



universität
wien

DIPLOMARBEIT

Titel der Diplomarbeit

Auch der Zufall spielt mit System - Grenzwertsätze der Wahrscheinlichkeitstheorie

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag.rer.nat.)

Verfasser:	Jürgen Steininger
Matrikel-Nummer:	0848618
Studienkennzahl:	A 190 406 313
Studienrichtung:	UF Mathematik und UF Geschichte, Sozialkunde und Politische Bildung
Betreuer:	Ao. Univ.-Prof. Dr. Peter Raith

Wien, Oktober 2013

Vorwort

Der Zufall ist in unserem Leben allgegenwärtig. Er führt Regie bei den wöchentlichen Ziehungen der Lottozahlen und entscheidet bei zahlreichen Gesellschafts- und Glücksspielen über Gewinnen und Verlieren. Obwohl wir den Zufall nicht beeinflussen können und ihn mit Glück und Pech assoziieren, stellen wir trotzdem gewisse Forderungen an ihn. So erwarten wir intuitiv, dass beim Roulette die Farben Rot und Schwarz etwa gleich häufig erscheinen oder ein fairer Münzwurf in der Hälfte der Fälle Kopf zum Ergebnis hat. Auch wenn wir kurzfristige Schwankungen oder Pechsträhnen akzeptieren, erwarten wir, dass der Zufall auf lange Zeit gesehen „fair“ handelt. Und tatsächlich: Wie wir sehen werden, gehorcht selbst der „unberechenbare“ Zufall gewissen mathematischen Gesetzen.

Das theoretische Fundament dieser langfristigen Gesetzmäßigkeiten bilden die Grenzwertsätze. Sie zählen zu den wichtigsten Resultaten der Wahrscheinlichkeitstheorie. Bekannte Beispiele dafür sind die Gesetze der Großen Zahlen oder der Zentrale Grenzwertsatz, die mittlerweile sogar in vereinfachten Versionen in jedem Schulbuch zu finden sind. Diese Grenzwertsätze haben nicht nur innerhalb der Stochastik eine fundamentale Bedeutung, sondern auch in zahlreichen Anwendungsbereichen wie in der Medizin, Biologie, Genetik oder Wirtschaft. Diese Tatsache unterstreicht die wachsende Bedeutung der Stochastik und zeugt von der fruchtbaren Wechselwirkung zwischen Theorie und Anwendung.

In der vorliegenden Diplomarbeit möchte ich mich allerdings weniger mit den Anwendungen beschäftigen, sondern die mathematischen Hintergründe dieser Grenzwertsätze ausarbeiten. Dazu möchte ich im ersten Kapitel zunächst wichtige Grundlagen und Hilfsmittel der Wahrscheinlichkeitstheorie erläutern. Neben den klassischen Begriffen wie Wahrscheinlichkeitsräumen, Zufallsvariablen oder Verteilungen werden auch die charakteristischen Funktionen eingeführt, die eines der wichtigsten Hilfsmittel darstellen. Im zweiten Kapitel werden die verschiedenen Konvergenzbegriffe der Stochastik vorgestellt, die wir später benötigen werden. Dabei gehen wir auch auf die Zusammenhänge dieser Konvergenzarten ein.

Mit diesem Wissen ausgestattet können wir uns anschließend im dritten und vierten Abschnitt mit den Gesetzen der Großen Zahlen bzw. den Zentralen Grenzwertsätzen beschäftigen. Dabei werden zahlreiche verschiedene Versionen mit unterschiedlich starken Voraussetzungen behandelt. Angefangen mit einfachen Spezialfällen werden die Annahmen immer mehr abgeschwächt, um schließlich zu sehr allgemeinen Resultaten zu gelangen.

Das fünfte Kapitel beschäftigt sich mit weniger bekannten Resultaten. In den bis dahin angeführten Theoremen setzen wir immer voraus, dass die zufälligen Ereignisse unabhängig voneinander erfolgen. Gerade für viele Anwendungsbereiche ist dies keine realistische Annahme. Deshalb führen wir eine spezielle Klasse von Stochastischen Prozessen ein, die Markov-Ketten genannt werden. Sie weisen ein einfaches Kurzzeitgedächtnis auf und ermöglichen damit weiterhin explizite Berechnungen. Wir interessieren uns dabei wieder für das Grenzverhalten. Davor werden jedoch grundlegende Eigenschaften eingeführt.

Im letzten Abschnitt werden Irrfahrten auf $\mathbb{Z}^d$ behandelt, die eine spezielle Klasse von Markov-Ketten darstellen. Augenmerk wird dabei wieder auf das Langzeitverhalten gelegt. Nach einer kurzen Einführung studieren wir vor allem die Rückkehrzeiten derartiger Prozesse in den verschiedenen Dimensionen. Anschließend untersuchen wir noch weitere Grenzwertaussagen der eindimensionalen Irrfahrt und geben einen Ausblick auf zeitstetige Prozesse.

Bedanken möchte ich mich bei meiner Familie, insbesondere meinen Eltern, die es mir ermöglicht haben, dieses Studium zu absolvieren und mich bei diversen Problemen immer voll unterstützt haben. Dank gebührt auch meinen Freunden und Studienkollegen, die mir immer mit Rat und Tat zu Seite standen.

Ein großer Dank gilt auch meinem Betreuer der Diplomarbeit, Peter Raith, der mir bei Fragen immer sofort behilflich war. Neben der ausgesprochen guten Betreuung gab er mir alle Freiheiten und ermöglichte mir somit ein selbständiges und ökonomisches Arbeiten.

Inhaltsverzeichnis

1	Wahrscheinlichkeitstheoretische Grundlagen	5
1.1	Wahrscheinlichkeitsräume	5
1.2	Zufallsvariablen	7
1.3	Erwartungswert und Varianz	12
1.4	Wichtige Verteilungen	14
1.5	Charakteristische Funktionen	16
1.5.1	Definition und Beispiele	16
1.5.2	Wichtige Eigenschaften	18
1.5.3	Eindeutigkeits- und Stetigkeitssatz	20
2	Konvergenzbegriffe	23
2.1	Konvergenz nach Verteilung	23
2.2	Stochastische Konvergenz	24
2.3	Fast sichere Konvergenz	24
2.4	Vergleich der Konvergenzbegriffe	25
3	Gesetze der Großen Zahlen	28
3.1	Ungleichungen	28
3.1.1	Tschebyscheff-Ungleichung	28
3.1.2	Kolmogorov-Ungleichung	29
3.2	Das Schwache Gesetz der Großen Zahlen	30
3.2.1	Schwaches Gesetz der Großen Zahlen von Bernoulli	31
3.2.2	Schwaches Gesetz der Großen Zahlen von Tschebyscheff	31
3.2.3	Schwaches Gesetz der Großen Zahlen ohne identische Verteilung	32
3.2.4	Schwaches Gesetz der Großen Zahlen von Chintschin	33
3.3	Das Starke Gesetz der Großen Zahlen	34
3.3.1	Das Lemma von Borel-Cantelli	34
3.3.2	Starkes Gesetz der Großen Zahlen von Borel	35
3.3.3	Starkes Gesetz der Großen Zahlen von Rajchman	36
3.3.4	Starkes Gesetz der Großen Zahlen ohne identische Verteilung	37
3.3.5	Starkes Gesetz der Großen Zahlen von Kolmogorov	39
4	Der Zentrale Grenzwertsatz	42
4.1	Konvergenz von diskreten Verteilungen	42
4.1.1	Binomialapproximation der hypergeometrischen Verteilung	42
4.1.2	Poisson-Approximation der Binomialverteilung	44
4.2	Der Zentrale Grenzwertsatz von de Moivre und Laplace	46
4.2.1	Stirling'sche Formel	46
4.2.2	Elementarer Beweis	47

4.2.3	Korrekturterme	50
4.2.4	Beispiele	51
4.3	Verallgemeinerungen	52
4.3.1	Zentraler Grenzwertsatz von Lindeberg und Lévy	52
4.3.2	Zentraler Grenzwertsatz von Lindeberg und Feller	53
4.3.3	Zentraler Grenzwertsatz von Ljapunov	54
5	Markov-Ketten	56
5.1	Grundlagen	56
5.1.1	Definition und Eigenschaften	56
5.1.2	Klassifikation von Zuständen	59
5.1.3	Rekurrenz und Transienz	61
5.1.4	Stationarität	65
5.2	Grenzverhalten von Markov-Ketten	68
5.2.1	Transiente Zustände	68
5.2.2	Rekurrente Zustände mit endlichem Zustandsraum	69
5.2.3	Rekurrente Zustände mit abzählbarem Zustandsraum	71
6	Einfache Irrfahrten auf $\mathbb{Z}^d$	76
6.1	Grundlagen	76
6.2	Rekurrenz und Transienz	77
6.2.1	Die 1-dimensionale Irrfahrt	77
6.2.2	Die 2-dimensionale Irrfahrt	79
6.2.3	Die 3-dimensionale Irrfahrt	80
6.3	Das Arkus-Sinus Gesetz	81
6.4	Ausblick: Die Brownsche Bewegung	87
	Literaturverzeichnis	90
	Anhang	94

1 Wahrscheinlichkeitstheoretische Grundlagen

In diesem Kapitel werden die grundlegenden Begriffe eingeführt, die für die Behandlung der weiteren Themen Voraussetzung sind. Hauptquellen für dieses Kapitel waren [1], [3], [5] und [7].

1.1 Wahrscheinlichkeitsräume

Zunächst müssen wir ein System $\mathcal{A}$ von Ereignissen festlegen, sodass man jedem Ereignis $A \in \mathcal{A}$ eine Wahrscheinlichkeit $\mathbb{P}[A]$ zuordnen kann. Solange der Ereignisraum Ω endlich oder abzählbar ist, kommt es zu keinen Schwierigkeiten, da wir dann für $\mathcal{A}$ die Potenzmenge $\mathcal{P}(\Omega)$ wählen können. Ist Ω jedoch überabzählbar, so kommt es zu Problemen. Der Ausweg aus dem Dilemma besteht darin, dass wir darauf verzichten, für alle A eine Wahrscheinlichkeit $\mathbb{P}[A]$ zu definieren. Stattdessen ordnen wir nur jenen Ereignissen Wahrscheinlichkeiten zu, die bestimmte Eigenschaften erfüllen.

Definition 1.1. Sei $\Omega \neq \emptyset$. Ein Mengensystem $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ mit den Eigenschaften

- (i) $\Omega \in \mathcal{A}$
- (ii) $A \in \mathcal{A} \Rightarrow A^c := \Omega \setminus A \in \mathcal{A}$
- (iii) $A_1, A_2, A_3, \dots \in \mathcal{A} \Rightarrow \bigcup_{i \geq 1} A_i \in \mathcal{A}$

heißt eine σ -Algebra in Ω . Das Paar $(\Omega, \mathcal{A})$ nennt man *Ereignisraum* oder *messbarer Raum*.

Bemerkung: Aus dieser Definition folgen sofort weitere Eigenschaften für die Ereignisse $A_1, A_2, \dots \in \mathcal{A}$:

- $\emptyset \in \mathcal{A}$, da $\emptyset = \Omega^c$.
- $\bigcap_{i \geq 1} A_i \in \mathcal{A}$, da $\bigcap_{i \geq 1} A_i = (\bigcup_{i \geq 1} A_i^c)^c$.
- $A_1 \cup A_2 \in \mathcal{A}$, da $A_1 \cup A_2 = A_1 \cup A_2 \cup \emptyset \cup \emptyset \cup \dots$

Satz 1.1. Sei $\Omega \neq \emptyset$ und $\mathcal{U} \subseteq \mathcal{P}(\Omega)$ beliebig. Dann gibt es genau eine kleinste σ -Algebra $\mathcal{A} = \sigma(\mathcal{U})$ in Ω , die $\mathcal{U}$ enthält. Dieses $\mathcal{A}$ heißt die von $\mathcal{U}$ erzeugte σ -Algebra.

Beweis. Sei Σ das System aller σ -Algebren $\mathcal{F}$ in Ω mit $\mathcal{F} \supseteq \mathcal{U}$. Σ ist nichtleer, da $\mathcal{P}(\Omega) \in \Sigma$. Daher können wir $\mathcal{A} := \bigcap_{\mathcal{F} \in \Sigma} \mathcal{F}$ setzen. Man kann nun leicht überprüfen, dass $\mathcal{A}$ die drei Eigenschaften einer σ -Algebra erfüllt. Nach Konstruktion ist diese auch die kleinste. $\square$

Dieser Satz ist ein reiner Existenzsatz. Nur in sehr speziellen Fällen, zum Beispiel für endliches $\mathcal{U}$, kann man $\sigma(\mathcal{U})$ explizit angeben. Für uns ist vor allem der Fall $\Omega = \mathbb{R}$ von Bedeutung. Da wir nicht allen Teilmengen von $\mathbb{R}$ eine Wahrscheinlichkeit zuordnen können, muss eine geeignete σ -Algebra gefunden werden. Eine explizite Angabe dieser gewünschten σ -Algebra ist jedoch nicht möglich. Auf jeden Fall sollte sie jedoch die Gesamtheit aller Intervalle beinhalten. Mithilfe von Satz 1.1 können wir die kleinste σ -Algebra finden, die das System der Intervalle umfasst. Dies führt zur folgenden Definition:

Definition 1.2. Sei $\Omega = \mathbb{R}$ und $\mathcal{U} = \{(a, b) : a, b \in \mathbb{R}, a < b\}$. Dann heißt $\mathcal{B} := \sigma(\mathcal{U})$ die *Borel'sche- σ -Algebra* auf $\mathbb{R}$ und jedes $A \in \mathcal{B}$ eine *Borel-Menge*.

- Bemerkung:**
1. Um die Borel-Mengen von $\mathbb{R}$ zu erhalten, beginnen wir sozusagen mit den offenen Intervallen $(a, b) \subseteq \mathbb{R}$ und fügen alle anderen Mengen hinzu, die nötig sind, um eine σ -Algebra zu erhalten.
 2. Hier wurden die offenen Intervalle als Erzeugendensystem für die Borel'sche σ -Algebra gewählt. Wir könnten dafür genauso halboffene oder abgeschlossene Intervalle heranziehen, da zum Beispiel jedes offene Intervall als Vereinigung von abzählbar vielen abgeschlossenen Intervallen dargestellt werden kann.
 3. Liegt der Ergebnisraum $\mathbb{R}$ vor, so betrachten wir als σ -Algebra stets die Borel'sche- σ -Algebra. Diese beinhaltet alle für uns relevanten Mengen von $\mathbb{R}$.

Damit können wir die Wahrscheinlichkeit nun wie folgt definieren:

Definition 1.3. Sei $(\Omega, \mathcal{A})$ ein Ereignisraum. Eine Abbildung $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ heißt *Wahrscheinlichkeit*, wenn folgende Eigenschaften erfüllt sind:

- (i) $\mathbb{P}[\Omega] = 1$.
- (ii) Für paarweise disjunkte $A_1, A_2, \dots \in \mathcal{A}$ ist $\mathbb{P}[\bigcap_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} \mathbb{P}[A_i]$.

Das Tripel $(\Omega, \mathcal{A}, \mathbb{P})$ heißt dann *Wahrscheinlichkeitsraum*.

Satz 1.2. Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, B, A_1, \dots, A_n \in \mathcal{A}$. Dann gilt:

- (i) $\mathbb{P}[\emptyset] = 0$.
- (ii) $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$.
- (iii) $\mathbb{P}[A \setminus B] = \mathbb{P}[A] - \mathbb{P}[A \cap B]$.
- (iv) $A \subseteq B \Rightarrow \mathbb{P}[A] \leq \mathbb{P}[B]$.

$$(v) \mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B].$$

$$(vi) \text{ Wenn } A_1 \subseteq A_2 \subseteq \dots, \text{ dann ist } \lim_{n \rightarrow \infty} \mathbb{P}[A_i] = \mathbb{P}[\bigcup_{i=1}^{\infty} A_n].$$

$$(vii) \text{ Wenn } A_1 \supseteq A_2 \supseteq \dots, \text{ dann ist } \lim_{n \rightarrow \infty} \mathbb{P}[A_i] = \mathbb{P}[\bigcap_{i=1}^{\infty} A_n].$$

Beweis. Die Beweise folgen aus einfachen mengentheoretischen Eigenschaften.

$$(i) \emptyset = \bigcup_{i=1}^{\infty} \emptyset. \Rightarrow \mathbb{P}[\emptyset] = \sum_{i=1}^{\infty} \mathbb{P}[\emptyset] \Rightarrow \mathbb{P}[\emptyset] = 0.$$

$$(ii) A \cap A^c = \emptyset \text{ und } A \cup A^c = \Omega. \Rightarrow 1 = \mathbb{P}[\Omega] = \mathbb{P}[A \cup A^c] = \mathbb{P}[A] + \mathbb{P}[A^c].$$

$$(iii) A = (A \setminus B) \cup (A \cap B) \text{ und } (A \setminus B) \cap (A \cap B) = \emptyset. \Rightarrow \mathbb{P}[A] = \mathbb{P}[A \setminus B] + \mathbb{P}[A \cap B].$$

$$(iv) A \subseteq B \Rightarrow B = A \cup (B \setminus A) \text{ und } A \cap (B \setminus A) = \emptyset. \Rightarrow \mathbb{P}[B] = \mathbb{P}[A \cup B \setminus A] = \mathbb{P}[A] + \mathbb{P}[B \setminus A] \geq \mathbb{P}[A].$$

$$(v) A \cup B = A \cup (B \setminus A) \text{ und } A \cap (B \setminus A) = \emptyset. \Rightarrow \mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B \setminus A] \stackrel{(iii)}{=} \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B].$$

(vi) Sei $A_0 = \emptyset$ und $C_n := A_n \setminus A_{n-1}$ für $n \in \mathbb{N}$. Dann ist $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} C_i$. Die Mengen C_n sind paarweise disjunkt und es gilt $A_n = C_1 \cup \dots \cup C_n$. Daraus folgt:

$$\begin{aligned} \mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i\right] &= \mathbb{P}\left[\bigcup_{i=1}^{\infty} C_i\right] = \sum_{i=1}^{\infty} \mathbb{P}[C_i] = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}[C_i] \\ &= \lim_{n \rightarrow \infty} \mathbb{P}[C_1 \cup \dots \cup C_n] = \lim_{n \rightarrow \infty} \mathbb{P}[A_n]. \end{aligned}$$

(vii) Da $A_1 \supseteq A_2 \supseteq \dots$ folgt $A_1^c \subseteq A_2^c \subseteq \dots$. Somit gilt nach (vi):

$$\begin{aligned} \mathbb{P}\left[\bigcap_{i=1}^{\infty} A_i\right] &= 1 - \mathbb{P}\left[\bigcup_{i=1}^{\infty} A_i^c\right] = 1 - \lim_{n \rightarrow \infty} \mathbb{P}[A_n^c] \\ &= 1 - \lim_{n \rightarrow \infty} (1 - \mathbb{P}[A_n]) = \lim_{n \rightarrow \infty} \mathbb{P}[A_n]. \end{aligned} \quad \square$$

1.2 Zufallsvariablen

Bei einem Zufallsexperiment interessiert man sich häufig nicht für alle Details, sondern nur für bestimmte Größen des Geschehens. Anstatt den direkten vom Zufall gesteuerten Ausgang $\omega \in \Omega$ zu betrachten, interessiert man sich für eine andere Größe $X(\omega)$.

Beispiel 1.1: Zweimaliges Würfeln

Registriert man die Augenzahl jedes einzelnen Würfels, so ist $\Omega = \{1, 2, 3, 4, 5, 6\}^2$. Interessiert man sich hingegen nur für die Augensumme, so ist der Ereignisraum

$\Omega' = \{2, 3, \dots, 12\}$. Dieser Vorgang kann durch eine Abbildung $X : \Omega \rightarrow \Omega'$, welche jedem $\omega = (\omega_1, \omega_2) \in \Omega$ die Summe $X(\omega) = \omega_1 + \omega_2$ zuordnet, beschrieben werden. Betrachtet man nun ein Ereignis $A = \{\omega : X(\omega) = k\}$ mit $k = 2, 3, \dots, 12$ und will diesem eine Wahrscheinlichkeit $\mathbb{P}[A]$ zuordnen, so ist es erforderlich, dass $A \in \mathcal{A}$ ist.

Die Überlegungen dieses Beispiels sollen nun verallgemeinert werden. Wir betrachten dazu einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ und eine Abbildung $X : \Omega \rightarrow \Omega'$. Im Bildraum Ω' liegen Ereignisse C vor, deren Gesamtheit eine σ -Algebra $\mathcal{C}$ bilden soll. Wir interessieren uns nun für Wahrscheinlichkeiten $\mathbb{P}[\{\omega : X(\omega) \in C\}]$ mit $C \in \mathcal{C}$. Da die Wahrscheinlichkeit nur für Ereignisse $A \in \mathcal{A}$ definiert ist, muss $X^{-1}(C) = \{\omega : X(\omega) \in C\} \in \mathcal{A}$ gelten. Dies führt zur folgenden Definition:

Definition 1.4. Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(\Omega', \mathcal{C})$ ein messbarer Raum. Eine Abbildung $X : \Omega \rightarrow \Omega'$ mit der Eigenschaft

$$C \in \mathcal{C} \Rightarrow X^{-1}(C) \in \mathcal{A}$$

heißt *Zufallsvariable*.

Bemerkung: 1. Die Definition einer Zufallsvariable ist völlig unabhängig vom Wahrscheinlichkeitsmaß.

2. Eine Abbildung, die die Eigenschaft der Definition erfüllt, bezeichnet man als messbar.
3. Hat man einen diskreten Wahrscheinlichkeitsraum und somit $\mathcal{A} = \mathcal{P}(\Omega)$, so ist jede Abbildung $X : \Omega \rightarrow \Omega'$ eine Zufallsvariable.
4. Statt $X^{-1}(C) = \{\omega : X(\omega) \in C\}$ schreibt man oft kurz $\{X \in C\}$. Bei Ausdrücken der Gestalt $\mathbb{P}[\{X \in C\}]$ werden zusätzlich die geschweiften Klammern weggelassen und nur $\mathbb{P}[X \in C]$ geschrieben.

Lemma 1.3. Sei $\mathcal{C}$ aus der Definition erzeugt durch ein Mengensystem $\mathcal{U}'$, d.h. es sei $\mathcal{C} = \sigma(\mathcal{U}')$. Dann ist $X : \Omega \rightarrow \Omega'$ bereits dann eine Zufallsvariable, wenn die Bedingung $X^{-1}(C) \in \mathcal{A}$ nur für alle $C \in \mathcal{U}'$ gilt.

Beweis. Das System $\mathcal{F}' = \{C \subseteq \Omega' : X^{-1}(C) \in \mathcal{A}\}$ ist eine σ -Algebra. Sie umfasst $\mathcal{U}'$, da wir vorausgesetzt haben, dass für alle $C \in \mathcal{U}'$ das Urbild $X^{-1}(C)$ in $\mathcal{A}$ liegt. Da nach Voraussetzung $\mathcal{C}$ die kleinste σ -Algebra ist, die $\mathcal{U}'$ umfasst, gilt auch $\mathcal{C} \subseteq \mathcal{F}'$. Somit gilt für alle $C \in \mathcal{C}$, dass $X^{-1}(C) \in \mathcal{A}$. $\square$

Da die Familie $\mathcal{U}'$ der Intervalle $(-\infty, x]$ im Fall $(\Omega', \mathcal{C}) = (\mathbb{R}, \mathcal{B})$ ein Erzeugenden-

system der Borel'schen σ -Algebra ist, kann eine reelle Zufallsvariable auch wie folgt definiert werden:

Definition 1.5. Sei $(\Omega', \mathcal{C}) = (\mathbb{R}, \mathcal{B})$. Eine Abbildung $X : \Omega \rightarrow \mathbb{R}$ ist eine *reelle Zufallsvariable*, wenn alle Mengen der Form $\{\omega : X(\omega) \leq x\} = \{X \leq x\} = X^{-1}((-\infty, x])$ zu $\mathcal{A}$ gehören.

Nun ist man oft nicht an den Wahrscheinlichkeiten der Ereignisse im Raum $(\Omega, \mathcal{A}, \mathbb{P})$ interessiert, sondern an den Wahrscheinlichkeiten der Ereignisse in $(\Omega', \mathcal{C})$. Dafür müssen wir das Wahrscheinlichkeitsmaß sozusagen in den Bildraum "transportieren". Ist C ein Ereignis des Bildraumes, so geht es darum mithilfe des Wahrscheinlichkeitsmaßes $\mathbb{P}$ des Urbildraums Ω und der Zufallsvariable X dem Ereignis C eine Wahrscheinlichkeit zuzuordnen. Dazu definiert man die Urbildabbildung X^{-1} für alle $C \in \mathcal{C}$. Durch Komposition von X^{-1} und $\mathbb{P}$ bekommen wir dann das sogenannte induzierte Wahrscheinlichkeitsmaß auf $(\Omega', \mathcal{C})$.

Satz 1.4. Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(\Omega', \mathcal{C})$ ein Messraum. Weiters sei $X : \Omega \rightarrow \Omega'$ eine Zufallsvariable. Durch die Abbildung

$$\mu_X(C) := (\mathbb{P} \circ X^{-1})(C) = \mathbb{P}[X^{-1}(C)] \text{ mit } C \in \mathcal{C}$$

wird ein Wahrscheinlichkeitsmaß auf $(\Omega', \mathcal{C})$ festgelegt.

Beweis. Die Abbildung erfüllt die beiden Eigenschaften:

- (i) $\mu_X(\Omega') = \mathbb{P}[X^{-1}(\Omega')] = \mathbb{P}[\Omega] = 1$.
- (ii) Seien $C_1, C_2, \dots \in \Omega'$ paarweise disjunkt. Dann sind auch deren Urbilder $X^{-1}(C_1), X^{-1}(C_2), \dots$ paarweise disjunkt und es gilt

$$\begin{aligned} \mu_X \left(\bigcup_{k \geq 1} C_k \right) &= \mathbb{P} \left[X^{-1} \left(\bigcup_{k \geq 1} C_k \right) \right] = \mathbb{P} \left[\bigcup_{k \geq 1} X^{-1}(C_k) \right] \\ &= \sum_{k \geq 1} \mathbb{P}[X^{-1}(C_k)] = \sum_{k \geq 1} \mu_X(C_k). \end{aligned}$$

Damit ist gezeigt, dass μ_X ein Wahrscheinlichkeitsmaß ist. □

Ist also X eine Zufallsvariable mit Werten in Ω' , so können wir für alle $C \in \mathcal{C}$ die Wahrscheinlichkeit $\mathbb{P}[X \in C] = \mathbb{P}[X^{-1}(C)]$ bilden. Wir erhalten somit eine Abbildung von $\mathcal{C}$ nach $[0, 1]$.

Definition 1.6. Sei X eine Zufallsvariable von $(\Omega, \mathcal{A}, \mathbb{P})$ nach $(\Omega', \mathcal{C})$. Dann heißt die Abbildung

$$\mu_X : \mathcal{C} \longrightarrow [0, 1], \mu_X(C) = \mathbb{P}[X \in C] = \mathbb{P}[X^{-1}(C)] = (\mathbb{P} \circ X^{-1})(C)$$

mit $C \in \mathcal{C}$ *Wahrscheinlichkeitsverteilung* von X .

Hat man es mit einer reellen Zufallsvariable zu tun, so kann man die Wahrscheinlichkeitsverteilung auch mithilfe der Verteilungsfunktion beschreiben, die wie folgt definiert ist:

Definition 1.7. Ist X eine reelle Zufallsvariable, so heißt die Funktion

$$F_X : \mathbb{R} \longrightarrow [0, 1], F_X(x) = \mu_X((-\infty, x]) = \mathbb{P}[X \leq x]$$

mit $x \in \mathbb{R}$ die Verteilungsfunktion von X .

Ist X eine reellwertige Zufallsvariable, so besagt ein Satz aus der Maßtheorie, dass die Wahrscheinlichkeitsverteilung von X durch die Verteilungsfunktion bereits eindeutig festgelegt ist. Während die Wahrscheinlichkeitsverteilung für allgemeine Wahrscheinlichkeitsräume definiert ist, bezieht sich der Begriff der Verteilungsfunktion nur auf den Fall $(\Omega', \mathcal{C}) = (\mathbb{R}, \mathcal{B})$. Ist klar bezüglich welcher Zufallsvariable die Verteilungsfunktion definiert ist, so schreiben wir $F_X(x) = F(x)$.

Satz 1.5. Sei $F(x)$ die Verteilungsfunktion von X . Dann gelten folgende Eigenschaften:

(i) Für $a < b \in \mathbb{R}$ folgt $\mathbb{P}[a < X \leq b] = F(b) - F(a)$ und $F(a) \leq F(b)$.

(d.h. F ist monoton steigend)

(ii) Für jede Folge $(x_n)_{n \in \mathbb{N}}$ mit $x_1 \geq x_2 \geq \dots$ und $\lim_{n \rightarrow \infty} x_n = x$ folgt

$\lim_{n \rightarrow \infty} F(x_n) = F(x)$. (d.h. F ist rechtsseitig stetig)

(iii) $\lim_{x \rightarrow \infty} F(x) = 1$ und $\lim_{x \rightarrow -\infty} F(x) = 0$.

Beweis. (i) Das Ereignis $\{X \leq b\}$ lässt sich in die disjunkten Ereignisse $\{X \leq a\}$ und $\{a < X \leq b\}$ zerlegen. Daraus folgt:

$$\begin{aligned} \mathbb{P}[X \leq b] &= \mathbb{P}[X \leq a] + \mathbb{P}[a < X \leq b] \\ \Leftrightarrow F(b) - F(a) &= \mathbb{P}[a < X \leq b]. \end{aligned}$$

Da Wahrscheinlichkeiten immer ≥ 0 sind, folgt $F(a) \leq F(b)$.

(ii) Die Ereignisse $\{X \leq x_n\}$ bilden eine monoton fallende Folge und es gilt

$\bigcap_{n \in \mathbb{N}} \{X \leq x_n\} = \{X \leq x\}$. Mit Satz 1.2(vii) folgt

$$F(x) = \mathbb{P}[X \leq x] = \lim_{n \rightarrow \infty} \mathbb{P}[X \leq x_n] = \lim_{n \rightarrow \infty} F(x_n).$$

(iii) Sei $(x_n)_{n \in \mathbb{N}} \geq 1$ eine monoton steigende Folge mit $\lim_{n \rightarrow \infty} x_n = \infty$. Dann bilden die Ereignisse $\{X \leq x_n\}$ eine monoton steigende Folge und es gilt $\bigcup_{n \in \mathbb{N}} \{X \leq x_n\} = \{X \leq \infty\}$. Mit Satz 1.2(vi) folgt

$$1 = \mathbb{P}[X \leq \infty] = \lim_{n \rightarrow \infty} \mathbb{P}[X \leq x_n] = \lim_{n \rightarrow \infty} F(x_n).$$

Damit ist auch $\lim_{x \rightarrow \infty} F(x) = 1$.

Für den Beweis von $\lim_{x \rightarrow -\infty} F(x) = 0$ wählt man eine monoton fallende Folge, die gegen $-\infty$ geht und verwendet Satz 1.2 sowie $\mathbb{P}[X \leq -\infty] = 0$. $\square$

Außerdem ist jede Funktion $F : \mathbb{R} \rightarrow [0, 1]$, die die drei Eigenschaften des Satzes erfüllt, eine Verteilungsfunktion. Es gilt also auch die Umkehrung des Satzes. Für uns sind vor allem zwei Fälle von Bedeutung:

Definition 1.8. Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \Omega'$ eine Zufallsvariable mit $B \subseteq \Omega'$.

1. X heißt *diskret*, wenn die Wertemenge $X(\Omega) = \{X(\omega) : \omega \in \Omega\}$ endlich oder abzählbar ist. Für die Wahrscheinlichkeitsverteilung gilt dann

$$\mu_X(B) = \sum_{x \in B} \mathbb{P}[X = x].$$

2. X heißt *stetig* (bzw. *kontinuierlich*), wenn es eine integrierbare Funktion $f : \mathbb{R} \rightarrow [0, \infty)$ mit der Eigenschaft $\int_{-\infty}^{\infty} f(x) dx = 1$ gibt, sodass für die Wahrscheinlichkeitsverteilung

$$\mu_X(B) = \int_B f(x) dx$$

gilt. Eine solche Funktion nennt man *Wahrscheinlichkeitsdichte*.

Bemerkung: Ist X eine reelle Zufallsvariable mit Dichte $f(x)$, so gilt für die Verteilungsfunktion $F(x) = \int_{-\infty}^x f(x) dx$. Wenn die Verteilungsfunktion differenzierbar ist, so ist ihre Ableitung also die Dichtefunktion.

Oft werden wir Folgen von Zufallsvariablen betrachten, die unabhängig und identisch verteilt sind. Deswegen definieren im Folgenden auch noch diese beiden Begriffe:

Definition 1.9. Zwei Zufallsvariablen heißen *identisch verteilt*, wenn sie dieselbe Wahrscheinlichkeitsverteilung haben.

Definition 1.10. Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $I \neq \emptyset$ eine beliebige Indexmenge. Eine Familie $(A_i)_{i \in I}$ von Ereignissen heißt *unabhängig*, wenn für alle endlichen nichtleeren Teilmengen $J \subseteq I$ gilt, dass $\mathbb{P}[\bigcup_{i \in J} A_i] = \prod_{i \in J} \mathbb{P}(A_i)$.

Definition 1.11. Seien $X_i : (\Omega, \mathcal{A}, \mathbb{P}) \longrightarrow (\Omega_i, \mathcal{A}_i)$ mit $i \in I$ Zufallsvariablen. Sie heißen *unabhängig*, wenn für alle $A_i \in \mathcal{A}_i$ die Ereignisse $\{X_i \in A_i\}$ unabhängig sind.

1.3 Erwartungswert und Varianz

Definition 1.12. Ist X eine diskrete Zufallsvariable und $\sum_{\omega \in \Omega} |X(\omega)| \mathbb{P}[\omega] < \infty$, so definiert man den Erwartungswert von X durch

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}[\omega].$$

Ist X eine stetige Zufallsvariable mit Wahrscheinlichkeitsdichte $f(x)$ und $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$, so definiert man den Erwartungswert von X durch

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx.$$

Bemerkung: Der Erwartungswert hängt ausschließlich von der Wahrscheinlichkeitsverteilung ab und kann als Schwerpunkt von dieser interpretiert werden. Im diskreten Fall können wir den Erwartungswert auch folgendermaßen schreiben:

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}[\omega] = \sum_{x \in X(\Omega)} x \sum_{\omega \in \{X=x\}} \mathbb{P}[\omega] = \sum_{x \in X(\Omega)} x \mathbb{P}[X=x].$$

Satz 1.6. Seien X und Y Zufallsvariablen, deren Erwartungswert existiere und $a \in \mathbb{R}$. Dann gilt $\mathbb{E}[X + aY] = \mathbb{E}[X] + a\mathbb{E}[Y]$. Das heißt der Erwartungswert ist ein linearer Operator.

Beweis. Wir zeigen den Satz für diskrete Zufallsvariablen, für stetige funktioniert er analog.

$$\begin{aligned} \mathbb{E}[X + aY] &= \sum_{\omega \in \Omega} (X + aY)(\omega) \mathbb{P}[\omega] \\ &= \sum_{\omega \in \Omega} (X(\omega) + aY(\omega)) \mathbb{P}[\omega] \\ &= \sum_{\omega \in \Omega} X(\omega) \mathbb{P}[\omega] + a \sum_{\omega \in \Omega} Y(\omega) \mathbb{P}[\omega] \\ &= \mathbb{E}[X] + a\mathbb{E}[Y]. \end{aligned} \quad \square$$

Definition 1.13. Seien X und Y Zufallsvariablen mit Erwartungswert. Sie heißen *unkorreliert*, wenn auch XY einen Erwartungswert besitzt und $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ gilt.

Bemerkung: Aus unabhängig folgt unkorreliert, die Umkehrung gilt jedoch nicht! Existiert $\mathbb{E}[X^k]$, so nennt man ihn das k -te Moment und schreibt $X \in \mathcal{L}^k$. Wichtig ist vor allem der Fall $k = 2$:

Definition 1.14. Für $X \in \mathcal{L}^2$ heißt $\text{Var}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2]$ die *Varianz* von X und $\sqrt{\text{Var}[X]}$ die *Standardabweichung* von X .

Während der Erwartungswert eine Maßzahl für den „mittleren“ Wert einer Zufallsvariable ist, dient die Varianz als Maß für die Streuung um den Erwartungswert.

Satz 1.7. Sei $X \in \mathcal{L}^2$ und $a \in \mathbb{R}$. Dann gilt:

$$(i) \text{ Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

$$(ii) \text{ Var}[aX] = a^2 \text{ Var}[X].$$

Beweis. Mit den Rechenregeln vom Erwartungswert erhalten wir

$$\begin{aligned} \text{Var}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] \\ &= \mathbb{E}[X^2] - \mathbb{E}[2X\mathbb{E}[X]] + \mathbb{E}[\mathbb{E}[X]^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2, \text{ und} \end{aligned}$$

$$\begin{aligned} \text{Var}[aX] &= \mathbb{E}[a^2 X^2] - \mathbb{E}[aX]^2 \\ &= a^2 \mathbb{E}[X^2] - a^2 \mathbb{E}[X]^2 \\ &= a^2 \text{ Var}[X] \end{aligned}$$

□

Satz 1.8. Gleichung von Bienayme

Seien $X_1, \dots, X_n$ Zufallsvariablen, die paarweise unkorreliert sind. Dann gilt:

$$\text{Var} \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n \text{Var}[X_i].$$

Beweis. Durch Verwendung der Unkorreliertheit erhalten wir:

$$\begin{aligned} \text{Var} \left[\sum_{i=1}^n X_i \right] &= \mathbb{E} \left[\left(\sum_{i=1}^n X_i \right)^2 \right] - \left(\sum_{i=1}^n \mathbb{E}[X_i] \right)^2 \\ &= \mathbb{E} \left[\sum_{i=1}^n X_i^2 + \sum_{i \neq j} X_i X_j \right] - \sum_{i=1}^n \mathbb{E}[X_i]^2 - \sum_{i \neq j} \mathbb{E}[X_i] \mathbb{E}[X_j] \end{aligned}$$

$$= \sum_{i=1}^n \mathbb{E}[X_i^2] + \underbrace{\sum_{i \neq j} \mathbb{E}[X_i X_j]}_{\sum_{i \neq j} \mathbb{E}[X_i] \mathbb{E}[X_j]} - \sum_{i=1}^n \mathbb{E}[X_i]^2 - \sum_{i \neq j} \mathbb{E}[X_i] \mathbb{E}[X_j].$$

Nach dem vorigen Satz folgt die Behauptung. $\square$

1.4 Wichtige Verteilungen

Hypergeometrische Verteilung

Eine Grundgesamtheit bestehe aus N Objekten, davon seien M vom Typ 1 und $N - M$ vom Typ 2. Es wird nun eine Stichprobe vom Umfang n ohne Zurücklegen gezogen, wovon $k \leq n$ Elemente vom Typ 1 sein sollen. Die gezogene Stichprobe ist eine Teilmenge vom Umfang n aus einer Menge von N Objekten, davon gibt es $\binom{N}{n}$. Um genau k Elemente vom Typ 1 zu ziehen, gibt es $\binom{M}{k} \binom{N-M}{n-k}$ Möglichkeiten. Diese Überlegung führt zur folgenden Definition:

Definition 1.15. Sei $N \in \mathbb{N}$ und $n, M < N$. Eine diskrete Zufallsvariable mit Wertebereich $\Omega' = \{0, 1, \dots, n\}$ und Einzelwahrscheinlichkeiten

$$\mathbb{P}[X = k] = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$$

mit $k \in \Omega'$ heißt *hypergeometrisch-verteilt* oder kurz $H(N, M, n)$ -verteilt.

Die hypergeometrische Verteilung beschreibt also die Wahrscheinlichkeit, dass beim Ziehen ohne Zurücklegen von n aus N Objekten, wovon M eine bestimmte Eigenschaft besitzen, genau k mit dieser Eigenschaft gezogen werden.

Bernoulli-Verteilung

Oft hat man es mit Situationen zu tun, in denen ein zufälliger Vorgang durchgeführt wird, bei dem nur zwei Ausgänge möglich sind (zum Beispiel ein Münzwurf), die durch 0 und 1 repräsentiert werden. Dabei wird der Ausgang 1 mit Wahrscheinlichkeit p meistens als Erfolg, der Ausgang 0 mit Wahrscheinlichkeit $q := 1 - p$ als Misserfolg gedeutet.

Definition 1.16. Sei $n \in \mathbb{N}$ und $p \in [0, 1]$. Eine diskrete Zufallsvariable mit Wertebereich $\{0, 1\}$ und Einzelwahrscheinlichkeiten $\mathbb{P}[X = 1] = p$ und $\mathbb{P}[X = 0] = q = 1 - p$ heißt *Bernoulli-verteilt* oder kurz $Ber(p)$ -verteilt.

Binomialverteilung

Führen wir einen Bernoulli-Versuch mehrmals hintereinander durch (zum Beispiel n -maliger Münzwurf) und wollen die Wahrscheinlichkeit dafür bestimmen, dass wir dabei genau k Erfolge erzielen, so gelangen wir direkt zur Binomialverteilung, die unter den diskreten Verteilungen eine besondere Stellung hat, da sie sehr häufig auftritt.

Definition 1.17. Sei $n \in \mathbb{N}$ und $p \in [0, 1]$. Eine diskrete Zufallsvariable mit Wertebereich $\Omega' = \{0, 1, \dots, n\}$ und Einzelwahrscheinlichkeiten

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

mit $k \in \Omega'$ heißt *binomialverteilt* oder kurz $B(n, p)$ -verteilt.

Betrachtet man das Modell am Anfang des Kapitels mit dem Unterschied, dass die gezogenen Objekte wieder in die Grundgesamtheit zurückgelegt werden, so erhält man die Binomialverteilung mit dem Parameter $p = \frac{M}{N}$.

Poisson-Verteilung

Modelliert man Bernoulli-Experimente mit geringer Erfolgswahrscheinlichkeit, so eignet sich eine andere Verteilung zur Berechnung. Später werden wir sehen, dass diese aus der Approximation der Binomialverteilung mit großem n und kleinem p entsteht.

Definition 1.18. Sei $\lambda > 0$. Eine diskrete Zufallsvariable mit Wertebereich $\Omega' = \{0, 1, 2, \dots\}$ und Einzelwahrscheinlichkeiten

$$\mathbb{P}[X = k] = \frac{\lambda^k}{k!} e^{-\lambda}$$

mit $k \in \Omega'$ heißt *Poisson-verteilt* oder kurz $P(\lambda)$ -verteilt.

Die Poisson-Verteilung kommt häufig zum Einsatz, wenn man Wahrscheinlichkeiten für „seltene“ Ereignisse innerhalb einer bestimmten Zeitspanne berechnen möchte.

Normalverteilung

Bisher haben wir nur Wahrscheinlichkeitsverteilungen von diskreten Zufallsvariablen betrachtet. Der wichtigste Fall ist jedoch, wenn X die Dichtefunktion einer Normalverteilung besitzt.

Definition 1.19. Sei $\mu \in \mathbb{R}$ und $\sigma^2 > 0$. Eine stetige Zufallsvariable mit Wertebereich

reich $\mathbb{R}$ und Wahrscheinlichkeitsdichte

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

für $x \in \mathbb{R}$ heißt *normalverteilt* oder kurz $N(\mu, \sigma^2)$ -verteilt.

Bemerkung: Ist $\mu = 0$ und $\sigma^2 = 1$, so spricht man von der Standardnormalverteilung. Ist eine Zufallsvariable X $N(\mu, \sigma^2)$ -verteilt, so gelangt man durch die Transformation $\frac{X-\mu}{\sigma}$ zu einer $N(0, 1)$ -verteilten Zufallsvariable. Aufgrund der besonderen Bedeutung der Standardnormalverteilung setzt man

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

und

$$\Phi(t) := \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^t \varphi(x) dx.$$

1.5 Charakteristische Funktionen

Charakteristische Funktionen sind ein wichtiges analytisches Hilfsmittel der Wahrscheinlichkeitstheorie. Sie sind neben den Verteilungsfunktionen eine weitere Beschreibungsmöglichkeit von Zufallsvariablen und sind bei vielen Beweisen von Grenzwertsätzen sehr hilfreich. Als Hauptquellen für diesen Abschnitt dienen [2] und [15].

1.5.1 Definition und Beispiele

Um charakteristische Funktionen betrachten zu können, benötigen wir komplexwertige Zufallsvariablen und deren Erwartungswert. Dabei ist $Z = X + iY$ eine komplexwertige Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$, wenn X und Y reellwertige Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathbb{P})$ sind. Beim Erwartungswert zerlegt man Z einfach in Real- und Imaginärteil, also $\mathbb{E}[Z] = \mathbb{E}[X] + i\mathbb{E}[Y]$.

Definition 1.20. Sei X eine reellwertige Zufallsvariable und $t \in \mathbb{R}$. Dann ist die *charakteristische Funktion* $\varphi_X : \mathbb{R} \rightarrow \mathbb{C}$ von X definiert durch

$$\varphi_X(t) = \mathbb{E}[e^{itX}].$$

Bemerkung: 1. Oft wird die bekannte Euler-Relation verwendet:

$$e^{itX} = \cos(tX) + i \sin(tX)$$

2. Ist X eine diskrete Zufallsvariable, so ist

$$\varphi_X(t) = \sum_{x \in X(\Omega)} e^{itx} \mathbb{P}[X = x].$$

3. Ist X eine stetige Zufallsvariable mit Dichte $f(x)$, so ist

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f(x) dx.$$

Im Folgenden werden die Beweise immer für stetige Zufallsvariablen mit Dichte $f(x)$ geführt, für diskrete funktionieren sie analog. Wegen $|e^{itX}| = 1$ ist

$$\int_{-\infty}^{\infty} |e^{itx}| f(x) dx = \int_{-\infty}^{\infty} f(x) dx = 1$$

und daher existiert zu jeder Zufallsvariable eine charakteristische Funktion. Außerdem sind charakteristische Funktionen gleichmäßig stetig. Einen Beweis dazu findet man in [2].

Beispiel 1.2: Charakteristische Funktionen von wichtigen Verteilungen

Sei X $Ber(p)$ -verteilt. Dann ist

$$\begin{aligned} \varphi_X(t) &= e^{it0} \mathbb{P}[X = 0] + e^{it1} \mathbb{P}[X = 1] \\ &= (1 - p) + e^{it} p. \end{aligned}$$

Sei X $B(n, p)$ -verteilt. Dann ist

$$\begin{aligned} \varphi_X(t) &= \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} e^{itk} \\ &= \sum_{k=0}^n \binom{n}{k} (pe^{it})^k (1-p)^{n-k} \\ &= (pe^{it} + (1-p))^n \\ &= (p(e^{it} - 1) + 1)^n. \end{aligned}$$

Sei X $P(\lambda)$ -verteilt. Dann ist

$$\begin{aligned} \varphi_X(t) &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} e^{itk} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} \\ &= e^{-\lambda} e^{\lambda e^{it}} \\ &= e^{\lambda(e^{it} - 1)}. \end{aligned}$$

Sei X $N(0, 1)$ -verteilt. Dann ist

$$\begin{aligned}
\varphi_X(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(tx) e^{-\frac{x^2}{2}} dx + \underbrace{\frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sin(tx) e^{-\frac{x^2}{2}} dx}_{=0} \\
\Rightarrow \varphi'_X(t) &= \int_{-\infty}^{\infty} -x \sin(tx) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \left(\underbrace{\left[-\cos(tx) e^{-\frac{x^2}{2}} \right]_{x=-\infty}^{\infty}}_{=0} + \int_{-\infty}^{\infty} t \cos(tx) e^{-\frac{x^2}{2}} dx \right) \\
&= \frac{1}{\sqrt{2\pi}} t \int_{-\infty}^{\infty} \cos(tx) e^{-\frac{x^2}{2}} dx \\
&= t \varphi_X(t).
\end{aligned}$$

Durch Lösen der Differentialgleichung $\varphi'_X(t) = t \varphi_X(t)$ mit $\varphi_X(0) = 1$ erhält man $\varphi_X(t) = e^{-\frac{t^2}{2}}$.

1.5.2 Wichtige Eigenschaften

Satz 1.9. Für eine charakteristische Funktion $\varphi_X(t)$ gilt:

- (i) $\varphi_X(0) = 1$.
- (ii) $|\varphi_X(t)| \leq 1$.
- (iii) $\varphi_X(-t) = \overline{\varphi_X(t)}$.
- (iv) $\forall a, b \in \mathbb{R} : \varphi_{aX+b}(t) = e^{ibt} \varphi_X(at)$.

Beweis. (i) $\varphi_X(0) = \int_{-\infty}^{\infty} f(x) dx = 1$.

$$(ii) \quad |\varphi_X(t)| = \left| \int_{-\infty}^{\infty} e^{itx} f(x) dx \right| \leq \int_{-\infty}^{\infty} |e^{itx}| f(x) dx = 1.$$

$$\begin{aligned}
(iii) \quad \varphi_X(-t) &= \int_{-\infty}^{\infty} e^{-itx} f(x) dx = \int_{-\infty}^{\infty} (\cos(-tx) + i \sin(-tx)) f(x) dx = \\
&= \int_{-\infty}^{\infty} (\cos(tx) - i \sin(tx)) f(x) dx = \overline{\varphi_X(t)}.
\end{aligned}$$

$$(iv) \quad \varphi_{aX+b}(t) = \mathbb{E}[e^{it(aX+b)}] = \mathbb{E}[e^{itb} e^{itaX}] = e^{itb} \mathbb{E}[e^{i(at)X}] = e^{itb} \varphi_X(at). \quad \square$$

Beispiel 1.3: Sei X $N(0, 1)$ -verteilt. Dann ist $\mu + \sigma X$ $N(\mu, \sigma^2)$ -verteilt. Mithilfe des vorigen Satzes erhalten wir die charakteristische Funktion einer $N(\mu, \sigma^2)$ -verteilten Zufallsvariable durch:

$$\varphi_{\mu+\sigma X}(t) = e^{it\mu} \varphi_X(\sigma t) = e^{it\mu} e^{-\frac{\sigma^2 t^2}{2}} = e^{t(i\mu - \frac{\sigma^2 t}{2})}.$$

Satz 1.10. Seien $X_1, \dots, X_n$ unabhängige Zufallsvariablen mit charakteristischen Funktionen $\varphi_{X_1}(t), \dots, \varphi_{X_n}(t)$. Dann gilt für $S_n := X_1 + \dots + X_n$:

$$\varphi_{S_n}(t) = \varphi_{X_1}(t) \varphi_{X_2}(t) \dots \varphi_{X_n}(t).$$

Beweis. Durch Einsetzen in die Definition und Verwendung der Unabhängigkeit erhalten wir:

$$\begin{aligned} \varphi_{S_n}(t) &= \mathbb{E}[e^{it(X_1 + \dots + X_n)}] \\ &= \mathbb{E}[e^{itX_1} e^{itX_2} \dots e^{itX_n}] \\ &= \mathbb{E}[e^{itX_1}] \mathbb{E}[e^{itX_2}] \dots \mathbb{E}[e^{itX_n}] \\ &= \varphi_{X_1}(t) \varphi_{X_2}(t) \dots \varphi_{X_n}(t). \end{aligned} \quad \square$$

Dieser Satz ist ein sehr nützliches Resultat, da die Addition unabhängiger Zufallsvariablen gewöhnlich zu komplizierten Operationen führt. Bei den charakteristischen Funktionen kann dies einfach durch die Multiplikation der einzelnen charakteristischen Funktionen ersetzt werden.

Im Weiteren beleuchten wir noch den Zusammenhang von charakteristischen Funktionen und den Momenten einer Zufallsvariable:

Satz 1.11. Sei X eine Zufallsvariable und $n \in \mathbb{N}$. Existiert $\mathbb{E}[X^n]$, so ist $\varphi_X(t)$ eine n -mal differentierbare Funktion und es gilt für $k \leq n$:

$$\mathbb{E}[X^k] = \frac{\varphi_X^{(k)}(0)}{i^k}.$$

Beweis. Sei $f(x)$ die Dichtefunktion von X . Differenzieren wir $\varphi_X(t)$ n -mal, so erhalten wir

$$\varphi_X^{(n)}(t) = i^n \int_{-\infty}^{\infty} x^n e^{itx} f(x) dx$$

Diese Differentiation ist gerechtfertigt, da nach Voraussetzung $\int_{-\infty}^{\infty} |x^n| f(x) dx < \infty$ gilt und daher

$$\left| \int_{-\infty}^{\infty} x^n e^{itx} f(x) dx \right| \leq \int_{-\infty}^{\infty} |x^n| f(x) dx < \infty$$

ist. Somit ist $\varphi_X^{(n)}(t)$ beschränkt und existiert daher. Für $k \leq n$ gilt daher:

$$\varphi_X^{(k)}(0) = i^k \int_{-\infty}^{\infty} x^k e^{i0x} f(x) dx = i^k \mathbb{E}[X^k]. \quad \square$$

Beispiel 1.4: Sei X $N(\mu, \sigma^2)$ -verteilt. Dann ist $\varphi'_X(t) = (i\mu - \sigma^2 t) e^{t(i\mu - \frac{\sigma^2 t}{2})}$. Daraus folgt:

$$\mathbb{E}[X] = \frac{1}{i} \varphi'_X(0) = \frac{1}{i} (i\mu) = \mu.$$

Satz 1.12. Für eine Zufallsvariable $X \in \mathcal{L}^n$ gilt:

$$\varphi_X(t) = \sum_{k=0}^n \frac{(it)^k \mathbb{E}[X^k]}{k!} + R_{n+1}(t)$$

mit $R_{n+1}(t) = o(t^n)$, das heißt $\lim_{t \rightarrow 0} \frac{R_{n+1}(t)}{|t|^n} = 0$.

Beweis. Entwicklung von e^{itX} in eine Taylorreihe um $t = 0$ ergibt

$$\begin{aligned} e^{itX} &= 1 + itX + \frac{(itX)^2}{2!} + \frac{(itX)^3}{3!} + \dots \\ &= \sum_{k=0}^n \frac{(itX)^k}{k!} + S_{n+1}(t) \end{aligned}$$

mit $S_{n+1}(t) = o(t^n)$.

Somit erhalten wir für die charakteristische Funktion

$$\begin{aligned} \varphi_X(t) &= \mathbb{E}[e^{itX}] = \mathbb{E}[1] + \mathbb{E}[itX] + \mathbb{E}\left[\frac{(itX)^2}{2!}\right] + \mathbb{E}\left[\frac{(itX)^3}{3!}\right] + \dots \\ &= 1 + it\mathbb{E}[X] + \frac{(it)^2}{2!} \mathbb{E}[X^2] + \frac{(it)^3}{3!} \mathbb{E}[X^3] + \dots \\ &= \sum_{k=0}^n \frac{(it)^k \mathbb{E}[X^k]}{k!} + R_{n+1}(t) \end{aligned}$$

mit $R_{n+1}(t) = o(t^n)$. $\square$

1.5.3 Eindeutigkeits- und Stetigkeitssatz

In diesem Abschnitt soll zunächst die zentrale Aussage, dass eine Wahrscheinlichkeitsverteilung *eindeutig* durch ihre charakteristische Funktion festgelegt ist, behandelt werden. Anschließend soll der für uns zentrale Stetigkeitssatz angeführt werden. Dieser ist das entscheidende Hilfsmittel bei den zentralen Grenzwertsätzen. Die Sätze werden in diesem Abschnitt nicht bewiesen, sondern nur auf die Literatur verwiesen, da sie sehr aufwendig und rein technisch sind. Bisher haben wir immer durch die Vorgabe einer Verteilung eine charakteristische Funktion gefunden. Der folgende Satz zeigt, dass auch die Umkehrung möglich ist:

Satz 1.13. *Umkehrformeln*

Sei X eine stetige Zufallsvariable mit charakteristischer Funktion $\varphi_X(t)$ und es gelte $\int_{-\infty}^{\infty} |\varphi_X(t)| dt < \infty$. Für die Wahrscheinlichkeitsdichte $f(x)$ von X gilt dann:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi_X(t) dt.$$

An Stetigkeitsstellen a und b mit $a < b$ gilt für die Verteilungsfunktion F außerdem:

$$F(b) - F(a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t) dt.$$

Beweis. Einen Beweis findet man in [2]. □

Beispiel 1.5: Sei X eine stetige Zufallsvariable mit $\varphi_X(t) = e^{-\frac{t^2}{2}}$. Mithilfe der Umkehrformel können wir die Dichte von X bestimmen:

$$\begin{aligned} f(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} e^{-\frac{t^2}{2}} dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{(t+ix)^2}{2}} e^{\frac{(ix)^2}{2}} dt \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(t+ix)^2}{2}} dt}_{=1} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \end{aligned}$$

Wir haben also eine eindeutige Zuordnung von Wahrscheinlichkeitsverteilung und charakteristischen Funktionen und erhalten somit folgenden bedeutenden Satz, der es rechtfertigt, von *der* charakteristischen Funktion einer Zufallsvariable zu sprechen.

Satz 1.14. *Eindeutigkeitssatz*

Sind X und Y Zufallsvariablen mit $\varphi_X(t) = \varphi_Y(t) \forall t \in \mathbb{R}$, so haben X und Y dieselbe Wahrscheinlichkeitsverteilung. Das heißt, die Wahrscheinlichkeitsverteilung einer Zufallsvariable ist durch ihre charakteristische Funktion eindeutig bestimmt.

Beweis. Einen Beweis dazu findet man in [2]. □

Beispiel 1.6: Seien X_1 und X_2 unabhängig und $N(\mu_1, \sigma_1^2)$ bzw. $N(\mu_2, \sigma_2^2)$ -verteilt. Wir suchen die Verteilung von $X_1 + X_2$. Mit Satz 1.10 erhalten wir

$$\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t) \varphi_{X_2}(t) = e^{t(i\mu_1 - \frac{\sigma_1^2 t}{2})} e^{t(i\mu_2 - \frac{\sigma_2^2 t}{2})}$$

$$= it(\mu_1 + \mu_2) - \frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}.$$

Dies ist die charakteristische Funktion einer $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ -verteilten Zufallsvariable und aufgrund des Eindeutigkeitssatzes ist $X_1 + X_2$ daher $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ -verteilt.

Nun können wir den für uns entscheidenden Satz formulieren, der einen Bezug zwischen der Konvergenz nach Verteilung (siehe 2. Kapitel) und der zugehörigen charakteristischen Funktion herstellt.

Satz 1.15. *Stetigkeitssatz von Lévy und Cramer*

Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen mit Verteilungsfunktionen $F_{X_n}(x)$ und charakteristischen Funktionen $\varphi_{X_n}(t)$. Dann sind folgende Aussagen äquivalent:

(i) $X_n \xrightarrow{d} X.$

(ii) $\varphi_{X_n}(t) \longrightarrow \varphi_X(t) \quad \forall t \in \mathbb{R}.$

Beweis. Der Beweis ist in [2] oder [15] zu finden. □

2 Konvergenzbegriffe

Bevor wir Grenzwertsätze und andere Approximationen studieren können, müssen wir zunächst definieren, wie die Konvergenz von Zufallsvariablen zu verstehen ist. In der Wahrscheinlichkeitstheorie sind dabei mehrere Konvergenzbegriffe gebräuchlich. Im Folgenden sollen diese behandelt und verglichen werden. Dazu betrachten wir eine Folge $(X_n)_{n \in \mathbb{N}}$ von Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Wir interessieren uns nun für das asymptotische Verhalten dieser Folge. Anders als bei den klassischen Folgen in der Analysis kann es in der Wahrscheinlichkeitstheorie keine „absolute“ Konvergenz geben. Der Grund dafür ist, dass wir bestimmte Realisationen nicht ausschließen können, auch wenn sie eine infinitesimale Wahrscheinlichkeit haben. Wenn wir beispielsweise eine Münze 100000 Mal werfen, so liegt auch das Ereignis, dass 100000 Mal Kopf erscheint im Ereignisraum, obwohl es „fast sicher“ nicht eintreten wird. Die Darstellung der Inhalte orientiert sich vor allem an den Werken [4] und [17].

2.1 Konvergenz nach Verteilung

Wie der Name schon besagt, betrachtet man hier Folgen von Verteilungsfunktionen. Der Konvergenzbegriff wird dabei über die punktweise Konvergenz dieser Funktionen eingeführt. Anwendung findet diese Konvergenzart vor allem in den Zentralen Grenzwertsätzen.

Definition 2.1. Sei $(F_n(x))_{n \in \mathbb{N}}$ die Folge der zugehörigen Verteilungsfunktionen von $(X_n)_{n \in \mathbb{N}}$. Dann konvergiert $(X_n)_{n \in \mathbb{N}}$ *nach Verteilung* gegen eine Zufallsvariable X , wenn für jedes $x \in \mathbb{R}$, in dem $F(x)$ stetig ist, gilt dass

$$\lim_{n \rightarrow \infty} F_n(x) = F(x).$$

Bemerkung: 1. Konvergiert X_n in Verteilung gegen X , so schreiben wir auch kurz $X_n \xrightarrow{d} X$.

2. Nur wenn $F(x)$ auch tatsächlich eine Verteilungsfunktion ist, sagt man, dass die Folge der Verteilungsfunktionen konvergiert. Es kann auch vorkommen, dass die Verteilungsfunktionen gegen eine Funktion konvergieren, die keine Verteilungsfunktion ist.
3. Streng genommen benötigt man keine Zufallsvariablen, sondern lediglich ihre Verteilungsfunktionen. Die Zufallsvariablen müssen demnach nicht einmal auf dem gleichen Wahrscheinlichkeitsraum definiert sein.

2.2 Stochastische Konvergenz

Dieser Konvergenzbegriff wird später beim Schwachen Gesetz der Großen Zahl verwendet werden.

Definition 2.2. Eine Folge $(X_n)_{n \in \mathbb{N}}$ von Zufallsvariablen konvergiert *stochastisch* gegen eine Zufallsvariable X , wenn $\forall \varepsilon > 0$ gilt:

$$\lim_{n \rightarrow \infty} \mathbb{P}[|X_n - X| > \varepsilon] = 0.$$

Bemerkung: Konvergiert X_n stochastisch gegen X , so schreiben wir auch kurz $X_n \xrightarrow{\mathbb{P}} X$. Diese Definition besagt nicht, dass die Zufallsvariablen X_n im Sinne des Grenzwertbegriffs der Analysis gegen X konvergiert. Es folgt daraus nicht, dass man zu jedem $\varepsilon > 0$ ein $N \in \mathbb{N}$ angeben kann, sodass $\forall n \geq N$ die Beziehung $|X_n - X| < \varepsilon$ gilt. Die Definition besagt nur, dass die Wahrscheinlichkeit, dass $|X_n - X| > \varepsilon$ ist, für $n \rightarrow \infty$ gegen 0 geht.

2.3 Fast sichere Konvergenz

Dieser Konvergenzbegriff macht Rückgriff auf die punktweise Konvergenz von Funktionenfolgen, ist jedoch etwas schwächer, da hier die Konvergenz nur für fast alle Elemente von Ω verlangt wird. Es handelt sich dabei um die stärkste Konvergenzart der Stochastik. Deshalb spricht man bei den Anwendungen auch vom Starken Gesetz der Großen Zahlen.

Definition 2.3. Eine Folge $(X_n)_{n \in \mathbb{N}}$ von Zufallsvariablen konvergiert fast sicher gegen eine Zufallsvariable X , wenn

$$\mathbb{P}[\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}] = \mathbb{P}[\lim_{n \rightarrow \infty} X_n = X] = 1.$$

Bemerkung: Konvergiert X_n fast sicher gegen X , so schreiben wir auch kurz $X_n \xrightarrow{f.s.} X$.

Wir wollen nun noch einerseits ein Kriterium für die fast sichere Konvergenz geben und andererseits einen wichtigen Konvergenzsatz anführen, der eine Aussage über die Konvergenz der Erwartungswerte macht.

Satz 2.1. $X_n \xrightarrow{f.s.} X \Leftrightarrow \sup_{k \geq n} |X_k - X| \xrightarrow{\mathbb{P}} 0.$

Beweis. ($\Rightarrow$) Für jedes ε sei

$$A_\varepsilon = \limsup_{n \rightarrow \infty} \{|X_n - X| > \varepsilon\} = \bigcap_n \bigcup_{k \geq n} \{|X_k - X| > \varepsilon\}.$$

Da $X_n \xrightarrow{f.s.} X$ folgt $\mathbb{P}[A_\varepsilon] = 0$. Außerdem gilt:

$$0 = \mathbb{P}[A_\varepsilon] = \lim_{n \rightarrow \infty} \mathbb{P} \left[\bigcup_{k \geq n} \{|X_k - X| > \varepsilon\} \right] = \lim_{n \rightarrow \infty} \mathbb{P} \left[\sup_{k \geq n} \{|X_k - X| > \varepsilon\} \right]$$

und somit $\sup_{k \geq n} |X_k - X| \xrightarrow{\mathbb{P}} 0$.

($\Leftarrow$) Falls $\sup_{k \geq n} |X_k - X| \xrightarrow{\mathbb{P}} 0$ gilt, folgt aus der obigen Gleichungskette $\mathbb{P}[A_\varepsilon] = 0$. Ist $\omega \in \{\omega \in \Omega : X_n(\omega) \xrightarrow{n \rightarrow \infty} X(\omega)\}$, so gilt:

$$\forall m \in \mathbb{N} \exists \varepsilon > \frac{1}{m} : \forall n \in \mathbb{N} \exists k \geq n : |X_k(\omega) - X(\omega)| > \varepsilon > \frac{1}{m}$$

und somit erhalten wir

$$\mathbb{P} \left[\{\omega \in \Omega : X_n(\omega) \xrightarrow{n \rightarrow \infty} X(\omega)\} \right] = \mathbb{P} \left[\bigcup_{m=1}^{\infty} A_{\frac{1}{m}} \right] \leq \sum_{m=1}^{\infty} \mathbb{P}[A_{\frac{1}{m}}] = 0.$$

Daher gilt $X_n \xrightarrow{f.s.} X$. □

Analog kann man zeigen, dass $(X_n)_{n \in \mathbb{N}}$ genau dann fast sicher konvergiert, wenn $\sup_{k \geq n} |X_k - X_n| \xrightarrow{\mathbb{P}} 0$.

Satz 2.2. Majorisierte Konvergenz

Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen. Gilt $X_n \xrightarrow{f.s.} X$ und existiert eine Zufallsvariable Y mit $\mathbb{E}[Y] < \infty$ und $|X_n| < Y$ fast sicher, dann ist

$$\lim_{n \rightarrow \infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Beweis. Einen Beweis findet man in [4], wo der Satz gleich in allgemeinen Maßräumen bewiesen wird. □

2.4 Vergleich der Konvergenzbegriffe

Zunächst muss festgehalten werden, dass sich die Konvergenz in Verteilung insofern von den beiden anderen Konvergenzarten unterscheidet, als dass die Folge $(X_n)_{n \in \mathbb{N}}$ von Zufallsvariablen nicht auf demselben Wahrscheinlichkeitsraum definiert sein muss. Außerdem ist die Verteilungskonvergenz der schwächste dieser drei Konvergenzbegriffe.

Satz 2.3. *Fast sichere Konvergenz impliziert stochastische Konvergenz. Die Umkehrung gilt nicht.*

Beweis. Aus $X_n \xrightarrow{f.s.} X$ folgt nach dem vorigen Satz $\sup_{k \geq n} |X_k - X| \xrightarrow{\mathbb{P}} 0$ und somit auch $|X_n - X| \xrightarrow{\mathbb{P}} 0$, also $X_n \xrightarrow{\mathbb{P}} X$.

Folgendes Gegenbeispiel zeigt, dass die Umkehrung nicht gilt:

Sei $\Omega = [0, 1)$ gleichverteilt bezüglich $\mathbb{P}$, das heißt $\mathbb{P}[X \in [a, b]] = b - a$.

Setze $A_1 = [0, \frac{1}{2})$, $A_2 = [\frac{1}{2}, 1)$, $A_3 = [0, \frac{1}{4}]$, $A_4 = [\frac{1}{4}, \frac{1}{2})$, $A_5 = [\frac{1}{2}, \frac{3}{4})$, $A_6 = [\frac{3}{4}, 1)$, usw.

Sei $X_n := \mathbf{1}_{A_n}$. Dann konvergiert X_n stochastisch gegen 0, denn für $0 < \varepsilon < 1$ gilt:

$$\mathbb{P}[|X_n - 0| \geq \varepsilon] = \mathbb{P}[A_n] \longrightarrow 0.$$

Andererseits findet man für jedes $\omega \in \Omega$ und jedes N Zahlen $n, m \geq N$, sodass $\omega \in A_m$ und $\omega \in A_n^c$. Somit gilt für jedes ω :

$$\liminf_{n \rightarrow \infty} X_n(\omega) = 0 \quad \text{und} \quad \limsup_{n \rightarrow \infty} X_n(\omega) = 1.$$

Daher konvergiert X_n nicht fast sicher. □

Satz 2.4. *Die stochastische Konvergenz impliziert die Konvergenz in Verteilung. Die Umkehrung gilt nicht.*

Beweis. Sei $F_n(x)$ die Verteilungsfunktion von X_n . Wähle zunächst drei reelle Zahlen $x < y < z$. Dann gilt:

$$\begin{aligned} \{X \leq x\} &= \{X \leq x \cap X_n \leq y\} \cup \{X \leq x \cap X_n > y\} \\ &\subseteq \{X_n \leq y\} \cup \{X \leq x \cap X_n > y\}. \end{aligned}$$

Daraus folgt $F(x) \leq F_n(y) + \mathbb{P}[|X_n - X| \geq y - x]$ und wegen $X_n \xrightarrow{\mathbb{P}} X$ folgt $F(x) \leq \liminf_{n \rightarrow \infty} F_n(y)$. Ebenso ist

$$\begin{aligned} \{X_n \leq y\} &= \{X_n \leq y \cap X \leq z\} \cup \{X_n \leq y \cap X > z\} \\ &\subseteq \{X \leq z\} \cup \{X_n \leq y \cap X > z\}. \end{aligned}$$

Daraus folgt $F_n(y) \leq F(z) + \mathbb{P}[|X_n - X| \geq z - y]$ und somit $\limsup_{n \rightarrow \infty} F_n(y) \leq F(z)$.

Insgesamt erhalten wir also

$$F(x) \leq \liminf_{n \rightarrow \infty} F_n(y) \leq \limsup_{n \rightarrow \infty} F_n(y) \leq F(z).$$

Ist y eine Stetigkeitsstelle, so folgt für $x \uparrow y$ und $z \downarrow x$ zunächst $F(x) \longrightarrow F(y)$ und $F(z) \longrightarrow F(y)$. Somit gilt

$$\lim_{n \rightarrow \infty} F_n(y) = F(y).$$

Folgendes Gegenbeispiel zeigt, dass die Umkehrung nicht gilt:

Sei X $N(0, 1)$ -verteilt und $X_n := -X$. Da die Verteilung von X symmetrisch ist, gilt $X_n = -X = X$ und damit $X_n \xrightarrow{d} X$. Es gilt jedoch

$$\mathbb{P}[|X_n - X| > \varepsilon] = \mathbb{P}[|2X| > \varepsilon] = \mathbb{P}\left[|X| > \frac{\varepsilon}{2}\right] = c \quad \text{mit } c \in \mathbb{R}.$$

Somit konvergiert X_n nicht stochastisch. □

Satz 2.5. *Die Umkehrung des vorigen Satzes gilt, wenn X konstant ist, das heißt: Falls $X_n \xrightarrow{d} X \equiv c$ für ein $c \in \mathbb{R}$, so gilt $X_n \xrightarrow{\mathbb{P}} X$.*

Beweis. Sei $X_n \xrightarrow{d} X \equiv c$. Dann gilt:

$$F_{X_n}(x) \xrightarrow{n \rightarrow \infty} F_c(x) = \begin{cases} 1, & x \geq c \\ 0, & x < c \end{cases}$$

Daraus folgt $\forall \varepsilon > 0$ und $\forall n \in \mathbb{N}$:

$$\begin{aligned} \mathbb{P}[|X_n - c| \geq \varepsilon] &= \mathbb{P}[X_n \leq c - \varepsilon] + \mathbb{P}[X_n \geq c + \varepsilon] \\ &= F_{X_n}(c - \varepsilon) + (1 - F_{X_n}(c + \varepsilon) - \mathbb{P}[X_n = c + \varepsilon]) \\ &\xrightarrow{n \rightarrow \infty} 0 + 1 - 1 + 0. \end{aligned}$$

Somit gilt $X_n \xrightarrow{\mathbb{P}} X$. □

Insgesamt gilt demnach folgende Beziehung der Konvergenzbegriffe:

$$\boxed{X_n \xrightarrow{f.s.} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X \Rightarrow X_n \xrightarrow{d} X}$$

3 Gesetze der Großen Zahlen

Wir betrachten eine Folge von unabhängigen Zufallsvariablen $X_1, X_2, \dots, X_n$ und fragen uns nach dem Konvergenzverhalten der arithmetischen Mittelwerte $\frac{1}{n} \sum_{i=1}^n X_i$. Man betrachtet dabei also eine oftmalige Versuchsserie eines Zufallsexperimentes, deren Ergebnisse addiert und anschließend durch n normiert werden. Wird beispielsweise eine faire Münze geworfen, so rechnen wir intuitiv damit, dass etwa die Hälfte der Ergebnisse Kopf sein wird. Genauso erwarten wir, dass wir mit einem Würfel auf lange Zeit in etwa einem Sechstel der Versuche die Zahl 6 werfen werden. In diesem Kapitel sollen diese intuitiven Aussagen präzise formuliert werden. Die Gesetze der Großen Zahlen besagen in etwa, dass sich dieser Mittelwert unter bestimmten Voraussetzungen asymptotisch einer konstanten Zahl, in den angeführten Beispielen der Erfolgswahrscheinlichkeit $\frac{1}{2}$ bzw. $\frac{1}{6}$, annähern wird. Die Aussagen haben eine fundamentale Bedeutung und bilden die Grundlage dafür, dass bei nicht-axiomatischen Zugängen zur Wahrscheinlichkeitstheorie die Wahrscheinlichkeit als relative Häufigkeit definiert wird. Je nach Konvergenzart unterscheidet man zwischen den Schwachen und den Starken Gesetzen der Großen Zahlen. Zuvor benötigen wir jedoch zwei wichtige Ungleichungen, die bei den Beweisen immer wieder verwendet werden. Die Sätze, Definitionen etc. wurden großteils in [4], [5], [7], [9] und [17] gefunden.

3.1 Ungleichungen

3.1.1 Tschebyscheff-Ungleichung

Lemma 3.1. *Markov-Ungleichung*

Sei $X \in \mathcal{L}^1$ eine nichtnegative Zufallsvariable. Dann gilt für jedes $a \in \mathbb{R}^+$:

$$\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}.$$

Beweis. Wir beweisen den Satz für stetige Zufallsvariablen, für diskrete funktioniert er analog. Sei $f(x)$ die Dichte von X . Dann gilt:

$$\begin{aligned} \mathbb{E}[X] &= \int_0^\infty x f(x) dx = \int_0^a x f(x) dx + \int_a^\infty x f(x) dx \\ &\geq \int_a^\infty x f(x) dx \\ &\geq a \int_a^\infty f(x) dx \\ &= a \mathbb{P}[X \geq a]. \end{aligned}$$

□

Satz 3.2. Tschebyscheff-Ungleichung

Sei $X \in \mathcal{L}^2$ und $\varepsilon > 0$. Dann gilt:

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq \varepsilon] \leq \frac{\text{Var}[X]}{\varepsilon^2}.$$

Beweis. Wir betrachten oBdA die Zufallsvariable $Y = (X - \mathbb{E}[X])^2$. Diese ist nicht-negativ und somit kann die Markov-Ungleichung angewendet werden:

$$\begin{aligned} \mathbb{P}[|X - \mathbb{E}[X]| \geq \varepsilon] &= \mathbb{P}[(X - \mathbb{E}[X])^2 \geq \varepsilon^2] \\ &\leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{\varepsilon^2} \\ &= \frac{\text{Var}[X]}{\varepsilon^2}. \end{aligned} \quad \square$$

Die Abschätzung der Tschebyscheff-Ungleichung ist sehr grob und liefert oft keine brauchbare Information. Der Wert liegt vor allem in der universellen Handhabbarkeit und im hohen theoretischen Nutzen. Sie kann auch angewendet werden, wenn man von einer Verteilung nur den Erwartungswert und die Varianz kennt. Aus der Tschebyscheff-Ungleichung geht auch die Bedeutung der Varianz als Maß für die Streuung hervor. Je kleiner die Varianz, desto größer die Wahrscheinlichkeit, dass X Werte innerhalb eines kleinen Intervalls um $\mathbb{E}[X]$ annimmt.

Beispiel 3.1: Aus langjähriger Erfahrung ist einem Schraubenhersteller bekannt, dass die Schraubenlänge einen Erwartungswert von 20mm und eine Varianz von 0.0225mm^2 hat. Eine Schraube, die mehr als 0.3mm vom Erwartungswert abweicht, gilt als Ausschuss. Wie groß ist der Ausschussanteil höchstens?

Sei X die Schraubenlänge. Nach der Tschebyscheff-Ungleichung gilt:

$$\mathbb{P}[|X - 20| \geq 0.3] \leq \frac{0.0225}{0.09} = 0.25$$

Obwohl man die Verteilung von X nicht kennt, ist der Ausschuss im ungünstigsten Fall höchstens 25 Prozent. Würde man die Verteilung kennen, wären natürlich genauere Abschätzungen möglich.

3.1.2 Kolmogorov-Ungleichung

Für die Starken Gesetze der Großen Zahlen benötigen wir manchmal noch eine stärkere Aussage als die Tschebyscheff-Ungleichung. Man spricht dabei von einer Maximalungleichung, weil eine Aussage über das Maximum von $\sum_{i=1}^n X_i$ gemacht wird.

Satz 3.3. Kolmogorov-Ungleichung

Seien $X_1, X_2, \dots, X_n$ unabhängige Zufallsvariablen in $\mathcal{L}^2$ und $\mathbb{E}[X_i] = 0$ für $i = 1, \dots, n$. Sei $S_k = \sum_{i=1}^k X_i$. Dann gilt $\forall \varepsilon > 0$:

$$\mathbb{P} \left[\max_{1 \leq k \leq n} |S_k| \geq \varepsilon \right] \leq \frac{1}{\varepsilon^2} \sum_{k=1}^n \text{Var}[X_k].$$

Beweis. Wir setzen $A_1 = \{|S_1| \geq \varepsilon\}$ und $A_k = \{|S_1| < \varepsilon, \dots, |S_{k-1}| < \varepsilon, |S_k| \geq \varepsilon\}$ für $k = 2, \dots, n$. Diese Ereignisse sind paarweise disjunkt und es gilt:

$$\bigcup_{k=1}^n A_k = \left\{ \max_{1 \leq k \leq n} |S_k| \geq \varepsilon \right\}.$$

Weiters gilt, dass für jedes k die Zufallsvariablen $\mathbb{1}_{A_k} S_k$ und $S_n - S_k$ unabhängig sind, da $\mathbb{1}_{A_k} S_k$ nur von $X_1, \dots, X_k$ und $S_n - S_k$ nur von $X_{k+1}, \dots, X_n$ abhängt. Außerdem ist nach Voraussetzung $\mathbb{E}[S_n - S_k] = 0$. Daraus folgt:

$$\mathbb{E}[\mathbb{1}_{A_k} S_k (S_n - S_k)] = \mathbb{E}[\mathbb{1}_{A_k} S_k] \mathbb{E}[(S_n - S_k)] = 0.$$

Insgesamt erhalten wir somit

$$\begin{aligned} \sum_{k=1}^n \text{Var}[X_k] &= \sum_{k=1}^n (\mathbb{E}[X_k^2] - (\mathbb{E}[X_k])^2) \\ &= \mathbb{E}[S_n^2] \\ &\geq \sum_{k=1}^n \mathbb{E}[\mathbb{1}_{A_k} S_n^2] \\ &= \sum_{k=1}^n \mathbb{E}[\mathbb{1}_{A_k} (S_k + (S_n - S_k))^2] \\ &\geq \sum_{k=1}^n \mathbb{E}[\mathbb{1}_{A_k} S_k^2] + 2\mathbb{E}[\mathbb{1}_{A_k} S_k (S_n - S_k)] \\ &= \sum_{k=1}^n \mathbb{E}[\mathbb{1}_{A_k} S_k^2] \\ &\geq \sum_{k=1}^n \varepsilon^2 \mathbb{P}[A_k] \\ &= \varepsilon^2 \mathbb{P} \left[\max_{1 \leq k \leq n} |S_k| \geq \varepsilon \right]. \end{aligned}$$

□

3.2 Das Schwache Gesetz der Großen Zahlen

Beim Schwachen Gesetz der Großen Zahlen untersucht man die stochastische Konvergenz von einer Folge von Zufallsvariablen. Es gibt dabei verschiedene Voraussetzungen unter denen das Gesetz formuliert werden kann. Einen ersten Spezialfall dieses Satzes kannte bereits Jakob Bernoulli (1654-1705).

3.2.1 Schwaches Gesetz der Großen Zahlen von Bernoulli

Wird eine faire Münze n Mal geworfen, so beträgt die Erfolgswahrscheinlichkeit $p = \frac{1}{2}$. Seien $X_1, X_2, \dots$ Zufallsvariablen, die wie üblich den Wert 1 bei einem Erfolg und den Wert 0 bei einem Misserfolg annehmen. Wir haben es somit mit einem Bernoulli-Versuchsschema zu tun. Durch $\frac{1}{n} \sum_{i=1}^n X_i$ wird dann die relative Häufigkeit der Erfolge angegeben. Wie in der Einleitung angeführt erwarten wir intuitiv, dass diese relative Häufigkeit in der Nähe der Erfolgswahrscheinlichkeit von $\frac{1}{2}$ liegen sollte. Der folgende Satz bestätigt diese Intuition und formalisiert diese Beobachtung, die auch oft Empirisches Gesetz der Großen Zahlen genannt wird.

Satz 3.4. *Schwaches Gesetz der Großen Zahlen von Bernoulli*

Seien $X_1, \dots, X_n$ unabhängige $\text{Ber}(p)$ -verteilte Zufallsvariablen. Dann gilt $\forall \varepsilon > 0$:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\left| \frac{1}{n} \sum_{i=1}^n X_i - p \right| \geq \varepsilon \right] = 0.$$

Beweis. Sei $S_n := X_1 + \dots + X_n$. Dann ist:

$$\mathbb{E} \left[\frac{1}{n} S_n \right] = \frac{1}{n} np = p \quad \text{und} \quad \text{Var} \left[\frac{1}{n} S_n \right] = \frac{1}{n^2} npq = \frac{pq}{n}.$$

Anwendung der Tschebyscheff-Ungleichung liefert:

$$\mathbb{P} \left[\left| \frac{1}{n} \sum_{i=1}^n X_i - p \right| \geq \varepsilon \right] = \frac{pq}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0. \quad \square$$

Für großes n ist also die Wahrscheinlichkeit, dass sich die relative Häufigkeit der Erfolge um mehr als ε von p unterscheidet, sehr klein. Somit haben wir ausgehend von einem axiomatischem Wahrscheinlichkeitsbegriff einen Zusammenhang zwischen Wahrscheinlichkeit und relativer Häufigkeit bewiesen ohne die Wahrscheinlichkeit als Grenzwert der relativen Häufigkeit definiert zu haben.

Bemerkung: Bei vielen Glücksspielen tritt eine Situation ein, die dem Satz entspricht. Ein Spieler spielt das gleiche Spiel immer wieder und es gibt nur die zwei Optionen des Erfolgs oder Misserfolgs. Auf lange Zeit gesehen wird er im Durchschnitt so oft gewinnen wie die Erfolgswahrscheinlichkeit einer einzelnen Runde ist.

3.2.2 Schwaches Gesetz der Großen Zahlen von Tschebyscheff

Die Voraussetzungen werden nun etwas abgeschwächt. Wir betrachten Zufallsvariablen, die nicht bernoulliverteilt, aber identisch verteilt sein müssen. Weiters ersetzen wir die Forderung der Unabhängigkeit durch die schwächere Annahme der Unkorreliertheit.

Satz 3.5. *Schwaches Gesetz der Großen Zahlen von Tschebyscheff*

Seien $X_1, \dots, X_n$ paarweise unkorrelierte und identisch verteilte Zufallsvariablen in $\mathcal{L}^2$ mit $\mathbb{E}[X_i] = \mu$ und $\text{Var}[X_i] = \sigma^2$. Dann gilt $\forall \varepsilon > 0$:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| \geq \varepsilon \right] = 0.$$

Beweis. Der Beweis verläuft analog wie vorher. Sei $S_n := X_1 + \dots + X_n$. Dann ist:

$$\mathbb{E} \left[\frac{1}{n} S_n \right] = \frac{1}{n} n \mu \quad \text{und} \quad \text{Var} \left[\frac{1}{n} S_n \right] = \frac{1}{n^2} n \sigma^2 = \frac{1}{n} \sigma^2.$$

Anwendung der Tschebyscheff-Ungleichung liefert wieder:

$$\mathbb{P} \left[\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| \geq \varepsilon \right] = \frac{\sigma^2}{n \varepsilon^2} \xrightarrow{n \rightarrow \infty} 0. \quad \square$$

3.2.3 Schwaches Gesetz der Großen Zahlen ohne identische Verteilung

Bisher waren die Zufallsvariablen immer identisch verteilt. Nun wollen wir noch eine Version betrachten, in der wir diese Bedingung nicht fordern. Stattdessen müssen wir eine andere Forderung an die Zufallsvariablen stellen mit Hilfe derer der Beweis wieder sehr leicht ist.

Satz 3.6. *Schwaches Gesetz der Großen Zahlen ohne identische Verteilung*

Seien $X_1, \dots, X_n$ paarweise unkorrelierte Zufallsvariablen in $\mathcal{L}^2$. Erfüllt die Folge die Bedingung

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = 0,$$

so gilt:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\left| \frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \right| \geq \varepsilon \right] = 0.$$

Beweis. Der Erwartungswert und die Varianz von $\frac{1}{n} \sum_{i=1}^n X_i$ können nun nicht mehr explizit angegeben werden. Trotzdem folgt mit der Tschebyscheff-Ungleichung und den Rechenregeln der Varianz sofort:

$$\begin{aligned} \mathbb{P} \left[\left| \frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \right| \geq \varepsilon \right] &\leq \frac{1}{\varepsilon^2} \text{Var} \left[\frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \right] \\ &= \frac{1}{n^2 \varepsilon^2} \sum_{i=1}^n \text{Var}[X_i] \xrightarrow{n \rightarrow \infty} 0. \quad \square \end{aligned}$$

Bemerkung: Der Satz stellt eine Verallgemeinerung des Schwachen Gesetzes der Großen Zahlen von Tschebyscheff dar. Sind die Zufallsvariablen zusätzlich identisch

verteilt mit $\text{Var}[X_i] = \sigma^2$, so gilt:

$$\frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i] = \frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0.$$

Alle bisherigen Versionen werden auch oft als $\mathcal{L}^2$ -Versionen des Schwachen Gesetzes der Großen Zahlen genannt, weil wir immer eine endliche Varianz gefordert haben.

3.2.4 Schwaches Gesetz der Großen Zahlen von Chintschin

Bisher war eine endliche Varianz Voraussetzung. Im folgenden Satz wird diese Voraussetzung fallen gelassen, jedoch fordern wir wieder die Unabhängigkeit der identisch verteilten Zufallsvariablen.

Satz 3.7. *Schwaches Gesetz der Großen Zahlen von Chintschin*

Seien $X_1, \dots, X_n$ unabhängige und identisch verteilte Zufallsvariablen in $\mathcal{L}^1$ mit $\mathbb{E}[X_i] = \mu$. Dann gilt $\forall \varepsilon > 0$:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| \geq \varepsilon \right] = 0.$$

Beweis. Unter diesen Voraussetzungen kann der Beweis nicht durch Anwendung der Tschebyscheff-Ungleichung geführt werden. Wir arbeiten stattdessen mit charakteristischen Funktionen. Sei dazu $\varphi_{X_i}(t)$ die charakteristische Funktion der Zufallsvariable X_i und $S_n = X_1 + \dots + X_n$. Mithilfe von Satz 1.10 folgt:

$$\begin{aligned} \varphi_{S_n}(t) &= \varphi_{S_n} \left(\frac{t}{n} \right) \\ &= \varphi_{X_1} \left(\frac{t}{n} \right) \cdots \varphi_{X_n} \left(\frac{t}{n} \right) \\ &= \left[\varphi_{X_1} \left(\frac{t}{n} \right) \right]^n. \end{aligned}$$

Nach Satz 1.12 können wir $\varphi_{X_1} \left(\frac{t}{n} \right)$ in eine Taylorreihe um $t = 0$ entwickeln und erhalten:

$$\begin{aligned} \left[\varphi_{X_1} \left(\frac{t}{n} \right) \right]^n &= \left[1 + \frac{\mu i t}{n} + o \left(\frac{t}{n} \right) \right]^n \\ &\xrightarrow{n \rightarrow \infty} e^{\mu i t} = \varphi_{\mu}(t). \end{aligned}$$

Also haben wir $\varphi_{S_n}(t) \rightarrow \varphi_{\mu}(t)$. Aus dem Stetigkeitssatz folgt daher $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{d} \mu$ und folglich nach Satz 2.5 $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \mu$. $\square$

Dieser Satz wird auch als $\mathcal{L}^1$ -Version des Schwachen Gesetzes der Großen Zahlen genannt, da wir lediglich einen endlichen Erwartungswert gefordert haben. Man kann

jedoch nicht behaupten, dass diese Version stärker ist, da wir hier andererseits wieder die Unabhängigkeit vorausgesetzt haben.

Zusammengefasst besagt das Schwache Gesetz der Großen Zahlen also intuitiv folgendes: Betrachten wir die unabhängige Wiederholung von n Versuchen $X_1, \dots, X_n$ als ein gesamtes Zufallsexperiment und wiederholen dieses immer wieder, so wird $\frac{X_1 + \dots + X_n}{n}$ für einen nahe bei 1 liegenden Anteil in der Nähe von $\mathbb{E}[X_i]$ für $1 \leq i \leq n$ liegen.

3.3 Das Starke Gesetz der Großen Zahlen

Eigentlich erwarten wir intuitiv noch mehr. Die normierte Zufallsvariable $\frac{X_1 + \dots + X_n}{n}$ soll nicht nur „meistens“ in der Nähe von $\mathbb{E}[X_i]$ liegen, sondern bei zunehmenden n gegen diesen konvergieren. Wir erwarten also für $\varepsilon > 0$, dass $\frac{X_1 + \dots + X_n}{n}$ von einem bestimmten $n_0 \in \mathbb{N}$ an im Intervall $[\mathbb{E}[X_i] - \varepsilon, \mathbb{E}[X_i] + \varepsilon]$ liegt und dort auch verbleibt. Aussagen dieser Art liefern die Starken Gesetze der Großen Zahlen. Wir haben es also mit der fast sicheren Konvergenz zu tun. Auch das Starke Gesetz der Großen Zahlen kann wieder unter verschiedenen Voraussetzungen formuliert werden. Bevor wir diese jedoch beweisen können, benötigen wir ein wichtiges Lemma.

3.3.1 Das Lemma von Borel-Cantelli

Obwohl es sich dabei um einen wichtigen Satz handelt, ist es üblich vom Lemma von Borel-Cantelli zu sprechen.

Lemma 3.8. *Borel-Cantelli*

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A_1, A_2, \dots$ eine Folge von Ereignissen aus $\mathcal{A}$. Sei $A^* = \{\omega \in \Omega : \omega \in A_k \text{ für unendlich viele } k\}$. Dann gilt:

- (i) Gilt $\sum_{k=1}^{\infty} \mathbb{P}[A_k] < \infty$, dann ist $\mathbb{P}[A^*] = 0$.
- (ii) Sind $A_1, A_2, \dots$ unabhängig und gilt $\sum_{k=1}^{\infty} \mathbb{P}[A_k] = \infty$, dann ist $\mathbb{P}[A^*] = 1$.

Beweis. (i) Ein $\omega \in \Omega$ gehört zu A^* , wenn es für jedes n ein $k(\omega) \geq n$ gibt, sodass $\omega \in A_k$. Also gilt:

$$A^* = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

Weiters gilt für jedes n :

$$\mathbb{P}[A^*] \leq \mathbb{P} \left[\bigcup_{k=n}^{\infty} A_k \right] \leq \sum_{k=n}^{\infty} \mathbb{P}[A_k].$$

Die Summe $\sum_{k=n}^{\infty} \mathbb{P}[A_k]$ ist wegen der Konvergenz für hinreichend großes n kleiner als eine beliebige vorgegebene Zahl $\varepsilon > 0$ und konvergiert somit gegen 0. Daraus folgt $\mathbb{P}[A^*] = 0$.

- (ii) Sei $B_i := \Omega \setminus A_i$ für $i \in \mathbb{N}$. Da die Ereignisse A_i unabhängig sind, sind somit auch die Ereignisse B_i unabhängig. Nach der Regel von de Morgan gilt dann $\Omega \setminus A^* = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} B_k$ und somit erhalten wir:

$$\begin{aligned} \mathbb{P} \left[\bigcap_{k=n}^m B_k \right] &= \prod_{k=n}^m \mathbb{P}[B_k] = \prod_{k=n}^m (1 - \mathbb{P}[A_k]) \\ &\leq e^{-\sum_{k=n}^m \mathbb{P}[A_k]} \xrightarrow{m \rightarrow \infty} 0. \end{aligned}$$

$$\Rightarrow \mathbb{P}[\Omega \setminus A^*] = \mathbb{P} \left[\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} B_k \right] = 0.$$

Daher ist $\mathbb{P}[A^*] = 1$. □

Beispiel 3.2: Wir betrachten eine Folge $(U_n)_{n \in \mathbb{N}}$ von Urnen, wobei die i -te Urne eine weiße und $i - 1$ schwarze Kugeln enthält. Aus jeder Urne wird zufällig eine Kugel gezogen. Sei A_i das Ereignis, dass die aus der i -ten Urne gezogene Kugel weiß ist. Dann ist $\mathbb{P}[A_i] = \frac{1}{i}$. Da die Ziehungen unabhängig sind und $\sum_{i=1}^{\infty} \mathbb{P}[A_i] = \infty$, werden mit Wahrscheinlichkeit 1 unendlich viele weiße Kugeln gezogen.

Beinhaltet die i -te Urne hingegen $i^2 - 1$ schwarze Kugeln, so erhalten wir $\sum_{i=1}^{\infty} \mathbb{P}[A_i] = \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty$ und somit werden nach dem Lemma von Borel-Cantelli mit Wahrscheinlichkeit 1 nur endlich viele weiße Kugeln gezogen.

3.3.2 Starkes Gesetz der Großen Zahlen von Borel

Das erste Starke Gesetz der Großen Zahlen wurde 1909 von Emile Borel (1871-1956) formuliert und ist eine direkte Übertragung von Bernoulli's schwachem Gesetz. Aufgrund des stärkeren Konvergenzbegriffs ist der Beweis jedoch etwas schwieriger.

Satz 3.9. *Starkes Gesetz der Großen Zahlen von Borel*

Seien $X_1, \dots, X_n$ unabhängige $\text{Ber}(p)$ -verteilte Zufallsvariablen. Dann gilt:

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{i=1}^n X_i - p \right| = 0 \right] = 1.$$

Beweis. Sei $S_n = X_1 + \dots + X_n$. Dann ist S_n $B(n, p)$ -verteilt und wir erhalten $\mathbb{E} \left[\frac{S_n}{n} - p \right] = 0$ und $\text{Var} \left[\frac{S_n}{n} - p \right] = \frac{pq}{n}$. Nach der Tschebyscheff-Ungleichung gilt

dann:

$$\mathbb{P} \left[\left| \frac{S_n}{n} - p \right| \geq \varepsilon \right] \leq \frac{pq}{n\varepsilon^2}.$$

Sei $A_i := \left\{ \left| \frac{S_{i^2}}{i^2} - p \right| \geq \varepsilon \right\}$. Dann gilt:

$$\sum_{i=1}^{\infty} \mathbb{P}[A_i] \leq \sum_{i=1}^{\infty} \frac{pq}{i^2\varepsilon^2} < \infty.$$

Nach dem Lemma von Borel-Cantelli folgt $\mathbb{P}[A_i \text{ unendlich oft}] = 0$. Also gilt:

$$\frac{S_{n^2}}{n^2} \xrightarrow{f.s.} p.$$

Die Aussage ist somit für die Teilfolge mit Quadratzahlenindizes bewiesen. Für die gesamte Folge war diese Argumentation nicht möglich, da sonst die Reihe $\sum_{i=1}^{\infty} \mathbb{P}[A_i]$ nicht konvergieren würde. Wir müssen also noch überprüfen wie sich $\frac{S_i}{i} - p$ zwischen aufeinander folgenden Indizes der Teilfolge verhält:

Für $i \in \mathbb{N}$ sei $n \in \mathbb{N}$ so gewählt, dass $n^2 \leq i < (n+1)^2$. Dann gilt folgende Abschätzung:

$$\begin{aligned} \left| \frac{S_i}{i} - p \right| &= \left| \frac{S_{n^2} - n^2p}{i} + \frac{S_i - S_{n^2}}{i} - \frac{(i - n^2)p}{i} \right| \\ &\leq \left| \frac{S_{n^2} - n^2p}{n^2} \right| + \left| \frac{S_{(n+1)^2} - S_{n^2}}{n^2} \right| + \left| \frac{((n+1)^2 - n^2)p}{n^2} \right| \\ &\leq \left| \frac{S_{n^2}}{n^2} - p \right| + \frac{2n+1}{n^2}(1+p). \end{aligned}$$

Für $i \rightarrow \infty$ ist auch $n \rightarrow \infty$. Da die rechte Seite fast sicher gegen 0 konvergiert folgt die Behauptung. $\square$

3.3.3 Starkes Gesetz der Großen Zahlen von Rajchman

Wie schon beim Schwachen Gesetz der Großen Zahlen müssen die Zufallsvariablen nicht bernoulliverteilt sein. Auch die Unkorreliertheit reicht wieder aus.

Satz 3.10. *Starkes Gesetz der Großen Zahlen von Rajchman*

Seien $X_1, \dots, X_n$ paarweise unkorrelierte Zufallsvariablen in $\mathcal{L}^2$ mit $\mathbb{E}[X_i] = \mu$ und $\text{Var}[X_i] = \sigma^2$. Dann gilt:

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| = 0 \right] = 1.$$

Beweis. Der Beweis erfolgt analog wie beim Starken Gesetz der Großen Zahlen von Borel, in dem man p durch μ und pq durch σ^2 ersetzt. $\square$

Dieser Satz stellt das Analogon zum Satz von Tschebyscheff dar und wird auch oft

als $\mathcal{L}^2$ -Version bezeichnet.

3.3.4 Starkes Gesetz der Großen Zahlen ohne identische Verteilung

Wir wollen nun wieder die Bedingung der identischen Verteilung fallen lassen. Um den Satz mit diesen Bedingungen beweisen zu können benötigen wir zwei Hilfssätze. Zunächst geben wir ein Kriterium für das Vorliegen fast sicherer Konvergenz. Anschließend formulieren wir ein Resultat aus der Analysis.

Lemma 3.11. *Seien $X_1, X_2, \dots$ Zufallsvariablen in $\mathcal{L}^2$ mit $\mathbb{E}[X_i] = 0$ für $i \in \mathbb{N}$. Sei $S_n = \sum_{i=1}^n X_i$ und es gelte*

$$\sum_{i=1}^{\infty} \text{Var}[X_i] < \infty.$$

Dann konvergiert S_n fast sicher.

Beweis. Sei $\varepsilon > 0$ und $m \in \mathbb{N}$. Dann gilt mit Hilfe der Kolmogorov-Ungleichung:

$$\begin{aligned} \mathbb{P} \left[\sup_{n \geq m} |S_n - S_m| > \varepsilon \right] &= \mathbb{P} \left[\sup_{n \geq m+1} \left| \sum_{i=m+1}^n X_i \right| > \varepsilon \right] \\ &= \lim_{k \rightarrow \infty} \mathbb{P} \left[\sup_{k \geq n \geq m+1} \left| \sum_{i=m+1}^n X_i \right| > \varepsilon \right] \\ &\leq \lim_{k \rightarrow \infty} \frac{1}{\varepsilon^2} \sum_{i=m+1}^k \text{Var}[X_i] \\ &= \frac{1}{\varepsilon^2} \underbrace{\sum_{i=m+1}^{\infty} \text{Var}[X_i]}_{\xrightarrow{m \rightarrow \infty} 0} \end{aligned}$$

Nach Satz 2.1 folgt die Behauptung. □

Lemma 3.12. Kronecker-Lemma

Es sei $(a_n)_{n \in \mathbb{N}}$ eine Folge reeller Zahlen und $(b_n)_{n \in \mathbb{N}}$ eine monoton steigende Folge positiver Zahlen mit $\lim_{n \rightarrow \infty} b_n = \infty$. Sei $s_n = \sum_{i=1}^n \frac{a_i}{b_i}$. Falls $(s_n)_{n \in \mathbb{N}}$ konvergiert, so folgt:

$$\lim_{n \rightarrow \infty} \frac{1}{b_n} \sum_{i=1}^n a_i = 0.$$

Beweis. Nach Voraussetzung $\exists s \in \mathbb{R}$ mit $s = \lim_{n \rightarrow \infty} s_n$. Aus $s_n - s_{n-1} = \frac{a_n}{b_n}$ folgt:

$$\sum_{i=1}^n a_i = \sum_{i=1}^n b_i (s_i - s_{i-1})$$

$$= b_n s_n - \sum_{i=1}^n (b_i - b_{i-1}) s_{i-1}.$$

Multiplizieren wir nun beide Seiten mit $\frac{1}{b_n}$, so erhalten wir

$$\frac{1}{b_n} \sum_{i=1}^n a_i = s_n - \frac{1}{b_n} \sum_{i=1}^n (b_i - b_{i-1}) s_{i-1}.$$

Aus $s = \lim_{n \rightarrow \infty} s_n$ folgt mithilfe des Cesaro-Limes $s = \lim_{n \rightarrow \infty} \frac{1}{b_n} (b_i - b_{i-1}) s_{i-1}$ und damit die Behauptung. $\square$

Bemerkung: Setzen wir im Kronecker-Lemma $b_n = \frac{1}{n}$, so sehen wir, dass es zum Nachweis von $\mathbb{P}[\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = 0] = 1$ reicht, $\mathbb{P}[\sum_{i=1}^n \frac{X_i}{i} \text{ konvergiert}] = 1$ zu zeigen.

Nun können wir das Starke Gesetz der Großen Zahlen beweisen, ohne eine identische Verteilung der Zufallsvariablen zu fordern. Stattdessen müssen wir wieder eine andere Bedingung an die Zufallsvariablen stellen, die auch als Kolmogorov-Kriterium bekannt ist.

Satz 3.13. *Starkes Gesetz der Großen Zahlen ohne identische Verteilung*

Seien $X_1, X_2, \dots$ unabhängige Zufallsvariablen in $\mathcal{L}^2$, die das Kolmogorov-Kriterium

$$\sum_{n=1}^{\infty} \frac{\text{Var}[X_n]}{n^2} < \infty$$

erfüllen. Dann gilt:

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \right| = 0 \right] = 1.$$

Beweis. Wir können oBdA annehmen, dass $\mathbb{E}[X_i] = 0$ für alle $i \in \mathbb{N}$. Sei $Y_i = \frac{X_i}{i}$. Dann gilt:

$$\sum_{n=1}^{\infty} \text{Var}[Y_n] = \sum_{n=1}^{\infty} \frac{\text{Var}[X_n]}{n^2} < \infty.$$

Weiters sei $S_n = \sum_{i=1}^n Y_i$. Aus dem Lemma 3.11 folgt dann, dass S_n fast sicher konvergiert. Wendet man nun noch das Kronecker-Lemma für $b_n = \frac{1}{n}$ an, so erhält man

$$\frac{1}{n} \sum_{i=1}^{\infty} X_i \xrightarrow{f.s.} 0$$

und damit die Behauptung. $\square$

In den vorigen Sätzen war nur die Unkorreliertheit gefordert. Liegt jedoch Unabhängigkeit vor, so ist dieser Satz eine Verallgemeinerung der bisherigen Sätze.

Sind die Zufallsvariablen nämlich zusätzlich identisch verteilt mit $\text{Var}[X_n] = \sigma^2$, so ist die Kolmogorov-Bedingung immer erfüllt, denn:

$$\sum_{n=1}^{\infty} \frac{\text{Var}[X_n]}{n^2} = \sigma^2 \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

Bei den bisherigen Sätzen spricht man wieder von $\mathcal{L}^2$ -Versionen des Starken Gesetzes der Großen Zahlen.

3.3.5 Starkes Gesetz der Großen Zahlen von Kolmogorov

Analog wie bei den Schwachen Gesetzen wollen wir nun noch eine $\mathcal{L}^1$ -Version herleiten. Dies benötigt einiges an Vorbereitung. Dabei betrachten wir wieder unabhängig und identisch verteilte Zufallsvariablen.

Wir benötigen dazu sogenannte gestutzte Zufallsvariablen: Ist $(X_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen, so betrachtet man

$$Y_n := X_n \mathbb{1}_{\{|X_n| \leq n\}}.$$

Y_n wird also gleich 0 gesetzt, wenn X_n außerhalb des Intervalls $[-n, n]$ liegt. Dadurch folgt automatisch, dass $|Y_n| \leq n$ und Y_n somit beschränkt ist. Weiters werden andere wichtige Ergebnisse von X_n auf Y_n übertragen, wie das folgende Lemma zeigt.

Lemma 3.14. *Seien $X_1, X_2, \dots$ unabhängig und identisch verteilte Zufallsvariablen mit $\mathbb{E}[X_i] = \mu$ und $Y_n := X_n \mathbb{1}_{\{|X_n| \leq n\}}$ mit $n \in \mathbb{N}$. Dann gilt:*

$$(i) \quad \mathbb{E}[Y_n] \longrightarrow \mu.$$

$$(ii) \quad \mathbb{P}[X_n = Y_n \text{ für fast alle } n] = 1.$$

$$(iii) \quad (Y_n)_{n \in \mathbb{N}} \text{ erfüllt die Kolmogorov-Bedingung, also } \sum_{n=1}^{\infty} \frac{\text{Var}[Y_n]}{n^2} < \infty.$$

Beweis. (i) Sei $Z_n = X_1 \mathbb{1}_{\{|X_1| \leq n\}}$. Da X_1 und X_n identisch verteilt sind, sind auch Z_n und Y_n identisch verteilt und somit gilt $\mathbb{E}[Z_n] = \mathbb{E}[Y_n]$. Andererseits gilt $|Z_n| \leq X_1$ für alle $n \in \mathbb{N}$ und $Z_n \xrightarrow{f.s.} X_1$. Nach dem Satz von der majorisierten Konvergenz folgt somit:

$$\mathbb{E}[Y_n] = \mathbb{E}[Z_n] \longrightarrow \mathbb{E}[X_1] = \mu.$$

(ii) Es gilt:

$$\begin{aligned} \sum_{n=1}^{\infty} \mathbb{P}[X_n \neq Y_n] &= \sum_{n=1}^{\infty} \mathbb{P}[|X_n| > n] \\ &= \sum_{n=1}^{\infty} \mathbb{P}[|X_1| > n] \end{aligned}$$

$$\begin{aligned}
&= \sum_{n=1}^{\infty} \sum_{k \geq n} \mathbb{P}[k \leq |X_1| < k+1] \\
&= \sum_{k=1}^{\infty} k \mathbb{P}[k \leq |X_1| < k+1] \\
&= \sum_{k=0}^{\infty} \mathbb{E}[k \mathbb{1}_{k \leq |X_1| < k+1}] \\
&\leq \sum_{k=0}^{\infty} \mathbb{E}[|X_1| \mathbb{1}_{k \leq |X_1| < k+1}] = \mathbb{E}[|X_1|] < \infty.
\end{aligned}$$

Nach dem Lemma von Borel-Cantelli folgt $\mathbb{P}[X_n = Y_n \text{ für fast alle } n] = 1$.

(iii) Für alle $n \in \mathbb{N}$ ist

$$\frac{1}{n^2} \leq \frac{2}{n(n+1)} = 2 \left(\frac{1}{n} - \frac{1}{n+1} \right)$$

und daher

$$\sum_{n \geq k} \frac{1}{n^2} \leq \frac{2}{k}$$

Wir erhalten somit:

$$\sum_{n=1}^{\infty} \frac{|X_1|^2 \mathbb{1}_{\{|X_1| \leq n\}}}{n^2} = \sum_{n \geq \max\{1, |X_1|\}} \frac{|X_1|^2}{n^2} \leq \frac{2|X_1|^2}{\max\{1, |X_1|\}} \leq 2|X_1|.$$

Mit Hilfe dieser Abschätzung folgt nun die Kolmogorov-Bedingung:

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{\text{Var}[Y_n]}{n^2} &\leq \sum_{n=1}^{\infty} \frac{\mathbb{E}[Y_n^2]}{n^2} \\
&= \sum_{n=1}^{\infty} \frac{\mathbb{E}[|X_1|^2 \mathbb{1}_{\{|X_1| \leq n\}}]}{n^2} \\
&\leq 2\mathbb{E}[|X_1|] < \infty.
\end{aligned}$$

□

Mit Hilfe dieses Lemmas können wir nun das Starke Gesetz der Großen Zahlen in der $\mathcal{L}^1$ -Version relativ einfach beweisen, das von Kolmogorov stammt.

Satz 3.15. *Starkes Gesetz der Großen Zahlen von Kolmogorov*

Seien $X_1, X_2, \dots$ unabhängige und identisch verteilte Zufallsvariablen in $\mathcal{L}^1$ mit $\mathbb{E}[X_i] = \mu$. Dann gilt:

$$\mathbb{P} \left[\lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| = 0 \right] = 1.$$

Beweis. Sei $Y_n = X_n \mathbb{1}_{\{|X_n| \leq n\}}$ und $S_n = X_1 + \dots + X_n$ für $n \in \mathbb{N}$. Wegen Lemma 3.14(ii) genügt es zu zeigen, dass $\frac{1}{n} \sum_{i=1}^n Y_i$ fast sicher gegen μ konvergiert. Dazu

schreiben wir $Y_i = \mathbb{E}[Y_i] + (Y_i - \mathbb{E}[Y_i])$ und erhalten:

$$\frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i] + \frac{1}{n} \sum_{i=1}^n (Y_i - \mathbb{E}[Y_i]).$$

Nach dem Lemma 3.14(i) gilt $\mathbb{E}[Y_n] \rightarrow \mu$. Nach dem Cauchy'schen Grenzwertsatz konvergiert also auch die erste Reihe auf der rechten Seite - das Cesaro-Mittel - gegen μ . Da $(Y_n)_{n \in \mathbb{N}}$ die Kolmogorov-Bedingung erfüllt, können wir auf die zweite Reihe das Starke Gesetz der Großen Zahlen vom vorigen Kapitel anwenden und erhalten, dass die Reihe fast sicher gegen 0 konvergiert. Insgesamt erhalten wir somit

$$\frac{1}{n} \sum_{i=1}^n Y_i \xrightarrow{f.s.} \mu. \quad \square$$

4 Der Zentrale Grenzwertsatz

In diesem Kapitel betrachten wir wieder eine Folge von Zufallsvariablen. Auch der Zentrale Grenzwertsatz macht eine Aussage über die Verteilung der Summe $S_n := X_1 + \dots + X_n$. Bei den Gesetzen der Großen Zahlen wurde dafür die Normierung $\frac{1}{n}$ gewählt, um die Konvergenz gegen den Erwartungswert $\mathbb{E}[X_i]$ zu beweisen. Wählt man stattdessen die Normierung $\frac{1}{\sqrt{n}}$, so sind andere Grenzwertaussagen möglich. Im Wesentlichen besagt der Zentrale Grenzwertsatz dann, dass diese normierte Summe für $n \rightarrow \infty$ unter bestimmten Voraussetzungen normalverteilt ist. Um diese Konvergenz beweisen zu können, muss davor eine Standardisierung dieser Summe vorgenommen werden, da sonst bei $\mathbb{E}[X_i] \neq 0$ die Dichte immer mehr verschoben und wegen $\text{Var}[X_i] > 0$ immer flacher werden würde. Dies ist ein äußerst bemerkenswertes Resultat, da es im Allgemeinen sehr schwierig wäre, eine exakte Wahrscheinlichkeitsverteilung von einer Summe von Zufallsvariablen zu finden. Der Zentrale Grenzwertsatz zeigt somit die Universalität der Normalverteilung. Streng genommen handelt es sich dabei nicht um einen einzigen Satz, sondern um eine Gruppe von Sätzen mit verschiedenen Voraussetzungen, die alle die Konvergenz von Summenverteilungsfunktionen gegen die Verteilungsfunktion einer normalverteilten Zufallsvariable zum Gegenstand haben. Man unterscheidet dabei zwischen lokalen und globalen Grenzwertsätzen, je nachdem ob man die Wahrscheinlichkeitsdichte oder die Verteilungsfunktion betrachtet. Der Konvergenzbegriff ist dabei immer die Konvergenz nach Verteilung. Es werden dabei also nur Aussagen über die Form der Verteilungen gemacht. Da es auch im diskreten Fall Grenzwertaussagen dieser Art gibt, sollen diese zunächst in einem kurzen Abschnitt betrachtet werden, ehe wir zu den Zentralen Grenzwertsätzen kommen. Als Quellen für dieses Kapitel wurden hauptsächlich die Werke [5], [7], [14] und [15] herangezogen. Einzelne Anregungen wurden auch in [3] und [9] gefunden.

4.1 Konvergenz von diskreten Verteilungen

4.1.1 Binomialapproximation der hypergeometrischen Verteilung

Für hinreichend große Grundgesamtheiten N spielt es fast keine Rolle, ob man eine kleine Stichprobe vom Umfang n mit oder ohne Zurücklegen zieht. Sind darin M Objekte mit einer bestimmten Eigenschaft enthalten und werden diese immer wieder zurückgelegt, so wird ein solches bei jedem der n Versuche mit konstanter Wahrscheinlichkeit $p = \frac{M}{N}$ gezogen. Zieht man ohne Zurücklegen, so erhalten wir für den 1. Zug dieselbe Wahrscheinlichkeit $p = \frac{M}{N}$, für den 2. Zug hingegen die Wahrscheinlichkeit $\frac{M-1}{N-1}$, wenn wir eines der M Objekte gezogen haben oder $\frac{M}{N-1}$, wenn wir eines der $N - M$ Objekte gezogen haben etc. Für große N weichen diese Werte

nur minimal voneinander ab. Wir können daher die hypergeometrische Verteilung bei bestimmten Voraussetzungen gut durch die Binomialverteilung approximieren.

Satz 4.1. Sei $(X_N)_{N \in \mathbb{N}}$ eine Folge von $H(N, M_N, n)$ -verteilten Zufallsvariablen und n fest. Wenn $N, M_N \rightarrow \infty$, sodass $\lim_{N \rightarrow \infty} \frac{M_N}{N} = p$, dann gilt für alle $k = 0, 1, \dots, n$:

$$\lim_{N \rightarrow \infty} \mathbb{P}[X_N = k] = \binom{n}{k} p^k (1-p)^{n-k}.$$

Beweis. Sei $M = M_N$. Es gilt:

$$\begin{aligned} \mathbb{P}[X_N = k] &= \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} = \frac{\frac{M!}{k!(M-k)!} \frac{(N-M)!}{(n-k)!(N-M-n+k)!}}{\frac{N!}{n!(N-n)!}} \\ &= \frac{\frac{M(M-1)\dots(M-k+1)}{k!} \frac{(N-M)(N-M-1)\dots(N-M-n+k+1)}{(n-k)!}}{\frac{N(N-1)\dots(N-n+1)}{n!}} \\ &= \frac{n!}{k!(n-k)!} \frac{M \dots (M-k+1)(N-M) \dots (N-M-n+k+1)}{N(N-1)\dots(N-n+1)} \frac{\frac{1}{N^n}}{\frac{1}{N^n}} \\ &= \binom{n}{k} \frac{\frac{M}{N} \dots \frac{M-k+1}{N} \frac{(N-M)}{N} \dots \frac{(N-M-n+k+1)}{N}}{\frac{N}{N} \frac{(N-1)}{N} \dots \frac{(N-n+1)}{N}} \\ &= \binom{n}{k} \frac{p \left(p - \frac{1}{N}\right) \dots \left(p - \frac{(k+1)}{N}\right) (1-p) \dots \left(1 - p - \frac{(n-k+1)}{N}\right)}{\left(1 - \frac{1}{N}\right) \dots \left(1 - \frac{(n+1)}{N}\right)} \\ &\xrightarrow{N \rightarrow \infty} \binom{n}{k} p^k (1-p)^{n-k}. \quad \square \end{aligned}$$

Faustregel: Ist $\frac{n}{N} \leq 0.05$, so ist die Binomialverteilung eine gute Approximation für die hypergeometrische Verteilung.

Bemerkung: Der Beweis hätte genauso mithilfe von charakteristischen Funktionen geführt werden können. Aufgrund der komplizierten Charakteristischen Funktion der Hypergeometrischen Verteilung wäre dies jedoch viel mehr Aufwand als der „rechnerische“ Beweis.

Beispiel 4.1: Einer Lieferung von N TV-Geräten, die unabhängig voneinander mit einer Wahrscheinlichkeit von 0.04 defekt sind, wird eine Stichprobe von 15 Stücken entnommen. Wie groß ist die Wahrscheinlichkeit, dass sich darunter nur einwandfreie Geräte befinden?

In der folgenden Tabelle werden die Wahrscheinlichkeiten für die hypergeometrisch-verteilten Zufallsvariablen X_N , die die Anzahl der einwandfreien Stück angeben, und verschieden großen N berechnet:

N	$\mathbb{P}[X_N = 15]$
50	0.4857
100	0.5164
500	0.5337
1000	0.5397
5000	0.5416

Berechnet man die Wahrscheinlichkeit mit der Binomialverteilung, die von N unabhängig ist, ergibt sich $0.96^{15} = 0.5421$.

4.1.2 Poisson-Approximation der Binomialverteilung

Die Binomialverteilung hat unter den diskreten Verteilungen eine besondere Bedeutung, da sie sehr häufig auftritt. Für großes n und k ist jedoch die Berechnung vom Binomialkoeffizienten $\binom{n}{k}$ sehr mühsam. Man versucht daher, Approximationen für die Binomialverteilung zu finden. Ist die Erfolgswahrscheinlichkeit p klein, so eignet sich die Poisson-Verteilung dafür.

Satz 4.2. Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von $B(n, p_n)$ -verteilten Zufallsvariablen. Wenn $n \rightarrow \infty$, sodass $\lambda = np_n$ konstant bleibt ($\Rightarrow p_n \rightarrow 0$), gilt für alle $k \in \mathbb{N}$:

$$\lim_{n \rightarrow \infty} \mathbb{P}[X_n = k] = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Beweis. Wir wollen dafür zwei Beweise anführen.

Rechnerisch:

$$\begin{aligned}
\mathbb{P}[X_n = k] &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} \frac{n(n-1) \dots (n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&= \frac{\lambda^k}{k!} \underbrace{\left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{k-1}{n}\right)}_{\xrightarrow{n \rightarrow \infty} 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\xrightarrow{n \rightarrow \infty} e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\xrightarrow{n \rightarrow \infty} 1} \\
&\xrightarrow{n \rightarrow \infty} \frac{\lambda^k}{k!} e^{-\lambda}.
\end{aligned}$$

Mit Charakteristischen Funktionen:

In Beispiel 1.2 wurden die Charakteristischen Funktionen der beiden Verteilungen berechnet.

$$\begin{aligned}\varphi_{X_n}(t) &= (p(e^{it} - 1) + 1)^n \\ &= \left(1 + \frac{\lambda e^{it} - 1}{n}\right)^n \\ &\xrightarrow{n \rightarrow \infty} e^{\lambda e^{it} - 1}\end{aligned}$$

Aus dem Stetigkeitssatz folgt die Behauptung. $\square$

Beispiel 4.2: Ein Unternehmen stellt pro Tag 1000 elektrische Bauteile her, von denen jedes unabhängig von den anderen mit einer Wahrscheinlichkeit von 0.0025 defekt ist. Wie groß ist die Wahrscheinlichkeit, dass an einem bestimmten Tag $k \in \{0, 1, 2, 3, 4\}$ defekt sind?

Sei X die Anzahl der defekten Bauteile. Dann ergeben sich für $n = 1000$, $p = 0.0025$ und $\lambda = np = 2.5$ folgende Werte:

	$P(\lambda)$	B(n,p)
k	$\mathbb{P}[X = k]$	$\mathbb{P}[X = k]$
0	0.0821	0.0818
1	0.2052	0.2051
2	0.2565	0.2567
3	0.2138	0.2141
4	0.1336	0.1337

Da in diesem Beispiel die Werte von n und p den Voraussetzungen einer Poissonapproximation sehr gut entsprechen, kommt es nur zu sehr minimalen Abweichungen.

Die Poissonapproximation ist jedoch nur für kleine Werte von p sinnvoll, wie folgendes Beispiel zeigt:

Beispiel 4.3: Wie groß ist die Wahrscheinlichkeit, dass beim 10-maligem Münzwurf genau 5 Mal Zahl erscheint?

Die Wahrscheinlichkeit dafür beträgt $\binom{10}{5}2^{-10} = 0.2461$. Würde man hier Poissonapproximation mit Parameter $\lambda = 5$ verwenden, so würde sich eine Wahrscheinlichkeit von $\frac{5^5}{5!}e^{-5} = 0.1755$ ergeben und somit deutlich von der tatsächlichen Wahrscheinlichkeit abweichen.

4.2 Der Zentrale Grenzwertsatz von de Moivre und Laplace

Zunächst betrachten wir den wichtigen Spezialfall von bernoulliverteilten Zufallsvariablen, der als Satz von de Moivre und Laplace bekannt ist und der mit elementaren Hilfsmitteln bewiesen werden kann. Für den Beweis des Satzes benötigen wir die Stirling'sche Formel.

4.2.1 Stirling'sche Formel

Für großes n und k ist die Berechnung des Binomialkoeffizienten $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ sehr mühsam. Wir benötigen eine gute Approximation für die auftretenden Fakultäten. In diesem Abschnitt zeigen wir vorerst nur, dass es eine Konstante $c > 0$ gibt, sodass $n! \sim ce^{-n}n^n\sqrt{n}$ gilt, wobei $a \sim b$ bedeutet, dass $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1$ ist. Später werden wir im Zuge des Beweises des Satzes von de Moivre und Laplace $c = \sqrt{2\pi}$ herleiten.

Satz 4.3. *Stirling'sche Formel, 1. Teil*

Es existiert ein $c > 0$ mit $n! \sim ce^{-n}n^n\sqrt{n}$.

Beweis. Sei $d_n := \log n! - (n + \frac{1}{2}) \log n + n$. Da $\log x$ konkav und monoton wachsend ist, gilt für $k = 1, \dots, n$:

$$\int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \log x \, dx \leq \log x \leq \int_k^{k+1} \log x \, dx.$$

Durch Summieren über $k = 1, \dots, n$ erhalten wir:

$$\int_{\frac{1}{2}}^{n+\frac{1}{2}} \log x \, dx \leq \log n! \leq \int_1^{n+1} \log x \, dx.$$

Da $x \log x - x$ eine Stammfunktion von $\log x$ ist, folgt:

$$(n + \frac{1}{2}) \log(n + \frac{1}{2}) - (n + \frac{1}{2}) - \frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \leq \log n! \leq (n+1) \log(n+1) - (n+1) - \log 1 - 1$$

$$\Leftrightarrow (n + \frac{1}{2}) \log(n + \frac{1}{2}) - n - \frac{1}{2} \log \frac{1}{2} \leq \log n! \leq (n+1) \log(n+1) - n.$$

$$\begin{aligned} \Rightarrow d_n &> (n + \frac{1}{2}) \log(n + \frac{1}{2}) - n - \frac{1}{2} \log \frac{1}{2} - (n+1) \log(n+1) + n \\ &= \underbrace{(n + \frac{1}{2}) \log\left(\frac{n+1/2}{n}\right)}_{>0} - \frac{1}{2} \log \frac{1}{2} > -\frac{1}{2} \log \frac{1}{2} > 0. \end{aligned}$$

Also ist d_n durch die positive Konstante $-\frac{1}{2} \log \frac{1}{2}$ nach unten beschränkt. Weiters gilt:

$$\begin{aligned} d_n - d_{n+1} &= \log n! - \log(n+1)! - (n + \frac{1}{2}) \log n + (n + \frac{3}{2}) \log(n+1) + n - (n+1) \\ &= (n + \frac{1}{2}) [\log(n+1) - \log n] - 1 \end{aligned}$$

$$= \int_n^{n+1} \log x \, dx - \frac{1}{2}[\log(n+1) - \log n] \geq 0.$$

Also ist d_n auch monoton fallend. Daraus folgt, dass $\lim_{n \rightarrow \infty} d_n =: d$ existiert und es gilt:

$$\begin{aligned} c := e^d &= e^{\lim_{n \rightarrow \infty} d_n} = \lim_{n \rightarrow \infty} e^{\log n! - (n+1/2) \log n + n} \\ &= \lim_{n \rightarrow \infty} \frac{n!}{n^{(n+1/2)} e^{-n}}. \end{aligned}$$

$$\Rightarrow n! \sim ce^{-n} n^n \sqrt{n}.$$

□

4.2.2 Elementarer Beweis

Wir werden in diesem Abschnitt den Beweis mit elementaren Hilfsmitteln durchführen, bei dem die Stirling'sche Formel eine zentrale Rolle spielt. Obwohl der Satz sofort aus den späteren Versionen folgen würde und der Beweis schneller geführt werden könnte, soll hier der elementare Beweis angeführt werden. Dieser ist zwar etwas länger, kommt jedoch ohne charakteristische Funktionen oder anderen technischen Hilfestellungen der Wahrscheinlichkeitstheorie aus. Deshalb ist dieser oft in einführenden Büchern zu finden, wie zum Beispiel in [3] oder [7].

Satz 4.4. *Zentraler Grenzwertsatz von de Moivre und Laplace*

Seien $X_1, \dots, X_n$ unabhängige $\text{Ber}(p)$ -verteilte Zufallsvariablen und $S_n = \sum_{i=1}^n X_i$. Somit ist S_n $B(n, p)$ -verteilt. Dann gilt für die Zufallsvariable $S_n^* := \frac{S_n - np}{\sqrt{npq}}$ und alle $a < b$:

$$\mathbb{P}[a \leq S_n^* \leq b] \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx = \Phi(b) - \Phi(a).$$

Beweis. Die Zufallsvariable S_n^* nimmt die Werte $x_n(k) = \frac{k - np}{\sqrt{npq}}$ für $k = 0, 1, \dots, n$ mit den Einzelwahrscheinlichkeiten $\mathbb{P}[S_n^* = x_n(k)] = \binom{n}{k} p^k q^{n-k}$ an. Die Werte liegen im Intervall $[x_n(0), x_n(n)] = \left[-\sqrt{\frac{np}{q}}, \sqrt{\frac{nq}{p}}\right]$ und die $x_n(k)$ liegen in einem Abstand von $\Delta_n := \frac{1}{\sqrt{npq}}$. Wir führen den Beweis in mehreren Schritten:

1. *Schritt:* Anwendung der Stirling'schen Formel

Es ist $\mathbb{P}[S_n^* = x_n(k)] = \frac{n!}{k!(n-k)!} p^k q^{n-k}$. Nach der Formel von Stirling gilt dann:

$$\begin{aligned} \frac{n!}{k!(n-k)!} &\sim \frac{c\sqrt{nn^n}e^{-n}}{c\sqrt{k}k^k e^{-k} c\sqrt{n-k}(n-k)^{(n-k)} e^{-(n-k)}} \\ &= \frac{\sqrt{nn^n}}{c\sqrt{k}k^k \sqrt{n-k}(n-k)^{(n-k)}} \\ &= \frac{1}{c} \sqrt{\frac{n}{k(n-k)}} \left(\frac{n}{k}\right)^k \left(\frac{n}{n-k}\right)^{n-k}. \end{aligned}$$

$$\Rightarrow \mathbb{P}[S_n^* = x_n(k)] \sim \frac{1}{c} \sqrt{\frac{n}{k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}.$$

2. Schritt: Grenzwverhalten der auftretenden Faktoren

Sei im Folgenden $x = x_n(k)$. Aus $x = \frac{k-np}{\sqrt{npq}}$ folgt:

$$\begin{aligned} k &= np + x\sqrt{npq} \quad \text{und} \quad n-k = nq - x\sqrt{npq}. \\ \Leftrightarrow \frac{k}{np} &= 1 + x\sqrt{\frac{q}{np}} \quad \text{und} \quad \frac{n-k}{nq} = 1 - x\sqrt{\frac{p}{nq}}. \end{aligned}$$

Betrachte nun $\frac{\sqrt{\frac{n}{npnq}}}{\sqrt{\frac{n}{k(n-k)}}} = \frac{\sqrt{k(n-k)}}{\sqrt{npnq}}$. Es gilt:

$$\begin{aligned} \frac{k}{np} \frac{n-k}{nq} &= \left(1 + x\sqrt{\frac{q}{np}}\right) \left(1 - x\sqrt{\frac{p}{nq}}\right) \\ &= 1 - \frac{1}{n}x^2 + x \left(\sqrt{\frac{q}{np}} - \sqrt{\frac{p}{nq}}\right) \\ &= 1 - \frac{1}{n}x^2 + \underbrace{\frac{q-p}{\sqrt{npq}}x}_{\xrightarrow{n \rightarrow \infty} 0} \xrightarrow{n \rightarrow \infty} 1. \\ \Rightarrow \sqrt{\frac{n}{k(n-k)}} &\sim \sqrt{\frac{n}{npnq}} = \frac{1}{\sqrt{npq}}. \end{aligned}$$

Für das Grenzwverhalten von $\left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k}$ betrachten wir die Reihenentwicklung vom Logarithmus: $\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$

$$\begin{aligned} \Rightarrow \log\left(\frac{np}{k}\right)^k &= -k \log\left(\frac{k}{np}\right) \\ &= -(np + x\sqrt{npq}) \log\left(1 + x\sqrt{\frac{q}{np}}\right) \\ &= -(np + x\sqrt{npq}) \left(x\sqrt{\frac{q}{np}} - \frac{x^2}{2}\frac{q}{np} + \frac{x^3}{3}\left(\sqrt{\frac{q}{np}}\right)^3 - \frac{x^4}{4}\frac{q^2}{n^2p^2} + \dots\right) \\ &= x(-\sqrt{npq}) + x^2(-q + \frac{q}{2}) + \underbrace{x^3\left(-\frac{1}{3}\sqrt{\frac{q^3}{np}} + \frac{1}{2}\sqrt{\frac{q^3}{np}}\right)}_{\xrightarrow{n \rightarrow \infty} 0} + \dots \end{aligned}$$

$$\Rightarrow \log\left(\frac{np}{k}\right)^k \sim -\sqrt{npq}x - \frac{q}{2}x^2.$$

Analog erhält man

$$\log\left(\frac{nq}{n-k}\right)^{n-k} \sim \sqrt{npq}x - \frac{p}{2}x^2.$$

Insgesamt gilt somit $\log\left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} \sim -\sqrt{npq}x - \frac{q}{2}x^2 + \sqrt{npq}x - \frac{p}{2}x^2 = -\frac{x^2}{2}$.

Daraus folgt $\forall x_n(k) \in [a, b]$:

$$\mathbb{P}[S_n^* = x_n(k)] \sim \frac{1}{c} \frac{1}{\sqrt{npq}} e^{\frac{[x_n(k)]^2}{2}}.$$

Dies ist die lokale Version des zentralen Grenzwertsatzes. Es handelt sich dabei um eine stochastische Konvergenz.

3. Schritt: Approximation durch Riemann-Summen

Um $\mathbb{P}[a \leq S_n^* \leq b]$ zu erhalten, müssen wir aufsummieren:

$$\begin{aligned}\mathbb{P}[a \leq S_n^* \leq b] &= \sum_{x_n(k) \in [a, b]} \mathbb{P}[S_n^* = x_n(k)] \\ &\sim \frac{1}{c} \sum_{x_n(k) \in [a, b]} e^{\frac{[x_n(k)]^2}{2}} \underbrace{\frac{1}{\sqrt{npq}}}_{=\Delta_n}.\end{aligned}$$

Diese Summe ist eine approximierende Riemann-Summe mit Zerlegung $x_n(k)$ für das Integral $\frac{1}{c} \int_a^b e^{-\frac{x^2}{2}} dx$, also

$$\mathbb{P}[a \leq S_n^* \leq b] \sim \frac{1}{c} \int_a^b e^{-\frac{x^2}{2}} dx.$$

4. Schritt: $c = \sqrt{2\pi}$ (2. Teil der Stirling'schen Formel)

Da Wahrscheinlichkeiten nicht größer als 1 sind, gilt $\forall a, b: \frac{1}{c} \int_a^b e^{-\frac{x^2}{2}} dx \leq 1$. Aus der Ungleichung von Tschebyscheff folgt $\forall \varepsilon > 0$:

$$\mathbb{P}[|S_n^*| \geq \varepsilon] \leq \frac{\text{Var}[S_n^*]}{\varepsilon^2} = \frac{1}{\varepsilon^2}.$$

Daraus folgt $\mathbb{P}[-\varepsilon \leq S_n^* \leq \varepsilon] \geq 1 - \frac{1}{\varepsilon^2}$. Also gilt für $n \rightarrow \infty$, $a = -\varepsilon$ und $b = \varepsilon$:

$$\frac{1}{c} \int_{-\varepsilon}^{\varepsilon} e^{-\frac{x^2}{2}} dx \geq 1 - \frac{1}{\varepsilon^2}.$$

Für $\varepsilon \rightarrow \infty$ erhalten wir also:

$$\begin{aligned}\frac{1}{c} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx &= 1. \\ \Rightarrow c^2 &= \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx dy \\ &= \int_0^{\infty} \int_0^{2\pi} e^{-\frac{r^2}{2}} r d\varphi dr = 2\pi \int_0^{\infty} e^{-\frac{r^2}{2}} r dr = 2\pi.\end{aligned}$$

Dabei wurde die Substitution $x = r \cos \varphi$ und $y = r \sin \varphi$ mit der Funktionaldeterminante r durchgeführt. Es ergibt sich also $c = \sqrt{2\pi}$ und somit insgesamt:

$$\mathbb{P}[a \leq S_n^* \leq b] \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx = \Phi(b) - \Phi(a). \quad \square$$

Der Satz von de Moivre und Laplace besagt also, dass die standardisierten Zufallsvariablen $S_n^* = \frac{S_n - np}{\sqrt{npq}}$ in Verteilung gegen eine standardnormalverteilte Zufallsvariable konvergieren.

Faustregel: Ist $npq > 9$, so ist die Normalverteilung eine gute Approximation für die Binomialverteilung. Umso näher p bei $\frac{1}{2}$, desto kleiner braucht n für gute Approximationen sein. Ist p hingegen sehr nahe bei 0 oder 1, so benötigt man große n .

4.2.3 Korrekturterme

Die Binomialverteilung ist eine diskrete Verteilung. Stellt man sie graphisch dar, so zeichnet man über jedem Intervall $[k - \frac{1}{2}, k + \frac{1}{2}]$ auf der x -Achse ein Rechteck mit der Höhe bzw. dem Flächeninhalt $\mathbb{P}[X = k]$. Da diese Histogramme mit steigendem n immer flacher werden und immer mehr nach rechts abwandern, geht man zur Standardisierung über und betrachtet $x_n(k)$ statt k . Da diese nun durch die stetige Standardnormalverteilung approximiert wird, kann die Näherung durch die sogenannte Stetigkeitskorrektur verbessert werden. Dabei wird bei der Normalverteilung die untere Grenze um 0.5 verkleinert und die obere Grenze um 0.5 vergrößert. Für große n spielen die Korrekturterme hingegen so gut wie keine Rolle.

Beispiel 4.4: Bedeutung der Stetigkeitskorrektur

Die Wahrscheinlichkeit bei 600-maligem Würfeln eines fairen Würfels zwischen 90 und 100 Mal eine Sechs zu werfen, beträgt:

$$\mathbb{P}[90 \leq S_n \leq 100] = \sum_{k=90}^{100} \binom{600}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{600-k} \approx 0,4025.$$

Eine solche Berechnung wäre ohne Computer kaum möglich. Deswegen verwendet man die Normalapproximation. Sei dazu $\sigma = \sqrt{600 \frac{1}{6} \frac{5}{6}}$.

Ohne Korrekturterme:

$$\begin{aligned} \mathbb{P}[90 \leq S_n \leq 100] &= \mathbb{P}\left[\frac{90 - 100}{\sigma} \leq S_n^* \leq \frac{100 - 100}{\sigma}\right] \\ &\approx \Phi(0) - \Phi\left(-\frac{10}{\sigma}\right) \\ &\approx 0.5 - 0.136758 \approx 0.3632. \end{aligned}$$

Mit Korrekturtermen:

$$\begin{aligned} \mathbb{P}[90 \leq S_n \leq 100] &= \mathbb{P}\left[\frac{90 - 0.5 - 100}{\sigma} \leq S_n^* \leq \frac{100 + 0.5 - 100}{\sigma}\right] \\ &\approx \Phi\left(\frac{0.5}{\sigma}\right) - \Phi\left(-\frac{10.5}{\sigma}\right) \approx 0.3968. \end{aligned}$$

Mit den Korrekturtermen ergeben sich also bessere Approximationen.

4.2.4 Beispiele

Aufgrund der zentralen Bedeutung des Satzes möchte ich noch zwei weitere Anwendungsbeispiele anführen.

Beispiel 4.5: Macht entschlossener Minderheiten

An einer Wahl zwischen zwei Parteien A und B nehmen 1000000 Wähler teil. Davon stimmen 2000 Wähler geschlossen für A, die restlichen 998000 Personen sind unentschieden und treffen ihre Wahl durch Werfen einer fairen Münze. Wie groß ist die Wahrscheinlichkeit, dass Partei A gewinnt?

S_n ist $B(998000, \frac{1}{2})$ -verteilt. Für einen Sieg benötigt A mindestens 498001 der restlichen Stimmen. Die Wahrscheinlichkeit dafür beträgt: (Korrekturterme spielen hier keine Rolle)

$$\begin{aligned}\mathbb{P}[S_n \geq 498001] &= 1 - \mathbb{P}[S_n \leq 498000] \\ &= 1 - \mathbb{P}\left[S_n^* \leq \frac{498000 - 499000}{\sqrt{249500}}\right] \\ &\approx 1 - \Phi\left(-\frac{1000}{\sqrt{249500}}\right) \approx 0.9774.\end{aligned}$$

Die Wahrscheinlichkeit für einen Sieg der Partei A ist also erstaunlich groß. Intuitiv glauben viele, dass diese 2000 Wähler der Partei A nicht wirklich einen Vorteil bringen, da sie ja lediglich 0.2 Prozent der gesamten Wählerschaft darstellen.

Beispiel 4.6: Überbuchen

Eine Fluggesellschaft nimmt - in der Praxis übliche - Überbuchungen vor, da sie aus Erfahrung weiß, dass 4 Prozent der Inhaber von Flugtickets nicht zum Abflug erscheinen. Pro Flug sind 400 Plätze vorhanden. Wie viele Buchungen können angenommen werden, damit die Wahrscheinlichkeit einer tatsächlichen Überbuchung höchstens 5 Prozent beträgt?

Es muss also $\mathbb{P}[S_n > 400] \leq 0.05$ gelten. Durch Normalapproximation erhalten wir:

$$\begin{aligned}\mathbb{P}[S_n \leq 400] &= \mathbb{P}\left[S_n^* \leq \frac{400 + 0.5 - 0.96n}{\sqrt{0.0384n}}\right] \\ &\approx \Phi\left(\frac{400.5 - 0.96n}{\sqrt{0.0384n}}\right) \geq 0.95.\end{aligned}$$

Es ist $\Phi(1.645) \approx 0.95$. Durch Lösen der Gleichung

$$\frac{400.5 - 0.96n}{\sqrt{0.0384n}} = 1.645$$

erhält man für $n \approx 410.39$. Es können also 410 Buchungen angenommen werden.

4.3 Verallgemeinerungen

Es ist eines bemerkenswertesten Resultate der Wahrscheinlichkeitstheorie, dass der Zentrale Grenzwertsatz nicht auf bernoulliverteilte Zufallsvariablen beschränkt bleibt, sondern unter sehr allgemeinen Bedingungen gilt.

4.3.1 Zentraler Grenzwertsatz von Lindeberg und Lévy

Zunächst betrachten wir eine Folge von unabhängigen und identisch verteilten Zufallsvariablen. Man beachte, dass dabei nichts über die spezielle Gestalt der Verteilung der Zufallsvariablen ausgesagt wird und das Grenzverhalten nur vom Erwartungswert und der Varianz abhängt. Insbesondere brauchen wir nicht wissen, ob es sich um diskrete oder stetige Zufallsvariablen handelt.

Satz 4.5. *Zentraler Grenzwertsatz von Lindeberg und Lévy*

Seien $X_1, \dots, X_n$ unabhängige und identisch verteilte Zufallsvariablen mit $\mathbb{E}[X_i] = \mu$ und $\text{Var}[X_i] = \sigma^2$. Sei $S_n = X_1 + \dots + X_n$ und $S_n^* := \frac{S_n - n\mu}{\sigma\sqrt{n}}$. Dann gilt $\forall t \in \mathbb{R}$:

$$\mathbb{P}[S_n^* \leq t] \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx.$$

Beweis. Ohne Beschränkung der Allgemeinheit sei $\mu = 0$, ansonsten betrachte die Zufallsvariablen $Y_i = X_i - \mu$. Wir arbeiten mit charakteristischen Funktionen. Sei $\varphi_{S_n^*}(t)$ die charakteristische Funktion von der standardisierten Zufallsvariable S_n^* . Dann gilt mit Hilfe von Satz 1.9 und Satz 1.10:

$$\begin{aligned} \varphi_{S_n^*}(t) &= \varphi_{\frac{1}{\sigma\sqrt{n}}(X_1 + \dots + X_n)}(t) = \varphi_{X_1 + \dots + X_n}\left(\frac{t}{\sigma\sqrt{n}}\right) \\ &= \varphi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right) \cdots \varphi_{X_n}\left(\frac{t}{\sigma\sqrt{n}}\right) \\ &= \left[\varphi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n. \end{aligned}$$

Nach Satz 1.12 können wir $\varphi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)$ in eine Taylorreihe um $t = 0$ entwickeln. Wir erhalten daher:

$$\begin{aligned} \left[\varphi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n &= \left[1 + \mathbb{E}[X_1] \frac{it}{\sigma\sqrt{n}} + \frac{1}{2} \mathbb{E}[X_1^2] \left(\frac{it}{\sigma\sqrt{n}}\right)^2 + o\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right]^n \\ &= \left[1 - \frac{1}{2} \sigma^2 \frac{t^2}{\sigma^2 n} + o\left(\frac{t^2}{\sigma^2 n}\right)\right]^n. \end{aligned}$$

Somit ist

$$\begin{aligned} \varphi_{S_n^*}(t) &= \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{\sigma^2 n}\right)\right]^n \\ &\xrightarrow{n \rightarrow \infty} e^{-\frac{t^2}{2}}. \end{aligned}$$

Also strebt die charakteristische Funktion von S_n^* gegen die charakteristische Funktion der $N(0, 1)$ -Verteilung. Nach dem Stetigkeitssatz folgt die Behauptung. $\square$

In Beispiel 1.6 haben wir gesehen, dass Linearkombinationen normalverteilter Zufallsvariablen wieder normalverteilt sind. Wir können somit die Standardisierung, die vorerst für einen Beweis nötig war, rückgängig machen und erhalten:

$$\sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2).$$

Daraus kann man folgendes erkennen:

1. Man sieht, dass die Normierung $\frac{1}{\sqrt{n}}$ sozusagen die Richtige war, da wegen

$$\text{Var} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right] = \frac{1}{n} \text{Var} \left[\sum_{i=1}^n X_i \right] = \frac{1}{n} n\sigma^2 = \sigma^2$$

die Varianz unabhängig von n wird.

2. Es lässt sich daraus das Schwache Gesetz der Großen Zahlen folgern: Wähle die Normierung $\frac{1}{n}$. Dann gilt:

$$\frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Wegen $\frac{\sigma^2}{n} \xrightarrow{n \rightarrow \infty} 0$ konzentriert sich die Wahrscheinlichkeitsverteilung von $\frac{1}{n} \sum_{i=1}^n X_i$ für zunehmendes n immer stärker bei ihrem Erwartungswert μ . Mit Hilfe von Satz 2.5 erhalten wir dann:

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{d} \mu \Rightarrow \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \mu.$$

4.3.2 Zentraler Grenzwertsatz von Lindeberg und Feller

Bis jetzt haben wir immer gefordert, dass die Zufallsvariablen unabhängig und identisch verteilt sind. Diese Annahme kann noch weiter abgeschwächt werden. Im folgenden Grenzwertsatz wird die Bedingung, dass die Zufallsvariablen identisch verteilt sein müssen, fallen gelassen. Lediglich die Unabhängigkeit wird noch gefordert. Jedoch müssen dann andere Bedingungen an die Folge der Zufallsvariablen gestellt werden, die im Wesentlichen sicherstellen, dass nicht ein Teil der Summanden alle anderen dominiert. Eine bekannte Bedingung dafür stammt von Lindeberg.

Definition 4.1. Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen in $\mathcal{L}^2$ mit $\mathbb{E}[X_i] = \mu_i$ und $\text{Var}[X_i] = \sigma_i^2$. Sei $s_n^2 := \sum_{i=1}^n \sigma_i^2$. Die Folge erfüllt die *Lindeberg-*

Bedingung, wenn $\forall \varepsilon > 0$ gilt:

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^n \mathbb{E} [(X_i - \mu_i)^2 \mathbb{1}_{\{|X_i - \mu_i| > \varepsilon s_n\}}] = 0.$$

Die Bedingung besagt intuitiv, dass außerhalb jeder ε -Kugel keine der Zufallsvariablen einen wesentlichen Beitrag zur Summe leisten darf. Der Beitrag jeder einzelnen Zufallsvariable muss mit großer Wahrscheinlichkeit klein sein.

Satz 4.6. *Zentraler Grenzwertsatz von Lindeberg und Feller*

Seien $X_1, X_2, \dots, X_n$ unabhängige Zufallsvariablen, die die Lindeberg-Bedingung erfüllen und sei $S_n^* := \frac{\sum_{i=1}^n (X_i - \mu_i)}{s_n}$. Dann gilt:

$$P(S_n^* \leq t) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx.$$

Beweis. Einen Beweis dazu findet man in [15]. □

Bemerkung: Die Umkehrung des Satzes gilt im Allgemeinen nicht, das heißt aus der Konvergenz von S_n^* gegen eine standardnormalverteilte Zufallsvariable folgt nicht unbedingt, dass die Lindeberg-Bedingung erfüllt ist. Stellt man an die Folge der Zufallsvariablen jedoch zusätzlich die Forderung der sogenannten Feller'schen Bedingung

$$\lim_{n \rightarrow \infty} \left(\max_{1 \leq j \leq n} \frac{\sigma_j}{s_n} \right) = 0,$$

so erhält man die Äquivalenz der beiden Aussagen.

4.3.3 Zentraler Grenzwertsatz von Ljapunov

Neben der Lindeberg-Bedingung gibt es noch eine andere hinreichende Bedingung für eine Folge von unabhängigen, aber nicht notwendigerweise identisch verteilten Zufallsvariablen, um den Zentralen Grenzwertsatz zu beweisen. Diese setzt voraus, dass die Zufallsvariablen Momente der Ordnung $(2 + \delta)$ für ein $\delta > 0$ besitzen.

Definition 4.2. Sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen mit $\mathbb{E}[X_i] = \mu_i$ und $\text{Var}[X_i] = \sigma_i^2$. Sei $s_n^2 := \sum_{i=1}^n \sigma_i^2$. Die Folge erfüllt die *Ljapunov-Bedingung*, wenn für ein $\delta > 0$ gilt:

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E} [|X_i - \mathbb{E}[X_i]|^{2+\delta}] = 0.$$

Satz 4.7. *Zentraler Grenzwertsatz von Ljapunov*

Seien $X_1, X_2, \dots, X_n$ unabhängige Zufallsvariablen, die die Ljapunov-Bedingung erfüllen und sei $S_n^* := \frac{\sum_{i=1}^n (X_i - \mu_i)}{s_n}$. Dann gilt:

$$\mathbb{P}[S_n^* \leq t] \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx.$$

Beweis. Für jedes $\varepsilon > 0$ und $n \in \mathbb{N}$ gilt die Abschätzung

$$[(X_i - \mu_i)^2 \mathbb{1}_{\{|X_i - \mu_i| > \varepsilon s_n\}}] \leq \left[\frac{|X_i - \mu_i|^{2+\delta}}{\varepsilon^\delta s_n^\delta} \mathbb{1}_{\{|X_i - \mu_i| > \varepsilon s_n\}} \right].$$

Daraus folgt:

$$\begin{aligned} \frac{1}{s_n^2} \sum_{i=1}^n \mathbb{E} [(X_i - \mu_i)^2 \mathbb{1}_{\{|X_i - \mu_i| > \varepsilon s_n\}}] &\leq \frac{1}{s_n^2} \sum_{i=1}^n \mathbb{E} \left[\frac{|X_i - \mu_i|^{2+\delta}}{\varepsilon^\delta s_n^\delta} \mathbb{1}_{\{|X_i - \mu_i| > \varepsilon s_n\}} \right] \\ &\leq \frac{1}{\varepsilon^\delta} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E} [|X_i - \mu_i|^{2+\delta}] \\ &\xrightarrow{n \rightarrow \infty} 0. \end{aligned}$$

Aus der Ljapunov-Bedingung folgt also die Lindeberg-Bedingung und aus dem Zentralen Grenzwertsatz von Lindeberg und Feller somit die Behauptung. $\square$

Bemerkung: Die Umkehrung, dass aus der Lindeberg-Bedingung die Ljapunov-Bedingung folgt, gilt nicht. Die Ljapunov-Bedingung ist also die stärkere der beiden Bedingungen. Diese wurden übrigens von den beiden Mathematikern unabhängig voneinander entdeckt.

Alle diese Sätze liefern das theoretische Fundament für die häufige Modellierung von unbekannten Größen durch normalverteilte Zufallsvariablen. Setzt sich eine Größe aus einer großen Anzahl von unabhängigen, unterschiedlich verteilten und nichtdominierenden Einzeleinflüssen zusammen, so ist diese nach dem Zentralen Grenzwertsatz annähernd normalverteilt. Eine solche Situation liegt in den Naturwissenschaften häufig vor.

5 Markov-Ketten

Bisher wurden alle wichtigen Aussagen über Folgen von Zufallsvariablen unter der Annahme der Unabhängigkeit getroffen. Sowohl bei den Gesetzen der Großen Zahlen, als auch bei den Zentralen Grenzwertsätzen wurden die Voraussetzungen immer weiter abgeschwächt - die Unabhängigkeit wurde jedoch stets gefordert. Für viele zufällige Geschehen ist diese Annahme nicht realistisch. Gerade reale Situationen sind voller Abhängigkeiten. Diese zusätzliche Voraussetzung führt jedoch dazu, dass die Berechnungen schwieriger werden. Es ist daher wichtig, ein Modell zu finden, dass die Abhängigkeit zwar berücksichtigt, jedoch weiterhin explizite Berechnungen möglich macht. Dafür eignet sich eine spezielle Klasse von stochastischen Prozessen, die Markov-Ketten genannt werden. Diese sind durch eine übersichtliche Form der Abhängigkeit charakterisiert und ermöglichen eine Interpretation der zeitlichen Entwicklung eines zufälligen Geschehens. Hauptquellen für dieses Kapitel waren [4], [9], [13] und [16]. Die Quellen [7] und [10] wurden zum Abgleich der Inhalte verwendet.

5.1 Grundlagen

Im ersten Abschnitt werden wichtige Grundlagen behandelt, die benötigt werden, um das Langzeitverhalten von Markov-Prozessen studieren zu können.

5.1.1 Definition und Eigenschaften

Definition 5.1. Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $I \neq \emptyset$ eine beliebige Indexmenge und $(\Lambda, \mathcal{C})$ ein messbarer Raum. Eine Familie $\{X_t, t \in I\}$ von Zufallsvariablen mit Werten in Λ heißt *stochastischer Prozess* mit Parameterbereich I und Zustandsraum Λ .

Bei den Markov-Ketten ist Λ endlich bzw. abzählbar und $I = \mathbb{N}_0 = \{0, 1, 2, \dots\}$. Wir haben es also mit einer Folge von diskreten Zufallsvariablen zu tun.

Definition 5.2. Eine *Markov-Kette* ist ein stochastischer Prozess $(X_n)_{n \in \mathbb{N}_0}$ mit abzählbarem Zustandsraum Λ , der die folgende Markov-Eigenschaft erfüllt: $\forall n \in \mathbb{N}_0$ und $\forall i, j, i_0, \dots, i_{n-1} \in \Lambda$ ist:

$$\mathbb{P}[X_{n+1} = j | X_n = i] = \mathbb{P}[X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0],$$

sofern $\mathbb{P}[X_n = i, X_n = i_{n-1}, \dots, X_0 = i_0] > 0$.

Wir interpretieren $X_n = i$ als Zustand des stochastischen Prozesses zur Zeit n . Die Markov-Eigenschaft besagt dann, dass die Wahrscheinlichkeit zur Zeit $n+1$ in einen

beliebigen Zustand zu gelangen nur vom Zustand zur Zeit n abhängt, nicht jedoch von den Zuständen davor. Kurz gesagt hängt die Zukunft von der Vergangenheit nur über die Gegenwart ab. Zur einfacheren Behandlung nehmen wir weiters an, dass die Wahrscheinlichkeit eines Überganges von einem Zustand zu einem anderen nicht von der Zeit n abhängt.

Definition 5.3. Eine Markov-Kette $(X_n)_{n \in \mathbb{N}_0}$ heißt *homogen* oder Kette mit *stationären Übergangswahrscheinlichkeiten*, wenn $\forall i, j \in \Lambda$ gilt, dass $\mathbb{P}[X_{n+1} = j | X_n = i]$ unabhängig von n ist. Wir definieren dann die (1-Schritt-)Übergangswahrscheinlichkeit durch

$$p_{ij} := \mathbb{P}[X_{n+1} = j | X_n = i].$$

Fassen wir diese p_{ij} zu einer $|\Lambda| \times |\Lambda|$ -Matrix zusammen, so erhalten wir eine stochastische Matrix $P = (p_{ij})_{i,j \in \Lambda}$, das heißt es gilt:

- (i) $p_{ij} \geq 0 \quad \forall i, j \in \Lambda$.
- (ii) $\sum_{j \in \Lambda} p_{ij} = 1 \quad \forall i \in \Lambda$.

Beispiel 5.1: Sei $\Lambda = \{1, 2, 3\}$, wobei die Zustände 1,2,3 für regnerisch, bewölkt und sonnig stehen. In der Matrix P kann man die Wahrscheinlichkeit ablesen, dass auf einen Tag mit Zustand $i \in \Lambda$ ein Tag mit Zustand $j \in \Lambda$ folgt:

$$P = \begin{pmatrix} 0.3 & 0.7 & 0 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.6 & 0.3 \end{pmatrix}$$

Nun könnte man sich fragen, wie das Wetter übermorgen oder in 5 Tagen aussehen wird.

Analog zu den 1-Schritt-Übergangswahrscheinlichkeiten kann man n -Schritt-Übergangswahrscheinlichkeiten definieren:

Definition 5.4. $p_{ij}^{(n)} := \mathbb{P}[X_{m+n} = j | X_m = i] \quad \forall n, m \in \mathbb{N}_0, \forall i, j \in \Lambda$.

Satz 5.1. *Chapman-Kolmogorov-Gleichung*

Ist $(X_n)_{n \in \mathbb{N}_0}$ eine Markov-Kette mit Zustandsraum Λ , so gilt für alle $i, j \in \Lambda$ und $m, n \in \mathbb{N}_0$:

$$p_{ij}^{(m+n)} = \sum_{k \in \Lambda} p_{ik}^{(m)} p_{kj}^{(n)}.$$

Beweis. Um von i nach j in $(m+n)$ Schritten zu kommen, müssen wir in m Schritten zu einem Zustand k und von dort in n Schritten weiter bis zu j gelangen:

$$\begin{aligned}
p_{ij}^{(m+n)} &= \mathbb{P}[X_{m+n} = j | X_0 = i] \\
&= \sum_{k \in \Lambda} \mathbb{P}[X_{m+n} = j, X_m = k | X_0 = i] \\
&= \sum_{k \in \Lambda} \mathbb{P}[X_{m+n} = j | X_m = k, X_0 = i] \mathbb{P}[X_m = k | X_0 = i] \\
&= \sum_{k \in \Lambda} \mathbb{P}[X_{m+n} = j | X_m = k] \mathbb{P}[X_m = k | X_0 = i] \\
&= \sum_{k \in \Lambda} p_{ik}^{(m)} p_{kj}^{(n)}. \quad \square
\end{aligned}$$

Bezeichnet $P^{(n)} = (p_{ij}^{(n)})_{i,j \in \Lambda}$ die Matrix der n -Schritt-Übergangswahrscheinlichkeiten, so besagt die Chapman-Kolmogorov-Gleichung:

$$P^{(m+n)} = P^{(m)} P^{(n)}.$$

Durch wiederholtes Anwenden erhalten wir:

$$P^{(n)} = P^{(n+1-1)} = P^{(n-1)} P = P^{(n-2)} P^2 = \dots = P^n.$$

Also erhalten wir die n -Schritt-Übergangsmatrix als n -te Potenz der 1-Schritt-Übergangsmatrix.

Beispiel 5.2: Das Wetter für Übermorgen erhalten wir somit einfach durch

$$P^{(2)} = P^2 = \begin{pmatrix} 0.3 & 0.56 & 0.14 \\ 0.26 & 0.58 & 0.16 \\ 0.24 & 0.55 & 0.21 \end{pmatrix}$$

Kennt man neben der Übergangsmatrix noch die sogenannte Startverteilung, so ist die Entwicklung der Markov-Kette vollständig spezifiziert.

Definition 5.5. Sei $(X_n)_{n \in \mathbb{N}_0}$ eine Markov-Kette mit Zustandsraum $\Lambda = \{i_1, \dots, i_n\}$. Der Zeilenvektor $\boldsymbol{\nu}^{(0)}$ mit

$$\boldsymbol{\nu}^{(0)} = (\nu_1^{(0)}, \dots, \nu_n^{(0)}) = (\mathbb{P}[X_0 = i_1], \dots, \mathbb{P}[X_0 = i_n])$$

heißt *Startverteilung* der Markov-Kette.

Die unbedingte Wahrscheinlichkeitsverteilung von $X_1, X_2, \dots$ fasst man entsprechend der Definition mit den Zeilenvektoren $\boldsymbol{\nu}^{(1)}, \boldsymbol{\nu}^{(2)}, \dots$ zusammen. Die Verteilung von X_n erhält man dann durch Multiplikation vom Vektor der Startverteilung mit der n -Schritt-Übergangsmatrix, wie der folgende Satz zeigt:

Satz 5.2. Sei $(X_n)_{n \in \mathbb{N}_0}$ eine Markov-Kette mit Zustandsraum Λ , Startverteilung $\boldsymbol{\nu}^{(0)}$ und Übergangsmatrix P . Dann gilt:

$$\boldsymbol{\nu}^{(n)} = \boldsymbol{\nu}^{(0)} P^n.$$

Beweis. Induktion nach n : Für $n = 1$ und $\forall j \in \Lambda$ gilt:

$$\begin{aligned} \nu_j^{(1)} &= \mathbb{P}[X_1 = j] \\ &= \sum_{i \in \Lambda} \mathbb{P}[X_0 = i] \mathbb{P}[X_1 = j | X_0 = i] \\ &= \sum_{i \in \Lambda} \nu_i^{(0)} p_{ij} \\ &= (\boldsymbol{\nu}^{(0)} P)_j, \end{aligned}$$

wobei im ersten Schritt der Satz von der totalen Wahrscheinlichkeit angewendet wurde und $(\boldsymbol{\nu}^{(0)} P)_j$ das j -te Element des Zeilenvektors $(\boldsymbol{\nu}^{(0)} P)$ bezeichnet. Daraus folgt $\boldsymbol{\nu}^{(1)} = \boldsymbol{\nu}^{(0)} P$.

Nach Induktionsvoraussetzung gelte nun $\boldsymbol{\nu}^{(n-1)} = \boldsymbol{\nu}^{(0)} P^{n-1}$. Dann gilt:

$$\begin{aligned} \nu_j^{(n)} &= \mathbb{P}[X_n = j] \\ &= \sum_{i \in \Lambda} \mathbb{P}[X_{n-1} = i] \mathbb{P}[X_n = j | X_{n-1} = i] \\ &= \sum_{i \in \Lambda} \nu_i^{(n-1)} p_{ij} \\ &= (\boldsymbol{\nu}^{(n-1)} P)_j. \end{aligned}$$

Also ist $\boldsymbol{\nu}^{(n)} = \boldsymbol{\nu}^{(n-1)} P = \boldsymbol{\nu}^{(0)} P^{n-1} P = \boldsymbol{\nu}^{(0)} P^n$. □

Da die Startverteilung im Weiteren oft eine Rolle spielen wird, führen wir noch eine praktische Schreibweise ein. Betrachten wir Wahrscheinlichkeiten von Ereignissen unter der Bedingung, dass die Kette in einem Zustand $i \in \Lambda$ gestartet wurde, so schreiben wir

$$\mathbb{P}_i[A] := \mathbb{P}[A | X_0 = i].$$

Analog bezeichnen wir mit $\mathbb{P}_\nu[A]$ die Wahrscheinlichkeit für das Eintreten von einem Ereignis A unter der Startverteilung $\boldsymbol{\nu}$.

5.1.2 Klassifikation von Zuständen

In diesem Abschnitt werden Eigenschaften von einzelnen Zuständen einer Markov-Kette behandelt. Sei dazu $(X_n)_{n \in \mathbb{N}_0}$ eine Markov-Kette mit Zustandsraum Λ .

Definition 5.6. Ein Zustand $i \in \Lambda$ heißt *absorbierend*, wenn $p_{ij} = 0 \ \forall j \neq i$. Das heißt, falls die Markov-Kette den Zustand i eingenommen hat, kann sie ihn nicht

mehr verlassen.

Definition 5.7. Ein Zustand $j \in \Lambda$ heißt von $i \in \Lambda$ aus *erreichbar* ($i \rightarrow j$), wenn $\exists n \in \mathbb{N}$ mit $p_{ij}^{(n)} > 0$. Ist $i \rightarrow j$ und $j \rightarrow i$, so sagen wir, dass i und j miteinander *kommunizieren* und schreiben $i \leftrightarrow j$.

Definition 5.8. $C \subseteq \Lambda$ heißt *abgeschlossen*, wenn C nicht verlassen werden kann, das heißt $p_{ij} = 0$, wenn $i \in C$ und $j \notin C$. ($\Leftrightarrow \sum_{k \in C} p_{ik} = 1$)

Satz 5.3. Die Relation des Kommunizierens ist eine Äquivalenzrelation.

Beweis. Die Reflexivität gilt wegen $p_{ii}^{(0)} = 1$ und die Symmetrie folgt aus der Definition. Es muss nur noch die Transitivität gezeigt werden:

Sei dazu $i \rightarrow j$ und $j \rightarrow k$, das heißt $\exists n, m \in \mathbb{N}$ mit $p_{ij}^{(n)} > 0$ und $p_{jk}^{(m)} > 0$. Mit der Chapman-Kolmogorov-Gleichung folgt

$$p_{ik}^{(n+m)} = \sum_{s \in \Lambda} p_{is}^{(n)} p_{sk}^{(m)} = p_{ij}^{(n)} p_{jk}^{(m)} + \sum_{s \neq j, s \in \Lambda} p_{is}^{(n)} p_{sk}^{(m)} \geq p_{ij}^{(n)} p_{jk}^{(m)} > 0.$$

Analog erhält man aus $k \rightarrow j$ und $j \rightarrow i$, dass $k \rightarrow i$ und somit $i \leftrightarrow k$. $\square$

Der Zustandsraum kann also in Äquivalenzklassen von miteinander kommunizierenden Zuständen zerlegt werden. Es kann dabei auch der Spezialfall eintreten, dass der gesamte Zustandsraum durch eine einzige Äquivalenzklasse ausgefüllt wird. In diesem Fall nennt man die Markov-Kette irreduzibel. Allgemein definiert man:

Definition 5.9. $C \subseteq \Lambda$ heißt *irreduzibel*, wenn $i \leftrightarrow j \forall i, j \in C$.

Beispiel 5.3: Sei $P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0.5 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Der Zustand 3 ist absorbierend. $\Lambda = \{1, 2, 3\}$ ist reduzibel, da nicht alle Zustände miteinander kommunizieren. $C = \{1, 2\} \subseteq \Lambda$ ist abgeschlossen und irreduzibel.

Definition 5.10. Für einen Zustand $i \in \Lambda$ heißt der größte gemeinsame Teiler $d(i) := \text{ggT}\{n \in \mathbb{N} : p_{ii}^{(n)} > 0\}$ die *Periode* von i . Gilt nicht $i \rightarrow i$, so sei $d(i) = \infty$.

Definition 5.11. Gilt $d(i) = 1$, so heißt der Zustand $i \in \Lambda$ *aperiodisch*. Eine Markov-Kette heißt aperiodisch, wenn alle Zustände aperiodisch sind und periodisch mit Periode d , wenn alle $d(i) = d \geq 2$ sind.

Beispiel 5.4: Sei $P_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ und $P_2 = \begin{pmatrix} 0 & 1 \\ 0.5 & 0.5 \end{pmatrix}$.

Die Markov-Kette mit Übergangsmatrix P_1 besitzt Periode 3, da jeder Zustand Periode 3 hat. Startet man bei der Markov-Kette mit Übergangsmatrix P_2 in einem

Zustand $\Lambda = \{1, 2\}$, so gibt es mehrere Möglichkeiten wieder zu diesem Zustand zu kommen. Betrachtet man zum Beispiel den Zustand 1, so ist $p_{11}^{(2)} = \frac{1}{2} > 0$ und $p_{11}^{(3)} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} > 0$ und somit $d = 1$. Der Zustand 1 ist also aperiodisch. Gleiches gilt für Zustand 2.

Satz 5.4. *Gilt $i \leftrightarrow j$, so ist $d(i) = d(j)$.*

Beweis. Nach Voraussetzung $\exists n, m \in \mathbb{N}$ mit $p_{ij}^{(n)} > 0$ und $p_{ji}^{(m)} > 0$. Für $k \in \mathbb{N}$ sei $p_{jj}^{(k)} > 0$. Dann gilt:

$$\begin{aligned} p_{ii}^{(n+m)} &\geq p_{ij}^{(n)} p_{ji}^{(m)} > 0 \quad \text{und} \\ p_{ii}^{(n+m+k)} &\geq p_{ij}^{(n)} p_{jj}^{(k)} p_{ji}^{(m)} > 0. \end{aligned}$$

Also gilt $d(i) \mid n+k+m$ und $d(i) \mid n+m$ und daher $d(i) \mid (n+k+m) - (n+m) = k$. Damit teilt $d(i)$ alle $k \in \mathbb{N}$ mit $p_{jj}^{(k)} > 0$ und $d(i) \leq d(j)$. Durch Vertauschen von i und j erhält man analog $d(j) \leq d(i)$. $\square$

5.1.3 Rekurrenz und Transienz

In diesem Abschnitt beschäftigen wir uns mit der Frage, ob man zu einem ausgehenden Zustand mit Wahrscheinlichkeit 1 wieder zurückkehrt oder nicht. Dazu definieren wir zunächst die Ersteintrittszeit und die erstmalige Übergangswahrscheinlichkeit.

Definition 5.12. Sei $i \in \Lambda$. Die *Ersteintrittszeit* von i ist definiert durch

$$T_i := \inf\{n \in \mathbb{N} : X_n = i\}.$$

Definition 5.13. Seien $i, j \in \Lambda$. Die Wahrscheinlichkeit, dass eine i gestartete Kette zum Zeitpunkt n den Zustand j zum ersten Mal erreicht, ist definiert durch

$$f_{ij}^{(n)} := \mathbb{P}_i[X_1 \neq j, X_2 \neq j, \dots, X_{n-1} \neq j, X_n = j] = \mathbb{P}_i[T_j = n].$$

Weiters sei

$$f_{ij}^* := \sum_{n=1}^{\infty} f_{ij}^{(n)}.$$

Bemerkung: Es gilt dann:

$$\mathbb{P}_i[T_j < \infty] = \mathbb{P}_i\left[\bigcup_{n=1}^{\infty} \{T_j = n\}\right] = \sum_{n=1}^{\infty} \mathbb{P}_i[T_j = n] = \sum_{n=1}^{\infty} f_{ij}^{(n)} = f_{ij}^*.$$

Das heißt f_{ij}^* entspricht der Wahrscheinlichkeit, dass eine in i gestartete Kette jemals den Zustand j erreicht.

Definition 5.14. Ein Zustand $i \in \Lambda$ heißt *rekurrent*, wenn $f_{ii}^* = 1$ und *transient*, wenn $f_{ii}^* < 1$.

Jeder Zustand ist also rekurrent oder transient. Ist der Zustand i rekurrent, so bedeutet das, dass man bei Start in i mit Sicherheit wieder zu i zurückkehrt. Ist i transient, so ist die Wahrscheinlichkeit für eine Rückkehr kleiner 1.

Satz 5.5.

- (i) Ist $i \in \Lambda$ rekurrent, so gilt $\mathbb{P}_i[X_n = i \text{ unendlich oft}] = 1$ und $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$.
- (ii) Ist $i \in \Lambda$ transient, so gilt $\mathbb{P}_i[X_n = i \text{ unendlich oft}] = 0$ und $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$.

Beweis. Sei $F_n := \{X_n = i, X_{n+1} \neq i, X_{n+2} \neq i, \dots\}$ das Ereignis, dass zur Zeit n der Zustand i zum letzten Mal besucht wird und $F_0 := \{X_k \neq i \ \forall k \in \mathbb{N}\}$. Dann sind die Mengen F_n disjunkt und es gilt $\{X_n = i \text{ unendlich oft}\}^c = \bigcup_{n=0}^{\infty} F_n$. Also gilt:

$$1 - \mathbb{P}_i[X_n = i \text{ unendlich oft}] = \sum_{n=0}^{\infty} \mathbb{P}_i[F_n]. \quad (5.1)$$

Außerdem gilt:

$$\mathbb{P}_i[F_n] = \mathbb{P}_i[X_n = i] \mathbb{P}_i[X_{n+1} \neq i, X_{n+2} \neq i, \dots | X_n = i] = p_{ii}^{(n)} \mathbb{P}_i[F_0] \quad (5.2)$$

und

$$\mathbb{P}_i[F_0] = 1 - f_{ii}^*.$$

- (i) $f_{ii}^* = 1 \Rightarrow \mathbb{P}_i[F_0] = 0$ und somit folgt aus (5.2), dass $\mathbb{P}_i[F_n] = 0$ für alle $n \in \mathbb{N}$. Aus (5.1) folgt dann $\mathbb{P}_i[X_n = i \text{ unendlich oft}] = 1$ und es ist

$$\sum_{n=1}^{\infty} \mathbb{P}_i[X_n = i] = \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty,$$

denn sonst würde $\mathbb{P}_i[X_n = i \text{ unendlich oft}] = 0$ nach dem Lemma von Borel-Cantelli gelten.

- (ii) $f_{ii}^* < 1 \Rightarrow \mathbb{P}_i[F_0] > 0$. Mit (5.1) und (5.2) folgt

$$1 - \mathbb{P}_i[X_n = i \text{ u.o.}] = \sum_{n=0}^{\infty} \mathbb{P}_i[F_n] = \underbrace{\mathbb{P}_i[F_0]}_{>0} \left(1 + \sum_{n=1}^{\infty} p_{ii}^{(n)} \right) < 1.$$

Daher ist $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$. Nach dem Lemma von Borel-Cantelli ist daher auch $\mathbb{P}_i[X_n = i \text{ unendlich oft}] = 0$. $\square$

Korollar 5.6. Ist $i \in \Lambda$ ein transienter Zustand, so gilt:

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \frac{1}{1 - f_{ii}^*}.$$

Beweis. Da i transient ist, gilt $\mathbb{P}_i[X_n = i \text{ unendlich oft}] = 0$. Nach dem 2. Teil des Beweises des vorigen Satzes und $\mathbb{P}_i[F_0] = 1 - f_{ii}^*$ gilt:

$$1 = (1 - f_{ii}^*) \left(1 + \sum_{n=1}^{\infty} p_{ii}^{(n)} \right)$$

Die erwartete Anzahl von Besuchen in i ab dem Startzeitpunkt ist daher:

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \frac{1}{1 - f_{ii}^*}. \quad \square$$

Man könnte die Aussage des Satzes auch anschaulich begründen: Wenn $f_{ii}^* = 1$, so kehrt man sicher zu i zurück. Ist man zurückgekehrt, hat man wieder dieselbe Situation und wird wieder mit Wahrscheinlichkeit 1 zurückkehren etc. Also wird man unendlich oft zurückkehren. Ist $f_{ii}^* < 1$, so gibt es bei jeder Rückkehr eine positive Wahrscheinlichkeit, nie mehr zurückzukehren. Da man beliebig viele Versuche hat, wird dies sicher passieren und damit ist die Wahrscheinlichkeit für unendlich viele Rückkehren 0. Obwohl diese Argumentation nicht ganz korrekt ist, da man annimmt, dass nach einer Rückkehr der Prozess neu gestartet wird, kann man sich den Satz damit gut veranschaulichen.

Da das Nachprüfen von Kriterien für Rekurrenz und Transienz sehr mühsam ist, stellt sich die Frage, ob man wirklich für jeden Zustand einzeln die Kriterien prüfen muss. Wie wir sehen werden, ist dies nicht der Fall - man kann die Zustände in Klassen einteilen, sodass alle Zustände innerhalb einer Klasse dieselbe Rekurrenzeigenschaft haben.

Satz 5.7. *Seien $i, j \in \Lambda$ und $i \leftrightarrow j$. Dann sind i und j entweder beide rekurrent oder beide transient.*

Beweis. Sei j rekurrent. Da $i \leftrightarrow j \Rightarrow \exists r, s \in \mathbb{N}$ mit $p_{ij}^{(r)} > 0$ und $p_{ji}^{(s)} > 0$. Es gilt:

$$p_{ii}^{(n+r+s)} \geq p_{ij}^{(r)} p_{jj}^{(n)} p_{ji}^{(s)} > 0.$$

Da j rekurrent ist, gilt:

$$\sum_{n=0}^{\infty} p_{jj}^{(n)} = \infty \Rightarrow \sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$$

und daher ist auch i rekurrent. Der Beweis funktioniert analog, wenn i rekurrent ist und ebenso für die Transienz. $\square$

Das heißt Rekurrenz und Transienz sind Klasseneigenschaften, sie hängen nur von den Kommunikationsklassen ab. Daher sind alle Zustände einer irreduziblen Menge $C \subseteq \Lambda$ entweder rekurrent oder transient. Dies kann man mit dem Konvergenzver-

halten von $\sum_{n=0}^{\infty} p_{ii}^{(n)}$ für ein beliebiges $i \in C$ feststellen. Wir wollen nun noch zeigen, wie wir Λ in rekurrente und transiente Zustände zerlegen können.

Lemma 5.8. *Seien $i, j \in \Lambda$. Ist i rekurrent und $i \rightarrow j$, dann folgt $j \rightarrow i$.*

Beweis. Da $i \rightarrow j \Rightarrow \exists s \in \mathbb{N}$ mit $p_{ij}^{(s)} > 0$. Wäre $j \nrightarrow i$, so wäre die Wahrscheinlichkeit, dass die Markov-Kette nicht mehr in i zurückkehrt, größer 0, also $f_{ii}^* < 1$. Das ist aber ein Widerspruch dazu, dass i rekurrent ist. $\square$

Satz 5.9. Zerlegungssatz

Der Zustandsraum Λ einer Markov-Kette kann folgendermaßen zerlegt werden:

$$\Lambda = T \cup R_1 \cup R_2 \cup \dots,$$

wobei T die Menge der transienten Zustände und R_n mit $n \in \mathbb{N}$ eine irreduzible, abgeschlossene Rekurrenzklasse ist.

Beweis. Zunächst hat Λ eine disjunkte Zerlegung in abzählbar viele Kommunikationsklassen:

$$\Lambda = \bigcup_{k=1}^{\infty} C_k,$$

wobei jedes dieser C_k ausschließlich transiente oder rekurrente Zustände enthält. Sei $T = \bigcup_{C_k \text{ transient}} C_k$ und R_n mit $n \in \mathbb{N}$ die nicht enthaltenen Klassen in T . Es folgt die Zerlegung $\Lambda = T \cup R_1 \cup R_2 \cup \dots$. Nach Konstruktion ist jedes R_n eine rekurrente Kommunikationsklasse und irreduzibel. Die Abgeschlossenheit von R_n folgt durch das vorige Lemma:

Wenn $i \in R_n$ und $i \rightarrow j \Rightarrow j \rightarrow i$, da i rekurrent. Daher gilt $i \leftrightarrow j$ und somit auch $j \in R_n$. $\square$

Ist $i \in \Lambda$ ein rekurrenter Zustand, so gilt nach Definition $f_{ii}^* = \mathbb{P}_i[T_i < \infty] = 1$. Wenn wir nun eine andere Startverteilung ν haben, gilt dann immer noch $\mathbb{P}_\nu[T_i < \infty] = 1$? Bei einer irreduziblen und rekurrenten Markov-Kette ist dies für alle Zustände der Fall:

Satz 5.10. *Sei $(X_n)_{n \in \mathbb{N}_0}$ eine irreduzible, rekurrente Markov-Kette mit Zustandsraum Λ und Startverteilung ν . Dann gilt $\forall i \in \Lambda$:*

$$\mathbb{P}_\nu[T_i < \infty] = 1.$$

Beweis. Es gilt:

$$\mathbb{P}_\nu[T_i < \infty] = \sum_{j \in \Lambda} \mathbb{P}_\nu[X_0 = j] \mathbb{P}_j[T_i < \infty].$$

Somit reicht es, $\mathbb{P}_j[T_i < \infty] = 1$ für alle $j \in \Lambda$ zu zeigen: Zu festem i und j wählen wir ein $m \in \mathbb{N}$, so dass $p_{ij}^{(m)} > 0$. Mit Satz 5.5 und der Markov-Eigenschaft erhalten wir:

$$\begin{aligned} 1 &= \mathbb{P}_i[X_n = i \text{ unendlich oft}] = \mathbb{P}_i[X_n = i \text{ für ein } n > m] \\ &= \sum_{k \in \Lambda} \mathbb{P}_i[X_m = k] \mathbb{P}_i[X_n = i \text{ für ein } n > m | X_m = k] \\ &= \sum_{k \in \Lambda} p_{ik}^{(m)} \mathbb{P}_k[T_i < \infty]. \end{aligned}$$

Aus $\sum_{k \in \Lambda} p_{ik}^{(m)} = 1$ folgt $\mathbb{P}_j[T_i < \infty] = 1$. □

5.1.4 Stationarität

Die Verteilung einer Markov-Kette hängt im Allgemeinen auf komplizierte Weise von der Startverteilung und natürlich vom Zeitpunkt n ab. Oft gibt es jedoch für eine Übergangsmatrix spezielle Verteilungen, sodass die mit dieser Verteilung gestartete Kette zu jedem Zeitpunkt dieselbe Verteilung besitzt.

Definition 5.15. Eine Wahrscheinlichkeitsverteilung $\pi = (\pi_1, \dots, \pi_n)$ heißt *stationäre Verteilung*, falls $\forall j \in \Lambda$ gilt:

$$\pi_j = \sum_{i \in \Lambda} \pi_i p_{ij}.$$

Oder in Matrixnotation:

$$\pi = \pi P.$$

Das heißt, wenn die Markov-Kette die Verteilung π zu einem bestimmten Zeitpunkt n besitzt, dann auch in allen nachfolgenden Zeitpunkten. Die Verteilung π ist ein linker Eigenvektor zum Eigenwert 1. Also kann man mithilfe der Linearen Algebra eine stationäre Verteilung bestimmen. Man bestimmt dazu den Eigenvektor zum Eigenwert 1 und normiert diesen, um eine Wahrscheinlichkeitsverteilung zu erhalten.

Wir beschäftigen uns nun mit der Frage, ob und unter welchen Bedingungen eine stationäre Verteilung existiert. Diese Frage ist eng verknüpft mit der erwarteten Rückkehrzeit.

Definition 5.16. Sei $i \in \Lambda$. Die *erwartete Rückkehrzeit* ist definiert durch

$$\mu_{ii} = \mathbb{E}_i[T_i] = \sum_{n=1}^{\infty} n \mathbb{P}_i[T_i = n] = \sum_{n=1}^{\infty} n f_{ii}^{(n)}.$$

Ist $\mathbb{P}_i[T_i = \infty] > 0$, so sei $\mu_{ii} = \infty$.

Falls $\mu_{ii} < \infty$, so ist natürlich $\mathbb{P}_i[T_i < \infty] = f_{ii}^* = 1$. Mithilfe von μ_{ii} können wir also rekurrente Zustände weiter unterteilen:

Definition 5.17. Sei $i \in \Lambda$ ein rekurrenter Zustand. Dann heißt i *positiv rekurrent*, falls $\mu_{ii} < \infty$ und *nullrekurrent*, falls $\mu_{ii} = \infty$.

Zunächst zeigen wir, dass eine stationäre Verteilung zu einer irreduziblen Matrix strikt positiv ist, falls sie existiert.

Lemma 5.11. Für eine irreduzible Markov-Kette mit Zustandsraum Λ und stationärer Verteilung π gilt:

$$\pi_i > 0 \quad \forall i \in \Lambda$$

Beweis. Wir teilen Λ in zwei disjunkte Mengen $\Lambda = \Lambda_+ \cup \Lambda_0$, so dass $\pi_i > 0 \quad \forall i \in \Lambda_+$. $\Lambda_+ \neq \emptyset$, da π eine Verteilung ist. Wir müssen nun zeigen, dass $\Lambda_0 = \emptyset$:

Sei $\Lambda_0 \neq \emptyset$. Dann folgt aus der Stationarität $\forall j \in \Lambda_0$:

$$\sum_{i \in \Lambda} \pi_i p_{ij} = \pi_j = 0.$$

Da alle Summanden ≥ 0 sind, folgt $\forall i \in \Lambda_+$ und $j \in \Lambda_0$:

$$p_{ij} = 0.$$

Da diese Markov-Kette irreduzibel ist, existiert zu jedem $i \in \Lambda_+$ und $j \in \Lambda_0$ ein n mit $p_{ij}^{(n)} > 0$, also existieren $i_1, \dots, i_{n-1} \in \Lambda$, sodass $p_{ii_1} p_{i_1 i_2} \dots p_{i_{n-1} j} > 0$. Da $i \in \Lambda_+$ und $j \in \Lambda_0$, muss wegen $p_{ij} = 0$ mindestens ein Faktor gleich 0 sein - Widerspruch. $\square$

Mit Hilfe dieses Lemmas können wir nun zeigen, dass positive Rekurrenz für die Existenz einer stationären Verteilung ein notwendiges Kriterium ist und die stationäre Verteilung dann durch $\pi_i = \frac{1}{\mu_{ii}}$ eindeutig gegeben ist.

Satz 5.12. Sei $(X_n)_{n \in \mathbb{N}_0}$ eine irreduzible Markov-Kette mit Zustandsraum Λ und π eine stationäre Verteilung. Dann sind alle $i \in \Lambda$ positiv rekurrent und die eindeutige stationäre Verteilung gegeben durch:

$$\pi_i = \frac{1}{\mu_{ii}} > 0 \quad \forall i \in \Lambda.$$

Beweis. Wir zeigen zunächst, dass $\pi_i \mu_{ii} = \mathbb{P}_\pi[T_i < \infty]$. Es gilt:

$$\begin{aligned} \{T_i \leq n\} &= \bigcup_{k=1}^n \{X_k = i, X_{k+1} \neq i, \dots, X_n \neq i\} \\ &= \bigcup_{k=0}^{n-1} \{X_{n-k} = i, X_{n-k+1} \neq i, \dots, X_n \neq i\}. \end{aligned}$$

Betrachten wir nun die Wahrscheinlichkeit, so können wir wegen der Stationarität um $n - k$ Zeitpunkte verschieben. Wir erhalten daher:

$$\begin{aligned} \mathbb{P}_\pi[T_i < \infty] &= \lim_{n \rightarrow \infty} \mathbb{P}_\pi[T_i \leq n] \\ &= \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} \mathbb{P}[X_{n-k} = i, X_{n-k+1} \neq i, \dots, X_n \neq i] \\ &= \lim_{n \rightarrow \infty} \sum_{k=0}^{n-1} \mathbb{P}[X_0 = i, X_1 \neq i, \dots, X_k \neq i] \\ &= \sum_{k=0}^{\infty} \mathbb{P}_i[T_i > k] \mathbb{P}_\pi[X_0 = i] \\ &= \mathbb{E}_i[T_i] \pi_i \\ &= \mu_{ii} \pi_i. \end{aligned}$$

Nach dem vorigen Lemma ist $\pi_i > 0 \ \forall i \in \Lambda$ und daher $\mu_{ii} \leq \frac{1}{\pi_i} < \infty$. Insbesondere gilt also $\mathbb{P}_i[T_i < \infty] = 1$, das heißt i ist rekurrent. Da die Markov-Kette irreduzibel ist, folgt nach Satz 5.10 $\mathbb{P}_\pi[T_i < \infty] = 1 \ \forall i \in \Lambda$, also $\mu_{ii} \pi_i = 1$. $\square$

Umgekehrt kann man zeigen, dass bereits die positive Rekurrenz eines Zustandes die Existenz einer stationären Verteilung zur Folge hat. Einen Beweis dazu findet man in [9]. Wir erhalten also insgesamt folgende Äquivalenz:

Satz 5.13. *Sei $(X_n)_{n \in \mathbb{N}_0}$ eine irreduzible Markov-Kette. Dann sind äquivalent:*

- (i) *Es existiert eine stationäre Verteilung.*
- (ii) *Es gibt einen positiv rekurrenten Zustand.*
- (iii) *Alle Zustände sind positiv rekurrent.*

Im Falle der Existenz der stationären Verteilung ist sie durch $\pi_i = \frac{1}{\mu_{ii}}$ eindeutig gegeben.

5.2 Grenzverhalten von Markov-Ketten

In diesem Kapitel beschäftigen wir uns mit dem Konvergenzverhalten der Übergangswahrscheinlichkeiten $p_{ij}^{(n)}$. Zunächst formulieren wir einen wichtigen Satz, der der Schlüssel zu vielen Resultaten ist.

Satz 5.14. *Sei $(X_n)_{n \in \mathbb{N}_0}$ eine Markov-Kette mit Zustandsraum Λ . Dann gilt für alle $i, j \in \Lambda$:*

$$p_{ij}^{(n)} = \sum_{k=1}^n f_{ij}^{(k)} p_{jj}^{(n-k)}.$$

Beweis. Wir spalten das Ereignis $\{X_n = j\}$ nach dem ersten Zeitpunkt, an dem die Kette j erreicht hat, auf. Somit erhalten wir:

$$\begin{aligned} p_{ij}^{(n)} &= \sum_{k=1}^n \mathbb{P}_i[T_k = j, X_n = j] \\ &= \sum_{k=1}^n \mathbb{P}_i[T_k = j] \mathbb{P}[X_n = j | X_k = j] \\ &= \sum_{k=1}^n f_{ij}^{(k)} p_{jj}^{(n-k)}. \end{aligned}$$

□

Diese Formel ist plausibel: Wenn man in i startet, muss man, um j in n Schritten zu erreichen, j ein erstes Mal in $k \leq n$ Schritten erreichen und dann in den restlichen $n - k$ Schritten von j wieder nach j zurückkehren.

5.2.1 Transiente Zustände

In diesem Abschnitt betrachten wir das Konvergenzverhalten von $p_{ij}^{(n)}$ für transiente Zustände j . Dieser Fall lässt sich sehr schnell erledigen.

Satz 5.15. *Sei $i, j \in \Lambda$, wobei j ein transienter und i ein beliebiger Zustand ist. Dann gilt:*

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0.$$

Beweis. Wenn j transient ist, erhalten wir mit Satz 5.5 $\sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty$ und somit $\lim_{n \rightarrow \infty} p_{jj}^{(n)} = 0$. Nach dem vorigen Satz folgt dann:

$$\begin{aligned} \sum_{n=1}^{\infty} p_{ij}^{(n)} &= \sum_{n=1}^{\infty} \sum_{k=1}^n f_{ij}^{(k)} p_{jj}^{(n-k)} \\ &= \sum_{k=1}^{\infty} f_{ij}^{(k)} \sum_{n=k}^{\infty} p_{jj}^{(n-k)} \end{aligned}$$

$$= f_{ij}^* \sum_{n=0}^{\infty} p_{jj}^{(n)} < \infty.$$

Somit gilt $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0$. □

5.2.2 Rekurrente Zustände mit endlichem Zustandsraum

In den nächsten beiden Abschnitten erhalten wir eines der wichtigsten Resultate über Markov-Ketten: Jede irreduzible, aperiodische und rekurrente Markov-Kette strebt gegen ihre stationäre Verteilung. Wir betrachten dieses Ergebnis zunächst für endliche Markov-Ketten und brauchen somit vorerst nicht zwischen positiver Rekurrenz und Nullrekurrenz unterscheiden. Dazu benötigen wir folgendes Lemma:

Lemma 5.16. *Eine endliche Markov-Kette $(X_n)_{n \in \mathbb{N}_0}$ ist genau dann irreduzibel und aperiodisch, wenn die L -Schritt-Übergangsmatrix $(p_{ij}^{(L)})_{i,j \in \Lambda}$ für ein $L \geq 1$ nur strikt positive Elemente enthält.*

Beweis. Der Beweis beruht auf zahlentheoretischen Argumenten und ist zum Beispiel in [12] zu finden. □

Satz 5.17. *Sei $(X_n)_{n \in \mathbb{N}_0}$ eine endliche, irreduzible und aperiodische Markov-Kette mit Zustandsraum Λ und stationärer Verteilung π . Dann gilt:*

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi_j = \frac{1}{\mu_{ii}}.$$

Beweis. Sei $j \in \Lambda$ beliebig, aber fest. Wir setzen

$$m_j^{(n)} := \min_{i \in \Lambda} p_{ij}^{(n)},$$

$$M_j^{(n)} := \max_{i \in \Lambda} p_{ij}^{(n)}.$$

Diese entsprechen den minimalen bzw. maximalen Werten der j -ten Spalte von P^n . Es gilt:

$$\begin{aligned} m_j^{(n+1)} &= \min_{i \in \Lambda} p_{ij}^{(n+1)} \\ &= \min_{i \in \Lambda} \sum_{k \in \Lambda} p_{ik} p_{kj}^{(n)} \\ &\geq \min_{i \in \Lambda} \sum_{k \in \Lambda} p_{ik} m_j^{(n)} \\ &= m_j^{(n)}. \end{aligned}$$

Analog erhält man $M_j^{(n+1)} \leq M_j^{(n)}$. Daher ist die Folge $(m_j^{(n)})_{n \in \mathbb{N}}$ monoton wachsend und $(M_j^{(n)})_{n \in \mathbb{N}}$ monoton fallend. Wir wollen nun zeigen, dass $M_j^{(n)} - m_j^{(n)} \xrightarrow{n \rightarrow \infty} 0$:

Da P irreduzibel und aperiodisch ist, folgt nach dem vorigen Lemma, dass es ein $L \geq 1$ und $\delta > 0$ gibt, sodass

$$p_{ij}^{(L)} \geq \delta \quad \forall i, j \in \Lambda.$$

Für festes $n \in \mathbb{N}$ wählen wir ein $i_0 \in \Lambda$ mit $p_{i_0j}^{(n+L)} = m_j^{(n+L)}$ und ein $i_1 \in \Lambda$ mit $p_{i_1j}^{(n+L)} = M_j^{(n+L)}$. Sei

$$\begin{aligned} I_+ &:= \{k \in \Lambda : p_{i_1k}^{(L)} \geq p_{i_0k}^{(L)}\}, \text{ und} \\ I_- &:= \Lambda \setminus I_+. \end{aligned}$$

Damit erhalten wir:

$$\sum_{k \in I_+} (p_{i_1k}^{(L)} - p_{i_0k}^{(L)}) + \sum_{k \in I_-} (p_{i_1k}^{(L)} - p_{i_0k}^{(L)}) = \sum_{k \in \Lambda} p_{i_1k}^{(L)} - \sum_{k \in \Lambda} p_{i_0k}^{(L)} = 1 - 1 = 0.$$

Damit gilt dann:

$$\begin{aligned} M_j^{(n+L)} - m_j^{(n+L)} &= p_{i_1j}^{(n+L)} - p_{i_0j}^{(n+L)} \\ &= \sum_{k \in \Lambda} p_{i_1k}^{(n)} p_{kj}^{(L)} - \sum_{k \in \Lambda} p_{i_0k}^{(n)} p_{kj}^{(L)} \\ &= \sum_{k \in \Lambda} (p_{i_1k}^{(L)} - p_{i_0k}^{(L)}) p_{kj}^{(n)} \\ &\leq \sum_{k \in I_+} (p_{i_1k}^{(L)} - p_{i_0k}^{(L)}) M_j^{(n)} + \sum_{k \in I_-} (p_{i_1k}^{(L)} - p_{i_0k}^{(L)}) m_j^{(n)} \\ &\leq \sum_{k \in I_+} (p_{i_1k}^{(L)} - p_{i_0k}^{(L)}) (M_j^{(n)} - m_j^{(n)}) \\ &\leq (M_j^{(n)} - m_j^{(n)}) \sum_{k \in I_+} (p_{i_1k}^{(L)} - \delta) \\ &\leq (1 - \delta) (M_j^{(n)} - m_j^{(n)}). \end{aligned}$$

Induktiv folgt nun für jedes $m \in \mathbb{N}$:

$$(M_j^{(mL)} - m_j^{(mL)}) \leq \underbrace{(M_j^{(0)} - m_j^{(0)})}_{\leq 1} \xrightarrow{m \rightarrow \infty} 0.$$

Wir setzen nun $\pi_j = \lim_{n \rightarrow \infty} m_j^{(n)} = \lim_{n \rightarrow \infty} M_j^{(n)}$. Da die Folgen $(m_j^{(n)})_{n \in \mathbb{N}}$ und $(M_j^{(n)})_{n \in \mathbb{N}}$ monoton und beschränkt sind, existieren diese Grenzwerte und wegen $m_j^{(n)} \leq M_j^{(n)}$ folgt sofort $\lim_{n \rightarrow \infty} = \pi_j \quad \forall j \in \Lambda$.

Nun müssen wir noch zeigen, dass π eine stationäre Verteilung und ein Wahrscheinlichkeitsvektor ist, der eindeutig bestimmt ist.

Da $(p_{1j}^{(n)})_{j \in \Lambda}$ ein Wahrscheinlichkeitsvektor ist und dieser für $n \rightarrow \infty$ gegen π konvergiert, ist auch π ein Wahrscheinlichkeitsvektor. Weiters gilt:

$$\sum_{k \in \Lambda} \pi_k p_{kj} \xleftarrow{n \rightarrow \infty} \sum_{k \in \Lambda} p_{ik}^{(n)} p_{kj} = p_{ij}^{(n+1)} \xrightarrow{n \rightarrow \infty} \pi_j.$$

Also ist $\sum_{k \in \Lambda} \pi_k p_{kj} = \pi_j$ und π somit eine stationäre Verteilung.

Sei nun π' eine weitere Wahrscheinlichkeitsverteilung mit $\pi' = \pi' P$. Dann ist auch $\pi' = \pi' P^n \forall n \in \mathbb{N}$. Also folgt $\forall j \in \Lambda$:

$$\pi'_j = \sum_{k \in \Lambda} \pi'_k p_{kj}^{(n)} \xrightarrow{n \rightarrow \infty} \sum_{k \in \Lambda} \pi'_k \pi_j \underbrace{\sum_{k \in \Lambda} \pi'_k}_{=1} = \pi_j$$

und daher $\pi' = \pi$. □

Der Satz besagt insbesondere auch, dass eine Markov-Kette mit den Bedingungen des Satzes ihre Anfangsverteilung vergisst, denn:

$$\lim_{n \rightarrow \infty} \mathbb{P}_\nu[X_n = j] = \lim_{n \rightarrow \infty} \sum_{i \in \Lambda} \nu_i p_{ij}^{(n)} = \lim_{n \rightarrow \infty} \pi_j \underbrace{\sum_{i \in \Lambda} \nu_i}_{=1} = \pi_j.$$

5.2.3 Rekurrente Zustände mit abzählbarem Zustandsraum

In endlichen Zustandsräumen konnte der Satz also mit elementaren Hilfsmitteln bewiesen werden. Des Weiteren mussten wir nicht zwischen positiv-rekurrent und nullrekurrent unterscheiden, da die Rekurrenz im endlichen Fall äquivalent mit positiver Rekurrenz ist. Wir wollen den Satz nun auch auf abzählbar unendliche Zustandsräume erweitern. Dabei arbeiten wir mit dem sogenannten Erneuerungssatz. Obwohl dieser durch die Markov-Ketten motiviert ist, macht er zunächst eine rein analytische Aussage. Wir werden dann seine Interpretation für Markov-Ketten betrachten. Dazu formulieren wir zunächst ein Lemma aus der Analysis.

Lemma 5.18. (i) Seien $(a_k)_{k \in \mathbb{N}_0}$ und $(b_k)_{k \in \mathbb{N}_0}$ reelle Folgen mit $\sum_{k=0}^{\infty} |a_k| < \infty$,

$a := \sum_{k=0}^{\infty} a_k$ und $b := \lim_{n \rightarrow \infty} b_n$. Dann gilt:

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n a_k b_{n-k} = ab.$$

(ii) Seien $(a_k)_{k \in \mathbb{N}_0}$ und $(b_k)_{k \in \mathbb{N}_0}$ reelle Folgen mit $a_k \geq 0 \forall k \in \mathbb{N}_0$, $\sum_{k=0}^{\infty} a_k = \infty$ und $0 \neq b = \lim_{n \rightarrow \infty} b_n$. Dann gilt:

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n a_k b_{n-k} = \infty.$$

Beweis. Der Beweis verwendet ausschließlich analytische Hilfsmittel und ist in diversen Analysis-Büchern zu finden. □

Setzen wir in Satz 5.14 $i = j$, $u_n := p_{jj}^{(n)}$ und $f_m := f_{jj}^{(m)}$ so erhalten wir die sogenannte **Erneuerungsgleichung**:

$$u_n = \sum_{m=1}^n f_m u_{n-m}$$

mit $u_0 = 1$. Da j rekurrent ist, ist $(f_m)_{m \in \mathbb{N}}$ ein Wahrscheinlichkeitsmaß auf $\mathbb{N}$. Wir formulieren nun den Erneuerungssatz ganz allgemein.

Satz 5.19. Erneuerungssatz

Seien $(u_n)_{n \in \mathbb{N}_0}$ und $(f_n)_{n \in \mathbb{N}}$ zwei Folgen nichtnegativer Zahlen für die die Erneuerungsgleichung gelte. Sei $0 \leq u_n \leq 1$, $\sum_{n=1}^{\infty} f_n = 1$, $d := \text{ggT}\{n \in \mathbb{N} : u_n > 0\} = 1$ und $m := \sum_{n=1}^{\infty} n f_n$. Dann gilt:

$$\lim_{n \rightarrow \infty} u_n = \frac{1}{m}.$$

Beweis. Wir führen den Beweis in mehreren Schritten:

1. *Schritt:* Für $n \in \mathbb{N}_0$ sei

$$r_n := \sum_{k=n+1}^{\infty} f_k.$$

Dann ist $r_0 = \sum_{k=1}^{\infty} f_k = 1$ und $\sum_{k=1}^{\infty} r_n = \sum_{k=1}^{\infty} k f_k = m$. Aus $r_k - r_{k-1} = -f_k$ folgt mit der Erneuerungsgleichung

$$u_n = - \sum_{k=1}^n (r_k - r_{k-1}) u_{n-k}.$$

Bringt man die negativen Terme alle auf die linke Seite, so erhält man

$$\sum_{k=0}^n r_k u_{n-k} = \sum_{k=0}^{n-1} r_k u_{n-1-k}.$$

Wenden wir diese Identität sukzessive auf kleiner werdendes n an, so ergibt sich $\forall n \in \mathbb{N}_0$:

$$\sum_{k=0}^n r_k u_{n-k} = 1. \quad (5.3)$$

2. *Schritt:* Sei $\lambda := \limsup_{n \rightarrow \infty} u_n$. Dann existiert eine Teilfolge $(n_k)_{k \in \mathbb{N}}$, für die $\lambda = \lim_{k \rightarrow \infty} u_{n_k}$ gilt. Mit der Erneuerungsgleichung gilt dann $\forall s \in \mathbb{N}$:

$$\begin{aligned} \lambda &= \lim_{k \rightarrow \infty} u_{n_k} = \lim_{k \rightarrow \infty} \left(f_s u_{n_k-s} + \sum_{m=1, m \neq s}^{n_k} f_m u_{n_k-m} \right) \\ &\leq \liminf_{k \rightarrow \infty} f_s u_{n_k-s} + \limsup_{k \rightarrow \infty} \sum_{m=1, m \neq s}^{n_k} f_m u_{n_k-m} \end{aligned}$$

$$\begin{aligned}
&\leq f_s \liminf_{k \rightarrow \infty} u_{n_k-s} + \sum_{m=1, m \neq s}^{n_k} f_m \limsup_{k \rightarrow \infty} u_{n_k-m} \\
&\leq f_s \liminf_{k \rightarrow \infty} u_{n_k-s} + (1 - f_s)\lambda
\end{aligned}$$

und daher folgt

$$\liminf_{k \rightarrow \infty} u_{n_k-s} \leq \lambda \leq \limsup_{k \rightarrow \infty} u_{n_k-s}.$$

Somit gilt $\lambda = \lim_{k \rightarrow \infty} u_{n_k-s}$. Wendet man dies für s_j mit f_{s_j} öfter an, so folgt

$$\lim_{k \rightarrow \infty} u_{n_k-m} = \lambda \text{ für } m = \sum_{j=1}^l c_j s_j \text{ mit } c_1, \dots, c_s \in \mathbb{N}.$$

Weiters gilt folgendes Resultat aus der Zahlentheorie:

Ist $ggT\{s_1, \dots, s_l\} = 1$, so $\exists N \in \mathbb{N}$, sodass $\forall n \geq N$ $c_1, \dots, c_l \in \mathbb{N}_0$ existieren mit $n = \sum_{j=1}^l c_j s_j$.

Also gilt $\forall n \geq N$:

$$\lim_{k \rightarrow \infty} u_{n_k-n} = \lambda. \quad (5.4)$$

3. *Schritt*: Mit der Gleichung (5.3) erhalten wir

$$\sum_{j=0}^{n_k-N} r_j u_{n_k-N-j} = 1.$$

Nun lassen wir k gegen unendlich gehen und benutzen Gleichung (5.4) und das vorherige Lemma:

1. Fall: $\sum_{j=0}^{\infty} r_j = \infty$. Wäre $\lambda \neq 0$, so würde man mit Gleichung (5.4) und dem 2. Teil des Lemmas "1 = ∞ " erhalten. Also ist $\lambda = 0$.
2. Fall: $\sum_{j=0}^{\infty} r_j < \infty$. Dann gilt $1 = \lambda \sum_{j=0}^{\infty} r_j$.

Insgesamt gilt also

$$\lambda = \frac{1}{\sum_{j=0}^{\infty} r_j}.$$

Mit der Definition von r_j folgt nun

$$\sum_{j=0}^{\infty} r_j = \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} f_k = \sum_{j=0}^{\infty} j f_j = m.$$

Somit ist $\lambda = \frac{1}{m}$.

4. *Schritt*: Führt man den vorangegangenen Beweis für $\limsup_{n \rightarrow \infty} u_n$ nun analog für $\liminf_{n \rightarrow \infty} u_n$ durch, so erhält man insgesamt

$$\liminf_{n \rightarrow \infty} u_n = \lambda = \frac{1}{m} \limsup_{n \rightarrow \infty} u_n. \quad \square$$

Diese allgemeine Aussage des Satzes, wenden wir nun für den Konvergenzsatz der Markov-Ketten an.

Satz 5.20. *Sei $(X_n)_{n \in \mathbb{N}_0}$ eine irreduzible und aperiodische Markov-Kette mit Zustandsraum Λ . Dann gilt:*

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \begin{cases} \frac{1}{\mu_{ii}} = \pi_j, & \text{falls } (X_n)_{n \in \mathbb{N}_0} \text{ positiv rekurrent} \\ 0, & \text{falls } (X_n)_{n \in \mathbb{N}_0} \text{ nullrekurrent.} \end{cases}$$

Beweis. Wir setzen im Erneuerungssatz $u_n = p_{ii}^{(n)}$ und $f_n = f_{ii}^{(n)}$. Da die Markov-Kette rekurrent ist, gilt $0 \leq u_n \leq 1$ und $\sum_{n=1}^{\infty} f_n = 1$. Offensichtlich ist $m = \mu_{ii}$. Aus der Aperiodizität folgt $d = 1$. Außerdem gilt mit Satz 5.14:

$$p_{ii}^{(n)} = \sum_{k=0}^n f_{ii}^{(k)} p_{ii}^{(n-k)}$$

und somit ist die Erneuerungsgleichung erfüllt. Es gelten somit alle Voraussetzungen des Erneuerungssatzes. Daher gilt:

$$\lim_{n \rightarrow \infty} p_{ii}^{(n)} = \frac{1}{\mu_{ii}}.$$

Außerdem gilt nach Satz 5.14 auch

$$p_{ij}^{(n)} = \sum_{k=0}^n f_{ij}^{(k)} p_{ii}^{(n-k)}.$$

Da $\sum_{k=0}^n f_{ij}^{(k)} = 1$ folgt mit Lemma 5.18:

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \frac{1}{\mu_{ii}}.$$

$\frac{1}{\mu_{ii}} = \pi_j$ folgt aus Satz 5.12 und im nullrekurrenten Fall aus $\mu_{ii} = \infty$. □

Bisher behandelten wir im rekurrenten Fall nur aperiodische Markov-Ketten. Wir wollen nun auch noch kurz den periodischen Fall charakterisieren. Um darüber Aussagen machen zu können benötigen wir folgendes technisches Lemma.

Lemma 5.21. *Sei $(X_n)_{n \in \mathbb{N}_0}$ eine irreduzible Markov-Kette mit Zustandsraum Λ . Wenn alle Zustände die Periode d besitzen, dann existiert eine Zerlegung von Λ in Mengen $C_0, C_1, \dots, C_{d-1}$, die periodisch durchlaufen werden, das heißt ist $i \in C_r$ und $p_{ij} > 0$, so folgt $j \in C_{r+1}$, wobei $C_d = C_0$ gilt.*

Beweis. Einen Beweis dazu findet man in [10]. □

Satz 5.22. *Sei $(X_n)_{n \in \mathbb{N}_0}$ eine irreduzible und rekurrente Markov-Kette mit Zustands-*

raum Λ und Periode $d \geq 2$. Dann gilt $\forall i \in \Lambda, j \in C_r(i)$ und $0 \leq r < d$:

$$\lim_{n \rightarrow \infty} p_{ij}^{(nd+r)} = \begin{cases} \frac{d}{\mu_{ii}} = \pi_j, & \text{falls } (X_n)_{n \in \mathbb{N}_0} \text{ positiv rekurrent} \\ 0, & \text{falls } (X_n)_{n \in \mathbb{N}_0} \text{ nullrekurrent.} \end{cases}$$

r drückt somit den Periodenabstand von i nach j aus. Ist $m \notin \{nd + r : n \in \mathbb{N}_0\}$, so ist $p_{ij}^{(m)} = 0$.

Beweis. Aus der Definition der $C_r(i)$ folgt sofort $p_{ij}^{(m)} = 0$ für $m \notin \{nd + r : n \in \mathbb{N}_0\}$. Anstelle von $(X_n)_{n \in \mathbb{N}_0}$ betrachten wir nun $(X_{dn})_{n \in \mathbb{N}_0}$ mit Übergangsmatrix $\tilde{P} = P^d$. Wir setzen im Erneuerungssatz $u_n = p_{ii}^{(nd)}$ und $f_n = f_{ii}^{(nd)}$. Dann gilt nach Satz 5.14 wieder

$$p_{ii}^{(nd)} = \sum_{k=1}^n f_{(jj)}^{(nd)} p_{jj}^{((n-k)d)}.$$

Sei m die mittlere Rückkehrzeit von $(X_{dn})_{n \in \mathbb{N}_0}$. Dann gilt:

$$m = \sum_{n=1}^{\infty} n f_{ii}^{(nd)} = \frac{1}{d} \sum_{n=1}^{\infty} n d f_{ii}^{(nd)} = \frac{1}{d} \sum_{l=1}^{\infty} l f_{ii}^{(l)} = \frac{\mu_{ii}}{d}.$$

Nach dem Erneuerungssatz gilt dann

$$\lim_{n \rightarrow \infty} p_{ii}^{(n)} = \frac{d}{\mu_{ii}}.$$

Analog wie im vorigen Satz folgt dann $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \frac{d}{\mu_{ii}}$. □

6 Einfache Irrfahrten auf $\mathbb{Z}^d$

6.1 Grundlagen

In diesem Kapitel betrachten wir ein grundlegendes stochastisches Modell, das eine Anwendung der Markov-Ketten ist. Man kann sich dabei die Bewegung eines Objekts vorstellen, bei dem jeder Schritt zufällig erfolgt. Wir werden uns hier auf sogenannte einfache Irrfahrten auf $\mathbb{Z}^d$ beschränken, die nur die Möglichkeit eines Schrittes der Länge 1 in eine bestimmte Richtung haben. Man spricht dabei auch von Zufallsbewegungen oder Random Walks. Als Quellen dienen vor allem [7], [9], [11] und [12].

Definition 6.1. Sei $Y_1, Y_2, \dots$ eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit Werten in $e \in \mathbb{Z}^d$, wobei $\|e\| = 1$. Dann heißt der durch

$$X_n = X_0 + \sum_{k=1}^n Y_k$$

definierte stochastische Prozess $(X_n)_{n \in \mathbb{N}_0}$ eine *einfache d -dimensionale Irrfahrt*.

Bemerkung: 1. Häufig wählt man $X_0 = \mathbf{0} \in \mathbb{Z}^d$.

2. In der Definition sind die Y_i und X_i d -komponentige Zufallsvektoren. Bei den Zufallsvariablen Y_i sind alle Koordinaten bis auf eine zufällig gewählte null, da wir pro Zeitschritt nur in eine Richtung gehen. Aufgrund der Annahme von einfachen Irrfahrten steht in dieser einen Koordinate ein Wert aus $\{-1, 1\}$.
3. Wegen $X_{n+1} = X_n + Y_{n+1}$ und wegen der Unabhängigkeit von Y_{n+1} von $Y_1, \dots, Y_n$ und somit auch von $X_0, \dots, X_n$ ist es klar, dass X_{n+1} nur von X_n abhängt, nicht jedoch von den Zuständen davor. Damit ist $(X_n)_{n \in \mathbb{N}_0}$ eine Markov-Kette.
4. Jeder Zustand in $\mathbb{Z}^d$ hat genau $2d$ "Nachbarn".

Definition 6.2. Eine Irrfahrt auf $\mathbb{Z}^d$ heißt *symmetrisch*, wenn $\forall i, j \in \mathbb{Z}^d$ gilt:

$$p_{ij} = \begin{cases} \frac{1}{2d} & \text{falls } \|i - j\| = 1 \\ 0, & \text{sonst.} \end{cases}$$

Man unterscheidet dabei noch zwischen beschränkten und unbeschränkten Irrfahrten. Bei der unbeschränkten Irrfahrt kann sich das Objekt im Gegensatz zur beschränkten Irrfahrt unendlich lange in alle Richtungen bewegen ohne an eine bestimmte Grenze zu stoßen. Festzulegen bleibt im beschränkten Fall noch das Verhalten an den Rändern. Dabei gibt es zwei wichtige Fälle:

- (i) Absorbierende Ränder
- (ii) Reflektierende Ränder

Während die beschränkte Irrfahrt mit absorbierenden Rändern gestoppt wird, sobald sie eine der Grenzen erreicht hat, läuft die Irrfahrt mit reflektierenden Rändern immer weiter, da sie die Grenze mit Wahrscheinlichkeit 1 wieder verlässt.

Wir interessieren uns vor allem wieder für das Langzeitverhalten von Random Walks und stellen uns die Frage, ob die Irrfahrt zu ihrem Ausgangspunkt mit Sicherheit zurückkehrt oder nicht, das heißt wir untersuchen die Rekurrenz bzw. Transienz von Irrfahrten. Die Frage nach der Rückkehr ist für beschränkte Irrfahrten schnell zu beantworten: Im Fall von absorbierenden Rändern kehren wir mit Wahrscheinlichkeit 1 nicht unendlich oft zum Ausgangspunkt zurück. Also ist bis auf die Grenzen jeder Zustand der beschränkten Irrfahrt mit absorbierenden Rändern transient. Hat die Irrfahrt jedoch reflektierende Ränder, so kehren wir sicher unendlich oft zu X_0 zurück, da die Irrfahrt ewig weiter läuft. Die beschränkte Irrfahrt mit reflektierenden Rändern ist also rekurrent (bzw. positiv rekurrent, da wir einen endlichen Zustandsraum vorliegen haben). Interessanter ist der Fall der unbeschränkten Irrfahrt, den wir jetzt in den verschiedenen Dimensionen untersuchen werden.

6.2 Rekurrenz und Transienz

Bei den einfachen unbeschränkten Irrfahrten auf $\mathbb{Z}^d$ ist es offensichtlich klar, dass alle Zustände miteinander kommunizieren. Wir haben es also mit irreduziblen Prozessen zu tun. Es reicht daher immer aus, einen speziellen Zustand der Irrfahrt auf Rekurrenz bzw. Transienz zu prüfen, um eine Aussage für alle Zustände zu erhalten.

6.2.1 Die 1-dimensionale Irrfahrt

In diesem Abschnitt kann man sich eine Gerade mit Werten aus $\mathbb{Z}$ vorstellen, auf der sich ein Objekt mit Schrittlänge 1 bewegt. Damit macht das Objekt in jedem Zeitschritt mit Wahrscheinlichkeit p einen Schritt vorwärts und mit Wahrscheinlichkeit $q = 1 - p$ einen Schritt rückwärts. Wir erhalten also folgende Übergangswahrscheinlichkeiten:

$$\begin{aligned} p_{i,i+1} &= p \\ p_{i,i-1} &= q \\ p_{i,j} &= 0 \text{ für } j \notin \{i-1, i+1\} \end{aligned}$$

Wir widmen uns nun der Frage, ob die eindimensionale Irrfahrt zu ihrem Ausgangspunkt mit Sicherheit zurückkehrt oder nicht.

Satz 6.1. *Die eindimensionale Irrfahrt auf $\mathbb{Z}$ ist für $p = q = \frac{1}{2}$ rekurrent und für $p \neq q$ transient.*

Beweis. Wir wählen oBdA $X_0 = 0$. Zunächst ist klar, dass die Kette nicht in einer ungeraden Anzahl von Zeitschritten zu 0 zurückkehren kann, das heißt $\forall n \in \mathbb{N}_0$ gilt:

$$p_{00}^{(2n+1)} = 0.$$

Für eine gerade Anzahl an Schritten erhalten wir die Wahrscheinlichkeit

$$p_{00}^{(2n)} = \binom{2n}{n} p^n q^n = \frac{(2n)!}{n!n!} p^n q^n.$$

Nach Satz 5.5 reicht es zu überprüfen, ob $\sum_{n=1}^{\infty} p_{00}^{(2n)}$ konvergiert oder divergiert. Da das Ersetzen von asymptotisch gleichen Ausdrücken nichts am Konvergenzverhalten der Reihe ändert, verwenden wir die Stirling'sche Formel und erhalten:

$$\begin{aligned} \sum_{n=1}^{\infty} p_{00}^{(2n)} &= \sum_{n=1}^{\infty} \frac{(2n)!}{n!n!} p^n q^n \\ &\sim \sum_{n=1}^{\infty} \frac{\sqrt{2\pi 2n} (2n)^{2n} e^{-2n} p^n q^n}{2\pi n^{2n+1} e^{-2n}} \\ &= \sum_{n=1}^{\infty} \frac{2^{2n} p^n q^n}{\sqrt{\pi n}} \\ &= \sum_{n=1}^{\infty} \frac{(4p(1-p))^n}{\sqrt{\pi n}}. \end{aligned}$$

Nun gilt

$$4p(1-p) = \begin{cases} 1 & \text{für } p = \frac{1}{2} \\ < 1, & \text{sonst.} \end{cases}$$

Für $p = q = \frac{1}{2}$ erhalten wir daher

$$\sum_{n=1}^{\infty} p_{00}^{(2n)} \sim \sum_{n=1}^{\infty} \frac{(4pq)^n}{\sqrt{\pi n}} = \frac{1}{\sqrt{\pi}} \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} = \infty.$$

Für $p \neq q$ erhalten wir hingegen

$$\sum_{n=1}^{\infty} p_{00}^{(2n)} \sim \sum_{n=1}^{\infty} \frac{(4pq)^n}{\sqrt{\pi n}} < \frac{1}{\sqrt{\pi}} \sum_{n=1}^{\infty} (4p(1-p))^n < \infty. \quad \square$$

Die eindimensionale Irrfahrt ist also nur im symmetrischen Fall rekurrent. Man kann sich nun noch die Frage stellen, ob die symmetrische Irrfahrt positiv rekurrent oder nullrekurrent ist. Diese Frage lässt sich schnell beantworten:

Satz 6.2. *Die unbeschränkte, symmetrische Irrfahrt auf $\mathbb{Z}$ ist nullrekurrent.*

Beweis. Aus $p_{00}^{(2n-1)} = 0$ und $p_{00}^{(2n)} \sim \frac{1}{\sqrt{\pi n}}$ folgt $\lim_{n \rightarrow \infty} p_{00}^{(n)} = 0$. Daraus folgt, dass $\mu_{00} = \infty$. $\square$

6.2.2 Die 2-dimensionale Irrfahrt

Im 2-dimensionalen Fall kann man sich einen Zufallswanderer in einem Gitter vorstellen, der bei jedem Gitterpunkt zufällig nach oben, unten, rechts oder links geht. Wie im vorigen Abschnitt fragen wir uns, ob dieser zu seinem Ausgangspunkt zurückkehrt. Ab dieser Dimension reicht es, für diese Fragestellung lediglich symmetrische Irrfahrten zu betrachten, denn aufgrund von Satz 6.1 ist es ohnehin klar, dass in höheren Dimensionen bei nicht vorhandener Symmetrie erst Recht Transienz vorliegen wird.

Satz 6.3. *Die symmetrische Irrfahrt in $\mathbb{Z}^2$ ist rekurrent.*

Beweis. Wieder wählen wir oBdA als Anfangspunkt $\mathbf{0} = (0, 0)$. Aufgrund der Symmetrie gilt

$$p_{ij} = \begin{cases} \frac{1}{4} & \text{falls } \|i - j\| = 1 \\ 0, & \text{sonst.} \end{cases}$$

Zunächst ist klar, dass die Schrittzahl wieder gerade sein muss. Um zu $(0, 0)$ zurückzukehren, müssen jeweils die gleiche Anzahl i von Links- und Rechtsschritten sowie j Bewegungen nach oben und unten stattfinden. Man erhält somit $2n = 2i + 2j$ Schritte. Somit kann man sich i Schritte in die eine Dimensionsrichtung und $(n - i)$ in die zweite bewegen. Dafür gibt es $\sum_{i=0}^n \frac{(2n)!}{i!(n-i)!(n-i)!}$ Möglichkeiten. Wir erhalten somit:

$$\begin{aligned} p_{00}^{(2n)} &= \left(\frac{1}{4}\right)^{2n} \sum_{i=0}^n \frac{(2n)!}{i!^2(n-i)!^2} \\ &= \left(\frac{1}{4}\right)^{2n} \sum_{i=0}^n \frac{(2n)!n!n!}{n!n!i!^2(n-i)!^2} \\ &= \left(\frac{1}{4}\right)^{2n} \binom{2n}{n} \underbrace{\sum_{i=0}^n \binom{n}{i} \binom{n}{n-i}}_{=\binom{2n}{n}} \\ &= \left(\frac{1}{4}\right)^{2n} \binom{2n}{n}^2. \end{aligned}$$

Somit erhalten wir mit der Stirling'schen Formel:

$$\begin{aligned} \sum_{n=1}^{\infty} p_{\mathbf{00}}^{(2n)} &= \sum_{n=1}^{\infty} \left(\frac{1}{4}\right)^{2n} \binom{2n}{n}^2 \\ &\sim \sum_{n=1}^{\infty} \left(\frac{1}{4}\right)^{2n} \left(\frac{4^n}{\sqrt{\pi n}}\right)^2 \\ &= \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} = \infty. \end{aligned}$$

Nach Satz 5.5 folgt die Rekurrenz. □

Analog wie im 1-dimensionalen Fall erhält man die Nullrekurrenz.

6.2.3 Die 3-dimensionale Irrfahrt

Hier hat jeder Kreuzungspunkt 6 Nachbarpunkte (Nord, Süd, Ost, West, Oben, Unten). Wir betrachten wieder das Rekurrenzverhalten von der symmetrischen Irrfahrt auf $\mathbb{Z}^3$.

Satz 6.4. *Die symmetrische und unbeschränkte Irrfahrt auf $\mathbb{Z}^3$ ist transient.*

Beweis. Sei oBdA $\mathbf{0} = (0, 0, 0)$. Wieder können wir nur in einer geraden Anzahl von Schritten zurückkehren und es gilt mit analogen Überlegungen wie im 2-dimensionalen:

$$\begin{aligned} p_{\mathbf{00}}^{(2n)} &= \sum_{\substack{i,j \\ 0 \leq i+j \leq n}} \frac{(2n)!}{i!j!(n-i-j)!(n-i-j)!} \left(\frac{1}{6}\right)^{2n} \frac{n!^2}{n!^2} \\ &= \left(\frac{1}{2}\right)^{2n} \binom{2n}{n} \sum_{\substack{i,j \\ 0 \leq i+j \leq n}} \left(\frac{n!}{i!j!(n-i-j)!}\right)^2 \left(\frac{1}{3}\right)^{2n}. \end{aligned}$$

Wir wollen nun eine Abschätzung vornehmen. Es gilt:

$$\sum_{\substack{i,j \\ 0 \leq i+j \leq n}} \frac{n!}{i!j!(n-i-j)!} \left(\frac{1}{3}\right)^n = 1$$

und daher

$$p_{\mathbf{00}}^{(2n)} \leq \left(\frac{1}{2}\right)^{2n} \binom{2n}{n} \left(\frac{1}{3}\right)^n \max_{\substack{i,j \\ 0 \leq i+j \leq n}} \left(\frac{n!}{i!j!(n-i-j)!}\right).$$

Nun betrachten wir das Maximum für große n . Dazu seien i_0 und j_0 Werte von i und j , die den Term $\frac{n!}{i!j!(n-i-j)!}$ maximieren, wobei $0 \leq i+j \leq n$ gelten muss. Wir erhalten somit folgende vier Ungleichungen:

$$\frac{n!}{(i_0-1)!j_0!(n-i_0+1-j_0)!} \leq \frac{n!}{i_0!j_0!(n-j_0-j_0)!},$$

$$\begin{aligned}
\frac{n!}{(i_0+1)!j_0!(n-i_0-1-j_0)!} &\leq \frac{n!}{i_0!j_0!(n-j_0-j_0)!}, \\
\frac{n!}{i_0!(j_0-1)!(n-i_0-j_0+1)!} &\leq \frac{n!}{i_0!j_0!(n-j_0-j_0)!}, \\
\frac{n!}{i_0!(j_0+1)!(n-i_0-j_0+1)!} &\leq \frac{n!}{i_0!j_0!(n-j_0-j_0)!}.
\end{aligned}$$

Diese vier Ungleichungen können wir auf zwei reduzieren:

$$\begin{aligned}
n - i_0 - 1 &\leq 2i_0 \leq n - i_0 + 1, \\
n - j_0 - 1 &\leq 2j_0 \leq n - j_0 + 1.
\end{aligned}$$

Betrachtet man diese Ungleichungen für große n , so ergibt sich $i_0 \approx \frac{n}{3}$ und $j_0 \approx \frac{n}{3}$. Damit gilt für alle n , die durch 3 teilbar sind:

$$\max_{\substack{i,j \\ 0 \leq i+j \leq n}} \left(\frac{n!}{i!j!(n-i-j)!} \right) \approx \frac{n!}{(\frac{n}{3})!(\frac{n}{3})!(\frac{n}{3})!}.$$

Ist n nicht durch 3 teilbar, so wird das Maximum bei $\frac{n!}{((n-1)/3)!((n-1)/3)!((n+2)/3)!}$ oder $\frac{n!}{((n-2)/3)!((n-2)/3)!((n+1)/3)!}$ angenommen. Wenden wir nun die Stirling'sche Formel oBdA für durch 3 teilbare n an, so erhalten wir:

$$\begin{aligned}
\sum_{n=1}^{\infty} p_{00}^{(2n)} &\leq \left(\frac{1}{2}\right)^{2n} \binom{2n}{n} \left(\frac{1}{3}\right)^n \frac{n!}{(\frac{n}{3})!(\frac{n}{3})!(\frac{n}{3})!} \\
&\approx \sum_{n=1}^{\infty} \frac{3\sqrt{3}}{2\pi^{3/2}n^{3/2}} < \infty.
\end{aligned}$$

Nach Satz 5.5 folgt die Transienz. □

Da die 3-dimensionale Irrfahrt nicht mehr rekurrent ist, gilt das natürlich auch für alle höheren Dimensionen. Im antisymmetrischen Fall haben wir ohnehin immer Transienz vorliegen. Die Resultate stammen übrigens von George Pólya (1887-1985). Der Legende nach soll dieser beim Spazieren auf dem Zürichberg an gleichen Nachmittagen mehrmals auf dieselben Personen getroffen sein, was ihm sehr verwunderlich vorkam. Daraufhin untersuchte er dieses Phänomen mit Hilfe von 2-dimensionalen Irrfahrten und entdeckte die Rekurrenz dieser. Dieses Resultat untersuchte er dann in den verschiedenen Dimensionen.

6.3 Das Arkus-Sinus Gesetz

Ein sehr interessantes und der Intuition widersprechendes Resultat ist das sogenannte Arkus-Sinus Gesetz. Die Idee dazu stammt von [6], die genaueren Ausführungen wurden zu einem großen Teil in [12] gefunden. Wir betrachten dabei die eindimensionale symmetrische Irrfahrt, die wir bei $X_0 = 0$ starten. Wir fragen uns nun nach dem Verhältnis der Zeit, die eine solche Irrfahrt im positiven bzw. negativen Bereich

verbringt. Zunächst illustrieren wir die Idee an einem Beispiel, das in [6] zu finden ist:

Beispiel 6.1: Angenommen, Spieler A wettet einen Euro, dass bei einem fairen Münzwurf Zahl erscheint. Spieler B wettet dagegen. Dieses Spiel wird nun 100 Mal wiederholt. Die Frage ist nun, wie lange Spieler A bei diesem Spiel im Gesamtgewinn geführt hat.

Die häufigste Antwort auf diese Frage ist, dass beide Spieler über die Hälfte aller Runden in Führung liegen. Das ist aber falsch. Dass bei diesem Glücksspiel einer der Spieler genau 50 Runden führt ist nicht der wahrscheinlichste, sondern sogar der unwahrscheinlichste Fall. Am Wahrscheinlichsten ist es nämlich, dass der letztendliche Sieger immer in Führung liegt und der Verlierer immer zurück liegt. Betrachten wir das Spiel für 4 Runden. Dann gibt es folgende 16 gleichwahrscheinliche Möglichkeiten. Bei Gleichstand geht die Runde an den, der vorher vorne lag. In der rechten Spalte steht immer die Zahl, die Spieler A in Führung lag:

1.	K K K K	0
2.	K K K Z	0
3.	K K Z K	0
4.	K K Z Z	0
5.	K Z K K	0
6.	K Z K Z	0
7.	K Z Z K	2
8.	K Z Z Z	2
9.	Z K K K	2
10.	Z K K Z	2
11.	Z K Z K	4
12.	Z K Z Z	4
13.	Z Z K K	4
14.	Z Z K Z	4
15.	Z Z Z K	4
16.	Z Z Z Z	4

Wie man sieht, liegt Spieler A bei 12 der möglichen Spielverläufe immer vorne oder hinten und nur bei 4 führt er genau die Hälfte der Runden. Man könnte also sagen:

„Sowohl Pech- als auch Glückssträhnen sind beharrliche Zeitgenossen.“

Wir wollen dieses Phänomen nun allgemein betrachten, wenn die Zahl der Runden immer größer wird: Seien dazu $Y_1, Y_2, \dots$ wieder unabhängige Zufallsvariablen mit $\mathbb{P}[Y_i = -1] = \mathbb{P}[Y_i = 1] = \frac{1}{2}$ und $X_n := \sum_{i=1}^n Y_i$. Für die symmetrische Irrfahrt $(X_n)_{n \in \mathbb{N}_0}$ führen wir zunächst noch eine weitere Variable ein:

$$Z_i = \begin{cases} 1, & \text{falls } X_i > 0 \text{ oder } X_i = 0 \text{ und } X_{i-1} > 0 \\ 0, & \text{sonst.} \end{cases}$$

Z_i ist dann ein Indikator dafür, im i -ten Spiel in der Gewinnzone zu liegen. Ist $X_i = 0$, so befindet man sich laut Definition von Z_i genau dann in der Gewinnzone, wenn man in der vorigen Runde in der Gewinnzone war. $\sum_{i=1}^{2n} Z_i$ gibt dann die Zeit an, die bis zur Zeit $2n$ in der Gewinnzone verbracht wurde. Sei weiters $p_{2k,2n}$ die Wahrscheinlichkeit bis zur Zeit $2n$ genau $2k$ -mal in der Gewinnzone zu sein, das

heißt:

$$p_{2k,2n} = \mathbb{P} \left[\sum_{i=1}^{2n} Z_i = 2k \right].$$

Des Weiteren definieren wir die Wahrscheinlichkeit nach $2n$ Schritten zu 0 zurückgekehrt zu sein durch

$$u_{2n} := \mathbb{P}[X_{2n} = 0] = \binom{2n}{n} \left(\frac{1}{2}\right)^{2n}$$

und die Wahrscheinlichkeit nach $2n$ Schritten das erste Mal zu 0 zurückgekehrt zu sein durch

$$f_{2n} := \mathbb{P}[X_1 \neq 0, X_2 \neq 0, \dots, X_{2n-1} \neq 0, X_{2n} = 0].$$

Mit diesen Definitionen können wir nun ein wichtiges Lemma formulieren:

Lemma 6.5. *Sei $(X_n)_{n \in \mathbb{N}_0}$ eine symmetrische Irrfahrt auf $\mathbb{Z}$. Dann gilt:*

$$f_{2n} = u_{2n-2} - u_{2n}$$

Beweis. Wir betrachten zunächst die Wahrscheinlichkeit

$$g_{2n} := \mathbb{P}[X_1 > 0, X_2 > 0, \dots, X_{2n-1} > 0, X_{2n} = 0].$$

Das heißt g_{2n} ist die Wahrscheinlichkeit, dass man sich immer im positiven Bereich aufhält und zum Zeitpunkt $2n$ zu 0 zurückkehrt. Daher gilt:

$$g_{2n} = \frac{1}{2} f_{2n}.$$

Außerdem gilt:

$$\begin{aligned} g_{2n} &= \frac{1}{2} \mathbb{P}[X_1 > 0, X_2 > 0, \dots, X_{2n-2} > 0, X_{2n-1} = 1] \\ &= \frac{1}{2} (\mathbb{P}[X_1 = 1, X_{2n-1} = 1] - \mathbb{P}[X_1 = 1, X_{2n-1} = 1, X_k \leq 0 \text{ für ein } k \in \{2, \dots, 2n-2\}]). \end{aligned}$$

Nun gilt für den ersten Term

$$\mathbb{P}[X_1 = 1, X_{2n-1} = 1] = \binom{2n-2}{n-1} \left(\frac{1}{2}\right)^{2n-2}.$$

Für den zweiten Term benutzen wir das sogenannte Reflexionsprinzip:

$$\begin{aligned} &\mathbb{P}[X_1 = 1, X_{2n-1} = 1, X_k \leq 0 \text{ für ein } k \in \{2, \dots, 2n-2\}] \\ &= \mathbb{P}[X_1 = 1, X_{2n-1} = -1, X_k \leq 0 \text{ für ein } k \in \{2, \dots, 2n-2\}] \\ &= \mathbb{P}[X_1 = 1, X_{2n-1} = -1] \\ &= \binom{2n-2}{n} \left(\frac{1}{2}\right)^{2n-2}. \end{aligned}$$

Hier ist es wichtig, dass der erste Schritt nach oben festgelegt ist, da dann alle Pfade, die zu der Wahrscheinlichkeit $\mathbb{P}[X_1 = 1, X_{2n-1} = -1]$ beitragen, durch 0

gehen müssen. Wir erhalten somit:

$$\begin{aligned}
f_{2n} &= 2g_{2n} = \binom{2n-2}{n-1} \left(\frac{1}{2}\right)^{2n-2} - \binom{2n-2}{n} \left(\frac{1}{2}\right)^{2n-2} \\
&= \left(\frac{1}{2}\right)^{2n-2} \left[\binom{2n-2}{n-1} - \binom{2n-2}{n} \right] \\
&= \left(\frac{1}{2}\right)^{2n-2} \frac{1}{2n} \binom{2n-2}{n-1} \\
&= \frac{1}{2n} u_{2n-2}.
\end{aligned}$$

Aus

$$\begin{aligned}
u_{2n} &= \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} = \frac{2n(2n-1)}{nn} \binom{2n-2}{n-1} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{2n-2} \\
&= \left(1 - \frac{1}{2n}\right) u_{2n-2} \\
&= u_{2n-2} - f_{2n}
\end{aligned}$$

folgt schließlich die Behauptung. $\square$

Damit können wir nun folgenden wesentlichen Satz formulieren:

Satz 6.6. *Sei $(X_n)_{n \in \mathbb{N}_0}$ eine symmetrische Irrfahrt auf $\mathbb{Z}$. Dann gilt:*

$$p_{2k,2n} = u_{2k} u_{2n-2k}.$$

Beweis. Sei $0 < k < n$. Da die eindimensionale Irrfahrt im symmetrischen Fall rekurrent ist, muss die Irrfahrt irgendwann die Nulllinie kreuzen, insbesondere ein erstes Mal. Die Wahrscheinlichkeit dafür sei f_{2r} , das heißt die erste Rückkehr findet zur Zeit $2r$ statt. Bis dahin muss die Irrfahrt stets positiv oder negativ sein. Beides tritt mit Wahrscheinlichkeit $\frac{1}{2}$ ein. Im Fall, dass sie bis zur Zeit $2r$ positiv ist, kann r nicht größer sein als k , denn sonst wäre man schon öfter als $2k$ -mal in der Gewinnzone gewesen. Im negativen Fall kann r nicht größer sein als $n-k$, da man sonst schon zu viel Zeit in der negativen Zone verbracht hätte, um genau $2k$ -mal in der Gewinnzone zu sein. Nach der ersten Rückkehr hat man dann noch $2n-2r$ Schritte zu tun. Im positiven Fall muss man dann noch $2k-2r$ Schritte in der Gewinnzone sein, im negativen Fall $2k$ Schritte. Es gilt daher:

$$p_{2k,2n} = \frac{1}{2} \sum_{r=1}^k f_{2r} p_{2k-2r,2n-2r} + \frac{1}{2} \sum_{r=1}^{n-k} f_{2r} p_{2k,2n-2r}. \quad (6.5)$$

Außerdem gilt:

$$u_{2n} = \mathbb{P}[X_{2n} = 0] = \sum_{r=1}^n f_{2r} \mathbb{P}[X_{2n-2r} = 0] = \sum_{r=1}^n f_{2r} u_{2n-2r}. \quad (6.6)$$

Nun können wir den Satz mit Induktion beweisen. Zunächst müssen wir zeigen, dass

$p_{0,2n} = p_{2n,2n} = u_{2n}$ gilt. Dabei folgt $p_{0,2n} = p_{2n,2n}$ sofort aus der symmetrischen Definition der Variablen Z_i . Da die Irrfahrt in einem ungeraden Zeitpunkt nicht in 0 sein kann, gilt

$$\mathbb{P}[X_1 > 0, X_2 > 0, \dots, X_{2n} > 0] = \mathbb{P}[X_1 > 0, \dots, X_{2n} > 0, X_{2n+1} > 0].$$

Andererseits gilt natürlich

$$\mathbb{P}[X_1 > 0, \dots, X_{2n} > 0, X_{2n+1} > 0] = \frac{1}{2} \mathbb{P}[X_1 \geq 0, \dots, X_{2n} \geq 0, X_{2n+1} \geq 0].$$

Daher gilt also:

$$\begin{aligned} p_{2n,2n} &= \mathbb{P}[X_1 \geq 0, \dots, X_{2n} \geq 0, X_{2n} \geq 0] \\ &= 2\mathbb{P}[X_1 > 0, \dots, X_{2n} > 0, X_{2n} > 0] \\ &= \mathbb{P}[\inf\{r \in \mathbb{N} : X_r = 0\} > 2n] \\ &= 1 - \sum_{r=1}^n f_{2r} \\ &= 1 - \sum_{r=1}^n u_{2r-2} - u_{2n}, \end{aligned}$$

wobei in der letzten Gleichung das vorige Lemma verwendet wurde. Da es sich bei dieser Summe um eine Teleskopsumme handelt und $u_0 = 1$ ist, erhält man

$$p_{2n,2n} = 1 - (u_0 - u_{2n}) = u_{2n}.$$

Nun kommen wir zum Induktionsschritt. Dazu nehmen wir an, dass $p_{2k,2m} = u_{2k}u_{2m-2k}$ für $m \leq n-1$ und $0 < k < m$ gelte. Dann folgt mit den Gleichungen (6.5) und (6.6) für $m < n$:

$$\begin{aligned} p_{2k,2n} &= \frac{1}{2} u_{2n-2k} \sum_{r=1}^k f_{2r} u_{2k-2r} + \frac{1}{2} u_{2k} \sum_{r=1}^{n-k} f_{2r} u_{2n-2k-2r} \\ &= \frac{1}{2} u_{2n-2k} u_{2k} + \frac{1}{2} u_{2k} u_{2n-2k} \\ &= u_{2k} u_{2n-2k}. \end{aligned}$$

□

Nun können wir das Arkus-Sinus-Gesetz formulieren, in dem wir das asymptotische Verhalten betrachten.

Satz 6.7. *Arkus-Sinus-Gesetz*

Sei $(X_n)_{n \in \mathbb{N}_0}$ eine symmetrische Irrfahrt auf $\mathbb{Z}$ und $R_{2n} := \sum_{i=1}^{2n} Z_i$ die bis zur Zeit $2n$ im positiven Bereich verbrachte Zeit. Dann gilt für alle $0 < a < b < 1$:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[a \leq \frac{R_{2n}}{2n} \leq b \right] = \frac{1}{\pi} \int_a^b \frac{1}{\sqrt{x(1-x)}} dx = \frac{2}{\pi} \left[\arcsin(\sqrt{b}) - \arcsin(\sqrt{a}) \right].$$

Beweis. Nach dem vorigen Satz gilt $\forall k \in \mathbb{N}_0$:

$$\mathbb{P}[R_{2n} = 2k] = p_{2k,2n} = u_{2k}u_{2n-2k} = \left(\frac{1}{2}\right)^{2n} \binom{2k}{k} \binom{2n-2k}{n-k}.$$

Daher gilt

$$\mathbb{P}\left[a \leq \frac{R_{2n}}{2n} \leq b\right] = \sum_{k \in J_n} \left(\frac{1}{2}\right)^{2n} \binom{2k}{k} \binom{2n-2k}{n-k}$$

mit $J_n := \{k \in \mathbb{N} : \frac{2k}{2n} \in [a, b]\}$. Wir untersuchen das Verhalten dieser Wahrscheinlichkeit nun für große n und wenden zunächst die Stirling'sche Formel an. Wir wissen bereits aus Satz 6.1, dass

$$\binom{2k}{k} \sim \frac{4^k}{\sqrt{\pi k}}.$$

Also erhalten wir insgesamt

$$\begin{aligned} \left(\frac{1}{2}\right)^{2n} \binom{2k}{k} \binom{2n-2k}{n-k} &\sim \frac{1}{4^n} \frac{4^k}{\sqrt{\pi k}} \frac{4^{n-k}}{\sqrt{\pi(n-k)}} \\ &= \frac{1}{\pi \sqrt{k(n-k)}} \end{aligned}$$

und somit gilt

$$\begin{aligned} \mathbb{P}\left[a \leq \frac{R_{2n}}{2n} \leq b\right] &\sim \sum_{k \in J_n} \frac{1}{\pi \sqrt{k(n-k)}} \\ &= \frac{1}{\pi} \frac{1}{n} \sum_{k \in J_n} \frac{1}{\sqrt{\frac{k}{n} \left(1 - \frac{k}{n}\right)}}. \end{aligned}$$

Dieser Ausdruck ist offensichtlich eine Riemannsumme für das Integral $\frac{1}{\pi} \int_a^b \frac{1}{\sqrt{x(1-x)}} dx$ für Unterteilung in Intervalle der Länge $\frac{1}{n}$. Also konvergiert dieser Ausdruck gegen das Integral und die Behauptung ist gezeigt. $\square$

Definition 6.3. Eine kontinuierliche Zufallsvariable X mit Wahrscheinlichkeitsdichte

$$f(x) = \frac{1}{\pi} \frac{1}{\sqrt{x(1-x)}}$$

heißt *Arkussinusverteilt*.

Der Name stammt dabei von der zugehörigen Verteilungsfunktion. Das Arkus-Sinus-Gesetz besagt also, dass $\frac{1}{2n} R_{2n}$ in Verteilung gegen die Arkussinusverteilung konvergiert. Anhand der Wahrscheinlichkeitsdichte kann man schon erkennen, dass die Wahrscheinlichkeit für $k = 0$ bzw. $k = n$ maximiert und für $k = \frac{n}{2}$ minimiert wird, wie wir auch im einleitenden Beispiel beobachten konnten.

6.4 Ausblick: Die Brownsche Bewegung

Im letzten Abschnitt dieser Arbeit soll noch ein kurzer Ausblick gegeben werden. Bisher wurden nur diskrete stochastische Prozesse behandelt. Natürlich gibt es zahlreiche zeitstetige Prozesse. Einer der bedeutendsten davon ist die Brownsche Bewegung. Um diesen Prozess zu konstruieren, beginnen wir mit einer eindimensionalen symmetrischen Irrfahrt. Der Großteil dieses Abschnitts findet sich in ähnlicher Form in [11].

Definition 6.4. Ein Prozess $(X_t)_{t \geq 0}$ hat *unabhängige Inkremente*, wenn die Zufallsvariablen $(X_{k_1} - X_{k_0}), (X_{k_2} - X_{k_1}), \dots, (X_{k_m} - X_{k_{m-1}})$ für alle $0 \leq k_0 < k_1 < \dots < k_m$ unabhängig sind.

Wenn $(X_n)_{n \in \mathbb{N}_0}$ eine symmetrische Irrfahrt ist, so ist $X_{k_{i+1}} - X_{k_i} = \sum_{j=k_i+1}^{k_{i+1}} Y_j$ ein Inkrement und beschreibt die Veränderung der Irrfahrt in der Zeit zwischen k_i und k_{i+1} . Da $\mathbb{E}[Y_j] = 0$ hat jedes Inkrement Erwartungswert 0 und wegen $\text{Var}[Y_j] = \mathbb{E}[Y_j^2] = 1$ erhalten wir $k_{i+1} - k_i$ für die Varianz eines Inkrements. Um eine Brownsche Bewegung zu approximieren, beschleunigen wir die Zeit und reduzieren die Schrittweite der Irrfahrt.

Definition 6.5. Für eine fixe Zahl $n \in \mathbb{N}$ definieren wir die *skalierte Irrfahrt* $W^{(n)}(t) = \frac{1}{\sqrt{n}} X_{nt}$, sofern $nt \in \mathbb{N}$. Ist $nt \notin \mathbb{N}$, so definieren wir $W^{(n)}(t)$ durch die lineare Interpolation zwischen den nächsten Werten s und u links und rechts von t , für die $ns \in \mathbb{N}$ und $nu \in \mathbb{N}$ gilt.

Die skalierte Irrfahrt $W^{(100)}(t)$ macht zum Beispiel zwischen den Zeitpunkten k und $k+1$ genau 100 Sprünge, jeweils mit einer Höhe von $\frac{1}{10}$ auf oder ab. Auch die skalierte Irrfahrt hat unabhängige Inkremente. Da $W^{(n)}(t) - W^{(n)}(s)$ die Summe von $n(t-s)$ unabhängigen Zufallsvariablen ist, jede davon mit Erwartungswert 0 und Varianz $\frac{1}{n}$, gilt $\mathbb{E}[W^{(n)}(t) - W^{(n)}(s)] = 0$ und $\text{Var}[W^{(n)}(t) - W^{(n)}(s)] = t - s$. Wir fixieren nun eine bestimmte Zeit t und fragen uns nach der Verteilung zu diesem Zeitpunkt.

Beispiel 6.2: Betrachten wir die skalierte Irrfahrt $W^{(100)}(0.25) = \frac{1}{10} X_{25}$. Wir können nun nach der Wahrscheinlichkeit des Ereignisses $A = \{\omega : 0 \leq W^{(100)}(0.25) \leq 0.2\}$ fragen. X_{25} kann jede ungerade Zahl zwischen -25 und 25 annehmen. Daraus folgt, dass $\frac{1}{10} X_{25}$ einen der Werte $-2.5, -2.3, -2.1, \dots, -0.1, 0.1, \dots, 2.3, 2.5$ annehmen muss. A tritt also nur dann ein, wenn $W^{(100)}(0.25)$ den Wert 0.1 annimmt. Dafür muss die Irrfahrt 13 Mal nach oben und 12 Mal nach unten gehen. Wir erhalten $\mathbb{P}[A] = \mathbb{P}[W^{(100)}(0.25) = 0.1] = \binom{25}{13} \left(\frac{1}{2}\right)^{25} = 0.1555$.

Diese Berechnungen können für große n und kompliziertere Ereignisse schnell sehr

mühsam werden. Mit Hilfe des Zentralen Grenzwertsatzes lässt sich dieses Problem jedoch leicht lösen.

Satz 6.8. Sei $W^{(n)}(t)$ für ein fixes $t \geq 0$ eine skalierte Irrfahrt. Dann gilt:

$$\mathbb{P}[a \leq W^{(n)}(t) \leq b] \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi t}} \int_a^b e^{-\frac{x^2}{2t}} dx.$$

Beweis. Die Zufallsvariable $X_{nt} = \sum_{j=1}^{nt} Y_j$ ist die Summe von unabhängigen und identisch verteilten Zufallsvariablen mit $\mathbb{E}[X_{nt}] = 0$ und $\text{Var}[X_{nt}] = nt$. Daher gilt $\mathbb{E}[W^{(n)}(t)] = 0$ und $\text{Var}[W^{(n)}(t)] = t$. Nach dem Zentralen Grenzwertsatz konvergiert $W^{(n)}(t)$ in Verteilung gegen eine $N(0, t)$ -Verteilung, deren Dichte $\frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$ ist. $\square$

Wir erhalten die Brownsche Bewegung durch diesen Grenzübergang. Die Eigenschaften der behandelten Irrfahrten werden für die Brownsche Bewegung zur Definition gemacht.

Definition 6.6. Ein stochastischer Prozess $(W_t)_{t \geq 0}$ heißt *Brownsche Bewegung*, falls folgende Bedingungen erfüllt sind:

- (i) $W_0 = 0$ fast sicher.
- (ii) $(W_t)_{t \geq 0}$ hat unabhängige Inkremente.
- (iii) Die Inkremente $W_t - W_s, 0 \leq s < t$ sind $N(0, t - s)$ -verteilt.
- (iv) $(W_t)_{t \geq 0}$ ist ein stetiger Prozess, d.h. fast alle Pfade von $(W_t)_{t \geq 0}$ sind stetig.

Eine Brownsche Bewegung wird oftmals einfach durch diese vier Punkte definiert. Man muss sich dann jedoch die Frage stellen, ob es überhaupt einen stochastischen Prozess gibt, der diese Bedingungen erfüllt. Ein Ansatz für diese Existenzaussage ist der oben behandelte Grenzübergang einer skalierten Irrfahrt. Ein Unterschied zwischen skaliertem Irrfahrt und der Brownschen Bewegung ist, dass die skalierte Irrfahrt eine natürliche Schrittzeit von $\frac{1}{n}$ hat und dazwischen linear verläuft, während es bei der Brownschen Bewegung keine linearen Teile gibt. Weiters ist die skalierte Irrfahrt für große n nur annähernd normalverteilt, während die Brownsche Bewegung per Definition exakt normalverteilt ist.

Beispiel 6.3: Sei $(W_t)_{t \geq 0}$ eine Brownsche Bewegung. Wir berechnen wieder die Wahrscheinlichkeit des Ereignisses $A = \{\omega : 0 \leq W^{(100)}(0.25) \leq 0.2\}$. Da $W(0.25)$ exakt $N(0, 0.25)$ -verteilt ist, erhalten wir $\mathbb{P}[A] = \frac{2}{\sqrt{2\pi}} \int_0^{0.2} e^{-2x^2} dx = 0.155422$. Die Werte der skalierten Irrfahrt $W^{(100)}(0.25)$ weichen also kaum vom exakten Ergebnis ab.

Die Brownsche Bewegung zählt zu den wichtigsten zeitstetigen stochastischen Prozessen und findet in zahlreichen Gebieten von Natur- und Wirtschaftswissenschaften bedeutende Anwendung. Daher gibt es auch eine sehr umfassende und interessante Theorie zur Brownschen Bewegung, die in vielen fortgeschrittenen Wahrscheinlichkeitstheoretischen Büchern und allen Werken zur sogenannten Stochastischen Analysis zu finden ist.

Literatur

- [1] H. Georgii: *Stochastik. Einführung in die Wahrscheinlichkeitstheorie und Statistik*, de Gruyter, Berlin, 2002.
- [2] B. Gnedenko: *Lehrbuch der Wahrscheinlichkeitsrechnung*, 4. Auflage, Akad.-Verl., Berlin, 1965.
- [3] N. Henze: *Stochastik für Einsteiger. Eine Einführung in die faszinierende Welt des Zufalls*, 7. Auflage, Vieweg, Wiesbaden, 2008.
- [4] C. Hesse: *Angewandte Wahrscheinlichkeitstheorie. Eine fundierte Einführung mit über 500 realitätsnahen Beispielen und Aufgaben*, Vieweg, Braunschweig/Wiesbaden, 2003.
- [5] A. Irle: *Wahrscheinlichkeitstheorie und Statistik. Grundlagen - Resultate - Anwendungen*, Teubner, Stuttgart/Leibzig/Wiesbaden, 2001.
- [6] W. Krämer: *Denkste!: Trugschlüsse aus der Welt der Zahlen und des Zufalls*, Piper Taschenbuch, Münchenn, 2011.
- [7] U. Krengel: *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, 8. Auflage, Vieweg, Wiesbaden, 2005.
- [8] K. Krickeberg, H. Ziezold: *Stochastische Methoden*, 4. Auflage, Springer, Paris/Kassel, 1994.
- [9] D. Meintrup, S. Schäffler: *Stochastik. Theorie und Anwendungen*, Springer, Berlin/Heidelberg, 2005.
- [10] P. Langrock, W. Jahn: *Einführung in die Theorie der Markovschen Ketten und ihre Anwendungen*, Teubner, Leipzig, 1979.
- [11] S. Shreve: *Stochastic Calculus for Finance II. Continious-Time Models*, Springer, New York, 2004.
- [12] <http://www-wt.iam.uni-bonn.de/ferrari/Lectures/WS11ProbaI/wt.pdf>
A. Bovier: *Einführung in die Wahrscheinlichkeitstheorie*, Universität Bonn, 2012. (abgerufen am 22.09.2012)
- [13] http://www-m12.ma.tum.de/foswiki/pub/M12/Allgemeines/GeburtsTodesProzesse/Geburts-und_Todesprozesse_WS2011-12.pdf
V. Hösel: *Geburts- und Todesprozesse*, TU München, 2011. (abgerufen am 01.10.2012)

- [14] <http://www.stochastik.uni-freiburg.de/Vorlesungen/vvSS2010/VorWTI I/skript200709.pdf>
H.R. Lerche: *Wahrscheinlichkeitstheorie II*, Universität Freiburg, 2007. (abgerufen am 10.09.2012)
- [15] <http://www.mathe.tu-freiberg.de/inst/stoch/Naether/stochii.pdf>
W. Näther: *Stochastik II. Skript zur Vorlesung*, TU Freiburg, 2005. (abgerufen am 25.08.2012)
- [16] <http://www3.math.tu-berlin.de/Vorlesungen/WS11/StoMo/stochmod.pdf>
M. Scheutzow: *Vorlesungsskript Stochastische Modelle*, TU Berlin, 2011. (abgerufen am 13.09.2012)
- [17] http://www.uni-ulm.de/fileadmin/website_uni_ulm/mawi.inst.110/lehre/ws07/wr/Wahrscheinlichkeitsrechnung.pdf
E. Spodarev: *Vorlesungsskript Wahrscheinlichkeitsrechnung*, Universität Ulm, 2008. (abgerufen am 27.07.2012)

Zusammenfassung

Die vorliegende Diplomarbeit setzt sich mit Grenzwertsätzen der Wahrscheinlichkeitstheorie auseinander. Es wird gezeigt, dass selbst der sprichwörtlich unberechenbare Zufall gewissen Gesetzen der Mathematik gehorcht. Die Resultate haben nicht nur für die theoretische Mathematik fundamentale Bedeutung, sondern spielen auch in zahlreichen Anwendungsgebieten eine große Rolle. Um den logischen Aufbau der späteren Inhalte nicht zu unterbrechen, werden in den ersten beiden Kapiteln grundlegende Begriffe der Stochastik behandelt, die für die Auseinandersetzung mit einem wahrscheinlichkeitstheoretischen Thema Voraussetzung sind. Neben dem üblichen Basiswissen werden auch charakteristische Funktionen eingeführt, die ein mächtiges Hilfsmittel der Wahrscheinlichkeitstheorie sind. Das dritte und vierte Kapitel beschäftigen sich mit Folgen von unabhängig und meist identisch verteilten Zufallsvariablen. Dabei werden die Schwachen und Starken Gesetze der Großen Zahlen sowie Zentrale Grenzwertsätze behandelt, die unter verschiedenen starken Voraussetzungen formuliert und bewiesen werden. Während die Gesetze der Großen Zahlen einen Zusammenhang zwischen arithmetischen Mitteln und Erwartungswerten (im Spezialfall zwischen relativen Häufigkeiten und Wahrscheinlichkeiten) herstellen, wird bei den Zentralen Grenzwertsätzen die Universalität der Normalverteilung gezeigt. Der fünfte Abschnitt befasst sich mit Markov-Ketten. Die Folge der Zufallsvariablen ist dabei nicht mehr unabhängig, sondern durch eine überschaubare Form der Abhängigkeit gegeben. Man könnte Markov-Ketten als stochastische Prozesse mit einem extremen Kurzzeitgedächtnis beschreiben, bei denen der nächste Schritt immer nur vom vorherigen, nicht jedoch von der restlichen Vergangenheit, abhängt. Das Hauptaugenmerk wird nach der Einführung der grundlegenden Begriffe wieder auf das Langzeitverhalten gelegt. Der letzte Abschnitt stellt eine interessante Anwendung der Markov-Ketten dar. Einfache Irrfahrten auf $\mathbb{Z}^d$ sind sehr spezielle stochastische Prozesse, die zu jedem Zeitpunkt genau einen Schritt mit konstanter Länge in eine bestimmte Richtung machen können. Das Langzeitverhalten solcher Irrfahrten wird in den verschiedenen Dimensionen studiert. Man interessiert sich dafür, ob die Irrfahrt zu ihrem Ausgangspunkt mit Sicherheit wieder zurückkehrt oder nicht. Abschließend wird der eindimensionalen Irrfahrt genauere Beachtung geschenkt. Das Arkus-Sinus-Gesetz und der Übergang zur Brownschen Bewegung werden dabei behandelt.

Abstract

This thesis deals with limit theorems of the probability theory. It is shown that even the “unpredictable randomness” follows certain laws of mathematics. These results have fundamental importance, not only in theoretical mathematics, but also in a large variety of applications. In order to not interrupt the following sections, the first two chapters contain basic knowledge about probability theory. Additionally to the usual basics also characteristic functions are introduced, which are a powerful appliance of probability theory. The third and fourth chapter deal with sequences of independent and mostly identically distributed random variables. The weak and strong laws of the large numbers are discussed and will be stated and proved under different conditions. While the laws of large numbers make a connection between arithmetic means and expected values (in special cases between relative frequencies and probabilities) the central limit theorems show the universality of the normal distribution. The fifth section deals with Markov chains. Therefore the sequences of random variables are not independent anymore but given by a clear form of dependence. One could describe Markov chains as stochastic processes with an extremely short time memory. Every step depends only on the former one but not on the rest of the past. After an introduction of the basic concepts the main attention is placed on the long term behaviour of such processes. The last section provides an interesting application of Markov chains. Simple random walks on $\mathbb{Z}^d$ are very special stochastic processes. At each time they make exactly one step with a constant length in a certain direction. The long term behaviour of such random walks is studied in different dimensions. We are interested in whether the random walk returns to his starting point with certainty or not. Finally, the one-dimensional random walk is given a more detailed analysis. The Arcus-Sinus-Law and the transition to the Brownian Motion are examined there as well.

Lebenslauf

Persönliche Daten

Name: Jürgen Steininger
Adresse: Waldberg 16, 3841 Windigsteig
Geburtsdatum: 28.12.1988
Staatsbürgerschaft: Österreich
Eltern: Franz und Renate Steininger
Geschwister: Andreas

Ausbildung

1995-1999: Volksschule Windigsteig
1999-2003: Hauptschule Waidhofen/Thaya
2003-2008: Bundeshandelsakademie Waidhofen/Thaya
13. Juni 2008: Reifeprüfung mit Auszeichnung bestanden
2008-2009: Präsenzdienst in Allentsteig
2009-2013: Studium an der Universität Wien:
UF Mathematik und UF Geschichte, Sozialkunde und Politische Bildung
2011-2013: Studium an der Universität Wien:
Bachelorstudium Mathematik
Ab 2013: Studium an der Universität Wien:
Masterstudium Mathematik