



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Punkt, Spiel, Satz, Sieg:
Gewinnwahrscheinlichkeiten im Tennis“

verfasst von / submitted by

Barbara Krivanek, BEd

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

Master of Education (MEd)

Wien, 2024 / Vienna, 2024

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

A 199 507 520 02

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

Masterstudium Lehramt Sek (AB) Lehrverbund
UF Englisch UF Mathematik

Betreut von / Supervisor:

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

Abstract Deutsch

In den vergangenen Jahren hat Tennis, insbesondere die vier großen Grand Slam Turniere, erheblich an Popularität gewonnen. Angesichts der Herausforderung, die Stochastik für viele Schüler*innen darstellt, gewinnt die Integration von Beispielen, die den Wert interdisziplinärer Aufgaben verdeutlichen, zunehmend an Bedeutung. Das wachsende Interesse bietet die Gelegenheit, Tennis als anschauliches Beispiel für die Verbindung von Sport und Mathematik zu nutzen.

Das Ziel dieser Masterarbeit ist es, zu untersuchen, inwiefern stochastische Modelle dazu dienen können, Sieger*innen von Tennisspielen vorherzusagen. Um diese Frage zu beantworten, werden verschiedene stochastische Modelle zur Berechnung der Gewinnwahrscheinlichkeit eines Games sowie eines Matches erstellt. Nach einführenden Kapiteln, welche die wesentlichen Aspekte der Kombinatorik und Wahrscheinlichkeitsrechnung behandeln, werden die stochastischen Modelle anhand eines Datensatzes der US Open-Finals mittels eines χ^2 -Anpassungstest getestet. Die Ergebnisse dieser statistischen Tests legen nahe, dass selbst in den US Open-Finals nicht von einer gleichwertigen Spielstärke der beiden Finalist*innen ausgegangen werden kann.

Im schulischen Kontext fühlen sich Schüler*innen oft von der Komplexität der Aufgaben überfordert, die zudem wenig Relevanz für sie haben. Um diesem Trend entgegenzuwirken, ist es entscheidend, den Schüler*innen Aufgaben zur Verfügung zu stellen, die ihre Interessen widerspiegeln. Diese Masterarbeit stellt daher einige Aufgaben vor, die auf einen fächerübergreifenden Ansatz zur Vermittlung von Stochastik in der Schule abzielen.

Keywords: Tennis, Gewinnwahrscheinlichkeiten, Statistische Verfahren, US Open

Abstract Englisch

In recent years, tennis has witnessed an unprecedented surge in popularity, particularly with the four major championships gaining widespread acclaim. Given the inherent challenges students face with stochastic concepts, there is a crucial need to design engaging assignments that underscore the value of interdisciplinary tasks. Due to this substantial interest in the sport, tennis proves to be a good example to illustrate how mathematics and sports are intertwined.

The purpose of this master's thesis is to explore how the application of stochastic models can be employed for predicting the outcome of a tennis match. To answer this question, numerous stochastic models for the winning probability and game winning probability are built. Following foundational chapters that establish the theoretical framework, these models are tested using a dataset derived from the US Open finals. The outcome of statistical tests reveals that even in the US Open finals the assumption of identical point-winning probabilities is not valid.

In the school context, students often feel overwhelmed by tasks that appear complex and lack personal relevance. To counteract this challenge, it is of utmost importance to provide students with material and tasks that align with their interests. Hence, this master's thesis introduces a set of tasks designed to foster an interdisciplinary approach to teaching stochastic concepts in schools.

Keywords: Tennis, winning probability, statistical hypothesis test, US Open

Inhaltsverzeichnis

Abstract Deutsch	i
Abstract Englisch	iii
1 Einleitung Tennissport	1
1.1 Tennissport im Wandel der Zeit	1
1.2 Grundlagen des Tennisspiels	2
1.3 Turniere im Überblick	5
2 Wahrscheinlichkeitsrechnung	7
2.1 Wahrscheinlichkeitsbegriff	7
2.2 Laplace Experimente	9
2.3 Kombinatorik	12
2.3.1 Geordnete Stichprobe	12
2.3.2 Ungeordnete Stichprobe	17
2.4 Diskrete Zufallsvariable	20
2.4.1 Definition einer diskreten Zufallsvariable	21
2.4.2 Verteilungsfunktion	23
2.4.3 Erwartungswert	26
2.4.4 Varianz und Standardabweichung	28
2.4.5 Binomialverteilung	30
2.4.6 Poisson Verteilung	35
2.5 Stetige Zufallsvariable	39
2.5.1 Erwartungswert, Varianz und Standardabweichung	41
2.5.2 Normalverteilung	43
2.5.3 Chi-Quadrat-Verteilung	53
3 χ^2- Anpassungstest	55
4 Modellierung eines Tennismatches	59
4.1 Gewinnwahrscheinlichkeit eines Games	59
4.2 Zu erwartende Dauer eines Games	62
4.3 Matchgewinnwahrscheinlichkeit in Abhängigkeit von der Satzgewinn- wahrscheinlichkeit	66

Inhaltsverzeichnis

4.4	Gewinnwahrscheinlichkeit eines Tie-Breaks	70
4.5	Gewinnwahrscheinlichkeit beim Aufschlag	74
5	Auswertung aller bisherigen US-Open Finalsple	81
5.1	Erwartete Anzahl gespielter Sätze im Herreneinzel	82
5.2	Erwartete Anzahl gespielter Sätze im Dameneinzel	84
6	Tennissport im Mathematikunterricht	89
6.1	Realitätsbezug im Mathematikunterricht	89
6.2	Stochastik im Lehrplan allgemeinbildende höherer Schulen	90
6.2.1	Sekundarstufe 1	91
6.2.2	Sekundarstufe 2	92
6.3	Aufgaben zum Thema Stochastik im Tennis	93
7	Fazit	99
	Literaturverzeichnis	101
	Abbildungsverzeichnis	106

1 Einleitung Tennissport

1.1 Tennissport im Wandel der Zeit

Heutzutage zählt der Tennissport nicht nur zu den bekanntesten und beliebtesten Freizeitsportarten, sondern zieht auch mit seinen Grand-Slam-Turnieren jährlich Millionen Menschen in seinen Bann. Entgegen der Erwartungen war der Tennissport in seinen Anfängen im 13. Jahrhundert alles andere als ein Volkssport [20, S.9]. Der einst elitäre Sport, der anfangs ausschließlich Geistlichen vorbehalten war, fand seinen Ursprung im nördlichen Frankreich, wo Mönche und Aristokraten eine aus Leder gefertigte und mit Wolle, Sand oder Haaren gefüllte Kugel mit der Handfläche hin und her spielten. Aufgrund dieser Spielweise wurde der Sport als *Jeu de Paume*, was aus dem Französischen übersetzt so viel wie ‚Spiel mit der Handinnenfläche‘ bedeutet, bezeichnet [19, S.18]. Überlieferungen zufolge wurde der Sport in Innenhöfen der Klöster ausgeübt, wobei das für die damalige Zeit charakteristische schräge Dach in das Regelwerk aufgenommen wurde. So war es beim Service notwendig, die Lederkugel an einer Längsseite des Daches zu platzieren [20, S.11].

Obwohl sich der Sport rasch auch unter den Bürger*innen verbreitete, stellte die Suche nach einem passenden Platz eine große Herausforderung dar. So wurde oft auf einem Kirchhof gespielt, was zur Folge hatte, dass nicht nur Geistliche von ihren Aufgaben und Tätigkeiten abgehalten wurden, sondern auch der Gottesdienst aufgrund des Lärms unterbrochen wurde. Diese Störungen sowie die Tatsache, dass Herrscher es bevorzugten ihre Untertanen bei Übungen mit Waffen zu sehen, führte zu Strafen und Spielverboten für Bürger*innen [20, S.13-14].

Trotz der anhaltenden Restriktionen gelang dem Tennissport im 16. und 17. Jahrhundert sein Durchbruch als Volkssport [20, S.20]. Den Grundstein dafür legte

1 Einleitung Tennissport

1555 das von dem Italiener Antonio Scaino veröffentlichte Werk *Trattato del gioco della palla*, in welchem er die Plätze, Bälle sowie Spielregeln detailliert erläuterte. Beispielsweise gab es in dieser Zeit des Aufschwungs zwei verschiedene Bälle. Entschied man sich für die ursprüngliche Version, bei der mit der Handfläche gespielt wurde, so wurde ein weicherer und größerer Ball verwendet. Wollte man jedoch die neu aufkommende Variante mit einem ovalen Racket aus Holz spielen, so kam ein festerer Ball zum Einsatz [20, S.48-49]. Die enorme Popularität hatte jedoch auch seine Schattenseiten. Da Sportstätten oft als ein gesellschaftlicher Treffpunkt fungierten, sahen vor allem arme Bürger*innen darin eine Chance Geld mit Sportwetten zu verdienen. So verlor der Tennissport zusehends seinen guten Ruf, da er nun mit exzessivem Alkoholkonsum sowie Diebstahl in Verbindung gebracht wurde. Dies hatte zur Folge, dass gegen Ende des 17. Jahrhunderts einige Tennisplätze geschlossen wurden [20, S.55-61].

Nach dieser turbulenten Phase gelang dem Tennissport Ende des 19. Jahrhunderts ein Neubeginn. So wird der Engländer Walter Clopton Wingfield als Begründer des modernen Tennissportes gesehen. Er entwickelte das *Lawn Tennis* und ließ dieses patentieren. Großes Interesse weckte auch das Tennisset, welches die wichtigsten Utensilien, wie Schläger, Bälle und Netz beinhaltete. Den endgültigen Durchbruch erreichte der Sport 1877 in Wimbledon, als er ins Vereinsprogramm der ‚All England Croquet Clubs‘ aufgenommen wurde und damit den Grundstein für das heutige Grand-Slam-Turnier in London legte [19, S.20-22].

1.2 Grundlagen des Tennisspiels

Tennis kann entweder zu zweit oder zu viert gespielt werden. Dadurch wird zwischen drei unterschiedlichen Varianten eines Matches unterschieden, nämlich dem Einzel, Doppel und Mixed. Während sich bei einem Einzelmatch auf jeder Seite ein*e Spieler*in befindet, spielen bei einem Doppel und Mixed jeweils zwei Spieler*innen auf einer Seite des Netzes. Während beim Doppel zwischen Herren- und Damendoppel unterschieden wird, spielen bei einem Mixed zwei Paare aus jeweils einer Dame und einem Herrn gegeneinander [19, S.150-151].

1.2 Grundlagen des Tennisspiels

Die Größe des Spielfeldes hängt von der durchgeführten Spielform ab und ist in den offiziellen Tennisregeln der International Tennis Federation (ITF) festgelegt. Das rechteckige Spielfeld hat stets eine Länge von 23,77 Metern und eine Breite von 8,23 Metern im Einzel und 10,97 Meter im Doppel sowie Mixed. Essentiell sind die Linienmarkierungen des Spielfeldes, welche in der nachstehenden Abbildung 1.1 dargestellt sind. Das gesamte Spielfeld ist durch Grundlinien, Seitenlinien sowie Aufschlaglinien begrenzt. Das 91 Zentimeter hohe Netz wird durch Pfosten mit einer Höhe von 107 Zentimeter gespannt und teilt das Spielfeld in zwei gleichgroße Hälften. Parallel im Abstand von 6,40 Metern zum Netz befindet sich die Aufschlaglinie. Die Aufschlagsmittellinie, welche auch als T-Linie bezeichnet wird, führt vom Netz mittig bis zur Aufschlaglinie. Im Regelwerk ist verankert, wenn ein Ball die Linie berührt, dass dieser weiterhin im Spiel ist [19, S.151].

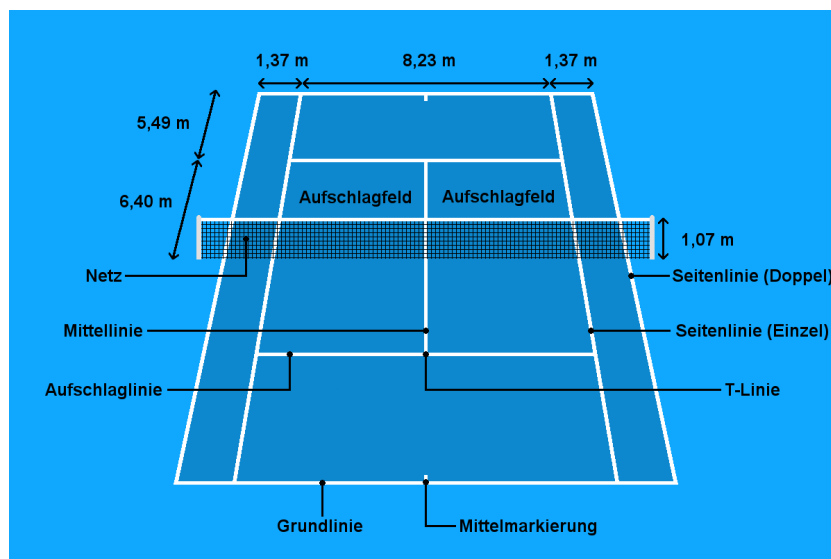


Abbildung 1.1: Abmaße eines Tennisfeldes (Quelle: <https://tennis-uni.com/tennisplatz-tennisfeld> - zuletzt zugegriffen am: 2.2.2024)

Bei Tennis handelt es sich um ein Rückschlagspiel, welches niemals in einem Unentschieden endet, sondern stets ein*e Gewinner*in und ein*e Verlierer*in vom Platz gehen. Ziel des Spiels ist es die für den Sieg benötigte Anzahl von Sätzen für sich zu entscheiden. Wie viele Sätze für einen Matchgewinn notwendig sind, hängt von dem Spielmodus ab. Es wird dabei zwischen Best-of-3 und Best-of-5

1 Einleitung Tennissport

unterschieden. Während bei Letzterem maximal fünf Sätze gespielt werden bis das Match entschieden ist, werden beim Best-of-3 nur maximal drei Sätze gespielt. Ein Satz ist dann entschieden, wenn sechs Spiele, welche auch Games genannt werden, mit einem zwei-Spiel Unterschied gewonnen wurden. Weiters gilt ein Game als gewonnen, wenn vier Punkte erreicht wurden [19, S.164-165].

Der bekannte Ausruf ‚Spiel, Satz, Sieg‘ ist auf die Struktur eines Tennisspieles zurückzuführen, welche im Folgenden erläutert wird:

Punkt: Ziel ist es den Ball mit Hilfe des Schlägers in der gegnerischen Hälfte zu platzieren, ohne dabei den Ball mit einem Körperteil zu berühren. Der Tennisball darf vor dem Schlag den Boden in der eigenen Spielhälfte maximal einmal berühren und muss anschließend über das Netz in die gegnerische Hälfte gebracht werden. Jene*r Spieler*in, der*die es öfters schafft den Ball regelkonform zurückzuspielen hat einen Punkt erzielt. Die gewonnenen Punkte werden mit der Zählweise 15, 30, 40 und Spiel gezählt [19, S.164].

Spiel oder Game: Ein Game ist dann gewonnen, wenn ein*e Spieler*in vier Punkte erreicht hat. Kommt es zu einem Einstand, also einem Spielstand von 40:40, muss ein*e Spieler*in zwei Punkte mehr als der*die Gegner*in gewinnen, um das Spiel für sich zu entscheiden. Angenommen, Spieler*in A gewinnt den ersten Punkt nach dem Einstand, so ist diese*r im ‚Vorteil‘. Bei einem weiteren Punktgewinn von Spieler*in A ist das Game gewonnen, bei einem Punktgewinn von Spieler*in B steht es erneut ‚Einstand‘ [19, S.164-165].

Satz: Für einen Satz müssen mindestens sechs Games gewonnen werden, wobei wieder ein Unterschied von zwei gewonnenen Spielen nötig ist. Aus diesem Grund gibt es folgende Möglichkeiten für einen Satzsieg 6:0, 6:1, 6:2, 6:3, 6:4 und 7:5. Der letztere Fall tritt bei einem Zwischenstand von 5:5 ein. Bei einem Spielstand von 6:6 kommt es zu einem ‚Tie-Break‘. Um dieses Tie-Break zu gewinnen, müssen mindestens sieben Punkte mit mindestens zwei Punkten Unterschied erzielt werden [19, S.164].

Sieg: Wie viele Sätze für einen Sieg notwendig sind hängt vom Spielmodus ab. Während meistens im Best-of-3 Modus gespielt wird, müssen bei den vier Grand-Slam-Turnieren im Herren Einzel drei Sätze gewonnen werden. Im Falle eines

Gleichstands in einem lang andauernden Match wird bei bestimmten Turnieren der entscheidende dritte bzw. fünfte Satz nicht ausgespielt, sondern es kommt stattdessen ein sogenanntes ‚Match-Tiebreak‘ zum Einsatz. Im Gegensatz zum Tie-Break wird hier bis zu zehn Punkten gespielt, wobei weiterhin ein Unterschied von zwei gewonnen Punkten notwendig ist [19, S.164-165].

1.3 Turniere im Überblick

Der Internationale Tennisverband (ITF) wurde 1913 von 15 Verbänden als ‚Federation Internationale de Lawn Tennis‘ (FILT) ins Leben gerufen und ist erst seit 1977 unter dem heutigen Namen bekannt. Bis dato sind 211 nationale Verbände der ITF beigetreten. Als höchstes Organ des Tennissports organisiert die ITF eine Vielzahl an Turnieren. Die bekanntesten sind hierbei die vier Grand-Slam-Turniere welche jährlich in Melbourne, Paris, London-Wimbledon und New York stattfinden. Für das Teilnehmer*innenfeld des zweiwöchigen Turniers qualifizieren sich 128 Einzelspieler*innen jeweils für die Damen- und Herrenbewerbe, sowie 64 Paare für die Doppelspiele. Im Gegensatz zu den meisten Turnieren, wird bei den Herren im Best-of-5 Modus gespielt. Des Weiteren ist die ITF für diverse Mannschaftsbewerbe wie dem Davis Cup, Billie Jean King Cup und Hopman Cup sowie den Wettbewerben bei den Olympischen und Paralympischen Spielen zuständig [19, S.36-37, 52].

Da Frauen und Männer größtenteils getrennte Turniere bestreiten, wurden zwei Vereinigungen gegründet. Die ATP ist die Spielervereinigung der professionellen männlichen Tennisspieler. Jährlich werden zwischen 60 und 70 Turniere im Zuge der ATP Tour organisiert [19, S.38-39]. Den Abschluss der ATP Tour bilden die ATP Finals, bei welchem die acht besten Einzelspieler sowie Doppelpaare starten dürfen. Die Turniere werden basierend auf ihrer Wertigkeit in folgende drei Kategorien aufgeteilt: ATP Masters 1000, ATP 500 sowie ATP 250 [19, S.53-54].

Das Pendant zu ATP ist die WTA, die Vereinigung der professionellen Tennisspielerinnen. Seit 1995 werden im Zuge der WTA Tour ebenfalls rund 60 Turniere

1 Einleitung Tennissport

ausgetragen [19, S.40]. Die Struktur der Turniere ist stark an jene der ATP angelehnt. Am Ende der WTA Tour finden einerseits die WTA Finals mit den acht punktebesten Einzelspielerinnen sowie Doppelpaaren und andererseits die WTA Elite Trophy statt. Ähnlich wie bei der ATP, werden auch bei der WTA die Turniere in folgende drei Kategorien unterteilt: WTA 1000, WTA 500 und WTA 250 [19, S.55-56].

Seit 1973 veröffentlicht die ATP und seit 1975 auch die WTA regelmäßig eine Weltrangliste. Beide Spieler*innenvereinigungen gehen bei der Berechnung nach dem gleichen Prinzip vor und veröffentlichen ihre Rankings wöchentlich während der Saison und während der Grand-Slam-Turniere alle zwei Wochen. Als Basis für die Erstellung der Weltrangliste dienen 16 Turniere bei den Damen und 18 Turniere bei den Herren der vergangenen 52 Wochen. Bei der ATP werden die vier Grand-Slam Turniere, acht der neun ATP Master 1000 Turniere sowie die sechs besten Ergebnisse aus ATP 500, ATP 250, ATP Challenger Tour, ITF Men's World Tennis Tour und dem Turnier in Monte Carlo für die Berechnung der Weltrangliste berücksichtigt. Für jene acht Spieler, welche die ATP Finals erreichen, wird dieses Ergebnis als 19. Turnier zusätzlich in der Berechnung berücksichtigt. Bei der WTA Weltrangliste werden nur 16 Turniere in die Berechnung miteinbezogen, für die acht Spielerinnen der WTA Finals fließt dieses auch analog zu den Herren, als zusätzliches Turnier in das Ranking ein [19, S.195-196].

2 Wahrscheinlichkeitsrechnung

Bevor die verschiedenen Gewinnwahrscheinlichkeiten im Tennissport betrachtet werden können, ist es notwendig die grundlegenden Begriffe zu definieren sowie essentielle Bereiche der Wahrscheinlichkeitsrechnung zu erarbeiten.

2.1 Wahrscheinlichkeitsbegriff

Definition 1. Bei einem *Zufallsexperiment* handelt es sich um ein Experiment, bei welchem schon vor der Durchführung die möglichen Ergebnisse bekannt sind, das tatsächlich eintretende Ergebnis jedoch vom Zufall abhängt [2, S.1]. Die Menge aller möglichen Ergebnisse eines Zufallsexperiments wird als *Grundraum* Ω bezeichnet.

$$\Omega = \{\omega \mid \omega \text{ ist ein mögliches Ergebnis des Zufallsexperiments}\}$$

Ein einzelnes Element ω des Grundraums wird *Ergebnis* genannt, während die Menge mehrerer Ergebnisse *Ereignis* heißt und mit großen lateinischen Buchstaben angeschrieben wird. Bei einem Ereignis, welches genau ein Element beinhaltet handelt es sich um ein *Elementarereignis* [5, S.154]. Die Menge aller Ereignisse wird als *Potenzmenge* $P(\Omega)$ bezeichnet [18, S.2].

Ein alltägliches Beispiel für ein Zufallsexperiment ist das Werfen eines Würfels. Während die möglichen Ergebnisse bereits vor dem Würfeln mit der Grundmenge $\Omega = \{1, 2, 3, 4, 5, 6\}$ bekannt sind, hängt die tatsächlich gewürfelte Augenzahl vom Zufall ab. In diesem Zusammenhang wäre die Menge

$$A = \{\text{Es wird eine ungerade Zahl gewürfelt}\} = \{1, 3, 5\}$$

2 Wahrscheinlichkeitsrechnung

ein mögliches Ereignis.

Während der Wahrscheinlichkeitsbegriff schnell abstrakt und wenig greifbar erscheinen mag, gelang es dem russischen Mathematiker Kolmogoroff 1933 diesen axiomatisch einzuführen [2, S.8]. Zuerst wird die Menge aller zulässigen Ereignisse mit Hilfe einer σ -Algebra festgelegt.

Definition 2. Ein System $\mathcal{A}$ von Teilmengen aus Ω heißt σ -Algebra, wenn gilt [18, S.3]:

1. $\emptyset \in \mathcal{A}$
2. aus $A \in \mathcal{A}$ folgt, dass $A' = \Omega \setminus A \in \mathcal{A}$
3. für abzählbar unendlich viele $A_1, A_2, \dots \in \mathcal{A}$ gilt $\bigcup_{j=1}^{\infty} A_j \in \mathcal{A}$

Definition 3. Sei $\mathcal{C} \subseteq P(\mathbb{R})$ die Menge aller abgeschlossenen endlichen Intervalle in $\mathbb{R}$, d.h.

$$\mathcal{C} = \{[a, b] : a < b\}, \quad a, b \in \mathbb{R}$$

so wird die von $\mathcal{C}$ auf $\mathbb{R}$ erzeugte σ -Algebra *Borelsche σ -Algebra* $B(\mathbb{R})$ genannt. Mengen aus $\mathcal{B}(\mathbb{R})$ heißen *Borelmengen* [18, S.5].

Definition 4. Mengen $A_i, i \in I$ heißen *paarweise disjunkt*, wenn gilt [6, S.52]:

$$A_i \cap A_j = \emptyset \quad \text{mit } i \neq j$$

Definition 5. Sei Ω ein beliebiger Grundraum mit einer σ -Algebra $\mathcal{A}$. Eine Abbildung P wird *Wahrscheinlichkeitsmaß* auf $(\Omega, \mathcal{A})$ bezeichnet, wenn gilt [18, S.7]:

1. $P : \mathcal{A} \rightarrow [0, 1]$ mit $P(\emptyset) = 0$ und $P(\Omega) = 1$

2. P ist σ -additiv, d.h. für abzählbar viele paarweise disjunkte Mengen $A_1, A_2, \dots \in \mathcal{A}$ gilt

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j)$$

Daraus lassen sich weitere Eigenschaften von Wahrscheinlichkeiten folgern [2, S.9]:

1. Für jedes Ereignis A gilt: $P(A') = 1 - P(A)$
2. Für das unmögliche Ereignis $\emptyset$ gilt $P(\emptyset) = 0$
3. Aus $A \subset B$ folgt $P(A) \leq P(B)$
4. Für beliebige Ereignisse A und B gilt $P(B \setminus A) = P(B \cap A') = P(B) - P(B \cap A)$
5. Für beliebige Ereignisse A und B gilt $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Definition 6. Ein Tripel $(\Omega, \mathcal{A}, P)$ mit dem Grundraum Ω , einer σ -Algebra $\mathcal{A}$ bestehend aus Teilmengen aus Ω und einem Wahrscheinlichkeitsmaß P wird *Wahrscheinlichkeitsraum* genannt [18, S.8].

2.2 Laplace Experimente

Ähnlich wie beim Werfen eines Würfels, wo es endlich viele mögliche Ausgänge gibt, welche als gleichwahrscheinlich angesehen werden, gibt es unzählige solcher Zufallsexperimente. Weitere Beispiele wären unter anderem das Werfen einer Münze oder das Drehen eines Glücksrades. Das folgende Kapitel ist an das Werk von Bosch [2, S.12-13] angelehnt.

Definition 7. Ein Zufallsexperiment, welches folgende zwei Eigenschaften erfüllt, wird *Laplace-Experiment* genannt.

2 Wahrscheinlichkeitsrechnung

1. Der Grundraum Ω besteht aus m verschiedenen Ereignissen

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_m\}$$

und besitzt die Elementarereignisse $\{\omega_1\}, \{\omega_2\}, \dots, \{\omega_m\}$.

2. Die Elementarereignisse besitzen alle die gleiche Wahrscheinlichkeit p , d.h. es gilt

$$P(\{\omega_1\}) = P(\{\omega_2\}) = \dots = P(\{\omega_m\}) = p$$

Das bedeutet, damit ein Zufallsexperiment als Laplace-Experiment klassifiziert werden kann, dürfen einerseits nur endlich verschiedene Ergebnisse möglich sein und andererseits müssen alle Elementarereignisse mit derselben Wahrscheinlichkeit eintreten.

Aus

$$\Omega = \{\omega_1\} \cup \{\omega_2\} \cup \dots \cup \{\omega_m\}$$

folgt

$$\begin{aligned} 1 &= P(\Omega) = P(\{\omega_1\}) + P(\{\omega_2\}) + \dots + P(\{\omega_m\}) \\ &= p + p + \dots + p = m \cdot p \end{aligned}$$

und durch Umformen erhält man

$$p = \frac{1}{m}$$

Somit gilt für die Wahrscheinlichkeit eines beliebigen Ereignisses

$A = \{\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_r}\}$, welches r verschiedene Ergebnisse besitzt, dass

$$P(A) = P(\{\omega_{i_1}\}) + \dots + P(\{\omega_{i_r}\}) = r \cdot p = \frac{r}{m}$$

Für die Berechnung der Wahrscheinlichkeit $P(A)$ gilt somit

$$P(A) = \frac{r}{m} = \frac{\text{Anzahl der für A günstigen Fälle}}{\text{Anzahl der insgesamt möglichen Fälle}} = \frac{|A|}{|\Omega|}$$

Beispiel: Ein alltägliches Beispiel für ein Laplace-Experiment wäre das Drehen eines Glücksrads mit drei gleich großen Sektoren, welche in den Farben Rot, Blau und Grün eingefärbt sind. Somit hat dieses Zufallsexperiment drei mögliche Fälle, nämlich $\Omega = \{Rot, Grün, Blau\}$. Ein Gewinn wird dann erzielt, wenn die Nadel des Glücksrades im blauen oder grünen Sektor stehen bleibt. Für dieses Ereignis $G = \{\text{Gewinn}\}$ gibt es zwei günstige Fälle, nämlich $A = \{Blau, Grün\}$. Somit kann aufgrund der Tatsache, dass alle Elementarereignisse mit der gleichen Wahrscheinlichkeit $p = \frac{1}{3}$ eintreten, die Wahrscheinlichkeit $P(G)$ folgendermaßen berechnet werden:

$$P(G) = \frac{|G|}{|\Omega|} = \frac{2}{3}$$

Anhand dieses Beispiels kann gezeigt werden, dass es teilweise effizienter ist, wenn anstelle der für ein Ereignis G günstigen Elementarereignisse die für G ungünstigen Elementarereignisse abgezählt werden. Anhand dieses dadurch bestimmten Gegenereignisses $\bar{G}$, kann die Wahrscheinlichkeit $P(G)$ folgendermaßen ermittelt werden:

$$P(G) = 1 - P(\bar{G}) = 1 - \frac{1}{3} = \frac{2}{3}$$

Die Verwendung des Gegenereignisses bei der Berechnung einer Wahrscheinlichkeit wird dann angewendet, wenn das Abzählen der ungünstigen Elementarereignisse weniger aufwendig als das Abzählen der günstigen Elementarereignisse ist. Die Berechnung der Wahrscheinlichkeit eines Laplace-Experimentes reduziert sich somit auf das Abzählen von Elementarereignissen.

2.3 Kombinatorik

Da offensichtlich Stichproben eine bedeutende Rolle bei der Berechnung von Wahrscheinlichkeiten von Laplace-Experimenten spielen, befasst sich die Kombinatorik, ein Teilgebiet der Mathematik, mit dem Abzählen. Für ein besseres Verständnis wird häufig das Urnenmodell herangezogen. Dabei befinden sich in einer Urne n nummerierte oder gefärbte Kugeln, aus welchen k -mal gezogen wird. Jeder Ziehvorgang des mehrstufigen Vorgangs entspricht im Baumdiagramm einem vom Startpunkt ausgehenden Pfad. In diesem Modell kann mit Hilfe der ersten Pfadregel die Wahrscheinlichkeit eines Pfades berechnet werden, indem das Produkt aller Wahrscheinlichkeiten entlang des Pfades ermittelt wird. Setzt sich ein Ereignis A aus mehreren Pfaden zusammen, so wird die Wahrscheinlichkeit mit Hilfe der zweiten Pfadregel berechnet, welche besagt, dass die Wahrscheinlichkeit $P(A)$ durch die Summe aller zu A gehörenden Pfade ermittelt wird [13, S.92].

Weitere geeignete Modelle für die Veranschaulichung sind das Mengenmodell, bei welchem aus einer n -elementigen Menge eine k -elementige Teilmenge ausgewählt werden soll, sowie das Fächermodell, bei welchem auf n nummerierte Fächer k Teilchen verteilt werden sollen [13, S.62].

2.3.1 Geordnete Stichprobe

Es sei eine Menge $M = \{1, \dots, n\}$ mit n Elementen, aus welcher k Elemente ausgewählt werden sollen. Spielt bei der Auswahl der k Elemente die Reihenfolge eine bedeutende Rolle, so wird diese geordnete Auswahl *Variation* bezeichnet. Weiters wird bei der Auswahl unterschieden, ob es sich um eine Auswahl mit oder ohne Wiederholung handelt [17, S.21].

Sei nun $M = \{1, \dots, n\}$ eine n -elementige Menge und $1 \leq k \leq n$, so kann die Anzahl der geordneten Stichproben vom Umfang k mit Wiederholung folgendermaßen berechnet werden [17, S.21]:

$$V(n, k) = n^k$$

Beispiel: Mit Hilfe des Urnenmodells kann diese Variation als Ziehen einer geordneten Stichprobe mit Zurücklegen veranschaulicht werden. In einer Urne befinden sich eine grüne und eine gelbe Kugel. Nun wird in die Urne gegriffen und eine Kugel gezogen. Nachdem die Farbe der Kugel notiert wurde, wird diese wieder in die Urne zurückgelegt. Berechne die Wahrscheinlichkeit, dass bei dreimaligem Ziehen die Farbreihenfolge Grün-Grün-Gelb gezogen wird.

Das nachstehende Baumdiagramm 2.1 veranschaulicht die möglichen Ziehungsverläufe.

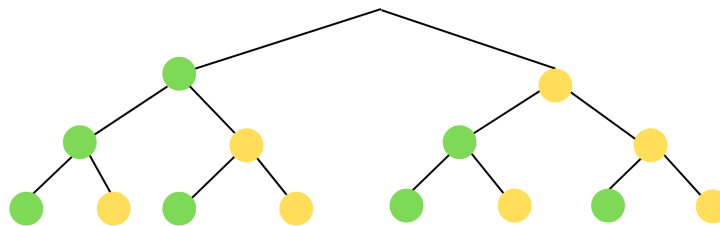


Abbildung 2.1: Baumdiagramm - Ziehen einer geordneten Stichprobe mit Zurücklegen

Anhand des Baumdiagramms kann erkannt werden, dass es insgesamt 8 mögliche Ereignisse gibt, jedoch nur einen günstigen Fall der zur gewünschten Farbreihenfolge Grün-Grün-Gelb führt. Da es sich hierbei um ein Laplace-Experiment handelt, wird die Wahrscheinlichkeit folgendermaßen berechnet:

$$P(\text{Grün-Grün-Gelb}) = \frac{\text{Anzahl günstigen Fälle}}{\text{Anzahl möglichen Fälle}} = \frac{1}{8}$$

Mit Hilfe der Formel der Kombinatorik kann die Anzahl der möglichen Fälle ebenfalls rasch berechnet werden. Da aus einer 2-elementigen Menge drei Elemente ausgewählt werden gilt:

$$n = 2 \quad k = 3 \quad \text{und somit} \quad 2^3 = 8$$

Bei einer Variation gibt es auch Fälle, bei welchen die Wiederholung nicht möglich

2 Wahrscheinlichkeitsrechnung

oder nicht wünschenswert ist. Sei nun $M = \{1, \dots, n\}$ eine n -elementige Menge und $1 \leq k \leq n$, so kann die Anzahl der geordneten Stichproben vom Umfang k ohne Wiederholung folgendermaßen berechnet werden [17, S.21]:

$$V(n, k) = n \cdot (n - 1) \cdots (n - k + 1) = \frac{n!}{(n - k)!} = k! \cdot \binom{n}{k}$$

Beispiel: Mit Hilfe des Urnenmodells kann diese Variation als Ziehen einer geordneten Stichprobe ohne Zurücklegen veranschaulicht werden. Wir gehen von einem ähnlichen Beispiel wie zuvor aus, wobei sich diesmal in einer Urne eine grüne, eine gelbe und eine rote Kugel befinden. Nun wird in die Urne gegriffen und eine Kugel gezogen. Nachdem die Farbe der Kugel notiert wurde, wird diese jedoch nicht mehr in die Urne zurückgelegt. Berechne die Wahrscheinlichkeit, dass bei zweimaligem Ziehen die Farbreihenfolge Rot-Gelb gezogen wird.

Das nachstehende Baumdiagramm 2.2 veranschaulicht die möglichen Ziehungsverläufe.

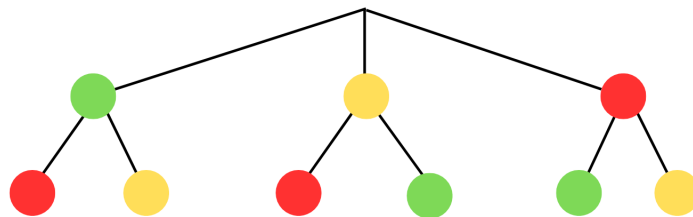


Abbildung 2.2: Baumdiagramm - Ziehen einer geordneten Stichprobe ohne Zurücklegen

Aus diesem Diagramm ist erschießbar, dass es insgesamt 6 mögliche Ereignisse gibt, jedoch nur einen günstigen Fall der zur gewünschten Farbreihenfolge Rot-Gelb führt. Da es sich hierbei um ein Laplace-Experiment handelt, wird die Wahrscheinlichkeit folgendermaßen berechnet:

$$P(\text{Rot-Gelb}) = \frac{\text{Anzahl günstigen Fälle}}{\text{Anzahl möglichen Fälle}} = \frac{1}{6}$$

Anhand dieses Beispiels ist ersichtlich, dass das Aufzeichnen aller Möglichkeiten schnell sehr zeitaufwendig werden kann. Mit Hilfe der Formel der Kombinatorik kann daher die Anzahl der möglichen Fälle rascher berechnet werden. Da aus einer 3-elementigen Menge 2 Elemente ausgewählt werden gilt:

$$n = 3 \quad k = 2 \quad \text{und somit} \quad 3! \cdot \binom{3}{2} = \frac{3!}{(3-2)!} = 3 \cdot 2 \cdot 1 = 6$$

Handelt es sich um eine Variation, bei der alle n verschiedenen Elemente der n -elementigen Menge $M = \{1, 2, \dots, n\}$ angeordnet werden, so wird dies als *Permutation* bezeichnet. Die Anzahl der Anordnungen kann folgendermaßen berechnet werden [17, S.11]:

$$P(n) = n! = 1 \cdot 2 \cdot \dots \cdot n$$

Mit Hilfe des Urnenmodells entspricht eine Permutation dem Ziehen einer geordneten Stichprobe vom Umfang n aus einer n -elementigen Menge ohne Zurücklegen.

Beispiel: Wir gehen von einem ähnlichen Beispiel wie zuvor aus, wobei sich diesmal in einer Urne eine grüne, eine gelbe und eine rote Kugel befinden. Nun wird in die Urne gegriffen und eine Kugel gezogen. Nachdem die Farbe der Kugel notiert wurde, wird diese jedoch nicht mehr in die Urne zurückgelegt. Berechne die Wahrscheinlichkeit, dass bei dreimaligem Ziehen die Farbreihenfolge Rot-Gelb-Grün gezogen wird.

Das nachstehende Baumdiagramm 2.3 veranschaulicht die möglichen Ziehungsverläufe.

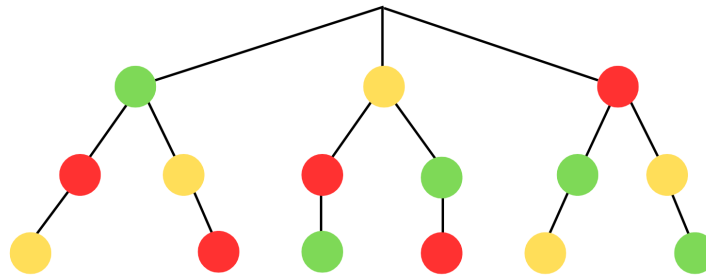


Abbildung 2.3: Baumdiagramm - Ziehen einer geordneten Stichprobe ohne Zurücklegen mit $n = k$

Anhand des Baumdiagramms kann erkannt werden, dass es insgesamt 6 mögliche Ereignisse gibt, jedoch nur einen günstigen Fall der zur gewünschten Farbreihenfolge Rot-Gelb-Grün führt. Aufgrund der Tatsache, dass es sich hierbei um ein Laplace-Experiment handelt, wird die Wahrscheinlichkeit folgendermaßen berechnet:

$$P(\text{Rot-Gelb-Grün}) = \frac{\text{Anzahl günstigen Fälle}}{\text{Anzahl möglichen Fälle}} = \frac{1}{6}$$

Es ist ersichtlich, dass das Aufzeichnen aller Möglichkeiten schnell sehr zeitaufwendig werden kann. Daher kann mit Hilfe der Formel der Kombinatorik die Anzahl der möglichen Fälle schneller ermittelt werden. Da aus einer 3-elementigen Menge 3 Elemente ausgewählt werden, handelt es sich um eine Permutation und es gilt:

$$n! = 3 \cdot 2 \cdot 1 = 6$$

Dies kann mit Hilfe folgender Vorstellung veranschaulicht werden. Beim ersten Zug gibt es noch drei mögliche Kugeln, da diese gezogene jedoch nicht mehr in die Urne zurückgelegt wird, gibt es bei der zweiten Ziehung nur noch zwei mögliche Kugeln. Beim letzten Zug befindet sich nur noch eine Kugel in der Urne.

Eine weitere Variante der Anordnung ist jener Fall, bei welchem alle n Elemente der n -elementigen Menge $M = \{1, \dots, n\}$ angeordnet werden sollen, wobei nicht

alle n Elemente verschieden sind. Von den n Elementen sind jeweils $n_1, n_2, \dots, n_r$ gleich und es gilt somit $n_1 + n_2 + \dots + n_r = n$. Die Anzahl der verschiedenen Anordnungsmöglichkeiten lässt sich folgendermaßen ermitteln [2, S.17]:

$$\frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_r!}$$

2.3.2 Ungeordnete Stichprobe

Im Gegensatz zur Variation, wo die Reihenfolge von Bedeutung ist, spielt diese bei sogenannten *Kombinationen* keine Rolle. Auch bei einer ungeordneten Stichprobe wird unterschieden ob Wiederholungen erlaubt sind oder nicht [17, S.17].

Es sei $M = \{1, \dots, n\}$ eine n -elementige Menge und $0 \leq k \leq n$, so kann die Anzahl der ungeordneten Stichproben vom Umfang k ohne Zurücklegen folgendermaßen berechnet werden [17, S.17]:

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

Mit Hilfe des Urnenmodells kann diese Kombination als ein Ziehen einer ungeordneten Stichprobe vom Umfangs k ohne Zurücklegen veranschaulicht werden. Wichtig ist hierbei zu beachten, dass zwei Ausgänge (a, b) und (b, a) , welche sich nur durch ihre Reihenfolge unterscheiden, als gleich erachtet werden, da die Reihenfolge keine Rolle spielt.

Beispiel: In einer Urne befinden sich eine grüne, eine rote und eine gelbe Kugel. Es wird zwei Mal in die Urne gegriffen und jeweils die Farbe der gezogenen Kugel notiert, wobei diese nicht mehr zurück in die Urne gelegt wird. Berechne die Wahrscheinlichkeit, dass eine grüne und eine gelbe Kugel gezogen werden.

Im nachstehenden Baumdiagramm 2.4 wurden alle möglichen Ziehungsverläufe dargestellt.

2 Wahrscheinlichkeitsrechnung

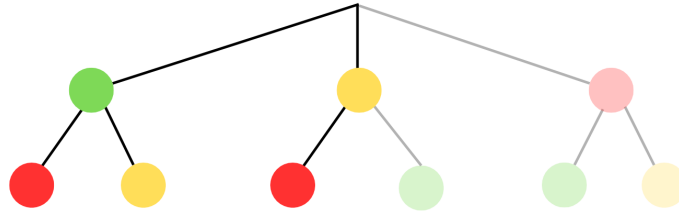


Abbildung 2.4: Baumdiagramm - Ziehen einer ungeordneten Stichprobe ohne Zurücklegen

Da es sich um das Ziehen einer ungeordneten Stichprobe handelt, sind folgende Ereignisse als äquivalent zu betrachten:

- $\{\text{Grün, Rot}\} = \{\text{Rot, Grün}\}$
- $\{\text{Rot, Gelb}\} = \{\text{Gelb, Rot}\}$
- $\{\text{Gelb, Grün}\} = \{\text{Grün, Gelb}\}$

Somit gibt es insgesamt drei mögliche Ereignisse, jedoch nur eines, welches zu der gewünschten Farbkombination führt. Mit Hilfe der Pfadregel lässt sich die Wahrscheinlichkeit für eine grüne und eine gelbe Kugel folgendermaßen ermitteln:

$$P(1 \times \text{Grün}, 1 \times \text{Gelb}) = \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

Mit Hilfe der Formel der Kombinatorik für Kombinationen ohne Wiederholungen kann die Anzahl der möglichen Fälle ebenfalls rasch ermittelt werden. Da aus einer 3-elementigen Menge zwei Elemente kombiniert werden, gilt:

$$n = 3 \quad k = 2 \quad \text{und somit} \quad \binom{3}{2} = \frac{3!}{2! \cdot (3-2)!} = \frac{3 \cdot 2 \cdot 1}{(2 \cdot 1) \cdot 1} = 3$$

Handelt es sich um jenen Fall, bei welchem die Wiederholung wünschenswert ist, so sei nun $M = \{1, \dots, n\}$ eine n -elementige Menge. Die Anzahl der ungeordneten

Stichproben vom Umfang k mit Wiederholung kann folgendermaßen berechnet werden [17, S.20]:

$$C(n, k) = \binom{n + k - 1}{k}$$

Beispiel: In einer Urne befinden sich eine grüne und eine gelbe Kugel. Es wird zwei Mal hintereinander gezogen, wobei jeweils die Farbe der Kugel notiert wird und diese anschließend wieder zurück in die Urne gelegt wird. Berechne die Wahrscheinlichkeit, dass zwei grüne Kugeln gezogen werden.

Das folgende Baumdiagramm 2.5 veranschaulicht alle möglichen Ziehungsverläufe.

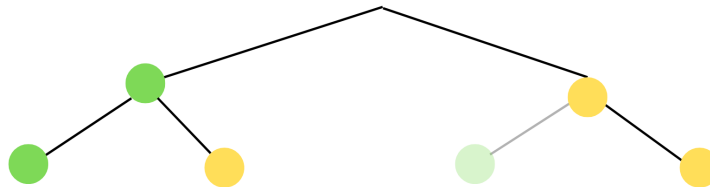


Abbildung 2.5: Baumdiagramm - Ziehen einer ungeordneten Stichprobe mit Zurücklegen

Aufgrund der Tatsache, dass die Reihenfolge hier keine Rolle spielt, sind die Fälle {Gelb, Grün} und {Grün, Gelb} als gleich zu betrachten. Somit gibt es drei mögliche Ereignisse, allerdings nur eines welches zu der gewünschten Farbkombination führt. Mit Hilfe der Formel der Kombination können die möglichen Fälle ebenfalls berechnet werden. Da aus 2-elementigen Menge zwei Elemente gezogen werden, gilt:

$$n = 2 \quad k = 2 \quad \text{und somit} \quad \binom{2 + 2 - 1}{2} = \binom{3}{2} = 3$$

Diese drei möglichen Ereignisse treten jedoch nicht mit gleicher Wahrscheinlichkeit auf, weshalb der womöglich intuitive Gedankengang, dass es sich somit um Wahrscheinlichkeit $p = \frac{1}{3}$ handelt, nicht zutrifft. Mit Hilfe der Pfadregeln, lässt

sich die Wahrscheinlichkeit für zwei grüne Kugeln folgendermaßen ermitteln:

$$P(2 \times \text{Grün}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

2.4 Diskrete Zufallsvariable

Anhand der Beispiele zum Urnenmodell konnte gezeigt werden, dass im Gegensatz zum Werfen eines Würfels, es auch Zufallsexperimente gibt, bei denen die Versuchsergebnisse nicht unmittelbar Zahlen sein müssen.

Allgemein sei Ω die Menge aller möglichen Ereignisse des Zufallsexperiments, so kann die Zufallsvariable X als eine Vorschrift veranschaulicht werden, durch welche jedem Ergebnis $\omega \in \Omega$ eine reelle Zahl $X(\omega)$ zugeordnet wird. Das bedeutet, dass auch für Versuchsergebnisse, welche keine Zahlen sind, nach jeder Durchführung des Zufallsexperimentes dem Ergebnis ω ein Zahlenwert $X(\omega)$ zugeordnet wird. Da das Eintreten eines Ergebnisses ω vom Zufall abhängt, gilt dies ebenfalls für die Zufallsvariable X . Mathematisch gesehen, ergibt sich daraus [2, S.55-56]:

Definition 8. Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Eine Abbildung

$$X : \Omega \rightarrow \mathbb{R}$$

wird als *Zufallsvariable* bezeichnet, wenn für jede Borelmenge $B \subseteq \mathbb{R}$ das vollständige Urbild $X^{-1}(B)$ Element der σ -Algebra $\mathcal{A}$ ist. Es gilt also [18, S.98]:

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A} \quad \text{für alle } B \in \mathcal{B}(\mathbb{R})$$

Die Menge aller möglichen Werte, welche eine Zufallsvariable X annehmen kann wird Wertevorrat W bezeichnet. Das bedeutet, damit eine Zahl x zum Wertevorrat W gehört, muss es mindestens ein Versuchsergebnis $\omega \in \Omega$ geben, für das $X(\omega) = x$ gilt. Zufallsvariablen werden stets mit lateinischen Großbuchstaben X , Y , Z etc. bezeichnet [2, S.56].

Seien X und Y zwei Zufallsvariablen auf demselben Grundraum Ω , und deren Summe, Differenz und Produkt folgendermaßen definiert:

$$\begin{aligned}(X + Y)(\omega) &:= X(\omega) + Y(\omega), & \omega \in \Omega \\(X - Y)(\omega) &:= X(\omega) - Y(\omega), & \omega \in \Omega \\(X \cdot Y)(\omega) &:= X(\omega) \cdot Y(\omega), & \omega \in \Omega\end{aligned}$$

so sind auch die Summe $X + Y$, Differenz $X - Y$ und das Produkt XY eine Zufallsvariable auf Ω . Sei nun $a \in \mathbb{R}$, so ist auch das a -Fache aX

$$(a \cdot X)(\omega) := a \cdot X(\omega), \quad \omega \in \Omega$$

eine Zufallsvariable auf Ω [13, S.11].

2.4.1 Definition einer diskreten Zufallsvariable

Definition 9. Eine Zufallsvariable X , bei welcher die möglichen Werte in einer endlich oder abzählbar unendlichen Menge W liegen, wird als *diskret* bezeichnet. Die Gesamtheit aller Zahlenpaare $(x_i, P(X = x_i))$ mit $x_i \in W$ wird *Verteilung* der diskreten Zufallsvariable X genannt [2, S.58].

Für eine diskrete Zufallsvariable X aus einem Intervall $I = [a, b]$ gilt [2, S.58]:

$$P(a \leq X \leq b) = \sum_{a \leq x_i \leq b} P(X = x_i)$$

Beispiel: Bei einem Gewinnspiel werden zwei gleiche Würfel geworfen. Die Zufallsvariable X bezeichnet die erhaltene Augensumme. Der Hauptgewinn wird erzielt, wenn die Augensumme 10, 11 oder 12 gewürfelt wird. In der nachstehenden Tabelle 2.1 werden alle Möglichkeiten für die jeweilige Augensumme dargestellt.

2 Wahrscheinlichkeitsrechnung

Augensumme	Würfelpaare
1	/
2	(1,1)
3	(1,2) ; (2,1)
4	(1,3) ; (3,1) ; (2,2)
5	(1,4) ; (4,1) ; (2,3) ; (3,2)
6	(1,5) ; (5,1) ; (2,4) ; (4,2) ; (3,3)
7	(1,6) ; (6,1) ; (5,2) ; (2,5) ; (4,3) ; (3,4)
8	(2,6) ; (6,2) ; (3,5) ; (5,3) (4,4)
9	(4,5) ; (5,4) ; (6,3) ; (3,6)
10	(4,6) ; (6,4) ; (5,5)
11	(5,6) ; (6,5)
12	(6,6)

Tabelle 2.1: Augensumme für zwei Würfel

Da es sich um ein Laplace's Zufallsexperiment handelt, gilt:

$$P(X = 10) = \frac{3}{36} \quad P(X = 11) = \frac{2}{36} \quad P(X = 12) = \frac{1}{36}$$

Und somit gilt für die Wahrscheinlichkeit des Hauptgewinns:

$$\begin{aligned} P(10 \leq X \leq 12) &= P(X = 10) + P(X = 11) + P(X = 12) \\ &= \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{6}{36} = \frac{1}{6} \end{aligned}$$

Die Werte der Zufallsvariable X können mit Hilfe der Tabelle 2.2 oder des Diagramms 2.6 dargestellt werden.

X	1	2	3	4	5	6	7	8	9	10	11	12
Wahrscheinlichkeit $P(X)$	0	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Tabelle 2.2: Werte der Zufallsvariable X

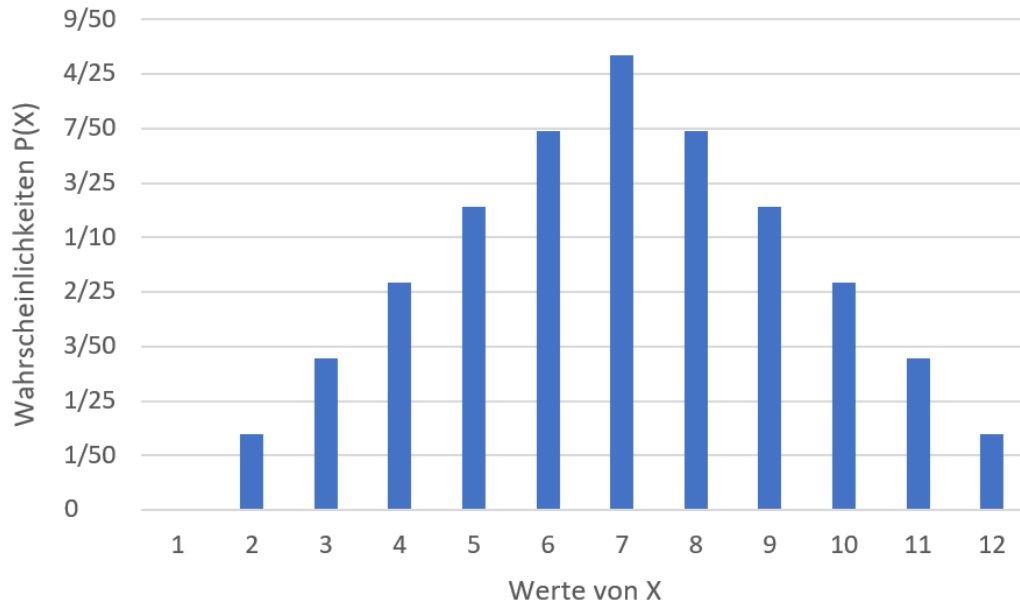


Abbildung 2.6: Wahrscheinlichkeitsfunktion

Weiters gilt:

$$P(X = 1) + P(X = 2) + \dots + P(X = 12) = 1$$

2.4.2 Verteilungsfunktion

Oft ist die Wahrscheinlichkeit dafür, dass eine Zufallsvariable nicht größer als ein fixer Wert x ist, von Interesse [2, S.58].

2 Wahrscheinlichkeitsrechnung

Definition 10. Ist X eine Zufallsvariable, so wird die reellwertige Funktion F

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$$

Verteilungsfunktion von X genannt.

Zu jeder diskreten Zufallsvariable X gibt es eine Verteilungsfunktion F , welche stets die Form einer Treppenfunktion hat. Die Sprünge der Treppen befinden sich an den Stellen x_i und haben eine Höhe von $P(X = x_i)$ [2, S.59]. Somit besitzt die Verteilungsfunktion F folgende Eigenschaften [13, S.293]:

- F ist monoton wachsend, d.h. für $a \leq b$ gilt stets $F(a) \leq F(b)$
- F ist rechtsseitig stetig
- $\lim_{x \rightarrow +\infty} F(x) = 1$
- $\lim_{x \rightarrow -\infty} F(x) = 0$

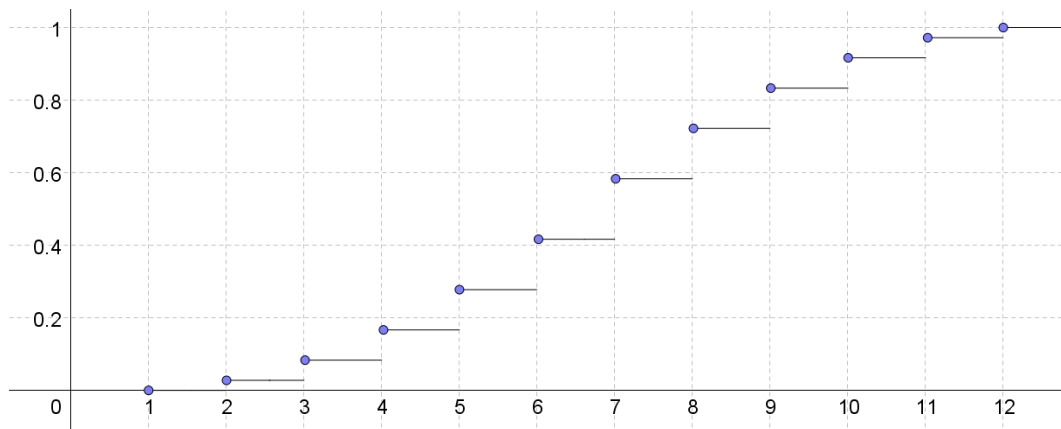
Satz 1. Sei X eine Zufallsvariable und F die Verteilungsfunktion dieser, so gelten nachstehende Gleichungen [2, S.60]:

1. $P(a < X \leq b) = F(b) - F(a)$
2. $P(a \leq X \leq b) = F(b) - \lim_{x \rightarrow a^-} F(x)$
3. $P(a < X) = 1 - F(a)$

Beispiel: Es werden erneut zwei gleiche Würfel geworfen und deren Augenzahl betrachtet. Die Zufallsvariable X bezeichnet die erhaltene Augensumme der beiden Würfel. Anhand der Wahrscheinlichkeitsverteilung, welche in Tabelle 2.2 bereits angeführt wurde, kann folgendes erkannt werden:

$$F(X) = \begin{cases} 0 & \text{falls } x < 2 \\ \frac{1}{36} & \text{falls } 2 \leq x < 3 \\ \frac{3}{36} & \text{falls } 3 \leq x < 4 \\ \frac{6}{36} & \text{falls } 4 \leq x < 5 \\ \frac{10}{36} & \text{falls } 5 \leq x < 6 \\ \frac{15}{36} & \text{falls } 6 \leq x < 7 \\ \frac{21}{36} & \text{falls } 7 \leq x < 8 \\ \frac{26}{36} & \text{falls } 8 \leq x < 9 \\ \frac{30}{36} & \text{falls } 9 \leq x < 10 \\ \frac{33}{36} & \text{falls } 10 \leq x < 11 \\ \frac{35}{36} & \text{falls } 11 \leq x < 12 \\ \frac{36}{36} = 1 & \text{falls } 12 \leq x \end{cases}$$

Die Verteilungsfunktion $F(X)$ ist untenstehend abgebildet. Anhand der Abbildung 2.7 ist die charakteristische Treppenform der Funktion ersichtlich. Die Sprünge befinden sich an den Stellen $i = 1, \dots, 12$ und haben jeweils als Höhe die zugehörige Wahrscheinlichkeit $P(X = i)$. Weiters ist zu erkennen, dass die Funktion stets monoton wachsend und zwischen den Stufen konstant ist.

Abbildung 2.7: Verteilungsfunktion $F(X)$

2.4.3 Erwartungswert

Definition 11. Sei X eine Zufallsvariable mit einem endlichen Wertevorrat $W = \{x_1, \dots, x_n\}$, so wird

$$\mu = E(X) = \sum_i x_i \cdot P(X = x_i)$$

der *Erwartungswert* von X genannt [2, S.63].

Definition 12. Sei X eine diskrete Zufallsvariable mit Werten in $\{x_1, x_2, \dots\}$ mit $x_i \geq 0$. Sei weiters $\sum_i |x_i| P(X = x_i)$ endlich. Der Erwartungswert einer diskreten Zufallsvariable ist dann definiert durch [18, S.182]:

$$E(X) = \sum_i x_i \cdot P(X = x_i)$$

Der Erwartungswert einer diskreten Zufallsvariable X kann als ein Durchschnittswert aller möglichen Werte von X verstanden werden. Bei der Ermittlung des Erwartungswertes wird jeder Wert x_i mit seiner jeweiligen Wahrscheinlichkeit $P(X = x_i)$ gewichtet.

Wird ein Zufallsexperiment n -mal durchgeführt, so wird mit großer Wahrscheinlichkeit der Mittelwert $\bar{x}$ für große n annähernd den Wert des Erwartungswertes $E(X)$ haben [2, S.62].

Sind X und Y zwei Zufallsvariablen auf Ω und $a \in \mathbb{R}$, so gelten folgende Eigenschaften [13, S.77]:

- $E(X + Y) = E(X) + E(Y)$
- $E(a \cdot X) = a \cdot E(X)$
- aus $X \leq Y$ folgt $E(X) \leq E(Y)$

Beispiel: Bei einer Tombola mit dem Einsatz von 3€ gibt es insgesamt 100 Lose die an der Ziehung teilnehmen. Der Hauptgewinn beträgt 100€, für drei weitere Lose werden je 10€ ausgezahlt und für 10 weitere jeweils 5€. Bei den restlichen Losen handelt es sich um Nieten. Die Zufallsvariable G gibt den tatsächlichen Gewinn in Euro an. Bevor der Erwartungswert $E(X)$ berechnet werden kann, ist es von Vorteil die Wahrscheinlichkeitsverteilung von G anzugeben.

g in Euro	0	5	10	100
Wahrscheinlichkeit $P(G=g)$	$\frac{86}{100}$	$\frac{1}{10}$	$\frac{3}{100}$	$\frac{1}{100}$

Tabelle 2.3: Wahrscheinlichkeitsverteilung

Für den Erwartungswert $E(X)$ gilt nun:

$$\mu = E(X) = 0 \cdot \frac{86}{100} + 5 \cdot \frac{1}{10} + 10 \cdot \frac{3}{100} + 100 \cdot \frac{1}{100} = 1,8$$

Somit beträgt der Erwartungswert des Gewinns bei der Tombola 1,80€ und es kann ein durchschnittlicher Gewinn von 1,80€ erwartet werden.

Definition 13. Seien X und Y zwei diskrete Zufallsvariablen, so werden diese stochastisch *unabhängig* genannt, wenn für alle Wertepaare (x_i, y_i)

$$P(X = x_i, Y = y_i) = P(X = x_i) \cdot P(Y = y_i)$$

gilt [2, S.74].

Für zwei diskrete Zufallsvariablen X und Y , welche stochastisch unabhängig sind, gilt für deren Erwartungswerte [2, S.77]:

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

2.4.4 Varianz und Standardabweichung

Ein beträchtlicher Nachteil des Erwartungswertes ist es, dass dieser sehr leicht durch Ausreißer beeinflusst wird, und somit unter Umständen zu fälschlichen Folgerungen führt. Aus diesem Grund wird meist ein Streuungsmaß angegeben, welches die zu erwartende Abweichung $X - \mu$ vom Erwartungswert $E(X) = \mu$ angibt.

Definition 14. Für eine diskrete Zufallsvariable X mit dem Erwartungswert $E(X) = \mu$, wird im Falle der Existenz der Zahlenwert

$$V(X) = \sigma^2 = E([X - \mu]^2) = \sum_i (x_i - \mu)^2 \cdot P(X = x_i)$$

die *Varianz* von X und die positive Quadratwurzel

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

Standardabweichung von X bezeichnet [2, S.70].

Alternativ kann die Varianz $V(X)$ einer diskreten Zufallsvariable auch folgendermaßen berechnet werden [2, S.70]:

$$V(X) = \sigma^2 = \sum_i x_i^2 \cdot P(X = x_i) - \mu^2 = E(X^2) - \mu^2$$

Dies ist der Fall, da aus

$$\begin{aligned} \sigma^2 &= E([X - \mu]^2) = \sum_i (x_i - \mu)^2 \cdot P(X = x_i) \\ &= \sum_i (x_i^2 - 2\mu x_i + \mu^2) \cdot P(X = x_i) \end{aligned}$$

durch Ausmultiplizieren

$$= \sum_i x_i^2 \cdot P(X = x_i) - 2\mu \sum_i x_i \cdot P(X = x_i) + \mu^2 \sum_i P(X = x_i)$$

und aufgrund der Definition des Erwartungswertes und der Tatsache, dass $\sum_i P(X = x_i) = 1$ gilt, folgt

$$\begin{aligned} &= \sum_i x_i^2 \cdot P(X = x_i) - 2\mu \cdot \mu + \mu^2 \cdot 1 \\ &= \sum_i x_i^2 \cdot P(X = x_i) - \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$

Beispiel: In einer Schulklasse mit 25 Kindern gab es bei der Mathematikshularbeit folgende Notenverteilung:

1. Sehr Gut (1): 3
2. Gut (2): 8
3. Befriedigend (3): 7
4. Genügend (4): 5
5. Nicht Genügend (5): 2

Der entsprechende Wahrscheinlichkeitsraum ist $(\Omega, \mathcal{A}, P)$ mit

$$\Omega = \{1, 2, 3, 4, 5\}, \quad \mathcal{A} = \mathcal{P}(\Omega)$$

und dem Wahrscheinlichkeitsmaß P in Tabelle 2.4. Die Zufallsvariable N gibt die Note in numerischer Darstellung an. Aus der obigen Notenverteilung ergibt sich folgende Wahrscheinlichkeitsverteilung:

2 Wahrscheinlichkeitsrechnung

Note	1	2	3	4	5
Wahrscheinlichkeit $P(N = n)$	$\frac{3}{25}$	$\frac{8}{25}$	$\frac{7}{25}$	$\frac{1}{5}$	$\frac{2}{25}$

Tabelle 2.4: Wahrscheinlichkeitsverteilung der Noten

Anhand dieser Wahrscheinlichkeitstabelle lassen sich der Erwartungswert $E(N)$, die Varianz $V(N)$ sowie die Standardabweichung σ schnell ermitteln.

$$E(N) = 1 \cdot \frac{3}{25} + 2 \cdot \frac{8}{25} + 3 \cdot \frac{7}{25} + 4 \cdot \frac{1}{5} + 5 \cdot \frac{2}{25} = \frac{14}{5} = 2,8$$

$$V(N) = \left(1 - \frac{14}{5}\right)^2 \cdot \frac{3}{25} + \left(2 - \frac{14}{5}\right)^2 \cdot \frac{8}{25} + \dots + \left(5 - \frac{14}{5}\right)^2 \cdot \frac{2}{25} = \frac{32}{25} = 1,28$$

$$\sigma = \sqrt{\frac{32}{25}} \approx 1,13$$

Seien X und Y zwei diskrete Zufallsvariablen die stochastisch unabhängig sind und deren Varianzen existieren, so gilt [2, S.79]:

$$V(X + Y) = V(X) + V(Y)$$

2.4.5 Binomialverteilung

Bei vielen Experimenten wird ein Zufallsversuch n -mal unter gleichbleibenden Bedingungen durchgeführt, wobei sich diese nicht gegenseitig beeinflussen.

Definition 15. Bei einem *Bernoulli-Versuch* handelt es sich es sich um einen Ver-

such welcher n -malig unter gleichbleibenden Bedingungen wiederholt wird und bei welchem genau zwei Versuchsausgänge möglich sind, nämlich der Treffer und die Niete [13, S.140].

Da es bei einem Bernoulli-Experiment nur zwei Ausgänge gibt, werden diese stets mit 1 und 0 codiert, wobei 1 für den gewünschten Fall, also den Treffer, und 0 für die Niete steht. Ein Bernoulli-Experiment besitzt somit einen Grundraum $\Omega = \{0, 1\}$, für welchen gilt [13, S.141]:

$$P(1) = p \quad P(0) = 1 - p$$

Definition 16. Sei $n \in \mathbb{N}$ und $p \in [0, 1]$. Eine Zufallsvariable X heißt *binomialverteilt* mit den Parametern n und p , falls

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k} \quad \text{mit} \quad q = 1 - p \quad k = 0, 1, \dots, n$$

Kurz schreibt man, dass die Zufallsvariable $B(n, p)$ -verteilt ist [2, S.86].

Satz 2. Sei A ein Ereignis welches mit der Wahrscheinlichkeit p eintritt. Eine diskrete Zufallsvariable X , die die Anzahl angibt wie oft ein Ereignis A bei n Versuchen auftritt, ist binomialverteilt [2, S.86].

Beweis. Die möglichen Werte für X , welche die Anzahl angeben wie oft A bei n Versuchen auftritt, sind $0, 1, \dots, n$. Für die Berechnung von $P(X = k)$ mit $k = 0, 1, \dots, n$ ist es hilfreich, zuerst alle Folgen mit n Elementen aufzuschreiben, welche genau k mal das Ereignis A und somit $n - k$ mal ein anderes Ereignis enthalten, um alle Teilereignisse darzustellen. Da das Ereignis A mit der Wahrscheinlichkeit p somit genau k mal eintritt, und das Gegenereignis mit der Wahrscheinlichkeit $1 - p$ genau $n - k$ mal eintritt, gilt somit für die Wahrscheinlichkeit

2 Wahrscheinlichkeitsrechnung

nach dem Multiplikationssatz:

$$p^k \cdot (1 - p)^{n-k}$$

Mit Hilfe des Binomialkoeffizienten $\binom{n}{k}$ kann die Anzahl der Teilereignisse ermittelt werden. Nach dem Additionssatz gilt somit für die Wahrscheinlichkeit

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

Die Zufallsvariable X ist somit binomialverteilt. □

Satz 3. Der Erwartungswert $E(X)$, die Varianz $V(X)$ und die Standardabweichung σ einer binomialverteilten Zufallsvariable X können folgendermaßen berechnet werden [2, S.87]:

$$E(X) = \mu = n \cdot p \qquad V(X) = \sigma^2 = n \cdot p \cdot q = n \cdot p \cdot (1 - p) \qquad \sigma = \sqrt{n \cdot p \cdot q}$$

Beweis. Der nachfolgende Beweis für den Erwartungswert ist angelehnt an die Beweisführung im Werk von Henze [13, S.380]. Sei X eine binomialverteilte Zufallsvariable X , so ist $\{0, 1, \dots, n\}$ der Wertebereich für X , sowie $P(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$. Nach Einsetzen in die Formel für den Erwartungswert für diskrete Zufallsvariablen erhält man folgenden Ausdruck:

$$\begin{aligned} E(X) &= \sum_{k=0}^n k \cdot \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k} \\ &= \sum_{k=1}^n k \cdot \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k} \\ &= \sum_{k=1}^n n \cdot \binom{n-1}{k-1} \cdot p^k \cdot (1 - p)^{n-k} \end{aligned}$$

Durch Herausheben von $n \cdot p$ folgt:

$$E(X) = n \cdot p \cdot \sum_{k=1}^n \binom{n-1}{k-1} \cdot p^{k-1} \cdot (1 - p)^{(n-1)-(k-1)}$$

Nach einer Indexverschiebung mit $j = k - 1$ gilt:

$$E(X) = n \cdot p \cdot \sum_{j=0}^{n-1} \binom{n-1}{j} \cdot p^j \cdot (1-p)^{n-1-j}$$

Aus dem binomischen Lehrsatz folgt schließlich:

$$\begin{aligned} E(X) &= n \cdot p \cdot (p + 1 - p)^{n-1} \\ &= n \cdot p \end{aligned}$$

Mit der gleichen Vorgehensweise kann auch die Formel für die Varianz sowie der Standardabweichung bewiesen werden. Dabei gilt $\mu = E(X) = n \cdot p$ sowie

$$P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

$$\begin{aligned} V(X) &= \sum_{k=0}^n (k - \mu)^2 \cdot P(X = k) \\ &= \sum_{k=0}^n k^2 \cdot P(X = k) - 2\mu \cdot \sum_{k=0}^n k \cdot P(X = k) + \mu^2 \cdot \sum_{k=0}^n P(X = k) \end{aligned}$$

Aus $\sum_{k=0}^n k \cdot P(X = k) = \mu$ und $\sum_{k=0}^n P(X = k) = 1$ folgt

$$V(X) = \sum_{k=1}^n k^2 \cdot P(X = k) - 2\mu^2 + \mu^2 = \sum_{k=1}^n k^2 \cdot P(X = k) - \mu^2$$

Durch Einsetzen von $P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$ sowie $\mu = np$ folgt:

$$V(X) = \sum_{k=1}^n k^2 \cdot \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} - (n \cdot p)^2$$

2 Wahrscheinlichkeitsrechnung

Da für Binomialkoeffizienten gilt, dass $\binom{n}{k} = \frac{n}{k} \cdot \binom{n-1}{k-1}$ folgt:

$$\begin{aligned} V(X) &= \sum_{k=1}^n k \cdot n \cdot \binom{n-1}{k-1} \cdot p^k \cdot (1-p)^{n-k} - n^2 \cdot p^2 \\ &= n \cdot p \cdot \left(\sum_{k=1}^n k \cdot \binom{n-1}{k-1} \cdot p^{k-1} \cdot (1-p)^{(n-1)-(k-1)} - n \cdot p \right) \end{aligned}$$

Durch eine Indexverschiebung mit $i = k - 1$ erhält man:

$$\begin{aligned} V(X) &= np \left(\sum_{i=0}^{n-1} (i+1) \cdot \binom{n-1}{i} \cdot p^i \cdot (1-p)^{n-1-i} - np \right) \\ &= np \left(\sum_{i=0}^{n-1} i \cdot \binom{n-1}{i} \cdot p^i \cdot (1-p)^{n-1-i} + \sum_{i=0}^{n-1} \binom{n-1}{i} p^i \cdot (1-p)^{n-1-i} - np \right) \end{aligned}$$

Aufgrund der Formel für den Erwartungswert sowie dem binomischen Lehrsatz folgt:

$$\begin{aligned} V(X) &= n \cdot p \cdot \left((n-1) \cdot p + (p+1-p)^{n-1} - n \cdot p \right) \\ &= n \cdot p \cdot (n \cdot p - p + 1 - n \cdot p) \\ &= n \cdot p \cdot (1-p) \end{aligned}$$

Folglich gilt für die Standardabweichung σ , dass

$$\sigma = \sqrt{n \cdot p \cdot (1-p)}$$

□

Beispiel: Einer Fluglinie ist aus langjähriger Erfahrung bekannt, dass 5% aller gebuchten Flugtickets nicht in Anspruch genommen werden. Aus diesem Grund verkauft die Fluglinie für einen Flug mit 120 Sitzplätzen insgesamt 125 Tickets, um möglichst alle Plätze zu belegen. Die Zufallsvariable X gibt die Anzahl jener Tickets an, die tatsächlich in Anspruch genommen werden.

1. Mit welcher Wahrscheinlichkeit bleibt genau ein Sitzplatz frei?

Ob ein Passagier sein Flugticket in Anspruch nimmt, ist ein Bernoulli-Experiment, bei welchem der Erfolg mit der Wahrscheinlichkeit $p = 95\%$ eintritt. Somit ist X binomialverteilt mit $n = 125$ und $p = 0,95$

$$P(X = 119) = \binom{125}{119} \cdot 0,95^{119} \cdot (1 - 0,95)^{125-119} \approx 0,1637 \hat{=} 16,37\%$$

2. Berechne mit welcher Wahrscheinlichkeit höchstens so viele Tickets in Anspruch genommen werden, wie es Sitzplätze gibt.

$$\begin{aligned} P(X \leq 120) &= 1 - P(X > 120) \\ &= 1 - (P(X = 121) + P(X = 122) + \dots + P(X = 125)) \\ &\approx 1 - (0,1221 + 0,0761 + 0,0353 + 0,0108 + 0,0016) \\ &\approx 0,7541 \hat{=} 75,41\% \end{aligned}$$

3. Wie groß ist die Wahrscheinlichkeit, dass mindestens ein Passagier aufgrund der Überbuchung am Boden bleibt und nicht mitfliegen kann?

$$\begin{aligned} P(X \geq 121) &= P(X = 121) + P(X = 122) + \dots + P(X = 125) \\ &\approx 0,1221 + 0,0761 + 0,0353 + 0,0108 + 0,0016 \\ &\approx 0,2459 \hat{=} 24,59\% \end{aligned}$$

4. Berechne den Erwartungswert $E(X)$ sowie die Standardabweichung σ von X .

$$\begin{aligned} E(X) &= 125 \cdot 0,95 = 118,75 \\ \sigma &= \sqrt{125 \cdot 0,95 \cdot 0,05} \approx 2,4367 \end{aligned}$$

2.4.6 Poisson Verteilung

Ein weiteres Beispiel einer diskreten Zufallsvariable ist die Poisson-Verteilung. Diese entsteht durch Approximation einer Binomialverteilung mit besonders großem

2 Wahrscheinlichkeitsrechnung

n , aber kleinem p [13, S.195].

Definition 17. Eine Zufallsvariable X besitzt eine *Poisson-Verteilung* mit $\lambda > 0$, falls [13, S.195]:

$$P(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \quad \text{für } k = 0, 1, \dots$$

Satz 4. Sei X_n eine Folge von $B(n, p)$ -verteilten Zufallsvariablen und ist $n \cdot p = \lambda$ konstant, so gilt [2, S.93]:

$$\lim_{n \rightarrow \infty} P(X_n = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \quad k = 0, 1, \dots$$

Beweis. Der folgende Beweis stammt aus dem Werk von Bosch [2, S.93]. Aus $n \cdot p = \lambda$ folgt $p = \frac{\lambda}{n}$. Somit folgt für ein konstantes k :

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} \cdot p^k \cdot q^{n-k} = \frac{n(n-1) \cdots (n-k+1)}{k!} \cdot \frac{\lambda^k}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n(n-1) \cdots (n-k+1)}{n \cdot n \cdots n} \cdot \frac{\lambda^k}{k!} \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= 1 \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \cdot \left(1 - \frac{\lambda}{n}\right)^{-k} \cdot \frac{\lambda^k}{k!} \cdot \left(1 - \frac{\lambda}{n}\right)^n \end{aligned}$$

Da k konstant ist gilt:

$$\lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) = 1$$

sowie

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1$$

Es gilt ebenfalls

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

Somit folgt die Behauptung:

$$\lim_{n \rightarrow \infty} P(X_n = k) = 1 \cdot 1 \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \quad k = 0, 1, \dots$$

□

Satz 5. Der Erwartungswert $E(X)$, die Varianz $V(X)$ und die Streuung σ einer Poisson-verteilten Zufallsvariable X mit Parameter λ kann folgendermaßen berechnet werden [2, S.94]:

$$E(X) = \lambda \quad V(X) = \lambda \quad \sigma = \sqrt{\lambda}$$

Beweis. Der nachstehende Beweis ist angelehnt an die Beweisführung im Werk von Henze [13, S.196]. Sei X eine Poisson-verteilte Zufallsvariable, so gilt $P(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$ mit $\lambda > 0$. Nach Einsetzen in die Formel für den Erwartungswert für diskrete Zufallsvariablen erhält man folgenden Ausdruck:

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} \\ &= \lambda \cdot e^{-\lambda} \cdot \sum_{k=0}^{\infty} k \cdot \frac{\lambda^{k-1}}{k!} \\ &= \lambda \cdot e^{-\lambda} \cdot \sum_{k=1}^{\infty} k \cdot \frac{\lambda^{k-1}}{k!} \\ &= \lambda \cdot e^{-\lambda} \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \end{aligned}$$

Da $\sum_{k=1}^{\infty} e^{-\lambda} \cdot \frac{\lambda^{k-1}}{(k-1)!} = 1$ ist folgt

$$E(X) = \lambda \cdot e^{-\lambda} \cdot e^{\lambda} = \lambda$$

2 Wahrscheinlichkeitsrechnung

Für die Berechnung der Varianz gilt:

$$V(X) = E(X^2) - (E(X))^2$$

Durch Einsetzen folgt für den ersten Summanden [18, S.198]:

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{\infty} k^2 \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} \\ &= \lambda \cdot \sum_{k=1}^{\infty} k \cdot e^{-\lambda} \cdot \frac{\lambda^{k-1}}{(k-1)!} \end{aligned}$$

Durch Indexverschiebung folgt:

$$\begin{aligned} E(X^2) &= \lambda \cdot \sum_{k=0}^{\infty} (k+1) \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} \\ &= \lambda \cdot \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} + \lambda \cdot \sum_{k=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^k}{k!} \end{aligned}$$

Da $E(X) = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \cdot \frac{\lambda^k}{k!} = \lambda$ und $\sum_{k=0}^{\infty} e^{-\lambda} \cdot \frac{\lambda^k}{k!} = 1$ gilt, folgt:

$$E(X^2) = \lambda^2 + \lambda$$

Durch Einsetzen in die Formel für die Varianz erhält man schließlich:

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned}$$

Folglich gilt für die Standardabweichung σ

$$\sigma = \sqrt{\lambda}$$

□

Beispiel: Von 100 Personen sind durchschnittlich 2 Personen farbenblind. Es werden 100 Personen zufällig ausgewählt. Mit welcher Wahrscheinlichkeit sind unter diesen genau 3 farbenblinde Personen?

$$n = 100 \quad p = \frac{2}{100} \quad \lambda = 100 \cdot \frac{2}{100} = 2$$

$$P(X = 3) = e^{-2} \cdot \frac{2^3}{3!} \approx 0,1804 \hat{=} 18,04\%$$

2.5 Stetige Zufallsvariable

Eine Zufallsvariable muss nicht zwingend nur diskrete Werte annehmen, sondern es gibt auch Fälle, in welchen sie jede reelle Zahl eines Intervalls $I = [a, b]$ annehmen kann.

Definition 18. Eine Zufallsvariable X wird *stetige Zufallsvariable* genannt, wenn der Wertevorrat W überabzählbar unendlich ist und eine nicht negative, integrierbare Funktion f existiert, so dass für die Verteilungsfunktion $F(x)$ gilt :

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) \, du$$

Dabei heißt die Funktion f *Dichte* der stetigen Zufallsvariable X [2, S.100].

Die Wahrscheinlichkeit $P(X \leq x)$ entspricht somit der Fläche, welche von dem Graphen der Funktion f und der x-Achse im Intervall $(-\infty; x]$ eingeschlossen wird. Während die Verteilungsfunktion F einer diskreten Zufallsvariable durch ihre Treppenform stets Sprungstellen besitzt, weist der Graph der Verteilungsfunktion F einer stetigen Zufallsvariable Y keine Sprungstellen auf [2, S.99]. In der nachstehen Abbildung 2.8 ist die Dichtefunktion einer stetigen Zufallsvariable abgebildet.

2 Wahrscheinlichkeitsrechnung

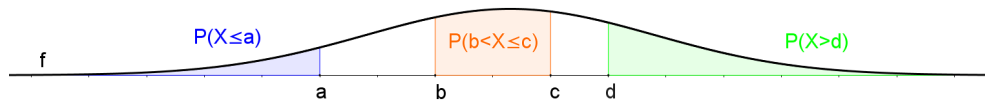


Abbildung 2.8: Dichtefunktion einer stetigen Zufallsvariable

Für eine Dichtefunktion f einer stetigen Zufallsvariable X gilt stets:

$$\int_{-\infty}^{+\infty} f(u) \, du = 1$$

Seien weiters $a, b, c \in \mathbb{R}$ beliebig mit $a < b$, so gilt [2, S.101]:

$$P(a < X \leq b) = F(b) - F(a) = \int_a^b f(u) \, du$$

$$P(X > c) = 1 - F(c) = \int_c^{+\infty} f(u) \, du$$

Beispiel: Die Funktionsfähigkeit eines technischen Gerätes ist eine stetige Zufallsvariable X und besitzt folgende Dichte

$$f : \mathbb{R} \rightarrow [0, 1] \quad f(x) = \begin{cases} 10x^{-2} & \text{falls } x > 10 \\ 0 & \text{sonst} \end{cases}$$

Wie groß ist die Wahrscheinlichkeit, dass dieses Gerät nach 100 Stunden noch funktionsfähig ist?

$$\begin{aligned}
P(X \geq 100) &= 1 - P(X < 100) = 1 - P(X \leq 100) = 1 - \int_{-\infty}^{100} f(x) \, dx \\
&= 1 - \int_{10}^{100} 10x^{-2} \, dx = 1 - (-10x^{-1}) \Big|_{10}^{100} \\
&= 1 + \frac{1}{10} - 1 = \frac{1}{10}
\end{aligned}$$

2.5.1 Erwartungswert, Varianz und Standardabweichung

Ist f die Dichtefunktion einer stetigen Zufallsvariable X und existiert das Integral $\int_{-\infty}^{+\infty} |x| \cdot f(x) \, dx$, so können der Erwartungswert $E(X)$, die Varianz $V(X)$ und die Standardabweichung σ der Zufallsvariable folgendermaßen berechnet werden [2, S.107]:

$$E(X) = \mu = \int_{-\infty}^{+\infty} x \cdot f(x) \, dx$$

$$V(X) = \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot f(x) \, dx$$

$$\sigma = \sqrt{V(X)} = \sqrt{\int_{-\infty}^{+\infty} (x - \mu)^2 \cdot f(x) \, dx}$$

Beispiel: Sei X eine Zufallsvariable mit der Dichte f mit

$$f : \mathbb{R} \rightarrow [0, 1] \quad f(x) = \begin{cases} 6x \cdot (1 - x) & \text{falls } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

2 Wahrscheinlichkeitsrechnung

Berechne den Erwartungswert $E(X)$, die Varianz $V(X)$ sowie die Standardabweichung σ der Zufallsvariable X .

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x \cdot f(x) \, dx = \int_{-\infty}^0 x \cdot 0 \, dx + \int_0^1 x \cdot (6x \cdot (1-x)) \, dx + \int_1^{+\infty} x \cdot 0 \, dx \\ &= \int_0^1 x \cdot (6x - 6x^2) \, dx = \int_0^1 6x^2 - 6x^3 \, dx \\ &= 2x^3 - \frac{6x^4}{4} \Big|_0^1 \\ &= \frac{1}{2} \end{aligned}$$

$$\begin{aligned} V(X) &= \int_{-\infty}^{+\infty} \left(x - \frac{1}{2}\right)^2 \cdot f(x) \, dx \\ &= \int_{-\infty}^0 \left(x - \frac{1}{2}\right)^2 \cdot 0 \, dx + \int_0^1 \left(x - \frac{1}{2}\right)^2 \cdot (6x \cdot (1-x)) \, dx + \int_1^{+\infty} \left(x - \frac{1}{2}\right)^2 \cdot 0 \, dx \\ &= \int_0^1 \left(x^2 - x + \frac{1}{4}\right) \cdot (6x - 6x^2) \, dx \\ &= \int_0^1 6x^3 - 6x^4 - 6x^2 + 6x^3 + \frac{6x}{4} - \frac{6x^2}{4} \, dx \\ &= \int_0^1 -6x^4 + 12x^3 - \frac{15x^2}{2} + \frac{6x}{4} \, dx \\ &= -\frac{6x^5}{5} + 3x^4 - \frac{5x^3}{2} + \frac{3x^2}{4} \Big|_0^1 \\ &= \frac{1}{20} \end{aligned}$$

Folglich gilt für die Standardabweichung:

$$\sigma = \sqrt{V(X)} = \sqrt{\frac{1}{20}}$$

2.5.2 Normalverteilung

Definition 19. Eine stetige Zufallsvariable X wird als *normalverteilt* bezeichnet, wenn sie eine Dichte f mit

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

besitzt. Die Zufallsvariable heißt kurz $N(\mu, \sigma^2)$ -verteilt, wobei $\mu \in \mathbb{R}$ und $\sigma > 0$ sind [2, S.135].

Definition 20. Der Spezialfall für $\mu = 0$ und $\sigma^2 = 1$ wird $N(0, 1)$ *Standardnormalverteilung* genannt. Eine Zufallsvariable Z , welche normalverteilt mit $\mu = 0$ und $\sigma^2 = 1$ ist, besitzt die Dichte

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$$

Die Verteilungsfunktion der Zufallsvariable Z wird Φ genannt [5, S.171].

Der Graph der Dichtefunktion einer $N(0, 1)$ - verteilten Zufallsvariable ist in der untenstehenden Abbildung 2.9 dargestellt und wird *Gaußsche Glockenkurve* genannt [2, S.132].

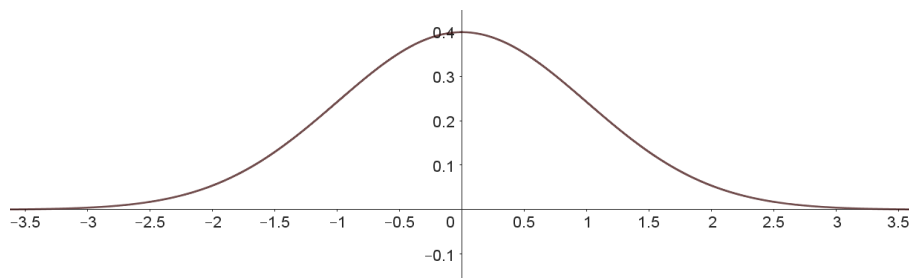


Abbildung 2.9: Gaußsche Glockenkurve

Anhand dieser Abbildung kann erkannt werden, dass die Dichtefunktion einer $N(0, 1)$ -verteilten Zufallsvariable X folgende Eigenschaften aufweist [5, S.171]:

2 Wahrscheinlichkeitsrechnung

- $\varphi(x)$ hat zwei Wendestellen bei $x = -1$ und $x = 1$ [2, S.132]
- φ ist achsensymmetrisch: $\varphi(x) = \varphi(-x)$
- $\Phi(x) = 1 - \Phi(-x), \quad x \in \mathbb{R}$

Beweis. Sei $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$ die Dichtefunktion einer $N(0, 1)$ -verteilten Zufallsvariable X . Die ersten drei Ableitungen, welche für die Untersuchung des Wendepunktes notwendig sind, können mit Hilfe der Produkt- und Kettenregel folgendermaßen ermittelt werden:

$$\begin{aligned}\varphi(x) &= \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \\ \varphi(x)' &= \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \cdot (-x) \\ \varphi(x)'' &= \frac{1}{\sqrt{2\pi}} \cdot \left[e^{-\frac{x^2}{2}} \cdot (-x) \cdot (-x) + e^{-\frac{x^2}{2}} \cdot (-1) \right] \\ &= \frac{1}{\sqrt{2\pi}} \cdot \left[e^{-\frac{x^2}{2}} \cdot x^2 - e^{-\frac{x^2}{2}} \right] \\ \varphi(x)''' &= \frac{1}{\sqrt{2\pi}} \cdot \left[e^{-\frac{x^2}{2}} \cdot (-x) \cdot x^2 + 2x \cdot e^{-\frac{x^2}{2}} - (e^{-\frac{x^2}{2}} \cdot (-x)) \right] \\ &= \frac{1}{\sqrt{2\pi}} \cdot \left[e^{-\frac{x^2}{2}} \cdot (-x)^3 + 2x \cdot e^{-\frac{x^2}{2}} \right]\end{aligned}$$

Durch Nullsetzen der zweiten Ableitung erhält man:

$$\begin{aligned}0 &= \frac{1}{\sqrt{2\pi}} \cdot \left[e^{-\frac{x^2}{2}} \cdot x^2 - e^{-\frac{x^2}{2}} \right] \\ 0 &= e^{-\frac{x^2}{2}} \cdot x^2 - e^{-\frac{x^2}{2}} \\ 0 &= e^{-\frac{x^2}{2}} \cdot (x^2 - 1)\end{aligned}$$

Da $e^{-\frac{x^2}{2}}$ nie Null wird, folgt Aufgrund des Satzes vom Nullprodukt, dass

$$0 = x^2 - 1 \quad \Rightarrow \quad x_1 = 1 \quad x_2 = -1$$

Weiters gilt für die dritte Ableitung an diesen Stellen:

$$\varphi(1)''' \neq 0 \quad \varphi(-1)''' \neq 0$$

Somit ist bewiesen, dass es sich bei den Stellen $x = 1$ und $x = -1$ um die Wendepunkte der Dichtefunktion einer $N(0, 1)$ -verteilten Zufallsvariable X handelt.

□

Um die gewünschten Wahrscheinlichkeiten berechnen zu können wird die von der x -Achse und dem Graphen der Dichtefunktion eingeschlossene Fläche betrachtet. Die Werte der Verteilungsfunktion

$$\Phi(z) = P(Z < z) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^z e^{-\frac{t^2}{2}} dt$$

können entweder mithilfe von Technologieeinsatz oder tabellarisch ermittelt werden. Aufgrund der Symmetrie von Φ gilt:

$$P(Z \leq -z) = \int_{-\infty}^{-z} \varphi(t) dt = \int_z^{\infty} \varphi(t) dt = P(Z \geq z)$$

Da allerdings auch gilt, dass

$$P(Z \leq z) + P(Z \geq z) = 1$$

folgt daraus für die Dichtefunktion Φ :

$$\Phi(-z) = 1 - \Phi(z) \quad \text{sowie} \quad P(-z \leq Z \leq z) = 2 \cdot \Phi(z) - 1$$

Dies hat zur Folge, dass es ausreichend ist die Funktion Φ für positive Werte zu tabellieren [2, S.133].

Handelt es sich um eine $N(0, 1)$ -verteilte Zufallsvariable, so kann die Fläche und somit die Wahrscheinlichkeit schnell mit Hilfe dieser Tabelle berechnet werden. Die Tabelle beinhaltet alle Werte von $\Phi(z)$ für $z \in [0, 3]$ aufgrund der Tatsache, dass

2 Wahrscheinlichkeitsrechnung

die Gaußsche Glockenkurve symmetrisch bezüglich der y -Achse ist und außerhalb des Intervalls $[-3, 3]$ annähernd Null ist und somit vernachlässigt werden kann.

Beispiel: Sei X eine $N(0, 1)$ verteilte Zufallsvariable. Berechne

1. $P(X < 1,72)$

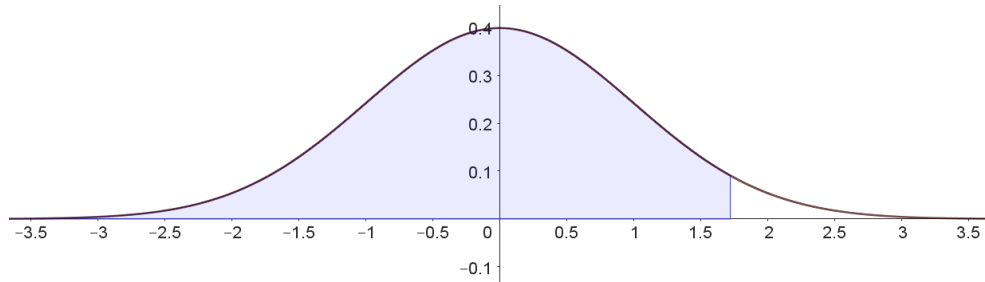


Abbildung 2.10: $P(X < 1,72)$

$$P(X < 1,72) = P(X \leq 1,72) = \Phi(1,72) = 0,95728$$

2. $P(X \leq -1,12)$

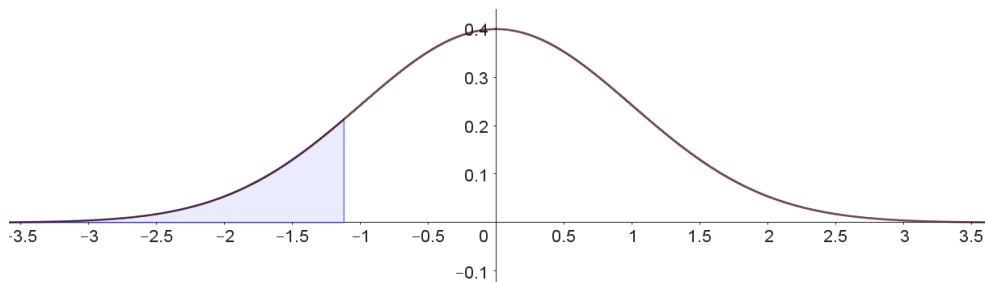


Abbildung 2.11: $P(X \leq -1,12)$

$$P(X \leq -1,12) = \Phi(-1,12) = 1 - \Phi(1,12) = 1 - 0,8686 = 0,1314$$

3. $P(-1,30 \leq X < 0,54)$

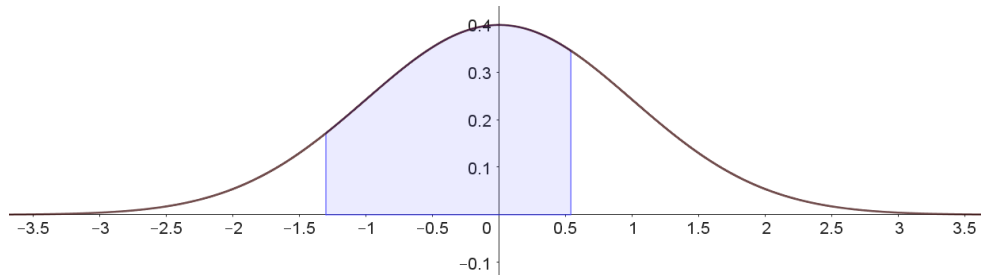


Abbildung 2.12: $P(-1,30 \leq X < 0,54)$

$$\begin{aligned} P(-1,30 \leq X < 0,54) &= \Phi(0,54) - \Phi(-1,30) = \Phi(0,54) - (1 - \Phi(1,30)) \\ &= 0,7054 - 1 + 0,9032 = 0,6086 \end{aligned}$$

Mit Hilfe dieser Tabelle können jedoch nur Integrale von $N(0,1)$ -verteilten Zufallsvariablen ermittelt werden. Um Wahrscheinlichkeiten einer Normalverteilung zu ermitteln, müssen diese auf die Standardnormalverteilung zurückgeführt werden.

Satz 6. Sei X eine normalverteilte Zufallsvariable mit den Parametern μ und σ . Dann hat $Y = \frac{X-\mu}{\sigma}$ die Standardnormalverteilung [2, S.135].

Beweis. Für den Beweis dieses Satzes wird zuerst der allgemeine Fall betrachtet, von welchem anschließend auf diesen Spezialfall übergegangen wird. Sei X eine Zufallsvariable mit der Wahrscheinlichkeitsdichte f und seien weiters $a \in \mathbb{R}$, $b > 0$ und $Y = \frac{X-a}{b}$. Sei $F(t) = P(X \leq t)$ die Verteilungsfunktion von X , sodass $F'(t) = f(t)$ gilt. Daraus folgt, dass

$$\begin{aligned} P(Y \leq t) &= P\left(\frac{X-a}{b} \leq t\right) = P(X \leq b \cdot t + a) \\ &= F(b \cdot t + a) \end{aligned}$$

2 Wahrscheinlichkeitsrechnung

Da die Ableitung dieser Funktion die Wahrscheinlichkeitsdichte g von Y ist, gilt:

$$g(t) = F'(b \cdot t + a) = b \cdot f(b \cdot t + a)$$

Wendet man diesen allgemeinen Fall auf eine $N(\mu, \sigma)$ -verteilte Zufallsvariable X mit der Wahrscheinlichkeitsdichte $f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ und $Y = \frac{X-\mu}{\sigma}$ an, so gilt für die Wahrscheinlichkeitsdichte g von Y :

$$\begin{aligned} g(x) &= \sigma \cdot \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(\sigma \cdot x + \mu - \mu)^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \end{aligned}$$

Dabei handelt es sich um die Wahrscheinlichkeitsdichte der $N(0, 1)$ -Verteilung [14, S.14]. $\square$

Beispiel: Beim Tennisspielen ist das Gewicht des Schlägers immer von großer Bedeutung. Bei der Herstellung eines bestimmten Modells ist das Gewicht des Schlägers $N(290, 5^2)$ -verteilt.

1. Berechne die Wahrscheinlichkeit, dass der Schläger maximal 300g wiegt.

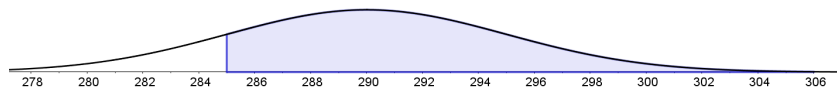


Abbildung 2.13: $P(X \leq 300)$

$$x = 300 \quad \Rightarrow \quad z = \frac{300 - \mu}{\sigma} = \frac{300 - 290}{5} = 2$$

$$P(X \leq 300) = P(Z \leq 2) = \Phi(2) = 0,97725$$

2. Wie groß ist die Wahrscheinlichkeit, dass das Gewicht um höchstens 7g vom Erwartungswert abweicht?

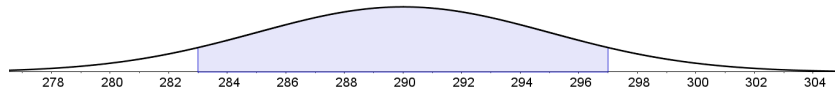


Abbildung 2.14: $P(283 \leq X \leq 297)$

$$P(\mu - 7 \leq X \leq \mu + 7) = P(283 \leq X \leq 297)$$

$$\begin{aligned} x_1 = 283 &\Rightarrow z_1 = \frac{283 - 290}{5} = -\frac{7}{5} = -1,4 \\ x_2 = 297 &\Rightarrow z_2 = \frac{297 - 290}{5} = \frac{7}{5} = 1,4 \end{aligned}$$

$$\begin{aligned} P(283 \leq X \leq 297) &= P(-1,4 \leq Z \leq 1,4) = \Phi(1,4) - (1 - \Phi(1,4)) \\ &= 2 \cdot \Phi(1,4) - 1 = 2 \cdot 0,9192 - 1 = 0,8385 \end{aligned}$$

3. Welchen Toleranzbereich $[\mu + t; \mu - t]$ um den Erwartungswert μ kann dieser Hersteller zu mindestens 95% seinen Kunden garantieren?

$$P(|X - \mu| \leq t) = \gamma \Rightarrow P(|X - 7| \leq t) = 0,95$$

$$0,95 = 2 \cdot \Phi(z) - 1 \Rightarrow \Phi(z) = 0,975 \Rightarrow z = 1,96$$

Nun muss die Standardisierung rückgängig gemacht werden indem t berechnet wird:

$$t = z \cdot \sigma = 1,96 \cdot 5 = 9,8$$

2 Wahrscheinlichkeitsrechnung

Somit kann der Toleranzbereich folgendermaßen berechnet werden:

$$x_1 = 290 - 9,8 = 280,2 \quad x_2 = 290 + 9,8 = 299,8 \quad \Rightarrow \quad [280,2; 299,8]$$

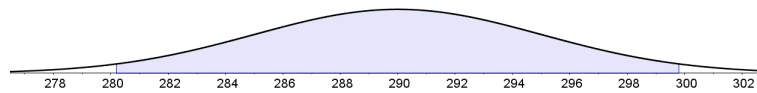


Abbildung 2.15: $P(|X - 7| \leq t) = 0,95$

Beispiel: Eine Produktionsfirma von Tennisschlägern behauptet, dass 90% ihrer Schläger mindestens 287g wiegen. Wie groß ist das durchschnittliche Gewicht der Schläger, wenn das Gewicht normalverteilt mit $\sigma = 4,3g$ ist?

$$\begin{aligned} P(X \geq 287) = 0,90 & \Rightarrow P(X \leq 287) = 0,1 \\ P(Z \leq z) = 0,1 & \Rightarrow \Phi(z) = 0,1 \Rightarrow z \approx -1,28 \end{aligned}$$

Nachdem zuerst zur Standardnormalverteilung übergegangen wurde, muss dieser Schritt nun rückgängig gemacht werden und μ mit Hilfe der Standardisierungsformel ermittelt werden:

$$z = \frac{x - \mu}{\sigma} \Rightarrow -1,28 = \frac{287 - \mu}{4,3} \Rightarrow \mu = 292,504g$$

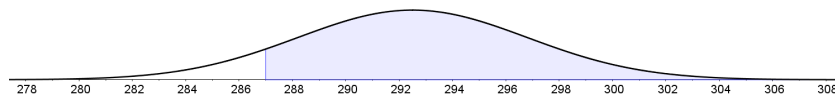


Abbildung 2.16: $P(X \geq 287) = 0,90$

Satz 7. Sei X eine normalverteilte Zufallsvariable, so gilt für den Erwartungswert

$E(X)$, die Varianz $V(X)$ und die Standardabweichung [2, S:105,106]:

$$E(X) = \mu \quad V(X) = \sigma^2 \quad \sigma = \sqrt{V(X)}$$

Beweis. Der nachstehende Beweis folgt der Beweisführung von Gebauer [10, S.3]. Sei X eine normalverteilte Zufallsvariable mit $f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Durch Einsetzen in die Formel für den Erwartungswert von stetigen Zufallsvariablen erhält man folgenden Ausdruck:

$$E(X) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} x \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Durch Substitution von $a = x - \mu$ gilt:

$$\begin{aligned} E(X) &= \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} (a + \mu) \cdot e^{-\frac{a^2}{2\sigma^2}} da \\ &= \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} a \cdot e^{-\frac{a^2}{2\sigma^2}} da + \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} \mu \cdot e^{-\frac{a^2}{2\sigma^2}} da \end{aligned}$$

Bei $f(a) = a \cdot e^{-\frac{a^2}{2\sigma^2}}$ handelt es sich um eine ungerade Funktion. Das Integral auf $[0, \infty)$ existiert, da

$$\begin{aligned} &\lim_{b \rightarrow +\infty} \int_0^b a \cdot e^{-\frac{a^2}{2\sigma^2}} da \\ &= \lim_{b \rightarrow +\infty} \left[-\sigma^2 \cdot e^{-\frac{a^2}{2\sigma^2}} \right]_0^b \\ &= \lim_{b \rightarrow +\infty} \left[-\sigma^2 \cdot e^{-\frac{b^2}{2\sigma^2}} + \sigma^2 \cdot e^{-\frac{0}{2\sigma^2}} \right] \\ &= 0 + \sigma^2 = \sigma^2 \end{aligned}$$

Wegen der Symmetrie existiert somit auch das Integral auf $(-\infty; 0]$. Somit ist die Funktion uneigentlich integrierbar auf $(-\infty; \infty)$, da die uneigentlichen Integrale

2 Wahrscheinlichkeitsrechnung

auf $(-\infty; 0]$ und $[0, \infty)$ existieren. Der Wert des uneigentlichen Integrals mit den Grenzen $-\infty$ und ∞ ist Null. Es folgt:

$$E(X) = 0 + \frac{\mu}{\sigma \cdot \sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{-\frac{a^2}{2\sigma^2}} da$$

Durch erneutes Substituieren von $b = \frac{a}{\sigma}$ und der Tatsache, dass es sich bei $f(b) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{b^2}{2}}$ db um die Dichte einer Standardnormalverteilung handelt, deren Integral mit den Grenzen $-\infty$ und ∞ gleich Eins ist, folgt:

$$E(X) = \frac{\mu}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{-\frac{b^2}{2}} db = \mu$$

Für den Nachweis der Varianz wird ähnlich vorgegangen. Durch das Einsetzen in die Formel für die Varianz von stetigen Zufallsvariablen erhält man folgenden Ausdruck:

$$V(X) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} (x - \mu)^2 \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Nach Substitution von $a = x - \mu$ folgt:

$$V(X) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} a^2 \cdot e^{-\frac{a^2}{2\sigma^2}} da$$

Nach erneuter Substitution von $b = \frac{a}{\sigma}$ gilt:

$$\begin{aligned} V(X) &= \frac{\sigma^2}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} b^2 \cdot e^{-\frac{b^2}{2}} db \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} b \cdot \left(b \cdot e^{-\frac{b^2}{2}} \right) db \end{aligned}$$

Nach partieller Integration mit $f'(b) = be^{-\frac{b^2}{2}}$ und $g(b) = b$ folgt:

$$V(X) = \sigma^2 \cdot \left(\frac{1}{\sqrt{2\pi}} \cdot \left[-b \cdot e^{-\frac{b^2}{2}} \right]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{-\frac{b^2}{2}} db \right)$$

Da $\left[-b \cdot e^{-\frac{b^2}{2}} \right]_{-\infty}^{\infty} = 0$ und $f(b) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{b^2}{2}}$ die Dichte einer Standardnormalverteilung ist, deren Integral mit den Grenzen $-\infty$ und ∞ gleich Eins ist, folgt:

$$\begin{aligned} V(X) &= \sigma^2 \cdot (0 + 1) \\ &= \sigma^2 \end{aligned}$$

Folglich gilt für die Standardabweichung einer normalverteilten Zufallsvariable:

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

□

2.5.3 Chi-Quadrat-Verteilung

Bei der Chi-Quadrat-Verteilung, oder kurz χ^2 -Verteilung, handelt es sich um eine weitere stetige Wahrscheinlichkeitsverteilung, welche vor allem in der Statistik ihre Anwendung findet.

Seien $X_1, \dots, X_n$ stochastisch unabhängige $N(0, 1)$ -verteilte Zufallsvariablen aus welchen die Quersumme

$$\chi^2 = X_1^2 + \dots + X_n^2 \quad \text{für } n = 1, 2, \dots$$

gebildet wird. Sei weiters $\tau(a)$ die *Gammafunktion*

$$\tau(a) = \int_0^{\infty} e^{-t} \cdot t^{a-1} dt$$

2 Wahrscheinlichkeitsrechnung

mit den Eigenschaften [2, S.154]:

1. $\tau(a+1) = a \cdot \tau(a)$
2. $\tau\left(\frac{1}{2}\right) = \sqrt{\pi}$ und $\tau(1) = 1$
3. $\tau(n) = (n-1)!$

Definition 21. Eine Zufallsvariable X besitzt eine *Chi-Quadrat-Verteilung* mit n Freiheitsgraden, falls für ihre Dichte f_n gilt [2, S.154]:

$$f_n(x) = \begin{cases} 0, & x \leq 0 \\ \frac{1}{2^{\frac{n}{2}} \cdot \tau(\frac{n}{2})} \cdot e^{-\frac{x}{2}} \cdot x^{\frac{n}{2}-1} & x > 0 \end{cases}$$

Satz 8. Sei X eine Chi-Quadrat-verteilte Zufallsvariable, so gilt für den Erwartungswert $E(X)$, die Varianz $V(X)$ und die Standardabweichung σ [2, S.155]:

$$E(\chi_n^2) = n \quad V(\chi_n^2) = 2n \quad \sigma = \sqrt{2n}$$

Beweis. Der Beweis kann in einem Videovortrag von Henze [12] gefunden werden. □

Satz 9. Für unabhängige $N(0,1)$ -verteilte Zufallsvariablen $X_1, \dots, X_n$ ist die Zufallsvariable $X_1^2 + \dots + X_n^2$ dann χ^2 -verteilt mit n Freiheitsgraden.

Beweis. Ein Beweis für diesen Satz kann im Skriptum von Hofbauer [14, S.24] nachgelesen werden. □

3 χ^2 – Anpassungstest

In der Statistik geht es oft darum Hypothesen aufzustellen und diese dann zu überprüfen. Dabei werden kleine Abweichungen von dem theoretisch berechneten Wert akzeptiert, bei größeren Abweichungen wird an der Korrektheit der Hypothese gezweifelt und diese schließlich verworfen. In der Praxis wird stets eine Irrtumswahrscheinlichkeit α angegeben, welche die Wahrscheinlichkeit angibt, dass eine Hypothese fälschlicherweise verworfen wird. Der Wert dieser Irrtumswahrscheinlichkeit trägt beträchtlich zur Zuverlässigkeit des Hypothesentests bei.

Um im weiteren Laufe der vorliegenden Masterarbeit die theoretisch berechneten Wahrscheinlichkeiten mit realen Daten vergleichen zu können, ist es notwendig ein statistisches Verfahren vorerst näher zu erläutern. Mit Hilfe eines sogenannten *Chi-Quadrat-Anpassungstest*, kurz χ^2 -Anpassungstest, kann getestet werden, ob Daten einer Stichprobe die Vermutungen einer bestimmten Verteilung der Zufallsvariable bestätigen.

Definition 22. Bei einem *Multinomial-Experiment* handelt es sich um ein Experiment, welches durch folgende Eigenschaften charakterisiert ist [9, S.277-278]:

1. Es gibt $r \geq 2$ mögliche Ausgänge $x_1, \dots, x_r$, welche als *Zellen* bezeichnet werden.
2. Die Wahrscheinlichkeiten für ein Ergebnis x_j ist p_j wobei $j = 1, \dots, r$ ist.

Definition 23. Sei X_0 eine Zufallsvariable mit einer vermuteten Verteilung F_0 . Sei weiters $X = (X_1, \dots, X_n)$ ein Stichprobenvektor, bei welchem die Koordinaten unabhängige Kopien von X_0 sind. Der Wertebereich der Zufallsvariable X_0 sei

3 χ^2 – Anpassungstest

in paarweise disjunkte Teilmengen S_k mit $k = 1, \dots, r$ zerlegt und n_k gibt die Häufigkeit an, mit welcher X mit den Werten in S_k und $p_k = P(X_0 \in S_k)$ auftritt. Man definiert die Testgröße χ^2 für die Ermittlung der Abweichung zwischen den theoretischen Häufigkeiten np_k und den tatsächlich auftretenden Häufigkeiten n_k durch [8, S.171]:

$$\chi^2 := \sum_{k=1}^r \frac{(n_k - np_k)^2}{np_k}$$

Im Allgemeinen wird bei einem χ^2 -Anpassungstest folgendermaßen vorgegangen [3, S.231]:

- Aufstellen der Hypothese: Dabei besitzen die r verschiedenen Ereignisse die Wahrscheinlichkeiten p_k mit $k = 1, \dots, r$. Die Summe dieser ergibt 1.
- Durchführung des Tests: Dabei wird das Zufallsexperiment n -mal unabhängig voneinander durchgeführt. Die dabei auftretenden Häufigkeiten werden mit n_k bezeichnet.
- Berechnung des Stichprobenwertes χ^2 : Dabei ist χ^2 Chi-Quadrat-verteilt mit $r - 1$ Freiheitsgraden.
- Ablehnung oder Zustimmung der Hypothese: Sei $\alpha \in (0, 1)$ und $k \in \mathbb{N}$. Dann nennen wir jenes $x \in \mathbb{R}$, für das $\int_0^x f_k(t) dt = 1 - \alpha$ gilt, wobei f_k die Dichte der χ^2 -Verteilung mit k Freiheitsgraden ist, das $(1 - \alpha)$ -Quantil der χ^2 -Verteilung mit k Freiheitsgraden, und bezeichnen dieses x mit $c_{1-\alpha, k}$.

Eine Hypothese H_0 wird dann verworfen, wenn die Stichprobe zu einem zu großen χ^2 Wert führt. Zuerst wird die Irrtumswahrscheinlichkeit α , also jene Wahrscheinlichkeit, dass die Hypothese fälschlicherweise verworfen wird, vorgegeben. Sei α das gewünschte Signifikanzniveau, so muss der Ablehnungsbereich $\mathbf{B}$ folgende Bedingung erfüllen, wobei c der kritische Wert ist [9, S.279]:

$$P(\chi^2 > c_0 \mid H_0) = \alpha$$

Dies ist äquivalent zu:

$$1 - P(\chi^2 \leq c_0 \mid H_0) = \alpha \quad \Leftrightarrow \quad P(\chi^2 \leq c_0 \mid H_0) = 1 - \alpha$$

Die obige Gleichung ist für $c_0 = c_{1-\alpha; k-1} = (1 - \alpha)$ -Quantil der $\chi^2(k - 1)$ Verteilung erfüllt, woraus für den Ablehnungsbereich folgt [9, S.279]:

$$\mathbf{B} =]c_{1-\alpha; k-1}; \infty[$$

Falls der berechnete Stichprobenwert größer als $c_{1-\alpha, k-1}$ ist und somit im Ablehnungsbereich liegt, wurde mit Sicherheit $1 - \alpha$ bewiesen, dass $P(X_0 \in S_k) = p_k$ für zumindest ein k falsch ist (also die behauptete Verteilung nicht stimmt). Wenn hingegen der berechnete Stichprobenwert höchstens $c_{1-\alpha, k-1}$ ergibt, dann sprechen die Daten mit Sicherheit $1 - \alpha$ nicht dagegen, dass $P(X_0 \in S_k) = p_k$ für alle k stimmt (also die behauptete Verteilung stimmt).

Beispiel: Ein Würfel wird 60 mal geworfen, wobei folgende Häufigkeiten, welche in Tabelle 3.1 dargestellt sind, für die jeweiligen Augenzahlen beobachtet werden. Es soll überprüft werden, ob es sich um einen fairen Würfel handelt.

Augenzahl	1	2	3	4	5	6
Aufgetretene Häufigkeit n_k	10	11	8	9	15	7
Berechnete Häufigkeit np_k	10	10	10	10	10	10

Tabelle 3.1: Augenzahl bei 60-maligen Würfeln

Bei diesem Experiment handelt es sich um einen Stichprobenumfang von $n = 60$, wobei es $r = 6$ verschiedene Ereignisse mit den relativen Häufigkeiten

$$h_1 = \frac{10}{60}, \quad h_2 = \frac{11}{60}, \quad h_3 = \frac{8}{60}, \quad h_4 = \frac{9}{60}, \quad h_5 = \frac{15}{60}, \quad h_6 = \frac{7}{60}$$

gibt. Da überprüft werden soll, ob es sich um einen fairen Würfel handelt lautet

3 χ^2 - Anpassungstest

die Hypothese:

$$H_0 : p_1 = \dots = p_6 = \frac{1}{6}$$

Für die Berechnung des Stichprobenwertes χ^2 folgt durch Einsetzen in die Formel:

$$\begin{aligned}\chi^2(x) &= \sum_{k=1}^r \frac{(n_k - np_k)^2}{np_k} \\ &= \frac{(10 - 10)^2}{10} + \frac{(11 - 10)^2}{10} + \frac{(8 - 10)^2}{10} + \frac{(9 - 10)^2}{10} + \frac{(15 - 10)^2}{10} + \frac{(7 - 10)^2}{10} \\ &= 0 + \frac{1}{10} + \frac{4}{10} + \frac{1}{10} + \frac{25}{10} + \frac{9}{10} \\ &= 4\end{aligned}$$

Für die Irrtumswahrscheinlichkeit wird $\alpha = 0,05$ vorausgesetzt. Für die Ermittlung des kritischen Wertes c wird das $(1 - 0,05)$ -Quantil mit $r - 1 = 5$ Freiheitsgraden aus der Tabelle der χ^2 - Verteilung für die Gleichung $F(c) = 1 - \alpha = 0,95$ abgelesen. Durch Ablesen aus der Tabelle ergibt sich $c = 11,07$. Somit gilt für den Ablehnungsbereich:

$$\mathbf{B} =]11,07; \infty[$$

Da sich der berechnete Stichprobenwert nicht in dem Ablehnungsbereich befindet, kann bei einer Irrtumswahrscheinlichkeit von $\alpha = 0,05$ von einem fairen Würfel gesprochen werden. Das bedeutet, dass mit Irrtumswahrscheinlichkeit 0,05 die Daten nicht dagegensprechen, dass es sich um einen fairen Würfel handelt.

4 Modellierung eines Tennismatches

Unter der Annahme einer konstanten Gewinnwahrscheinlichkeit für einen Punkt oder ein Spiel lässt sich ein Tennismatch modellieren und verschiedene Gewinnwahrscheinlichkeiten ermitteln. Obwohl in der Realität nicht von einer konstanten Gewinnwahrscheinlichkeit für einen Punkt oder Satz ausgegangen werden kann, da beispielsweise Resignation eine Folge eines verlorenen Punktes bzw. Satzes sein kann, können mögliche Spielverläufe mit ihren Wahrscheinlichkeiten näherungsweise modelliert werden.

4.1 Gewinnwahrscheinlichkeit eines Games

Dieses Unterkapitel ist angelehnt an das Werk von Stewart [21, S.35-37].

Spieler*in A und Spieler*in B entscheiden sich dazu ein Tennismatch zu spielen. Es wird angenommen, dass Spieler*in A einen Punkt mit der konstanten Wahrscheinlichkeit p gewinnt. Dabei ist p unabhängig davon wer gerade Aufschlag hat. Des Weiteren werden jegliche weitere Einflussfaktoren auf die sportliche Leistung außer Acht gelassen. Mit Hilfe der Gegenwahrscheinlichkeit kann ermittelt werden, dass Spieler*in B einen Punkt mit der Wahrscheinlichkeit $q = 1 - p$ gewinnt.

Wie bereits in Kapitel 1.2 erwähnt, ist ein Game dann gewonnen, wenn ein*e Spieler*in vier Punkte erreicht hat. Bei einem Einstand, also bei einem Spielstand von 40:40 muss ein*e Spieler*in zwei Punkte mehr als sein*ihre Gegner*in gewinnen, um das Spiel für sich zu entscheiden. In der nachstehenden Abbildung 4.1 sind alle möglichen Spielverläufe eines Games ersichtlich. Dabei bezeichnet Game A den Gewinn des Games für Spieler*in A und Game B den Gewinn des Games für Spieler*in B.

4 Modellierung eines Tennismatches

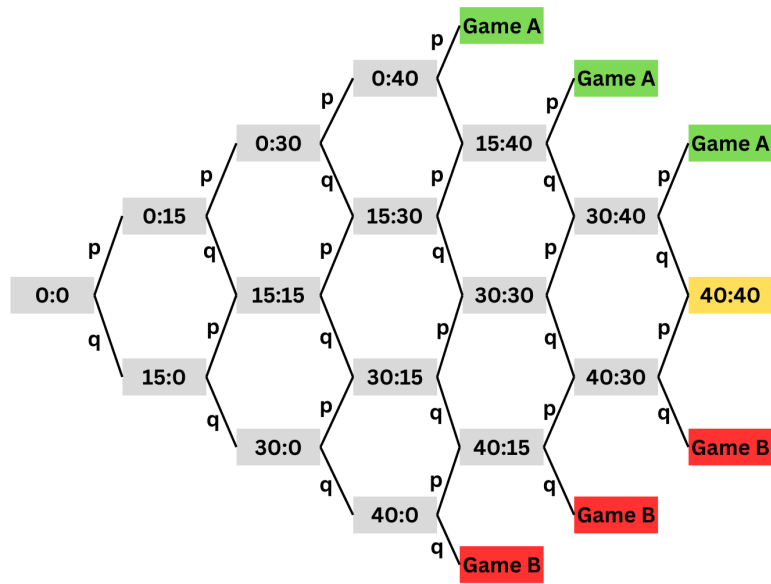


Abbildung 4.1: Spielverlauf eines Games

Beim Spielstand 40:40 handelt es sich um einen Einstand. Die möglichen Spielverläufe bei einem Einstand sind in der untenstehenden Abbildung 4.2 dargestellt.

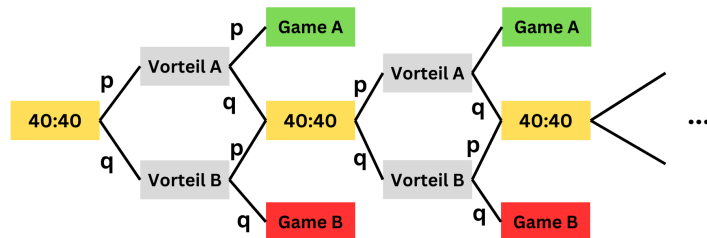


Abbildung 4.2: Spielverlauf eines Games bei Einstand

Anhand der obigen Abbildungen kann erkannt werden, dass Spieler*in A, abgesehen von dem Fall eines Einstands, drei verschiedene Möglichkeiten hat das Game zu gewinnen. Dazu müssen entweder vier, fünf oder sechs Punkte gespielt werden. Die jeweiligen Gewinnwahrscheinlichkeiten für Spieler*in A bei drei, vier oder fünf gespielten Punkten können folgendermaßen ermittelt werden:

4.1 Gewinnwahrscheinlichkeit eines Games

$$P(\text{A gewinnt nach 4 Punkten}) = p^4$$

$$P(\text{A gewinnt nach 5 Punkten}) = \binom{4}{1} \cdot p^4 \cdot q = 4 \cdot p^4 \cdot q$$

$$P(\text{A gewinnt nach 6 Punkten}) = \binom{5}{2} \cdot p^4 \cdot q^2 = 10 \cdot p^4 \cdot q^2$$

Dabei stehen die Binomialkoeffizienten für die Anzahl der Möglichkeiten um den jeweils gewünschten Spielstand zu erreichen. Ist das Game nach sechs Punkten nicht entschieden, so kommt es zu einem Einstand. Die Wahrscheinlichkeit, dass es nach sechs gespielten Punkten 40:40 steht kann folgendermaßen berechnet werden:

$$P(\text{Einstand}) = \binom{6}{3} \cdot p^3 \cdot q^3 = 20 \cdot p^3 \cdot q^3$$

Da ein Game nicht nach sieben, neun, usw. gespielten Punkten entschieden werden kann, müssen acht, zehn, usw. Punkte gespielt werden. Da ein Einstand theoretisch unendlich oft auftreten kann, erhält man für die Gewinnwahrscheinlichkeit eines Games bei Einstand folgende unendliche Reihe:

$$\begin{aligned} P(\text{A gewinnt nach Einstand}) &= 20p^3q^3 \cdot p^2 + 20p^3q^3 \cdot 2pq \cdot p^2 + 20p^3q^3 \cdot 2pq \cdot 2pq \cdot p^2 + \dots \\ &= 20 \cdot p^5 \cdot q^3 + 40 \cdot p^6 \cdot q^4 + 80 \cdot p^7 \cdot q^5 + \dots \\ &= 20 \cdot p^5 \cdot q^3 \cdot [1 + 2pq + 4p^2q^2 + \dots] \\ &= 20 \cdot p^5 \cdot q^3 \cdot [1 + 2pq + (2pq)^2 + (2pq)^3 \dots] \end{aligned}$$

Werden nun die Wahrscheinlichkeiten für einen Gewinn nach vier, fünf und sechs gespielten Punkten sowie nach einem Einstand summiert, erhält man für die Gewinnwahrscheinlichkeit eines Games folgenden Ausdruck mit einer geometrischen Reihe:

4 Modellierung eines Tennismatches

$$\begin{aligned} P(\text{Game A}) &= p^4 + 4p^4q + 10p^4q^2 + 20 \cdot p^5 \cdot q^3 \cdot [1 + 2pq + (2pq)^2 + (2pq)^3 \dots] \\ &= p^4 + 4p^4q + 10p^4q^2 \cdot [1 + 2pq + (2pq)^2 + (2pq)^3 \dots] \end{aligned}$$

Da es sich bei dem Ausdruck in der Klammer um eine geometrische Reihe mit $|2pq| < 1$ handelt, kann folgende Summenformel der geometrischen Reihe angewendet werden:

$$\sum_{k=0}^{\infty} (2pq)^k = 1 + 2pq + (2pq)^2 + \dots = \frac{1}{1 - 2pq}$$

Somit kann die Gewinnwahrscheinlichkeit eines Games bei konstanter Punktgewinnwahrscheinlichkeit p folgendermaßen berechnet werden:

$$\begin{aligned} P(\text{Game}) &= p^4 + 4p^4q + 10p^4p^2 \cdot \frac{1}{1 - 2pq} \\ &= p^4 + 4p^4q + \frac{10p^4p^2}{1 - 2pq} \end{aligned}$$

4.2 Zu erwartende Dauer eines Games

Die folgende Ausarbeitung folgt dem Werk von Croucher [7, S.82-84].

Die Dauer eines gespielten Games hängt stets von der Anzahl der gespielten Punkte ab. Ähnlich wie die Wahrscheinlichkeit ein Game zu gewinnen, lässt sich ebenfalls die durchschnittliche Dauer eines Games ermitteln. Angenommen Spieler*in A gewinnt einen Punkt mit der konstanten Wahrscheinlichkeit p und folglich sein*ihre Gegner*in mit der Wahrscheinlichkeit $q = 1 - p$.

Sei X eine diskrete Zufallsvariable, die angibt wie viele Punkte ausgespielt werden, bis ein*e Spieler*in das Game für sich entschieden hat. Somit kann die zu erwartende Dauer eines Games folgendermaßen angegeben werden:

4.2 Zu erwartende Dauer eines Games

$$E(X) = \sum_{k \geq 1}^{+\infty} k \cdot P(\text{A oder B siegt nach } k \text{ Punkten})$$

Sei P_k die Wahrscheinlichkeit, dass ein Game nach k Punkten entschieden ist. Da für ein Game mindestens vier Punkten gespielt werden müssen, und nach einem Einstand jeweils zwei Punkte Differenz erzielt werden müssen, gilt:

$$P_1 = P_2 = P_3 = P_7 = P_9 = P_{11} = \dots = 0$$

Folglich gilt für die durchschnittliche Dauer eines Games:

$$E(X) = 4 \cdot P_4 + 5 \cdot P_5 + 6 \cdot P_6 + 8 \cdot P_8 + 10 \cdot P_{10} + \dots$$

Bei der Berechnung der jeweiligen Wahrscheinlichkeiten P_k ist es nun im Vergleich zum Kapitel 4.1 zu berücksichtigen, dass sowohl Spieler*in A als auch Spieler*in B das Game gewinnen können. Folglich gilt für die Gewinnwahrscheinlichkeiten:

$$\begin{aligned} P_4 &= p^4 + q^4 \\ P_5 &= 4 \cdot (p^4 \cdot q + p \cdot q^4) \\ P_6 &= 10 \cdot (p^4 \cdot q^2 + p^2 \cdot q^4) \\ P_8 &= 20 \cdot (p^5 \cdot q^3 + p^3 \cdot q^5) = 20p^3q^3 \cdot (p^2 + q^2) \\ P_{10} &= P_8 \cdot 2pq = 20p^3q^3 \cdot (p^2 + q^2) \cdot (2pq) \\ P_{12} &= 20p^3q^3 \cdot (p^2 + q^2) \cdot (2pq)^2 \\ &\dots \end{aligned}$$

Zu Beginn wird die unendliche Summe der Gewinnwahrscheinlichkeiten nach einem Einstand betrachtet. Dabei gilt:

4 Modellierung eines Tennismatches

$$\begin{aligned} & 8 \cdot P_8 + 10 \cdot P_{10} + 12 \cdot P_{12} + \dots \\ &= 8 \cdot 20p^3q^3(p^2 + q^2) + 10 \cdot 20p^3q^3(p^2 + q^2)(2pq) + 12 \cdot 20p^3q^3(p^2 + q^2)(2pq)^2 + \dots \\ &= 20p^3q^3 \cdot (p^2 + q^2) \cdot (8 + 10 \cdot 2pq + 12 \cdot (2pq)^2 + \dots) \\ &= 40p^3q^3 \cdot (p^2 + q^2) \cdot (4 + 5 \cdot 2pq + 6 \cdot (2pq)^2 + \dots) \\ &= 40p^3q^3 \cdot (p^2 + q^2) \cdot \left(\sum_{i=1}^{\infty} (i+3) \cdot (2pq)^{i-1} \right) \end{aligned}$$

Für die weitere Berechnung ist es von Vorteil zu beachten, dass für die Summe einer geometrischen Reihe folgendes gilt:

Betrachtet man zur Vereinfachung die Funktion

$$\frac{1}{1-x} = \sum_{i=0}^{\infty} x^i \quad \text{für } |x| < 1 \text{ mit } x = 2pq$$

Für diese gilt:

$$\frac{1}{1-x} = \sum_{i=0}^{\infty} x^i = 1 + \sum_{i=1}^{\infty} x^i \quad \text{falls } |x| < 1$$

Dabei handelt es sich um eine Potenzreihe mit Konvergenzradius 1, welche gliedweises Differenzieren erlaubt, ohne dass sich der Konvergenzradius dabei ändert. Durch Differenzieren beider Seiten nach x ergibt sich:

$$\left(1 + \sum_{i=1}^{\infty} x^i \right)' = \sum_{i=1}^{\infty} i \cdot x^{i-1} = \frac{1}{(1-x)^2}$$

Somit gilt für die zuvor ermittelte geometrische Reihe:

$$\sum_{i=1}^{\infty} (i+3) \cdot x^{i-1} = \frac{1}{(1-x)^2} + 3 \cdot \frac{1}{1-x}$$

4.2 Zu erwartende Dauer eines Games

Einsetzen von $x = 2pq$ führt nun zu:

$$\sum_{i=1}^{\infty} (i+3) \cdot (2pq)^{i-1} = \frac{1}{(1-2pq)^2} + 3 \cdot \frac{1}{1-2pq}$$

Weiter gilt $p + q = 1$ und somit:

$$1 - 2pq = (p + q)^2 - 2pq = p^2 + q^2$$

Durch Einsetzen aller Werte kann die durchschnittliche Dauer eines Games folgendermaßen ermittelt werden:

$$\begin{aligned} E(X) &= 4 \cdot (p^4 + q^4) + 20 \cdot (p^4q + pq^4) + 60 \cdot (p^4q^2 + p^2q^4) \\ &\quad + 40p^3q^3 \cdot (p^2 + q^2) \cdot \left(\frac{1}{(p^2 + q^2)^2} + \frac{3}{p^2 + q^2} \right) \\ &= 4 \cdot (p^4 + q^4) + 20 \cdot (p^4q + pq^4) + 60 \cdot (p^4q^2 + p^2q^4) + 40p^3q^3 \cdot \left(\frac{1}{p^2 + q^2} + 3 \right) \end{aligned}$$

In der nachstehenden Tabelle 4.1 ist die zu erwartende Dauer in Abhängigkeit der Punktgewinnwahrscheinlichkeit eines*r Spielenden angegeben. Es ist ersichtlich, dass die längste Spieldauer bei einer Punktgewinnwahrscheinlichkeit von $p = 0,5$ zu erwarten ist. In diesem Fall treffen zwei Spieler*innen aufeinander und müssen durchschnittlich 6,75 Punkte durchspielen, bis eine*r von beiden das Game für sich entscheiden kann.

4 Modellierung eines Tennismatches

Punktgewinnwahrscheinlichkeit p	Durchschnittlich gespielte Punkte
0,10	4,46
0,20	5,09
0,30	5,83
0,40	6,48
0,50	6,75
0,60	6,48
0,70	5,83
0,80	5,09
0,90	4,46

Tabelle 4.1: Durchschnittlich gespielte Punkte

4.3 Matchgewinnwahrscheinlichkeit in Abhängigkeit von der Satzgewinnwahrscheinlichkeit

Dieses Unterkapitel folgt dem Kapitel *Tennis auf zwei Gewinnsätze* des Werkes von Bosch [1, S.75-78].

Einige Tennismatche werden auf zwei gewonnene Sätze gespielt. Angenommen Spieler*in A gewinnt einen Satz mit der konstanten Wahrscheinlichkeit p . Wie zuvor werden auch hier alle beeinflussenden Faktoren außer Acht gelassen und es wird angenommen, dass diese Wahrscheinlichkeit konstant ist. Folglich gewinnt Spieler*in B einen Satz mit der Wahrscheinlichkeit $q = 1 - p$.

Die verschiedenen Spielverläufe eines Best-of-3 Tennismatches sind in der nachstehenden Abbildung 4.3 veranschaulicht.

4.3 Matchgewinnwahrscheinlichkeit in Abhängigkeit von der Satzgewinnwahrscheinlichkeit

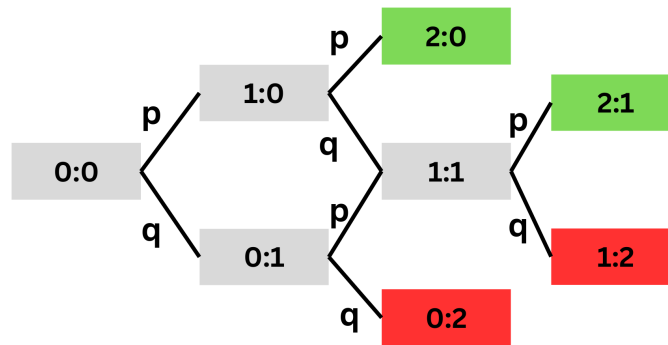


Abbildung 4.3: Spielverlauf eines Matches mit Best-Of-3 Modus

Anhand der Grafik kann erkannt werden, dass es auf insgesamt sechs verschiedene Weisen verlaufen kann. Weiters ist es bei zwei Spielverläufen der Fall, dass das Match bereits nach zwei gespielten Sätzen entschieden ist. Die Matchgewinnwahrscheinlichkeit in Abhängigkeit von der konstanten Satzgewinnwahrscheinlichkeit p kann folgendermaßen ermittelt werden:

$$\begin{aligned} P(\text{Matchgewinn}) &= p^2 + p \cdot q \cdot p + q \cdot p \cdot p \\ &= p^2 \cdot (1 + 2q) \end{aligned}$$

Wird $q = 1 - p$ eingesetzt, so erhält man für die Matchgewinnwahrscheinlichkeit:

$$\begin{aligned} P(\text{Matchgewinn}) &= p^2 \cdot [1 + 2 \cdot (1 - p)] \\ &= p^2 \cdot (3 - 2p) \end{aligned}$$

In der folgenden Tabelle 4.3 sind die Matchgewinnwahrscheinlichkeiten für bestimmte Satzgewinnwahrscheinlichkeiten angegeben.

4 Modellierung eines Tennismatches

Satzgewinnwahrscheinlichkeit	Matchgewinnwahrscheinlichkeit
0,30	0,216
0,35	0,282
0,40	0,352
0,45	0,425
0,50	0,500
0,55	0,575
0,60	0,648
0,65	0,718
0,70	0,784
0,75	0,844
0,80	0,896

Tabelle 4.2: Matchgewinnwahrscheinlichkeit im Best-of-3 Modus

Anhand dieser Tabelle ist ersichtlich, dass für $p > 0,5$ die Wahrscheinlichkeit dieses Match zu gewinnen größer als die Satzgewinnwahrscheinlichkeit selbst ist. Das bedeutet, dass beispielsweise ein*e Spieler*in einen Satz mit einer Wahrscheinlichkeit von 0,75 aber das Match mit einer Wahrscheinlichkeit von 0,844 gewinnt. Somit profitiert der*die bessere Spieler*in von dem Best-of-3 Modus, da diese die Siegeschancen des*der stärkeren Tennisspieler*in vergrößert.

Bei den vier Grand-Slam Turnieren wird im Herren Einzel im Best-of-5 Modus gespielt. Das bedeutet, dass drei Sätze gewonnen werden müssen um das Match für sich entscheiden zu können. In der nachstehenden Abbildung 4.4 sind die möglichen Spielverläufe dargestellt. Es kann erkannt werden, dass es zwei Verläufe gibt bei welchen das Match bereits in drei Sätzen entschieden ist.

4.3 Matchgewinnwahrscheinlichkeit in Abhängigkeit von der Satzgewinnwahrscheinlichkeit

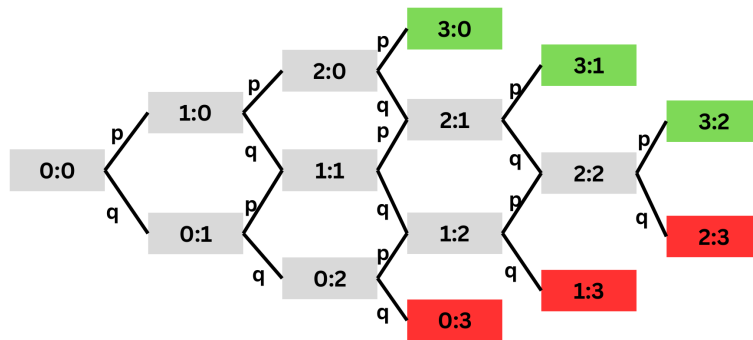


Abbildung 4.4: Spielverlauf eines Matches mit Best-of-5 Modus

Die Matchgewinnwahrscheinlichkeit in Abhängigkeit von der konstanten Satzgewinnwahrscheinlichkeit p im Best-of-5 Modus kann folgendermaßen berechnet werden:

$$\begin{aligned} P(\text{Matchgewinn}) &= p^3 + 3 \cdot p^3 \cdot q + 6 \cdot p^3 \cdot q^2 \\ &= p^3 \cdot (1 + 3p + 6q^2) \end{aligned}$$

Wird $q = 1 - p$ eingesetzt so erhält man für die Matchgewinnwahrscheinlichkeit:

$$\begin{aligned} P(\text{Matchgewinn}) &= p^3 \cdot [1 + 3 \cdot (1 - p) + 6 \cdot (1 - p)^2] \\ &= p^3 \cdot (1 + 3 - 3p + 6 - 12p + 6p^2) \\ &= p^3 \cdot (6p^2 - 15p + 10) \end{aligned}$$

In der nachstehenden Tabelle 4.3 sind Matchgewinnwahrscheinlichkeiten für bestimmte Satzgewinnwahrscheinlichkeiten bei einem Best-of-5 Modus angegeben.

4 Modellierung eines Tennismatches

Satzgewinnwahrscheinlichkeit	Matchgewinnwahrscheinlichkeit
0,30	0,163
0,35	0,235
0,40	0,317
0,45	0,407
0,50	0,500
0,55	0,593
0,60	0,683
0,65	0,764
0,70	0,837
0,75	0,896
0,80	0,942

Tabelle 4.3: Matchgewinnwahrscheinlichkeit im Best-Of-5 Modus

Anhand dieser Tabelle und dem Vergleich mit der Tabelle 4.3 für die Matchgewinnwahrscheinlichkeiten im Best-of-3 Modus ist ersichtlich, dass das Spielen auf drei gewonnene Sätze dem*der besseren Spieler*in noch mehr entgegenkommt, da die Siegeschancen weiter vergrößert werden. So gewinnt beispielsweise ein*e Spieler*in einen Satz mit einer Wahrscheinlichkeit von 0,75, das Match im Best-of-3 Modus mit einer Wahrscheinlichkeit von 0,844 und ein Match im Best-of-5 Modus sogar mit einer Wahrscheinlichkeit von 0,896.

4.4 Gewinnwahrscheinlichkeit eines Tie-Breaks

Die nachstehende Erläuterung folgt, falls nicht anders gekennzeichnet, dem Werk von Stewart [21, S.38-41].

Wie in Kapitel 1.2 erwähnt kann es bei einem Tennismatch auch zu einem sogenannten Tie-Break kommen. Dieser Fall tritt bei einem Spielstand von 6:6 ein, das heißt, wenn beide Spieler*innen sechs Games gewonnen haben. Um dieses Tie-Break zu gewinnen müssen sieben Punkte mit einem Unterschied von mindestens zwei Punkten erzielt werden. Im Gegensatz zu Games wird beim Tie-Break mit

4.4 Gewinnwahrscheinlichkeit eines Tie-Breaks

Ziffern gezählt. [19, S.164].

Angenommen Spieler*in A gewinnt einen Punkt mit der konstanten Wahrscheinlichkeit p und Spieler*in B mit der Wahrscheinlichkeit $q = 1 - p$. Das nachstehende Diagramm 4.6 veranschaulicht die möglichen Verläufe eines Tie-Breaks.

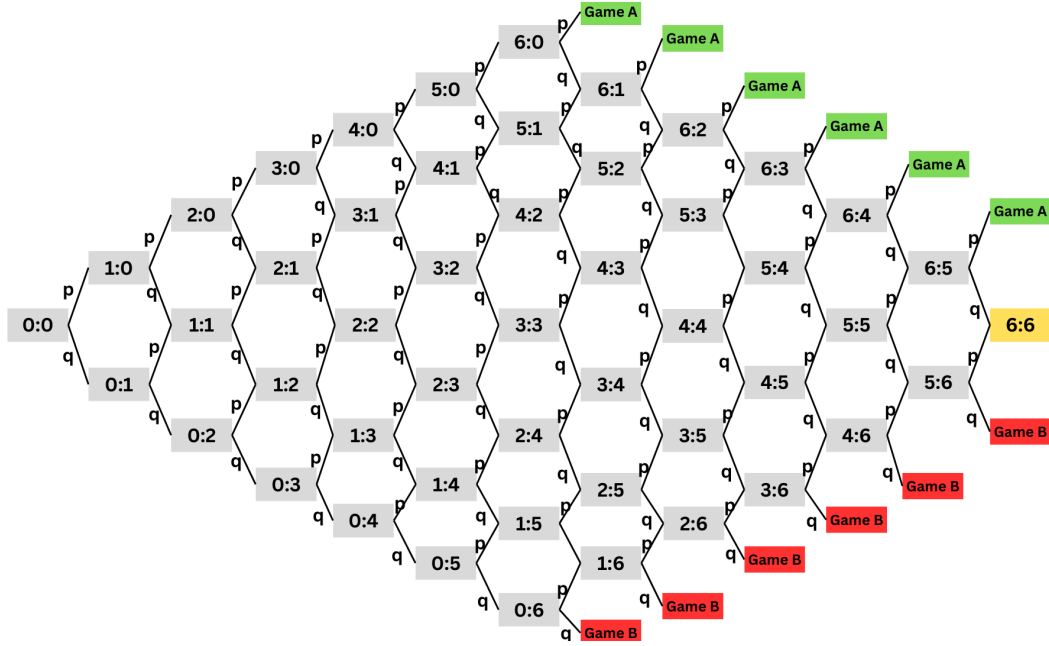


Abbildung 4.5: Spielverlauf eines Tie-Breaks

Mit Hilfe dieser Abbildung kann abgelesen werden, dass die Wahrscheinlichkeit ein Tie-Break vor einem Einstand gewonnen wird, folgendermaßen berechnet werden kann:

$$\begin{aligned} P(\text{Game vor Einst.}) &= p^7 + \binom{7}{1} p^7 q + \binom{8}{2} p^7 q^2 + \binom{9}{3} p^7 q^3 + \binom{10}{4} p^7 q^4 + \binom{11}{5} p^7 q^5 \\ &= p^7 + 7p^7 q + 28p^7 q^2 + 84p^7 q^3 + 210p^7 q^4 + 462p^7 q^5 \end{aligned}$$

Diese Gewinnwahrscheinlichkeit kann verkürzt auch mit Hilfe der Summenschreibweise angegeben werden:

4 Modellierung eines Tennismatches

$$P(\text{Game A vor Einstand}) = \sum_{i=0}^5 \binom{6+i}{i} \cdot p^7 \cdot q^i$$

Analog zur Gewinnwahrscheinlichkeit eines Games handelt es sich bei dem Stand 6:6 um einen Einstand. Die möglichen Spielverläufe sind analog zu jenen eines Games in der nachstehenden Abbildung dargestellt:

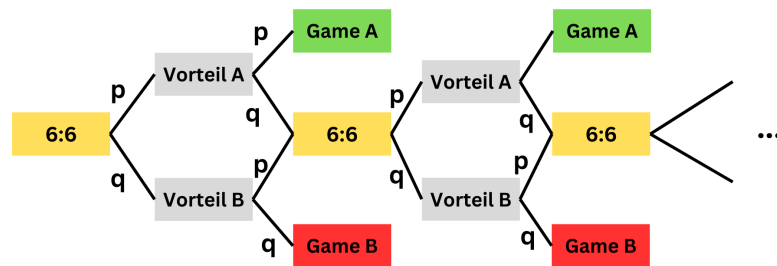


Abbildung 4.6: Spielverlauf eines Einstands im Tie-Break

Anhand dieser Abbildung kann erkannt werden, dass die Wahrscheinlichkeit, dass es zu einem Einstand kommt folgendermaßen ermittelt werden kann:

$$P(\text{Einstand}) = \binom{12}{6} \cdot p^6 \cdot q^6$$

Da ein Einstand theoretisch unendlich oft auftreten kann, erhält man für die Gewinnwahrscheinlichkeit eines Tie-Breaks nach Einstand folgende unendliche Reihe:

4.4 Gewinnwahrscheinlichkeit eines Tie-Breaks

$$\begin{aligned}
 P(\text{Game nach Einstand}) &= \binom{12}{6} p^6 q^6 \cdot p^2 + \binom{12}{6} p^6 q^6 \cdot 2pq \cdot p^2 \\
 &\quad + \binom{12}{6} p^6 q^6 \cdot 4p^2 q^2 \cdot p^2 + \dots \\
 &= \binom{12}{6} p^8 q^6 + \binom{12}{6} p^9 q^7 \cdot 2 + \binom{12}{6} p^{10} q^8 \cdot 4 + \dots \\
 &= \binom{12}{6} p^8 q^6 \cdot [1 + 2pq + 4p^2 q^2 + \dots] \\
 &= \binom{12}{6} p^8 q^6 \cdot [1 + 2pq + (2pq)^2 + (2pq)^3 + \dots]
 \end{aligned}$$

Werden nun die Wahrscheinlichkeiten für einen Gewinn vor einem Einstand sowie nach einem Einstand summiert, so erhält man für die Gewinnwahrscheinlichkeit eines Tie-Breaks folgenden Ausdruck mit einer geometrischen Reihe:

$$\begin{aligned}
 P(\text{Tie Break}) &= \sum_{i=0}^5 \binom{6+i}{i} \cdot p^7 \cdot q^i + \binom{12}{6} p^8 q^6 \cdot [1 + 2pq + (2pq)^2 + (2pq)^3 + \dots] \\
 &= p^7 + 7p^7 q + 28p^7 q^2 + 84p^7 q^3 + 210p^7 q^4 + 462p^7 q^5 \\
 &\quad + 924p^8 q^6 \cdot [1 + 2pq + \dots] \\
 &= p^7 + 7p^7 q + 28p^7 q^2 + 84p^7 q^3 + 210p^7 q^4 + 462p^7 q^5 \cdot [1 + 2pq + \dots]
 \end{aligned}$$

Da es sich bei dem Ausdruck in der Klammer um dieselbe geometrische Reihe wie zuvor mit $|2pq| < 1$ handelt, kann folgende Summenformel der geometrischen Reihe angewendet werden:

$$\sum_{k=0}^{\infty} (2pq)^k = 1 + 2pq + (2pq)^2 + \dots = \frac{1}{1 - 2pq}$$

Somit kann die Gewinnwahrscheinlichkeit eines Tie-Breaks bei konstanter Punktgewinnwahrscheinlichkeit p folgendermaßen berechnet werden:

$$\begin{aligned} P(\text{Game Tie Break}) &= \sum_{i=0}^5 \binom{6+i}{i} \cdot p^7 \cdot q^i + \binom{12}{6} p^8 q^6 \cdot \frac{1}{1-2pq} \\ &= p^7 + 7p^7 q + 28p^7 q^2 + 84p^7 q^3 + 210p^7 q^4 + \frac{462p^7 q^5}{1-2pq} \end{aligned}$$

4.5 Gewinnwahrscheinlichkeit beim Aufschlag

Die Ausarbeitung dieses Kapitels folgt bis auf weiteres dem Werk von Havil [11, S.9-15].

Das in dem Kapitel 1.2 erklärte Zählsystem im Tennis mag für Unerfahrene nicht nur verwirrend wirken, sondern weist auch eine gewisse Anomalie auf. So ist es beispielsweise der Fall, dass ein*e sehr gute*r Tennisspieler*in, welche*r mit einer Wahrscheinlichkeit von nahezu 1 als Aufschläger*in auch den Punkt gewinnt, bei einem Stand von 40:30 und 30:15 eine geringere Chance hat das Game zu gewinnen als zu Beginn bei einem Stand von 0:0. Um diesen Sachverhalt nachweisen zu können sei p die konstante Wahrscheinlichkeit, dass ein*e Spieler*in als Aufschläger*in einen Punkt gewinnt. Demzufolge ist $q = 1 - p$ die Wahrscheinlichkeit, dass diese*r Spieler*in den Punkt verliert. Sei weiters $P(a, b)$ die Wahrscheinlichkeit, dass der*die Spieler*in als Aufschläger*in das Spiel gewinnt, wobei er*sie a Punkte und sein*ihre Gegner*in b Punkte erzielt hat. Um diese Aussage nachweisen zu können müssen zuerst $P(40, 30)$ sowie $P(30, 15)$ ermittelt werden und diese anschließend mit $P(0, 0)$ verglichen werden.

Wahrscheinlichkeit $P(40, 30)$:

Zu Beginn ist es wichtig anzumerken, dass der Spielstand 'Vorteil A' jenem vom 40:30 und 'Vorteil B' jenem von 30:40 entspricht. Somit gilt für die Situation des Einstandes folgender Zusammenhang, wenn dieser Fall in den Gewinn sowie Verlust des folgenden Punktes unterteilt wird:

4.5 Gewinnwahrscheinlichkeit beim Aufschlag

$$\begin{aligned}P(40, 40) &= p \cdot \text{Vorteil A} + q \cdot \text{Vorteil B} \\&= p \cdot P(40, 30) + q \cdot P(30, 40)\end{aligned}$$

Durch Anwendung der Vorgehensweise des Aufteilens in den Gewinn und Verlust des nächsten Punktes und des anschließenden Multiplizierens mit der Wahrscheinlichkeit, dass der Satz mit dem jeweils neuen Stand gewonnen wird, erhält man:

$$P(30, 40) = p \cdot P(40, 40) \quad P(40, 30) = p + q \cdot P(40, 40)$$

Durch Einsetzen der beiden Zusammenhänge in die obige Gleichung ergibt sich für $P(40, 40)$:

$$\begin{aligned}P(40, 40) &= p \cdot (p + q \cdot P(40, 40)) + q \cdot (p \cdot P(40, 40)) \\&\Leftrightarrow P(40, 40) = p^2 + 2pq \cdot P(40, 40) \\&\Leftrightarrow P(40, 40) \cdot (1 - 2pq) = p^2 \\&\Leftrightarrow P(40, 40) = \frac{p^2}{1 - 2pq}\end{aligned}$$

Aufgrund der Tatsache, dass $p = 1 - q$ und somit $1 = p + q$, gilt für den Nenner:

$$1 - 2 \cdot p \cdot q = (p + q)^2 - 2 \cdot p \cdot q = p^2 + q^2$$

Durch Einsetzen erhält man:

$$P(40, 40) = \frac{p^2}{p^2 + q^2}$$

Setzt man diesen Ausdruck in die Gleichungen für $P(30, 40)$ sowie $P(40, 30)$ ein,

4 Modellierung eines Tennismatches

so liefert dies:

$$\begin{aligned} P(30, 40) &= p \cdot P(40, 40) \\ &= \frac{p^3}{p^2 + q^2} \end{aligned}$$

$$\begin{aligned} P(40, 30) &= p + q \cdot P(40, 40) \\ &= p + \frac{p^2 \cdot q}{p^2 + q^2} \end{aligned}$$

Dieser Ausdruck für $P(40, 30)$ wird in weiterer Folge mit jenen von $P(30, 15)$ sowie $P(0, 0)$ verglichen, um den zu untersuchenden Sachverhalt nachzuweisen.

Wahrscheinlichkeit $P(30,15)$:

Für die Ermittlung der Wahrscheinlichkeit $P(30, 15)$ ist es hilfreich die möglichen Spielverläufe mit Hilfe eines Diagramms darzustellen. Da es sich bei $P(30, 15)$ um die Wahrscheinlichkeit handelt, dass der*die Aufschläger*in das Spiel bei einem aktuellen Stand von 30:15 gewinnt, beginnt das Diagramm bei diesem Spielstand. Die nachstehende Abbildung 4.7 veranschaulicht die möglichen Verläufe.

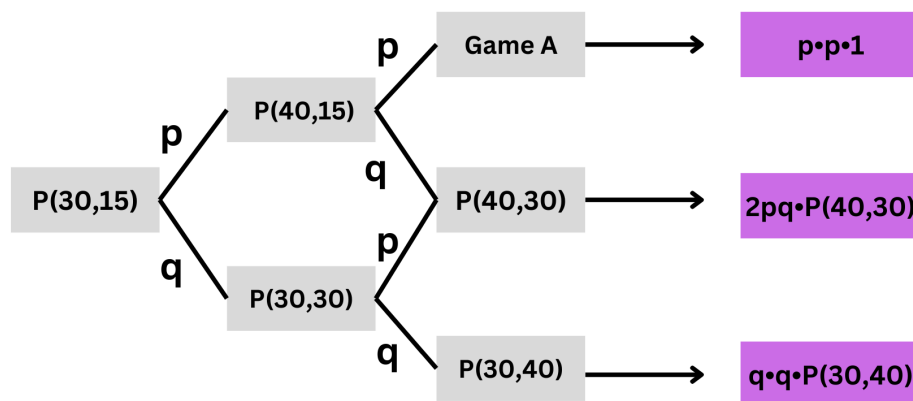


Abbildung 4.7: Spielverlauf für den Fall $P(30,15)$

4.5 Gewinnwahrscheinlichkeit beim Aufschlag

Mit Hilfe dieses Baumdiagramms und dem Einsetzen der zuvor berechneten Wahrscheinlichkeiten kann die Wahrscheinlichkeit $P(30, 15)$ folgendermaßen ermittelt werden:

$$\begin{aligned} P(30, 15) &= p^2 + 2pq \cdot P(40, 30) + q^2 \cdot P(30, 40) \\ &= p^2 + 2pq \cdot \left(p + \frac{p^2 \cdot q}{p^2 + q^2} \right) + q^2 \left(\frac{p^3}{p^2 + q^2} \right) \\ &= p^2 + 2p^2q + \frac{2p^3q^2}{p^2 + q^2} + \frac{p^3q^2}{p^2 + q^2} \\ &= p^2 \cdot (1 + 2q) + \frac{3p^3q^2}{p^2 + q^2} \end{aligned}$$

Wahrscheinlichkeit $P(0,0)$:

Die Anfangswahrscheinlichkeit $P(0, 0)$ wurde bereits im Kapitel 4.1 hergeleitet und festgelegt durch:

$$P(0, 0) = p^4 + 4p^4q + \frac{10p^4q^2}{1 - 2pq}$$

Durch Umformungen und der aus dem Kapitel 4.1 bereits bekannten Eigenschaft, dass es sich um eine geometrische Reihe mit $|2pq| < 1$ handelt, erhält man folgenden äquivalenten Ausdruck für die Wahrscheinlichkeit $P(0, 0)$:

$$P(0, 0) = p^4 \cdot (1 + 4q + 10q^2) + \frac{20p^5q^3}{1 - 2pq}$$

Aufgrund der Tatsache, dass $p = 1 - q$ und somit $1 = p + q$, gilt für den Nenner:

$$1 - 2 \cdot p \cdot q = (p + q)^2 - 2 \cdot p \cdot q = p^2 + q^2$$

Durch Einsetzen erhält man:

4 Modellierung eines Tennismatches

$$P(0,0) = p^4 \cdot (1 + 4q + 10q^2) + \frac{20p^5q^3}{p^2 + q^2}$$

Nachdem nun die drei Wahrscheinlichkeiten $P(0,0)$, $P(30,15)$ sowie $P(40,30)$ ermittelt wurden, gilt es diese zu vergleichen um den zu untersuchenden Sachverhalt nachzuweisen. In den untenstehenden Abbildungen 4.8, 4.9 und 4.10 sind die Graphen der jeweiligen Wahrscheinlichkeiten in Abhängigkeit von p dargestellt.

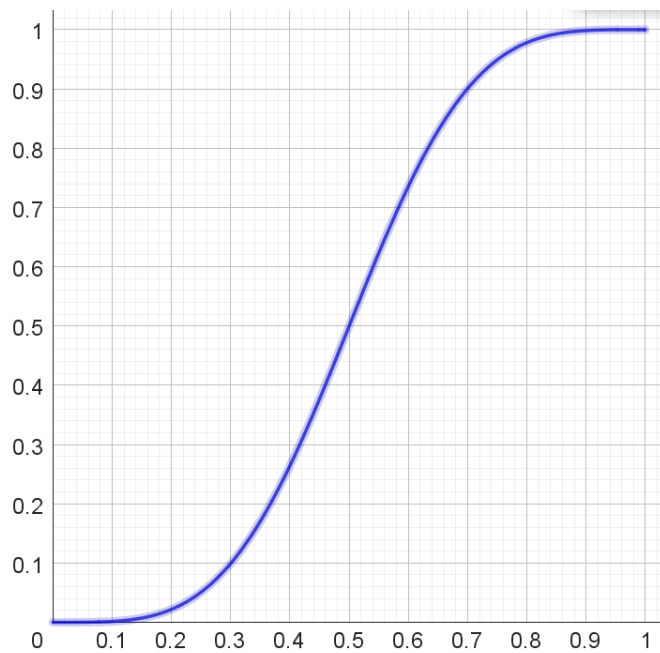


Abbildung 4.8: $P(0,0)$ in Abhängigkeit von p

4.5 Gewinnwahrscheinlichkeit beim Aufschlag

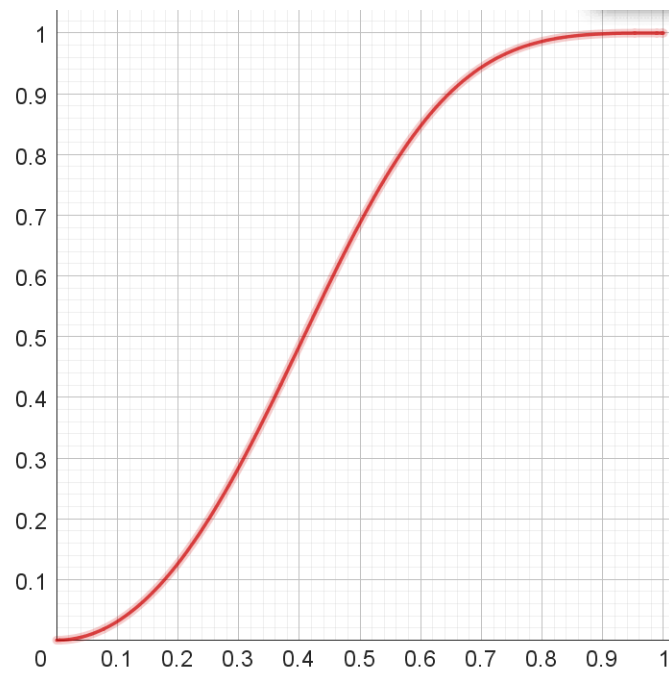


Abbildung 4.9: $P(30, 15)$ in Abhängigkeit von p

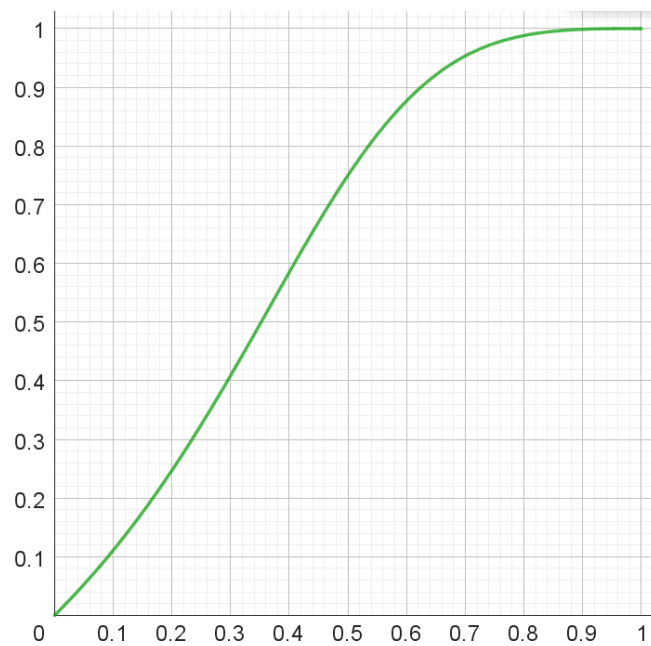


Abbildung 4.10: $P(40, 30)$ in Abhängigkeit von p

4 Modellierung eines Tennismatches

Schnittpunkt von $P(0,0)$ und $P(30,15)$: Für die Berechnung des Schnittpunktes von $P(0,0)$ und $P(30,15)$ muss folgende Gleichung gelöst werden:

$$p^4 \cdot (1 + 4q + 10q12) + \frac{20p^5 \cdot q^3}{p^2 + q^2} = p^2 \cdot (1 + 2q) + \frac{3p^3 \cdot q^2}{p^2 + q^2}$$

Der Einsatz einer mathematischen Software zum Lösen dieser Gleichung führt neben den beiden trivialen Lösungen $p = 0$ und $p = 1$ zu folgenden Lösungen:

$$p_1 = \frac{1}{4} \cdot (1 + \sqrt{7}) \quad p_2 = \frac{1}{4} \cdot (1 - \sqrt{7})$$

Die einzige positive und somit relevante Lösung in diesem Kontext ist

$p_1 = \frac{1}{4} \cdot (1 + \sqrt{7}) \approx 0,911438$. Somit wurde nachgewiesen, dass für $p > 0,911438$ gilt, dass $P(0,0) > P(30,15)$.

Schnittpunkt von $P(0,0)$ und $P(40,30)$: Für die Berechnung des Schnittpunktes von $P(0,0)$ und $P(40,30)$ ist es nun notwendig folgende Gleichung zu lösen:

$$p^4 \cdot (1 + 4q + 10q12) + \frac{20p^5 \cdot q^3}{p^2 + q^2} = p + \frac{p^2 \cdot q}{p^2 + q^2}$$

Durch erneutes Einsetzen einer mathematischen Software erhält man neben den beiden trivialen Lösungen $p = 0$ und $p = 1$ folgende Lösung für p :

$$p = \frac{1}{6} + \frac{1}{24} \cdot \sqrt[3]{1216 - 192 \cdot \sqrt{33}} + \frac{1}{6} \cdot \sqrt[3]{19 + 3 \cdot \sqrt{33}} \approx 0,919643$$

Somit wurde nachgewiesen, dass für jedes $p > 0,919643$ gilt, dass

$P(0,0) > P(40,30)$.

Zusammenfassend lässt sich somit sagen, dass wenn zwei gleichstarke Tennisspieler*innen aufeinandertreffen und der*die Aufschläger*in mit einer Wahrscheinlichkeit von $p > 0,92$ den Punkt gewinnt, diese*r zu Beginn des Games eine höhere Gewinnchance hat, als bei einem Spielstand von 30:15 bzw. 40:30.

5 Auswertung aller bisherigen US-Open Finalsspiele

Im folgenden Kapitel werden alle bisherigen Finalsspiele der US-Open im Damen- sowie Herreneinzel analysiert und bezüglich der erwarteten Anzahl der gespielten Sätze betrachtet. Mit Hilfe der Anwendung des Chi-Quadrat-Anpassungstests soll somit untersucht werden, ob die theoretischen Werte mit der Realität übereinstimmen und ob angenommen werden kann, dass im Finale zwei Spieler*innen mit der gleichen Spielstärke aufeinandertreffen. Alle Spielergebnisse wurden der Seite der US Open [22] entnommen.

Das Tennisturnier, welches heutzutage in New York ausgetragen wird, wurde 1881 unter dem Namen „US National Championships“ gegründet. Zu Beginn wurde noch in Newport auf Rhode Island gespielt, bevor 1915 das Turnier nach New York umzog. Während seit 1881 jedes Jahr das Turnier des Herreneinzels durchgeführt wurde, durften Frauen erst 1887 an dem Turnier teilnehmen [15]. Somit gab es im Herreneinzel bisher 143 ausgespielte Finalentscheidungen, während es bei den Damen nur 135 gab, da es bei zwei Finalspielen zu einer Aufgabe einer Finalistin kam.

Da es sich bei den vorliegenden Daten um die Finalsspiele der US Open handelt, wird angenommen, dass die beiden Spieler*innen jeweils mit einer Wahrscheinlichkeit $p = \frac{1}{2}$ einen Satz für sich entscheiden. Während ein Herreneinzel im Best-of-5 Modus gespielt wird, spielen die Frauen wie gewohnt im Best-of-3 Modus, mit den Ausnahmen von 1891, 1892 und 1894 bis 1901. Diese Ausnahmen waren und sind bis heute einmalig und wurden bis auf diese Zeitspannen bis dato nicht mehr wiederholt. Die nachstehende Tabelle 5.1 gibt einen Überblick darüber, wie viele

5 Auswertung aller bisherigen US-Open Finalsplele

Matches der Damen und Herren in zwei, drei, vier bzw. fünf Sätzen entschieden wurden.

Anzahl Sätze	Herren 1881 - 2023	Damen best-of-3	Damen best-of-5
2	/	88	0
3	60	37	3
4	53	0	2
5	30	0	5

Tabelle 5.1: Anzahl Sätze US Open Finale [22]

5.1 Erwartete Anzahl gespielter Sätze im Herreneinzel

Im folgenden Abschnitt werden die Ergebnisse der Herren betrachtet und analysiert. Unter der Annahme, dass die beiden Finalisten die gleiche Spielstärke aufweisen und somit mit einer Wahrscheinlichkeit von $p = 0,5$ einen Satz gewinnen und der Tatsache, dass das Herreneinzel stets im Best-of-5 Modus gespielt wird, können die Wahrscheinlichkeiten für einen Matchgewinn nach drei, vier bzw. fünf Sätzen ermittelt werden. Im Kapitel 4.3 wurden bereits die jeweiligen Gewinnwahrscheinlichkeiten für einen Spieler hergeleitet. Durch Einsetzen der Werte $p = \frac{1}{2}$ und $q = \frac{1}{2}$ sowie Multiplikation mit Zwei, erhält man folgende Wahrscheinlichkeiten:

$$P(\text{genau 3 Sätze}) = 2 \cdot \left(\frac{1}{2}\right)^3 = \frac{1}{4}$$

$$P(\text{genau 4 Sätze}) = 6 \cdot \left(\frac{1}{2}\right)^4 = \frac{3}{8}$$

$$P(\text{genau 5 Sätze}) = 12 \cdot \left(\frac{1}{2}\right)^5 = \frac{3}{8}$$

Mit Hilfe dieser Wahrscheinlichkeiten kann nun die theoretisch berechnete Häufig-

5.1 Erwartete Anzahl gespielter Sätze im Herreneinzel

keit an Matches, welche in drei, vier oder fünf Sätzen entschieden wurden, ermittelt werden. Die nachstehende Tabelle 5.2 fasst dies für den Fall der US Open mit 143 Entscheidungen im Herreneinzel zusammen.

Sätze	Wahrscheinlichk. p_k	berechnete Häufigk. np_k	tatsächliche Häufigk. n_k
3	0,25	35,75	60
4	0,375	53,625	53
5	0,375	53,625	30

Tabelle 5.2: Anzahl Sätze US Open Finale Herreneinzel

Mit Hilfe des Chi-Quadrat-Anpassungstest soll nun überprüft werden, ob mit dem theoretischen Ansatz eine aussagekräftige Modellierung vorliegt. Bei diesem Experiment handelt es sich um einen Stichprobenumfang von $n = 143$ mit $r = 3$ verschiedenen Ereignisse mit folgenden relativen Häufigkeiten:

$$h_1 = \frac{60}{143}, \quad h_2 = \frac{53}{143}, \quad h_3 = \frac{30}{143}$$

Die Hypothese dieses Tests lautet:

$$H_0 : p_1 = \frac{1}{4}, \quad p_2 = \frac{3}{8}, \quad p_3 = \frac{3}{8}$$

Für die Berechnung des Stichprobenwertes χ^2 folgt durch Einsetzen in die Formel, dass

$$\begin{aligned} \chi^2(x) &= \sum_{k=1}^3 \frac{(n_k - np_k)^2}{np_k} \\ &= \frac{(60 - 35,75)^2}{35,75} + \frac{(53 - 53,625)^2}{53,625} + \frac{(30 - 53,625)^2}{53,625} \\ &\approx 23,7675 \end{aligned}$$

Für die Irrtumswahrscheinlichkeit wird $\alpha = 0,05$ vorausgesetzt. Durch Ablesen

5 Auswertung aller bisherigen US-Open Finalsspiele

aus der Tabelle für die χ^2 -Verteilung für die Gleichung $F(c) = 1 - \alpha = 0,95$ mit $r - 1 = 2$ Freiheitsgraden folgt für den Ablehnungsbereich:

$$\mathbf{B} =]5,991; \infty[$$

Da sich der berechnete Stichprobenwert mit 23,7675 in dem Ablehnungsbereich befindet, muss die Hypothese verworfen werden. Somit kann erkannt werden, dass die beiden Finalisten meist nicht die gleiche Spielstärke aufweisen, und somit selbst im Finale nicht von einer Satzgewinnwahrscheinlichkeit von $p = 0,5$ ausgegangen werden kann. Mit Irrtumswahrscheinlichkeit 0,05 wurde bewiesen, dass die Satzgewinnwahrscheinlichkeit der Finalisten im Herreneinzel bei den US-Open nicht jeweils 0,5 ist.

5.2 Erwartete Anzahl gespielter Sätze im Dameneinzel

Analog zum Herreneinzel wird von der Annahme ausgegangen, dass die beiden Finalistinnen die gleiche Spielstärke aufweisen und somit mit einer Wahrscheinlichkeit von $p = 0,5$ einen Satz gewinnen. Im Gegensatz zu der Vorgehensweise bei dem Herreneinzel ist es nun notwendig die Daten der Frauen aufzuteilen, da im Zeitraum von 1891 bis 1892 sowie 1894 bis 1901 mit dem Best-of-5 Modus gespielt wurde, während die anderen Jahre der gewohnte Best-of-3 Modus zum Einsatz kam. Betrachtet man zunächst jene Jahre in welchen, analog zu den Männern, mit dem Best-of-5 Modus gespielt wurde, so ergibt sich folgende Tabelle 5.3 für jene zehn Finals:

Sätze	Wahrscheinlichk. p_k	berechnete Häufigk. np_k	tatsächliche Häufigk. n_k
3	0,25	2,5	3
4	0,375	3,75	2
5	0,375	3,75	5

Tabelle 5.3: Anzahl Sätze US Open Finale Damenfinale Best-of-5

5.2 Erwartete Anzahl gespielter Sätze im Dameneinzel

Bei diesem Experiment handelt es sich um einen Stichprobenumfang von $n = 10$ mit $r = 3$ verschiedenen Ereignissen mit folgenden relativen Häufigkeiten:

$$h_1 = \frac{3}{10}, \quad h_2 = \frac{2}{10}, \quad h_3 = \frac{5}{10}$$

Da überprüft werden soll, ob es sich dabei um eine ausreichend gute Modellierung handelt, lautet die Hypothese:

$$H_0 : p_1 = \frac{1}{4}, \quad p_2 = \frac{3}{8}, \quad p_3 = \frac{3}{8}$$

Durch Einsetzen in die Formel folgt für den Stichprobenwert χ^2 :

$$\begin{aligned} \chi^2(x) &= \sum_{k=1}^3 \frac{(n_k - np_k)^2}{np_k} \\ &= \frac{(3 - 2,5)^2}{2,5} + \frac{(2 - 3,75)^2}{3,75} + \frac{(5 - 3,75)^2}{3,75} \\ &= \frac{4}{3} \approx 1,3333 \end{aligned}$$

Für die Irrtumswahrscheinlichkeit wird erneut der Wert $\alpha = 0,05$ vorausgesetzt, weshalb bei gleichbleibenden Freiheitsgraden der Wert von zuvor übernommen werden kann. Somit ist auch hier der Verwerfungsbereich:

$$\mathbf{B} =]5,991; \infty[$$

In diesem Fall liegt der berechnete Stichprobenwert nicht im Ablehnungsbereich, folglich wird die Hypothese nicht verworfen. Um eine sinnvolle Aussage tätigen zu können werden nun noch jene Finalspiele betrachtet, welche im Best-of-3 Modus gespielt wurden.

Im Kapitel 4.3 wurden ebenfalls die jeweiligen Gewinnwahrscheinlichkeiten für einen Matchgewinn nach zwei bzw. drei Sätzen hergeleitet. Durch erneutes Ein-

5 Auswertung aller bisherigen US-Open Finalsplele

setzen der Werte $p = 0,5$ und $q = 0,5$ sowie Multiplizieren mit Zwei erhält man folgende Wahrscheinlichkeiten:

$$P(\text{genau 2 Sätze}) = 2 \cdot \left(\frac{1}{2}\right)^2 = \frac{1}{2}$$

$$P(\text{genau 3 Sätze}) = 4 \cdot \left(\frac{1}{2}\right)^3 = \frac{1}{2}$$

Mit Hilfe dieser Wahrscheinlichkeiten kann nun eine Tabelle mit der theoretisch berechneten Anzahl an Matches welche in zwei bzw. drei Sätzen gewonnen wurden, erstellt werden. Die nachstehende Tabelle 5.4 fasst dies für die 125 Finalsplele der US Open im Dameneinzel im Best-of-3 Modus zusammen.

Sätze	Wahrscheinlichk. p_k	berechnete Häufigk. np_k	tatsächliche Häufigk. n_k
2	0,5	62,5	88
3	0,5	62,5	37

Tabelle 5.4: Anzahl Sätze US Open Finale Damenfinale Best-of-3

Im Gegensatz zu dem Spielverlauf im Best-of-5 Modus handelt es sich mit $n = 125$ nun um ein Experiment mit einem wesentlich größeren Stichprobenumfang. Weiters gibt es hier nur $r = 2$ verschiedene Ereignisse. Die Hypothese die es zu überprüfen gilt lautet:

$$H_0 : p_1 = \frac{1}{2}, \quad p_2 = \frac{1}{2}$$

Durch erneutes Einsetzen in die Formel für den Stichprobenwert χ^2 folgt:

5.2 Erwartete Anzahl gespielter Sätze im Dameneinzel

$$\begin{aligned}\chi^2(x) &= \sum_{k=1}^2 \frac{(n_k - np_k)^2}{np_k} \\ &= \frac{(88 - 62,5)^2}{62,5} + \frac{(37 - 62,5)^2}{62,5} \\ &= 20,808\end{aligned}$$

Durch Ablesen des kritischen Wertes mit der Irrtumswahrscheinlichkeit $\alpha = 0,05$ und $r - 2 = 1$ Freiheitsgraden erhält man folgenden Ablehnungsbereich:

$$\mathbf{B} =]3,841; \infty[$$

Auch hier muss aufgrund des großen Stichprobenwerts χ^2 die Hypothese verworfen werden. Betrachtet man nun beide Fälle der Damenfinalspiele lässt sich sagen, dass nicht von der gleichen Spielstärke der beiden Finalistinnen ausgegangen werden kann. Da wesentlich mehr Matches bereits nach zwei Sätzen entschieden wurden, erkennt man, dass es meist eine dominierende Spielerin gab, die das Match in der Mindestanzahl von Sätzen für sich entscheiden konnte. Mit Irrtumswahrscheinlichkeit 0,05 wurde bewiesen, dass die Satzgewinnwahrscheinlichkeit der Finalistinnen im Dameneinzel bei den US-Open nicht jeweils 0,5 ist.

6 Tennissport im Mathematikunterricht

6.1 Realitätsbezug im Mathematikunterricht

Häufige Kritik an der Mathematik und vor allem dem Mathematikunterricht in der Sekundarstufe ist, dass die Mathematik viel zu abstrakt und realitätsfern sei. Dies hat dazu geführt, dass Kinder häufig Schwierigkeiten haben ihr erlerntes Wissen in praktischen Anwendungen umzusetzen bzw. die Verbindungen zur Realität herzustellen. Aus diesem Grund wird vermehrt versucht Realitätsbezüge im Mathematikunterricht einzubauen, um einerseits dieser Kritik entgegenzuwirken und andererseits die Verbindung von Mathematik und Alltagsgeschehen näherbringen zu können.

Im Allgemeinen wird zwischen zwei verschiedenen Arten der Einbindung von realitätsnahen Situationen unterschieden, nämlich den sogenannten „eingekleideten Aufgaben“ und den „Modellierungsaufgaben“. Bei den meisten Schulaufgaben handelt es sich um Textaufgaben bei welchen nicht das Lösen des Sachproblems im Vordergrund steht, sondern das Üben und Festigen des zuvor Erlernten. Diese Übungsaufgaben haben zum Vorteil, dass sie in kurzer Zeit das Übersetzen von Texten in die mathematische Sprache festigen. Im Gegensatz zu den eingekleideten Aufgaben haben Modellierungsaufgaben das Strukturieren und Analysieren von realitätsnahen Situationen im Fokus. Eine grafische Darstellung des Modellierungsprozesses ist in der nachstehenden Abbildung 6.1 ersichtlich [16, S.109].

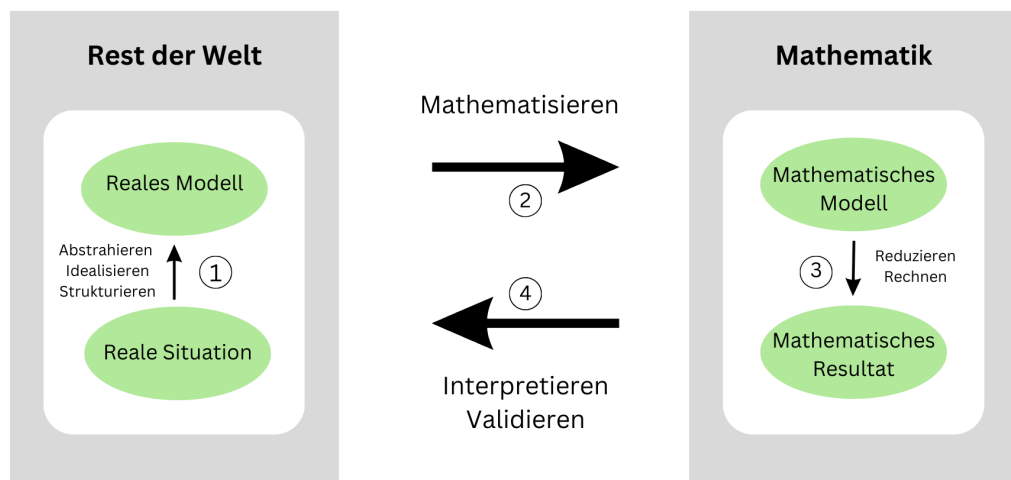


Abbildung 6.1: Modellierungsprozess

Während es viele verschiedene Darstellungen mit unterschiedlich vielen Zwischenschritten für den Modellierungsprozess gibt, so ist der Unterschied zwischen der Realität und der Mathematik stets Gegenstand eines jeden Prozesses [16, S.109].

6.2 Stochastik im Lehrplan allgemeinbildende höherer Schulen

Die allgemeinen Bildungsziele haben als Ziel die Förderung eines fächerübergreifenden und fächerverbindenden Unterrichts. So argumentiert der Lehrplan für allgemeinbildende höhere Schulen beispielsweise, dass sowohl fachspezifische als auch fächerübergreifende Bezüge essentiell für das integrative Denken junger Menschen sind [4, S.9].

6.2.1 Sekundarstufe 1

In allgemeinbildenden höheren Schulen (AHS) kommt das Kompetenzmodell zu tragen, welches die Themengebiete der Mathematik in der Sekundarstufe 1 in vier inhaltliche Kompetenzbereiche unterteilt, nämlich:

1. *Zahlen und Maße*
2. *Variablen und Funktionen*
3. *Figuren und Körper*
4. *Daten und Zufall* [4, S.83]

Stochastik und somit die beschreibende Statistik sowie der Wahrscheinlichkeitsbegriff fallen in den Kompetenzbereich der Daten und Zufälle.

Folgende Inhalte werden in der **1. Klasse** in Bezug auf Daten und Zufall gelehrt:

- *Daten erheben, ordnen, darstellen und aus unterschiedlichen Darstellungsformen ablesen*
- *einfache statistische Kennzahlen ermitteln und interpretieren* [4, S.85]

In der **2. Klasse** erweitern sich die Kompetenzen durch:

- *relative Häufigkeiten ermitteln, grafisch darstellen und grafische Darstellungen interpretieren* [4, S.86]

In der **3. Klasse** ist erstmals der Wahrscheinlichkeitsbegriff im Lehrplan verankert. Nach Absolvierung der 3. Klasse sollen die Schüler*innen

- *statistische Darstellungen erstellen und nutzen; Manipulationen in statistischen Darstellungen erkennen*
- *aufbauend auf einem intuitiven Wahrscheinlichkeitsbegriff Wahrscheinlichkeiten in einfachen Zufallsexperimenten ermitteln, vergleichen und interpretieren* [4, S.86]

Die **4.Klasse** in einer AHS befasst sich in diesem Kompetenzbereich mit

- *Kreuztabellen erstellen und interpretieren*
- *Wahrscheinlichkeiten bei ein- und zweistufigen Zufallsexperimenten ermitteln und interpretieren [4, S.87]*

6.2.2 Sekundarstufe 2

Der Lehrstoff, welcher in der Unterstufe gelehrt wird, fokussiert sich stark auf beschreibende Statistik und hat relativ wenig mit jenen Bereichen der Stochastik zu tun, welche in der Oberstufe behandelt werden. Mit Ausnahme der 5. Klasse (9. Schulstufe) ist die Stochastik in jedem Jahr im Lehrplan verankert und wird dementsprechend gelehrt. Der Lehrplan der AHS Oberstufe ist ein semestrierter und kompetenzorientierter Lehrplan.

Nach dem **4. Semester**, also der 6. Klasse (10. Schulstufe) sollen die Schüler*innen folgende Kompetenzen erreicht haben:

- *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 [...]*
- *Den Zusammenhang zwischen relativen Häufigkeiten und Wahrscheinlichkeiten kennen*
- *Mit Wahrscheinlichkeiten rechnen können [...]*
- *Bedingte Wahrscheinlichkeiten und (stochastische) Unabhängigkeiten [...] kennen*
- *Satz von Bayes kennen und anwenden können [4, S.216]*

6.3 Aufgaben zum Thema Stochastik im Tennis

Das **6. Semester** befasst sich mit den Kompetenzen der diskreten Wahrscheinlichkeitsverteilung

- *Die Begriffe diskreten Zufallsvariable und diskreten Wahrscheinlichkeitsverteilung kennen*
- *Den Zusammenhang zwischen relativen Häufigkeiten und Wahrscheinlichkeiten kennen*
- *Erwartungswert, Varianz und Standardabweichung einer diskreten Zufallsvariablen [...] kennen und deuten können*
- *Den Binomialkoeffizienten und seine wichtigsten Eigenschaften kennen*
- *Mit diskreten Verteilungen [...] in anwendungsorientierten Bereichen arbeiten können [4, S.217]*

Im **7. Semester** sollen die Schüler*innen

- *Die Begriffe stetige Zufallsvariable und „stetige Verteilung“ kennen*
- *Die Normalverteilung zur Approximation der Binomialverteilung einsetzen können*
- *Die Normalverteilung in anwendungsorientierten Bereichen verwenden können*
- *Konfidenzintervalle ermitteln und interpretieren können*
- *Einfache statistische Hypothesentests durchführen und deren Ergebnisse interpretieren können [4, S.218]*

6.3 Aufgaben zum Thema Stochastik im Tennis

Aufgrund der anhaltenden Kritik, dass die Mathematik realitätsfern sei und dem Wunsch nach einem fächerübergreifenden Arbeiten, befasst sich das nachstehende Kapitel mit möglichen Aufgaben zum Thema Stochastik im Tennis. Ziel dieser

Aufgaben ist es, den Kindern die Verbindung und vor allem die Bedeutung von Mathematik im Sport näherzubringen. Ein paar dieser Aufgaben könnten beispielsweise in einem fächerübergreifenden Projekt mit dem Sportunterricht durchgeführt werden.

Aufgabe 1: Novak Djokovic und Rafael Nadal sind zwei weltberühmte Tennisspieler. Seit ihrer ersten Begegnung 2006 sind die beiden Sportler mittlerweile insgesamt 59 mal in einem Tennismatch aufeinandergetroffen. Insgesamt konnte Novak Djokovic bis dato 30 Siege einspielen.

1. Mit welcher Wahrscheinlichkeit Gewinnt Novak Djokovic ein Match gegen Rafael Nadal?
2. Angenommen die beiden Sportler behalten ihr spielerisches Niveau bei. Wie viele Siege könnte Rafael Nadal erwarten, wenn er noch 15 mal gegen Novak Djokovic spielt? Diskutiert, inwiefern solche Vorhersagen zuverlässig sind.
3. Recherchiere die aktuelle Siegesquote von Novak Djokovic und vergleiche sie mit jener gegen Nadal. Wie unterscheiden sie sich? Welche Gründe könnte es dafür geben?

Lösungserwartung:

1. $P(\text{Sieg Djokovic gegen Nadal}) = \frac{30}{59} \approx 0,5085$
2. Es gilt: $n = 15$ sowie $p = \frac{29}{59}$, dass Nadal gegen Djokovic gewinnt. Sei X die Anzahl der Siege von Nadal gegen Djokovic, so gilt:

$$E(X) = n \cdot p = 15 \cdot \frac{29}{59} = 7,37$$

Diese Berechnung würde vorhersagen, dass Rafael Nadal mit 7,37 Siegen bei 15 zukünftigen Spielen rechnen könnte. Solche Voraussagen können bei gleichbleibender Spielstärke beider Sportler einen ungefähren Einblick geben was erwartet werden kann. Da die Leistung von Sportlern ebenfalls von der

6.3 Aufgaben zum Thema Stochastik im Tennis

täglichen Verfassung, der Umgebung, dem Bodenbelag etc. beeinflusst werden kann, ist es wichtig zu beachten, dass es sich um eine reine Modellierung handelt.

3. Aktuell hält Novak Djokovic bei einer Siegesquote von 83,25%. Im Vergleich zu seiner Siegesquote gegen Rafael Nadal ist zu erkennen, dass diese wesentlich höher ist, woraus geschlossen werden kann, dass Rafael Nadal ein starker Gegner für Novak Djokovic ist und im Vergleich zu anderen Spielern mehr Spiele gegen ihn gewinnen konnte.

Aufgabe 2: Gemeinsam mit eurem*eurer Sportlehrer*in startet ihr als Klasse dieses Semester das Projekt „Tennis spielen leicht gemacht“. Ihr werdet über das gesamte Semester in eurem Sportunterricht Tenniseinheiten finden. In den letzten beiden Einheiten notiert ihr wie viele Aufschläge ihr insgesamt durchgeführt habt und wie viel von diesen ins Netz bzw. in das Feld gegangen sind.

1. Berechne die Trefferquote für jede*n Schüler*in sowie den Durchschnitt der Klasse
2. Stelle eure Ergebnisse in geeigneten Diagrammen grafisch dar.

Aufgabe 3: Gemeinsam mit eurem*eurer Sportlehrer*in startet ihr als Klasse dieses Semester das Projekt „Tennis spielen leicht gemacht“. Ihr werdet über das gesamte Semester in eurem Sportunterricht selbst Tenniseinheiten spielen. Außerdem werdet ihr die aktuelle Tennissaison verfolgen.

1. Suche dir zwei Sportler*innen der Top 50 der WTA oder ATP Liste und recherchiere Daten zu deren bisherigen Aufeinandertreffen
2. Modelliere mit Hilfe deiner Daten ein Tennismatch im Best-of-3 oder Best-of-5 Modus eurer ausgewählten Personen.

Tipp: Ermittle die Satzgewinnwahrscheinlichkeit jedes*jeder Sportler*in und modelliere anschließend die Matchgewinnwahrscheinlichkeit in Abhängigkeit der Satzgewinnwahrscheinlichkeit.

6 Tennissport im Mathematikunterricht

Lösungserwartung:

$$\begin{aligned}P(\text{Matchgewinn Best-of-3}) &= p^2 + p \cdot q \cdot p + q \cdot p \cdot p \\&= p^2 \cdot (1 + 2q)\end{aligned}$$

Wird $q = 1 - p$ eingesetzt, so erhält man für die Matchgewinnwahrscheinlichkeit:

$$\begin{aligned}P(\text{Matchgewinn Best-of-3}) &= p^2 \cdot [1 + 2 \cdot (1 - p)] \\&= p^2 \cdot (3 - 2p)\end{aligned}$$

Die Matchgewinnwahrscheinlichkeit in Abhängigkeit von der konstanten Satzgewinnwahrscheinlichkeit p bei dem Best-of-5 Modus kann folgendermaßen berechnet werden:

$$\begin{aligned}P(\text{Matchgewinn}) &= p^3 + 3 \cdot p^3 \cdot q + 6 \cdot p^3 \cdot q^2 \\&= p^3 \cdot (1 + 3p + 6q^2)\end{aligned}$$

Wird $q = 1 - p$ eingesetzt so erhält man für die Matchgewinnwahrscheinlichkeit:

$$\begin{aligned}P(\text{Matchgewinn}) &= p^3 \cdot [1 + 3 \cdot (1 - p) + 6 \cdot (1 - p)^2] \\&= p^3 \cdot (1 + 3 - 3p + 6 - 12p + 6p^2) \\&= p^3 \cdot (6p^2 - 15p + 10)\end{aligned}$$

Aufgabe 4: Iga Swiatek und Coco Gauff sind wohl zwei der besten Tennisspielerinnen der Gegenwart. Die beiden spielen ein Match auf zwei gewonnene Sätze.

1. Recherchiere die bisherigen Aufeinandertreffen der beiden Sportlerinnen und

6.3 Aufgaben zum Thema Stochastik im Tennis

ermittle deren Satzgewinnwahrscheinlichkeit.

2. Wie groß ist die Wahrscheinlichkeit, dass Swiatek ohne einen Satzverlust das Match gewinnt?
3. Wie wahrscheinlich ist es, dass Swiatek das Match in 3 Sätzen gewinnt?

Lösungserwartung:

1. Bisher sind die beiden Spielerinnen 10 mal aufeinander getroffen. Von den insgesamt 21 gespielten Sätzen hat Swiatek 19 für sich entscheiden können. Somit gilt für ihre Satzgewinnwahrscheinlichkeit:

$$p = \frac{19}{21} \approx 0,9048$$

2. Für den Sieg in 2 Sätzen gilt:

$$P(\text{Sieg in 2 Sätzen}) = p^2 = \left(\frac{19}{21}\right)^2 = \frac{361}{441} \approx 0,8186$$

3. Für einen Sieg in 3 Sätzen gilt:

$$P(\text{Sieg in 3 Sätzen}) = 2 \cdot p^2 \cdot q = 2 \cdot \left(\frac{19}{21}\right)^2 \cdot \frac{2}{21} \approx 0,1559$$

Aufgabe 5: Eine Firma produziert Tennisschläger und verpackt diese in einen Karton. Die Firma gibt an, dass 97% aller Tennisschläger fehlerfrei produziert und verpackt werden. Die Zufallsvariable X gibt die Anzahl der fehlerhaften Tennisschläger bei 1000 produzierten Schlägern an.

1. Unter der Annahme einer Binomialverteilung berechne den Erwartungswert sowie die Standardabweichung der Zufallsvariable X .
2. Würdest du bei dieser Firma deinen Tennisschläger herstellen lassen? Begründe!

6 Tennissport im Mathematikunterricht

Lösungserwartung: Es sei $n = 1000$ und $p = 0,03$, somit gilt:

$$E(X) = n \cdot p = 1000 \cdot 0,03 = 30$$

$$\sigma = \sqrt{n \cdot p \cdot (1 - p)} = \sqrt{1000 \cdot 0,03 \cdot 0,97} \approx 5,39$$

7 Fazit

Die vorliegende Masterarbeit befasst sich mit den Modellierungen von Tennisspielen durch die Anwendung der Wahrscheinlichkeitsrechnung. Durch den Vergleich von theoretischen Berechnungen mit der Realität konnte gezeigt werden, dass im Sport nicht nur der Zufall, sondern vor allem auch der*die Sportler*in eine bedeutende Rolle spielt.

Nach einer kurzen Einführung in die Geschichte des Tennissports wurden die wichtigsten Aspekte der Kombinatorik und Wahrscheinlichkeitsrechnung behandelt. Diese ermöglichten eine detaillierte Modellierung eines Tennismatches und der Berechnung diverser Gewinnwahrscheinlichkeiten. So stellte sich bei der genaueren Betrachtung der Zählweise heraus, dass bei einem Zusammentreffen zweier gleichstarker Tennisspieler*innen, wobei der*die Aufschläger*in mit einer Wahrscheinlichkeit von $p > 0,92$ den Punkt gewinnt, diese*r zu Beginn des Games eine höhere Gewinnchance hat als bei einem Spielstand von 30:15 bzw. 40:30. Des weiteren wurde gezeigt, dass die Zählweise sowie der Spielmodus dem*der besseren Sportler*in mehr entgegenkommt.

Nachdem die zuvor erarbeitete Theorie nun für den Tennissport angewendet wurde, befasst sich diese Masterarbeit mit dem Vergleich der Theorie mit der Praxis. Hierfür wurden die Ergebnisse aller bisherigen US Open Finalsspiele des Damen- und Herreneinzels herangenommen und untersucht, inwiefern dies aussagekräftige Modellierungen sind und inwiefern davon ausgegangen werden kann, dass sich im Finale zwei Spieler*innen mit vergleichbarer Spielstärke gegenüberstehen. Mit Hilfe eines Chi-Quadrat-Anpassungstest wurde gezeigt, dass selbst im Finale nicht von zwei gleichstarken Sportler*innen ausgegangen werden kann.

7 Fazit

Das letzte Kapitel befasste sich mit verschiedenen Aufgaben zum Thema Stochastik im Tennis. Ziel war es, mit Hilfe von verschiedenen Aufgabestellungen Möglichkeiten zu schaffen, wie die Verbindung von Mathematik und Sport einerseits in gewohnter Weise aber auch fächerübergreifend dargestellt und geübt werden kann.

Abschließend lässt sich sagen, dass diese Masterarbeit einen guten Überblick über die Verbindung von Sport und Mathematik biete, aber definitiv nur einen kleinen Anteil der Möglichkeiten beleuchtet. So könnte bei einer detaillierteren Analyse der Gewinnwahrscheinlichkeiten der Bodenbelag berücksichtigt werden. Einige Tennisspieler*innen bevorzugen einen gewissen Belag, weshalb die Gewinnwahrscheinlichkeiten bei verschiedenen Turnieren unterschiedlich ausfallen könnten. Der Tennissport bietet auch abgesehen von der Wahrscheinlichkeitsrechnung die Möglichkeit andere mathematische Teilbereiche zu kombinieren. So könnte in einer weiteren Modellierung beispielsweise die Flugbahn eines Schlags, die Symmetrie eines Tennisballes oder die strategischen Vorgehensweisen betrachtet und untersucht werden.

Literaturverzeichnis

- [1] BOSCH, K. *Lotto und andere Zufälle: Wie man die Gewinnquoten erhöht!*, 2 ed. Oldenbourg Verlag, München, 1999.
- [2] BOSCH, K. *Elementare Einführung in die Wahrscheinlichkeitsrechnung*, 11 ed. Vieweg Teubner, Wiesbaden, 2011.
- [3] BOSCH, K. *Statistik für Nichtstatistiker: Zufall und Wahrscheinlichkeiten*, 6 ed. Oldenbourg Verlag München, Berlin, 2012.
- [4] BUNDESMINISTERIUM FÜR BILDUNG, WISSENSCHAFT UND FORSCHUNG. Lehrplan der allgemeinbildenden höheren schulen. <https://www.ris.bka.gv.at/GeltendeFassung/Bundesnormen/10008568/Lehrpl%c3%a4ne%20e2%80%93%20allgemeinbildende%20h%c3%b6here%20Schulen%2c%20Fassung%20vom%2003.01.2024.pdf>, 2024. zuletzt zugegriffen am 3.1.2024.
- [5] CRAMER, E., AND KAMPS, U. *Grundlagen der Wahrscheinlichkeitsrechnung und Statistik*, 3 ed. Springer Spektrum, Berlin Heidelberg, 2014.
- [6] CRAMER, E., AND NESLEHOVA, J. *Vorkurs Mathematik: Arbeitsbuch zum Studienbeginn in Bachelor-Studiengängen*, 7 ed. Springer-Verlag, Berlin Heidelberg, 2018.
- [7] CROUCHER, J. Changing the rules of tennis: Who has the advantage? *Teaching Statistics* 7 (1985), 82–84.
- [8] FALK, M., HAIN, J., MAROHN, F., FISCHER, H. AND MICHEL, R. *Statistik in Theorie und Praxis*. Springer Spektrum, Berlin Heidelberg, 2014.
- [9] FROST, I. *Statistik für Wirtschaftswissenschaftler*, 3 ed. Tübingen, 2018.

- [10] GEBAUER, H. Grundbegriffe der stochastik ii. <https://www.math.uni-bielefeld.de/~rkruse/files/proseminar0910/GrundbegriffeStoch2.pdf>, 2009. zuletzt zugegriffen am 23.12.2023.
- [11] HAVIL, J. *Verblüfft?!: Mathematische Beweise unglaublicher Ideen*. Springer-Verlag, Berlin, 2009.
- [12] HENZE, N. Die chi-quadrat verteilung. <https://doi.org/10.5445/IR/1000137323>, 2021. zuletzt zugegriffen am 20.1.2024.
- [13] HENZE, N. *Stochastik für Einsteiger: Eine Einführung in die faszinierende Welt der Zufalls*, 13 ed. Springer Spektrum, Wiesbaden, 2021.
- [14] HOFBAUER, F. Wahrscheinlichkeitstheorie und statistik. <https://www.mat.univie.ac.at/~bruin/wkth.pdf>. zuletzt zugegriffen am 20.1.2024.
- [15] HUBBARD, L. How the u.s open changed the history of tennis. <https://www.townandcountrymag.com/leisure/sporting/a28832053/us-open-tennis-history/>. zuletzt zugegriffen am 31.12.2023.
- [16] HUMENBERGER, H. Modellierungsaufgaben im unterricht - selbst erfahrungen sammeln. In *Neue Materialien für einen realitätsbezogenen Mathematikunterricht 3*. Springer Fachmedien Wiesbaden, 2017.
- [17] KRAUS, M. *Kompaktkurs Kombinatorik: Gezählt, verteilt und wohlgeordnet*. Springer Spektrum, 2023.
- [18] LINDE, W. *Stochastik für das Lehramt*. De Gruyter Oldenbourg, Berlin, 2014.
- [19] REISNER, D. *Tennis: Alles, was man wissen muss*. Meyer & Meyer Verlag, Aachen, 2021.
- [20] STEMMLER, T. *Kleine Geschichte des Tennisspiels: Von Jeu de paume zum Tennis*. Insel Verlag, Frankfurt am Main, 1995.
- [21] STEWART, I. *Spiel, Satz und Sieg für die Mathematik: vergnügliche Ausflüge in die Welt der Zahlen*. Insel Verlag, Frankfurt am Main, 1997.

- [22] US OPEN. Past us open champions. https://www.usopen.org/en_US/visit/history/mschamps.html. zuletzt zugegriffen am 26.1.2023.

Abbildungsverzeichnis

1.1	Abmaße eines Tennisfeldes (Quelle: https://tennis-uni.com/tennisplatz-tennisfeld - zuletzt zugegriffen am: 2.2.2024)	3
2.1	Baumdiagramm - Ziehen einer geordneten Stichprobe mit Zurücklegen	13
2.2	Baumdiagramm - Ziehen einer geordneten Stichprobe ohne Zurücklegen	14
2.3	Baumdiagramm - Ziehen einer geordneten Stichprobe ohne Zurücklegen mit $n = k$	16
2.4	Baumdiagramm - Ziehen einer ungeordneten Stichprobe ohne Zurücklegen	18
2.5	Baumdiagramm - Ziehen einer ungeordneten Stichprobe mit Zurücklegen	19
2.6	Wahrscheinlichkeitsfunktion	23
2.7	Verteilungsfunktion $F(X)$	25
2.8	Dichtefunktion einer stetigen Zufallsvariable	40
2.9	Gaußsche Glockenkurve	43
2.10	$P(X < 1, 72)$	46
2.11	$P(X \leq -1, 12)$	46
2.12	$P(-1, 30 \leq X < 0, 54)$	47
2.13	$P(X \leq 300)$	48
2.14	$P(283 \leq X \leq 297)$	49
2.15	$P(X - 7 \leq t) = 0, 95$	50
2.16	$P(X \geq 287) = 0, 90$	50
4.1	Spielverlauf eines Games	60
4.2	Spielverlauf eines Games bei Einstand	60
4.3	Spielverlauf eines Matches mit Best-Of-3 Modus	67

Abbildungsverzeichnis

4.4	Spielverlauf eines Matches mit Best-of-5 Modus	69
4.5	Spielverlauf eines Tie-Breaks	71
4.6	Spielverlauf eines Einstands im Tie-Break	72
4.7	Spielverlauf für den Fall $P(30,15)$	76
4.8	$P(0,0)$ in Abhängigkeit von p	78
4.9	$P(30,15)$ in Abhängigkeit von p	79
4.10	$P(40,30)$ in Abhängigkeit von p	79
6.1	Modellierungsprozess	90