernd gleich sein muss. Damit das überhaupt passieren kann, muss aber nach einer Serie, in der nur „Kopf“ geworfen wurde, die Wahrscheinlichkeit, dass „Zahl“ kommt größer werden, ansonsten wäre die Anzahl für das Ergebnis „Kopf“ zu hoch. Die Spieler wissen auf der anderen Seite natürlich, dass die Münze sich nicht merkt, welches Ergebnis im letzten Wurf bzw. in denen davor gekommen ist. Daher wird im nächsten Wurf wieder „Kopf“ bzw. „Zahl“ mit einer Wahrscheinlichkeit von jeweils 0,5 geworfen. Dies ist jedoch ein Widerspruch zum Gesetz der großen Zahlen. Das ist zumindest die Theorie hierzu.

Die Erklärung dazu ist, dass für ein großes n zwar die Anzahl von „Kopf“ bzw. „Zahl“ annähernd gleich sein soll, doch die Betonung hier liegt auf „annähernd gleich“. Wenn bei $n=10000$ Würfen das Ergebnis „Zahl“ 4930 Mal kommt, dann ist die relative Häufigkeit für das Ereignis gleich:

$$r_n = \frac{4930}{10000} = 0,493$$

Das liegt annähernd bei der tatsächlichen Wahrscheinlichkeit von $p=0,5$ und somit in der Erklärreichweite vom Gesetz der großen Zahlen. Der Unterschied zwischen Kopf und Zahl liegt in absoluten Zahlen jedoch trotzdem bei 140, was ein nicht vernachlässigbarer Teil ist [Szekely 1990, S. 42-43].

5.3.2 Bemerkung

Es spielt keine Rolle, wie oft eine Münze geworfen wurde und wie oft hintereinander ein bestimmtes Ergebnis geworfen wurde. Beim nächsten Wurf ist die Wahrscheinlichkeit für „Kopf“ genau so groß wie für „Zahl“ und zwar jeweils 0,5. Die Münze merkt sich nicht, welches Ergebnis vor diesem Wurf eingetreten ist.

Dass sich bei n Würfeln auch mal eine Serie von gleichen Ergebnissen bilden kann, wurde von Paul Erbos und Alfred Renyi gezeigt [Szekely 1990, S. 43]. Sie haben bewiesen, dass bei n Würfeln mit einer Münze die längste zu erwartende Serie von gleichen Ergebnissen mit $\log_2(n)$ angegeben werden kann. Das ist ein Erwartungswert. Es kann natürlich auch zufällig eine längere Serie eintreten. Das würde bedeuten, dass bei 100 Würfeln eine Serie von sechs oder sieben gleichen Ergebnissen in Folge nicht ungewöhnlich wäre. Bei einer Serie von 1 Million Würfeln wäre es eine Serie zwischen 19 und 20 gleichen Ergebnissen in Folge.

Ein weiterer, fast paradoxer, Zusammenhang findet sich, wenn der Quotient der Anzahl der zwei Ergebnisse und die Wahrscheinlichkeit, dass „Kopf“ und „Zahl“ genau gleich oft vorkommen betrachtet wird. Der Quotient

$$\frac{\text{Anzahl „Kopf“}}{\text{Anzahl „Zahl“}}$$

strebt für n gegen unendlich gegen 1. Bei zehn Würfeln ist die Wahrscheinlichkeit für genau fünf Mal „Kopf“ gegeben als

$$P(\text{fünf Mal Kopf}) = \frac{\binom{10}{5}}{2^{10}} \approx 0,25$$

Bei 100 Würfeln genau 50 Mal Kopf zu Würfeln ist durch die Wahrscheinlichkeit

$$P(50 \text{ Mal Kopf}) = \frac{\binom{100}{50}}{2^{100}} \approx 0,08$$

gegeben. Im Fall von 1000 Würfeln ist die Wahrscheinlichkeit für 500 Mal „Kopf“ dann

$$P(500 \text{ Mal Kopf}) = \frac{\binom{1000}{500}}{2^{1000}} \approx 0,002$$

Die allgemeine Formel für diesen Sachverhalt mit n Würfeln bildet sich aus (4) mit den günstigen Fällen (10) und den möglichen Fällen (14)

$$P(k \text{ Mal Kopf}) = \frac{\binom{n}{k}}{2^n} = p$$

Die Wahrscheinlichkeit für n gegen unendlich geht gegen Null während der oben beschriebene Quotient gegen Eins konvergiert [Szekely 1990, S. 46]. Einen Beweis dafür hat der Mathematiker Abraham de Moivre gefunden. Auf den Beweis dazu wird nicht näher eingegangen, dieser kann in [Szekely 1990, S. 46-49] nachgelesen werden.

Obwohl die Wahrscheinlichkeit, dass bei 100 Würfeln genau 50 Mal „Kopf“ kommt ungefähr 8 % ist, ist die Wahrscheinlichkeit, dass „Kopf“ mindestens 60 Mal kommt, geringer. Diese Wahrscheinlichkeit liegt nämlich bei ungefähr 3 %. Die Wahrscheinlichkeit für den Fall, dass „Kopf“ mindestens 55 Mal kommt ist wiederum 16 %. Das zeigt, dass zwar leichte Schwankungen um, in diesem Fall, 50 möglich sind, die Wahrscheinlichkeit, dass es große Ausreißer gibt, ist jedoch eher gering.

5.4 Roulette bzw. Geburtstagsparadoxon

5.4.1 Geburtstagsparadoxon

Bei einer Feier sitzen 365 Personen in einem Saal. Die Chance, dass jeder an einem anderen Tag Geburtstag hat, ist sehr gering, aber in der Theorie möglich, da es auch 365 verschiedene Möglichkeiten gibt, an denen die Personen Geburtstag haben könnten. Sobald auch nur eine Person mehr anwesend ist, müssen zwei Personen am selben Tag Geburtstag haben. Soweit ist an diesem Sachverhalt nichts paradoxes.

In einer Schulklasse sitzen 25 Schüler*innen. Sie kennen nur ihren eigenen Geburtstag, jedoch nicht den Geburtstag der Anderen. Sie wollen sich jetzt die Wahrscheinlichkeit ausrechnen, mit welcher mindestens zwei Schüler*innen am selben Tag Geburtstag haben. Sie nummerieren sich selbst von Eins bis 25 durch und stellen sich in einer Reihe auf, um sich das besser vorzustellen. Um die Wahrscheinlichkeit zu bestimmen, dass mindestens zwei Geburtstage auf den selben Tag fallen, wird zuerst die Wahrscheinlichkeit berechnet, dass alle Geburtstage auf verschiedene Tage fallen und dann die Gegenwahrscheinlichkeit (1) berechnet.

Die erste Person hat garantiert an einem Tag Geburtstag, an dem noch keine andere Person Geburtstag hat, da noch niemand anderes ein Datum belegt. Das Datum des Geburtstages von Person Eins ist jetzt belegt. Wenn die möglichen Fälle die 365 Tage des Jahres sind (ohne Schaltjahr) und die günstigen Fälle die Tage sind, die noch nicht durch eine Person belegt worden sind, dann wird diese Anzahl mit jeder Person, die dazu kommt, um eins weniger. Im Fall von Person Eins gab es 365 günstige Fälle bevor sie dazu gekommen ist und danach sind es noch 364.

Kommt jetzt die zweite Person dazu, dann hat diese eine Wahrscheinlichkeit von $\frac{364}{365}$, dass sie einen Geburtstag hat, welcher noch nicht gewählt wurde. Das geht immer so weiter, bis die letzte Person an der Reihe ist. Diese hat dann eine Wahrscheinlichkeit

von $\frac{341}{365}$, dass sie einen Geburtstag hat, welcher noch nicht gewählt wurde.

Dieser Sachverhalt kann jetzt als Baumdiagramm dargestellt werden. Die zwei möglichen Äste bei jeder Entscheidung sind „hat einen einzigartigen Geburtstag“ und „hat keinen einzigartigen Geburtstag“. Nach der ersten Pfadregel müssen die einzelnen Wahrscheinlichkeiten zusammen multipliziert werden. Das Ereignis A beschreibt, dass alle Personen einen einzigartigen Geburtstag haben. Dies schaut wie folgt aus:

$$P(A) = \frac{365}{365} \cdot \frac{364}{365} \cdot \dots \cdot \frac{341}{365} = \frac{365 \cdot 364 \cdot \dots \cdot 341}{365^{25}} \approx 0,4313 \quad (23)$$

Das bedeutet, dass die Wahrscheinlichkeit, dass alle 25 Schüler*innen einen einzigartigen Geburtstag haben bei ungefähr 43,13 % liegt. Wenn A das Ereignis „alle haben einen einzigartigen Geburtstag“ ist, dann soll B das Gegenereignis sein. Also gilt $B = \bar{A}$. In diesem Fall ist dann

$$P(B) = 1 - P(A)$$

Also ist $P(B) = 1 - 0,4313 \approx 0,5687$. Die Wahrscheinlichkeit, dass mindestens zwei Schüler*innen am selben Tag Geburtstag haben, ist daher 56,87 %. Für viele Personen sorgt diese Zahl auf den ersten Blick für Verwunderung.

Aus (21) folgt die allgemeine Schreibweise für $n = 365$ Tage und k Personen mit:

$$P(A) = \frac{n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-k+1)}{n^k} \quad (24)$$

In Abbildung 18 können die Wahrscheinlichkeiten für eine bis 50 Personen betrachtet werden. In dieser Abbildung ist auch zu sehen, dass bei 50 Personen die Wahrscheinlichkeit, dass zumindest zwei Personen am selben Tag Geburtstag haben, schon recht nahe bei 1 ist. Es reichen schon 55 Personen, damit zumindest eine 99% Wahrschein-

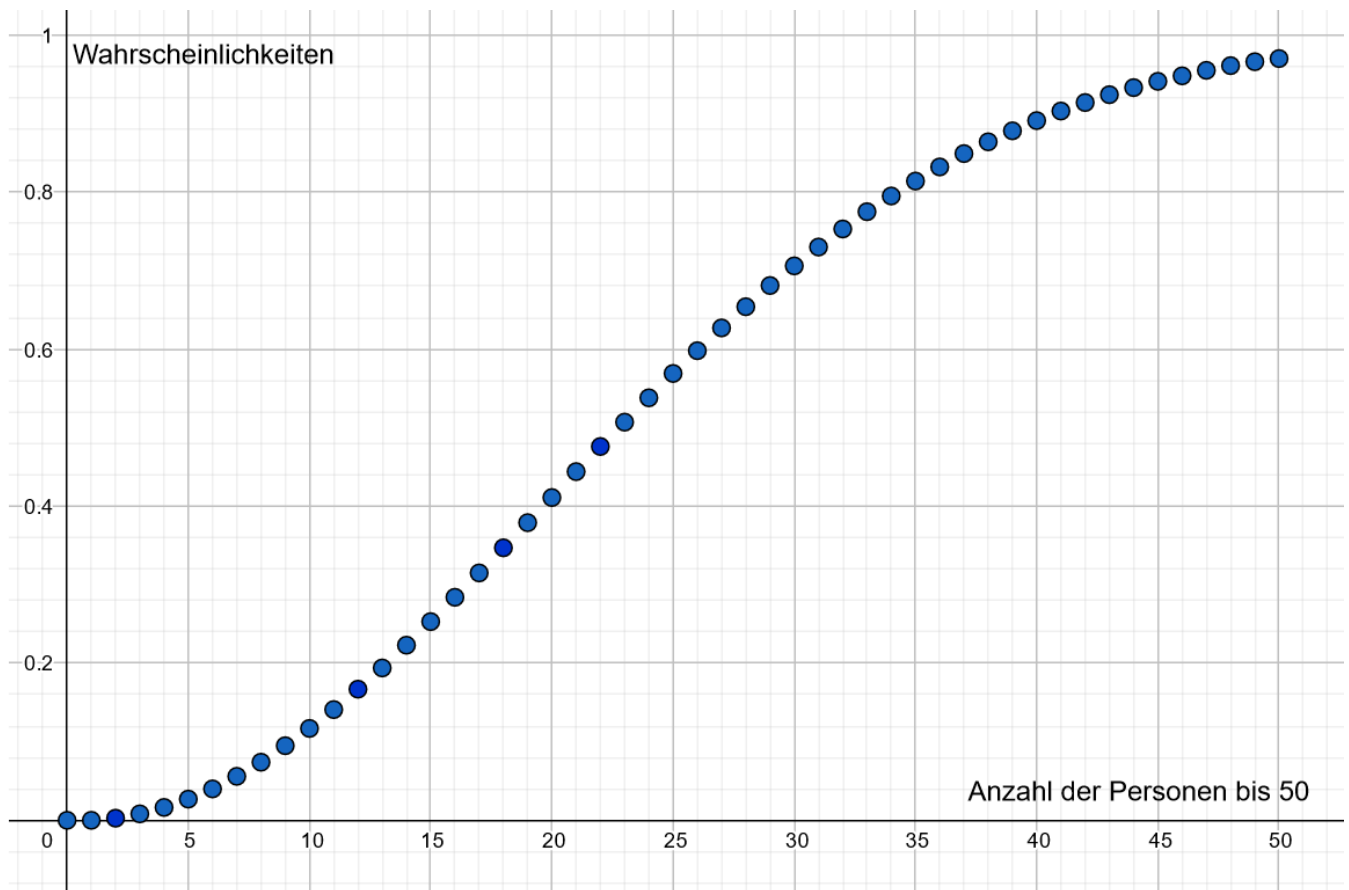


Abbildung 18: Wahrscheinlichkeiten für mindestens ein Geburtstag am selben Tag für 1 bis 50 Personen [<https://www.geogebra.org/m/MWB7JATF> abgerufen am 27.03.2021 um 16:57]

lichkeit erreicht wird, und bei 68 Personen beträgt die Wahrscheinlichkeit schon mehr als 99,9%. [Szekely 1990, S. 67]

Bemerkenswert ist außerdem auch, dass 13 Personen nötig sind, um die Lücke von 0,9% von 99% auf 99,9% zu schließen. Bei genauerer Betrachtung ist dies aber nicht mehr so. Wird zum Beispiel ein Ereignis A hergenommen, welches an einem bestimmten Tag mit einer Wahrscheinlichkeit von $P(A)=0,99$ eintritt und vergleicht dieses mit einem Ereignis B, welches mit einer Wahrscheinlichkeit von $P(B)=0,999$ auftritt. Wird jetzt ein ganzes Jahr betrachtet, also 365 Tage, dann beträgt die Wahrscheinlichkeit, dass A an allen 365 Tagen eintritt etwas mehr als 2,5%. Wird das Ereignis B auf 365 Tage betrachtet, so erhält man die Wahrscheinlichkeit von ungefähr 69,4%.

Der Unterschied zwischen den Wahrscheinlichkeiten von A und B liegt zwar nur bei 0,9%, dieser Unterschied ist aber aus dem Grund so relevant, da diese Differenz bei fast 100% liegt. Ein Ereignis, welches mit $P(A)=0,99$ eintritt, hätte einen Erwartungswert von 990 bei 1000 Versuchen, also würde in 10 Versuchen erwartungsgemäß Ereignis A nicht eintreten. Bei einer Wahrscheinlichkeit von $P(B)=0,999$ wäre der Erwartungswert 999 von 1000 Versuchen. Also würde bei nur einem einzigen von den 1000 Versuchen das Ereignis B erwartungsgemäß nicht eintreten. Anders gesagt, misslingt Ereignis A zehn Mal häufiger als Ereignis B, was wiederum ein großer Unterschied ist, welcher aber durch die Wahrscheinlichkeiten $P(A)=0,99$ und $P(B)=0,999$ auf den ersten Blick so nicht ersichtlich ist. [Szekely 1990, S. 60]

5.4.2 Rouletteparadoxon

Eine ähnliche Frage wie beim Geburtstagsparadoxon stellt sich beim Roulette. Beim europäischen Roulette kommt eine bestimmte Zahl mit einer Wahrscheinlichkeit von ca. 2,7%. Man wählt sich eine beliebige Zahl zwischen 1 und 36. Jetzt wird zehn Mal gedreht. Was passiert mit einer größeren Wahrscheinlichkeit? Die Zahl die ausgesucht wurde, kommt mindestens einmal (A), oder mindestens eine beliebige Zahl kommt doppelt vor (B).

Für den ersten Fall berechnet man die Gegenwahrscheinlichkeit zum Ereignis „Ausgewählte Zahl kommt mindestens einmal“

$$P(A) = 1 - \left(\frac{36}{37}\right)^{10} \approx 0,2397$$

Bei zehn Runden würde mit einer Wahrscheinlichkeit von ca. 23,97% die Zahl kommen, welche man ausgewählt hat. Berechnet man jetzt die Wahrscheinlichkeit für das Ereignis „Eine Zahl kommt doppelt“ benutzt man aus rechnerischen Gründen wieder die Gegenwahrscheinlichkeit.

$$P(B) = 1 - \left(1 \cdot \frac{36}{37} \cdot \frac{35}{37} \cdot \dots \cdot \frac{28}{37}\right) \approx 0,2629$$

Die Wahrscheinlichkeit, dass bei zehn Runden Roulette eine Zahl doppelt vorkommt liegt bei ca. 26,29%. Es ist daher wahrscheinlicher, dass eine Zahl doppelt vorkommt, als dass eine bestimmte, vom Spieler ausgesuchte, Zahl einmal kommt.

Eine mögliche Begründung dafür liegt darin, dass man bei der Auswahl von einer bestimmten Zahl darauf vertraut werden, dass eben diese Zahl kommt. Beim Ereignis „Eine Zahl kommt doppelt“ spielt es keine Rolle, welche dieser Zahlen doppelt vorkommt. Es ist nur wichtig, dass eine Zahl doppelt vorkommt. Geht man von dem Ereignis aus, dass schon neun einzigartige Zahlen gekommen sind, liegt die Wahrscheinlichkeit, dass die zehnte Zahl eine der neun vorherigen trifft bei ca. 24,32%. Das ist eine relativ hohe Wahrscheinlichkeit, und dieser Fall beschäftigt sich mit nur einem der Ergebnisse, die für „mindestens eine Zahl ist doppelt“ möglich sind.

5.5 Der Fall Sally Clark

Zum Abschluss soll noch ein tragischer Fall aus Großbritannien beleuchtet werden, welcher sich in den Jahren 1999 bis 2003 ereignet hat. Die Paar Clark, bestehend aus Sally Clark und Steve Clark waren verheiratet. Sie haben 1996 einen Sohn bekommen, welcher nur Wochen nach der Geburt plötzlich verstorben ist. Ihr zweiter Sohn ist ebenfalls ein paar Wochen nach der Geburt im Jahr 1998 verstorben. Das britische Gericht hat Sally Clark vorgeworfen, ihre Kinder durch starke physische Misshandlung ermordet zu haben. Diese Theorie wurde durch die Aussagen eines Gutachters besonders gestärkt. Auf eben diese Aussage soll jetzt eingegangen werden.

Der Gutachter argumentierte so, dass die Chance, dass ein Kind an einem plötzlichem Kindstod stirbt, für eine nicht rauchende Familie, in Großbritannien bei ungefähr 1 zu 8500 liegt. Dass in der selben Familie zwei Kinder hintereinander am plötzlichen

Kindstod sterben, so die Aussage des Gutachters, liegt daher bei 1 zu 73 Millionen. Daher ist es fast unmöglich, dass beide Kinder an plötzlichem Kindstod gestorben sind. Es liegt daher sehr nahe, dass zumindest eines der beiden Kinder von der Mutter ermordet wurde. Dieser Sachverhalt wurde, benannt nach dem Gutachter, „Meadow’s law“ genannt. Dieser besagt sinngemäß, dass ein plötzlicher Kindstod tragisch, zwei plötzliche Kindstode in der selben Familie suspekt und drei plötzliche Kindstode Mord seien. Mithilfe dieses Gutachtens wurde Sally Clark im Jahre 1999 zu lebenslanger Haft verurteilt [Aus dem Gerichtsprotokoll 2003, 94-110].

Das Urteil wurde zwei Mal angefochten. Beim ersten Mal im Jahre 2000 wurde das Urteil bestätigt, beim zweiten Mal im Jahr 2003 wurde dieses Urteil dann aber aufgehoben. Auch die Mithilfe von Professoren, welche Mathematik oder Statistik lehrten war bei der zweiten Anfechtung des Urteils wichtig. Die Argumente gegen die Aussage des Gutachters stützten sich dabei auf drei Säulen.

Das erste Problem liegt in der Wahrscheinlichkeit 1 zu 8500. Diese Zahl entspricht nicht der Wahrscheinlichkeit, dass ein Kind am plötzlichen Kindstod stirbt. Sie beinhaltet die Faktoren „Eltern sind Nichtraucher“, „Eltern sind wohlhabend“ und „Eltern leben in einem stabilen Haushalt“. Alle diese Faktoren machen es weniger wahrscheinlich, dass ein Kind am plötzlichem Kindstod stirbt. Die Zahl sollte, ohne die Berücksichtigung dieser Faktoren, eher bei 1 zu 1300 liegen. Außerdem wurden Faktoren, die einen plötzlichen Kindstod begünstigen, nicht berücksichtigt, wie zum Beispiel „Kind ist ein Junge“.

Das zweite Problem liegt in der Unabhängigkeit der beiden Ereignisse. Der Gutachter hat die Wahrscheinlichkeiten so behandelt, als ob sie unabhängig voneinander wären. Diese Aussagen können so nicht getätigt werden, da in der Forschung zum plötzlichen Kindstod noch keine wirkliche Ursache dafür herausgefunden werden konnte. Es ist nicht geklärt, ob es nicht zum Beispiel ein Gen gibt, welches die Wahrscheinlichkeit für den plötzlichen Kindstod stark anhebt. Allgemein kann es sein, dass bisher unbekannte,

noch nicht erforschte Faktoren innerhalb einer Familie die Wahrscheinlichkeit für einen plötzlichen Kindstod stark beeinflussen. Außerdem liegt der Schluss nahe, wenn das erste Baby stirbt, auch das zweite Baby eine stark erhöhte Chance hat am plötzlichen Kindstod zu sterben. Es ist daher anzunehmen, dass bei einem plötzlichen Kindstod die Wahrscheinlichkeit für eine Veranlagung oder Einfluss von negativen Faktoren in eben dieser Familie erhöht sein muss.

Das dritte Problem besteht in den bedingten Wahrscheinlichkeiten. In der Aussage des Gutachters wurde davon ausgegangen, dass eine geringe Wahrscheinlichkeit, dass das Kind an plötzlichem Kindstod stirbt, automatisch bedeutet, dass wenn das Kind stirbt, die Wahrscheinlichkeit für Mord hoch sein muss[Thompson, Shumann 1987].

Angenommen das Ereignis A repräsentiert den Fall „zwei Babys sterben hintereinander in einer Familie“ und das Ereignis B repräsentiert den Fall „Sally Clark ist unschuldig“. Die Wahrscheinlichkeit $P(A|B)$ ist sehr gering, da allgemein die Wahrscheinlichkeit von plötzlichem Kindstod, vor allem hintereinander, sehr gering ist. Soweit hat der Gutachter auch argumentiert. Das Problem liegt in seiner Schlussfolgerung. Er sagt, dass dadurch, dass $P(A|B)$ sehr gering ist, die Wahrscheinlichkeit, dass Sally Clark unschuldig ist auch sehr gering sein muss. Er setzt hier $P(A|B) = P(B|A)$ gleich, was aber eine falsche Schlussfolgerung ist. Nur weil $P(A|B)$ sehr gering ist, muss $P(B|A)$ nicht zwangsläufig auch gering sein. Das ist klarer verständlich, wenn man bedenkt, dass die Wahrscheinlichkeit von dem Ereignis C „Mutter tötet beide Kinder“ auch sehr gering ist, da doppelter Kindermord prinzipiell ein sehr seltenes Ereignis ist. Man muss daher die Wahrscheinlichkeit $P(C)$ mit den Wahrscheinlichkeiten $P(A)$ und $P(\text{zwei Kinder sterben am plötzlichen Kindstod})$ vergleichen. Vergleicht man alle Faktoren miteinander, erhält man, je nachdem welche Faktoren wie gewichtet werden, eine Wahrscheinlichkeit zwischen 5% und 50%, dass Sally Clark einen Doppelmord an ihren Kindern begangen hat. Daher kann statistisch absolut nicht bewiesen werden, dass Sally Clark ihre Kin-

der ermordet hat. Die Zahlen würden eher ihre Unschuld belegen. Es gilt daher, sich ebenfalls auf die anderen Beweise zu beziehen und die statistische Aufarbeitung der Zahlen als einen kleinen Teil der Beweisführung zu sehen und nicht als Hauptargument für eine Verurteilung von Clark [vgl. Watkins 2000].

Der oben beschriebene Fall bezieht sich auch auf eine typische Fehlvorstellung bei Schüler*innen (und auch bei Erwachsenen) im Mathematikunterricht. Die Wahrscheinlichkeiten $P(A|B) = P(B|A)$ werden fälschlicherweise gleich gesetzt. Dies führt in der Schule zu einer Fehleinschätzung, die nicht weiter schlimm ist, hier war sie jedoch tragisch. Sally Clark wurde verurteilt und war drei Jahre lang im Gefängnis. Sie dürfte dabei psychische Probleme entwickelt haben, die am Ende zu ihrem Tod im Jahre 2007 durch einer Alkoholvergiftung geführt haben. Dieser Fall zeigt, dass der Umgang mit bedingten Wahrscheinlichkeiten in speziellen Fällen sehr wichtig ist.

6 Resümee

Diese Masterarbeit beschäftigt sich mit bedingten Wahrscheinlichkeiten und deren Anwendung in der Schule, sowie weiteren stochastischen Paradoxa, welche ein besseres Verständnis für (bedingte) Wahrscheinlichkeiten liefern sollen. Bedingte Wahrscheinlichkeiten sind für Menschen oft etwas „paradoxes“, da sie teilweise gegen die Intuition des Menschen gehen und dadurch oft zu Verwunderung führen. Dass ein positiver Test eines Ungeborenen auf Trisomie 21 bei einer schwangeren Frau Panik auslöst, ist der bedingten Wahrscheinlichkeit geschuldet.

Der Schwerpunkt bei dieser Masterarbeit liegt auf dem Präsentieren der mathematischen Sachverhalte dieser Paradoxa auf Schulniveau sowie eine fachdidaktische Beleuchtung der bedingten Wahrscheinlichkeit. Besonders bei der Einführung der bedingten Wahrscheinlichkeit wurde darauf geachtet, dass das Konzept so in der Schule ebenfalls durchführbar ist. Auf die Verwendung des Satzes von Bayes wurde weitestgehend verzichtet. Ich bin mir persönlich bewusst, dass nicht alle Paradoxa in der Schule be-

trachtet werden können, vor allem aus Zeitgründen. Diese Masterarbeit soll jedoch eine Auswahl an Beispielen liefern, die für die Schule geeignet ist.

Aus meiner Sicht ist das gesamte Themengebiet der Stochastik und insbesondere die bedingte Wahrscheinlichkeit eine Materie, die auch von Lehrer*innen häufiger nicht ganz verstanden wird. Dadurch kommt es in Folge zwangsläufig auch im Unterricht zu Schwierigkeiten beim Verständnis bei den Schüler*innen (die durch ein besseres Verständnis auf der Seite des/der Lehrenden so nicht notwendig sind).

Mit dieser Masterarbeit wurde dem/der Leser*in gezeigt, dass bedingte Wahrscheinlichkeiten potentiell wichtig für den Schulunterricht sind und sich den Platz in den Lehrplänen der Mathematik zurecht verdient haben.

7 Literaturverzeichnis

- [1] Bewersdorff, Jörg. *STATISTIK - Wie und warum sie funktioniert: Ein mathematisches Lesebuch*. Vieweg+Teubner, Wiesbaden. 2011
- [2] Bosch, Karl; *Elementare Einführung in die Wahrscheinlichkeitsrechnung*. 10. Auflage. Vieweg+Teubner, Wiesbaden. 2010
- [3] Bourier, Günther. *Statistik-Übungen: Beschreibende Statistik - Wahrscheinlichkeitsrechnung - Schließende Statistik*. 5. Auflage. Springer Fachmedien, Wiesbaden. 2014
- [4] Bundesministerium für Bildung. *Lehrplan für die Allgemeinbildende höhere Schule*
<https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=BundesnormenGesetz&esnummer=10008568>; zuletzt aufgerufen am 23.03.2021
- [5] Fischer, Gerd; Lehner, Matthias; Puchert, Angela. *Einführung in die Stochastik: Die grundlegenden Fakten mit zahlreichen Erläuterungen, Beispielen und Übungsaufgaben*. Springer Spektrum, Wiesbaden. 2015
- [6] *Gerichtsprotokoll zum Fall Sally Clark, 3 Gerichtsverhandlung*. Zeilen 94 - 110.
<http://www.bailii.org/ew/cases/EWCA/Crim/2003/1020.html>; zuletzt aufgerufen am 23.04.2021
- [7] Glaser, Christian. *Risiko im Management*. Seiten 205-208. Springer Fachmedien, Wiesbaden. 2019
- [8] Hauer-Typpelt, Petra. *Unbedingt „Bedingte Wahrscheinlichkeiten? Der Satz von Bayes im Schulunterricht* in: zur Didaktik der Mathematik der Höheren Schulen der Österreichischen Mathematischen Gesellschaft,, Heft 39. S.54-65, Wien. 2007

- [9] Hofer, Elisabeth; Abels, Simone; Lembens, Anja. *Forschendes Lernen und das 5E-Modell* in Plus Lucis Naturwissenschaftsdidaktik, Heft 1/2016
Kampf Kampf, Jürgen; Kunze Markus. *Stochastik für Wirtschaftswissenschaftler*. Universität Ulm, Ulm 2016
Schwach Kütting, Herbert; Sauer Martin J. *Elementare Stochastik*. Springer Verlag, Berlin Heidelberg. 2013
- [10] Mittag, Hans-Joachim; Schüller Katharina. *Statistik: Eine Einführung mit interaktiven Elementen*. 6. Auflage. Springer Verlag, Berlin, Heidelberg. 2020
- [11] Mittelbach, Henning. *Statistik*. S. 65 - 76. Walter de Gruyter GmbH Co. KG, Berlin. 2015
- [12] Rosenhouse, A.P.M.J. ; Rosenhouse, J . *The Monty Hall Problem: The Remarkable Story of Math's Most Contentious Brain Teaser*. Oxford University Press, Oxfors. 2009
- [13] Strick, Heinz Klaus. *Stochastische Paradoxien*. Springer Fachmedien Imprint: Springer Spektrum, Wiesbaden. 2020
- [14] Statistik-hagen. *Interaktive Seite für relative Häufigkeit beim Münzwurf*.
<https://stat.iks-hagen.de/app/muenzwurf/>; zuletzt besucht am 23.03.2021
- [15] Székely, Gábor J. *Paradoxa:klassische und neue Überraschungen aus Wahrscheinlichkeitsrechnung und mathematischer Statistik*. Verlag Harri Deutsch, Thun und Frankfurt am Main. 1990
- [16] Thompson, W.C.; Shumann, E.L. *Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy*. Law and Human Behaviour, Kapitel 2. 1987
- [17] Tietze, Uwe-Peter; Klika, Manfred; Wolpers, Hans (Hrsg.). *Mathematikunterricht in der Sekundarstufe II*. Band 3: Didaktik der Stochastik. Vieweg Verlag, Braunschweig. 2002

- [18] Vogt, Peter. *Kontingenz und Zufall: Eine Ideen- und Begriffsgeschichte*. Akademie Verlag, Berlin. 2006
- [19] Watkins, Stephen J. *Conviction by mathematical error? Doctors and lawyers should get probability theory right*.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1117305/>; zuletzt aufgerufen am 23.04.2021
- [20] Weitz, Edmund. *Konkrete Mathematik (nicht nur) für Informatiker*. Springer Spektrum, Wiesbaden. 2018
- [21] Zwerenz, Karlheinz. *Statistik : Einführung in die computergestützte Datenanalyse*. De Gruyter Oldenbourg, München. 2015