



universität
wien

MAGISTERARBEIT / MASTER THESIS

Titel der Magisterarbeit / Title of the Master Thesis

„Galton-Watson-Verzweigungsprozesse – Theorie & Anwendung“

verfasst von / submitted by

Rafael Jochum, BSc

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

Magister der Sozial- und Wirtschaftswissenschaften

(Mag. rer. soc. oec.)

Wien, 2023 / Vienna, 2023

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

UA 066 951

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

Magisterstudium Statistik

Betreut von / Supervisor:

ao. Univ.-Prof. Dipl.-Ing. Dr. Werner Schachinger

Kurzfassung

Die Einsatzpalette der Galton-Watson-Prozesse, welche die Grundlage für viele weitere Verzweigungsprozesse bilden, ist weit gefächert. Diese stochastischen Modelle wurden bereits in der Vergangenheit ergiebig erforscht. Aber auch gegenwärtig sind sie Gegenstand vieler Untersuchungen. In der vorliegenden Masterarbeit wird in erster Linie die elementare mathematische Theorie sowohl für das klassische Modell als auch für Galton-Watson-Prozesse mit mehreren unterschiedlichen Typen neu aufbereitet. Dabei wird mit einer Detailliertheit vorgegangen, welche über jene in der gängigen, dazu zu findenden Literatur hinausgeht. Zudem wird ein mathematischer Appendix zur Verfügung gestellt, in welchem ggf. viele notwendige Begrifflichkeiten und Resultate nachgeschlagen werden können. Insgesamt war die Zielsetzung, dass die Herleitungen der Erkenntnisse möglichst niederschwellig und vollumfänglich nachvollziehbar sind.

Abgedeckt werden jeweils theoretische Analysen der stochastischen Eigenschaften der Populationsgrößen – vielfach wird mit wahrscheinlichkeitserzeugenden Funktionen vorgegangen. Weiters ist eine Betrachtung der Prozesse als Markov-Ketten bis hin zur Bestimmung der Aussterbewahrscheinlichkeiten beinhaltet. Bzgl. der traditionellen Galton-Watson-Prozesse werden darüber hinaus erste asymptotische Grenzwertresultate gegeben und u.a. die Verteilung des Aussterbezeitpunkts und der Gesamtpopulation der Prozesse untersucht.

Eine Vertiefung anhand konkreter Beispielprozesse wird mit der Statistiksoftware **R** unterstützt. Dies bildet den zweiten Schwerpunkt dieser Arbeit. Hierzu werden ergänzend zu bestehenden Paketen eigene Funktionen bereitgestellt. Diese ermöglichen Analysen der zuvor theoretisch erzielten Ergebnisse, was abschließend anhand eines weiteren expliziten Beispiels eines Galton-Watson-Prozesses mit mehreren verschiedenen Typen demonstriert wird. Um diesbezüglich einen potentiellen Anknüpfungspunkt zu bieten, findet sich der Quellcode der R-Routinen im Anhang.

Abstract

The scope of application of Galton-Watson processes, which constitute the basis for many other branching processes, is comprehensive. Having been extensively researched in the past, these stochastic models continue to be the subject of many current investigations. In the present master thesis, primarily the elementary mathematical theory for the classical model as well as for Galton-Watson processes with multiple types is reworked. In this respect, particular attention is paid to the detailed approach, which exceeds what one usually finds in the standard literature. Moreover, a mathematical appendix is provided which serves as a reference work for various necessary terms and results. Overall, the derived findings should be accessible and fully comprehensible.

Theoretical analyses of the stochastic properties of population sizes are covered, predominantly using probability generating functions. In addition, both a consideration of the processes as Markov chains and the determination of extinction probabilities are included. With regard to traditional Galton-Watson processes, initial asymptotic limit theorems are given. Furthermore, the distribution of extinction time and total population of processes are examined.

Based on precise example processes a deepening of understanding is supported by the statistics software R. This forms the second aim of the present master thesis. Complementing the existing packages, custom functions are presented which facilitate the analyses of the previous theoretically obtained results. These analyses are finally demonstrated by including another explicit example of a Galton-Watson process with multiple types. In order to provide a potential connecting point, the source code of the R functions is incorporated in the appendix.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Abgrenzung der Themenstellung und Methodik	1
1.2	Mathematische Vorbemerkungen	3
1.2.1	Grundlegende Notation, Symbole und Begrifflichkeiten	3
2	Einfache Galton-Watson-Verzweigungsprozesse	5
2.1	Informelle Beschreibung des GW-Verzweigungsprozesses	5
2.2	Weitere mathematische Vorüberlegungen und Werkzeuge	7
2.2.1	Zufallsvariablen mit Werten in den natürlichen Zahlen	7
2.2.2	Wahrscheinlichkeitserzeugende Funktionen	9
2.3	Definition des GW-Verzweigungsprozesses	25
2.3.1	Grundlegende Definitionen	25
2.3.2	GWP als Markov-Kette und Übergangswahrscheinlichkeiten	26
2.4	Wahrscheinlichkeitsverteilung und Momente der Populationsgrößen	31
2.4.1	Wahrscheinlichkeitserzeugende Funktionen der individuellen Nachkommenszahl und der Populationsgrößen	31
2.4.2	Momente der individuellen Nachkommenszahl und der Populationsgröße	35
2.4.3	Basis-Annahmen bzw. triviale Fälle	40
2.5	Aussterbe- bzw. Überlebenswahrscheinlichkeiten	41
2.5.1	Aussterben des Prozesses	41
2.5.2	Aussterbe- und Überlebenswahrscheinlichkeit	41
2.5.3	Bestimmung der Aussterbewahrscheinlichkeit und Kritikalität des GWP	42
2.5.4	Instabilität des Prozesses	46
2.6	Asymptotische bzw. Grenzwertresultate	48
2.6.1	Konvergenz der normalisierten Populationsgrößen	48
2.6.2	Verteilung von W	49
3	Galton-Watson-Verzweigungsprozesse mit mehreren unterschiedlichen Typen	55
3.1	Vorbemerkungen	55
3.1.1	Informelle Beschreibung des Multitype Galton-Watson-Prozesses	55
3.1.2	Multivariate Notation und Konventionen	56
3.1.3	Zufallsvektoren und -matrizen	58
3.1.4	Multivariate wahrscheinlichkeitserzeugende Funktionen	61
3.2	Definition des Multitype GW-Verzweigungsprozesses	65
3.2.1	Grundlegende Definitionen	65

3.2.2	Exemplarische Realisierung	68
3.3	Verteilung, Markov-Eigenschaft und Momente	70
3.3.1	Wahrscheinlichkeitserzeugende Funktionen für den Multitype GWP	70
3.3.2	Momente der Reproduktionszahlen und Populationsgrößen	74
3.3.3	Eigenschaften in Abhängigkeit von $\mathbf{M}^n$	82
3.3.4	Multitype GWP als Markov-Kette	89
3.4	Aussterben und Aussterbewahrscheinlichkeit	93
3.4.1	Aussterben des Multitype Galton-Watson-Prozesses	93
3.4.2	Kritikalität des MGWP – positiv regulärer Fall	95
4	Anwendungsbeispiele	99
4.1	Einfache GW-Verzweigungsprozesse – geometrischer Fall	99
4.1.1	Definition und allgemeine Analyse des geometrischen GWP	100
4.1.2	Asymptotisches Verhalten superkritischer geometrischer GWP	106
4.1.3	Exkurs: Gesamtpopulation und Aussterbezeitpunkt – allgemein	113
4.1.4	Gesamtpopulation geometrischer GW-Prozesse	120
4.1.5	Der auf das Aussterben bedingte Prozess	127
4.2	Multitype GW-Verzweigungsprozesse mit $\mathbf{R}$	130
4.2.1	$\mathbf{R}$ -Funktionen zur Analyse eines Multitype GWP	130
4.2.2	Anwendung: Numerisches Beispiel	139
4.2.3	Anwendung: Weitere Multitype GWP	148
5	Zusammenfassung und Ausblick	151
A	Appendix – Mathematisch	153
A.1	Maß- und wahrscheinlichkeitstheoretische Grundbegriffe	153
A.1.1	Mengentheorie	153
A.1.2	Maßtheorie	158
A.1.3	Wahrscheinlichkeitstheorie – Allgemein	164
A.1.4	Wahrscheinlichkeitstheorie - Diskret	174
A.2	Weitere Resultate und Begrifflichkeiten – Wahrscheinlichkeitstheorie	180
A.2.1	Diskrete Faltung der Verteilungen $\mathbb{N}_0$ -wertiger Zufallsvariablen	180
A.2.2	Erwartungswerte, Varianzen und Kovarianzen	184
A.2.3	Bedingte Wahrscheinlichkeit und Erwartung	186
A.2.4	Markov-Ketten	193
A.2.5	Konvergenzarten	198
A.2.6	Laplace-Transformation und charakteristische Funktionen	199
A.2.7	Martingale in diskreter Zeit	201
A.3	Resultate und Begrifflichkeiten – Analysis	203
A.3.1	Funktionen	203
A.3.2	Reihen und Potenzreihen	204
A.3.3	Doppelfakultät und Catalan-Zahlen	210

B Appendix	213
B.1 Nebenrechnungen	213
B.2 Quellcodes	224
B.2.1 R-Quellcode für den Abschnitt 4.2 – MGWP mit R	224
B.2.2 R-Quellcode für verwendete Funktionen und Abhängigkeiten	227
Verzeichnisse	233
Abkürzungsverzeichnis	233
Abbildungsverzeichnis	234
Tabellenverzeichnis	234
Verzeichnis dokumentierter R-Funktionen	234
Verzeichnis der R-Ausgaben	234
Literaturverzeichnis	235
Software- und Paketressourcen	238

1 Einleitung

In der zweiten Hälfte des 19. Jahrhunderts interessierten sich der britische Naturforscher Sir FRANCIS GALTON¹ und der ebenfalls britische Kleriker sowie Mathematiker Rev. HENRY W. WATSON für die Entwicklung von vorwiegend aristokratischen Familienstammbäumen. Es wurde beobachtet, dass Adelshäuser bzw. ihre Nachnamen oft dazu tendieren, früher oder später unterzugehen. GALTON und WATSON veröffentlichten im Jahre 1875 den Artikel [45] (*'On the probability of the extinction of families.'*), in welchem sie ein mathematisches Modell zur Prognose der Wahrscheinlichkeit des Aussterbens solcher Familiennamen beschrieben. Aus diesem Grund fungieren beide als Namensgeber für die sogenannten *Galton-Watson-Verzweigungsprozesse*², kurz nur *GW-Prozesse*.

Bis in die Gegenwart bilden diese die Grundlage für eine Vielzahl an *Verzweigungsprozessen*, mit welchen 'Populationsdynamiken' modelliert werden. Deren Anwendungsgebiete beschränken sich dabei keineswegs auf die Ahnenforschung (auch Genealogie genannt). Sie können in etlichen Bereichen zuverlässige Resultate liefern – sei es in der Demographie, Physik, Biologie und Medizin bis hin zur Epidemiologie oder aber auch in der Ökonomie sowie Informatik (vgl. etwa [2]).

Deshalb soll die zugrundeliegende mathematische Theorie über die GW-Prozesse im Rahmen der vorliegenden Arbeit neu aufbereitet werden. Hierbei wird auch auf computergestützte Analysen des Erarbeiteten gesetzt. Etwa ab der Mitte des vergangenen Jahrhunderts sind mehrere umfangreiche Standardwerke zu dieser Thematik (d.h. zu Verzweigungsprozessen im weiteren Sinne) entstanden – rechnergestützte Methoden standen zu den Veröffentlichungszeitpunkten meist noch nicht (oder noch nicht in der Form von heute) zur Verfügung. Besonders genannt seien die Monographien [19] von THEODORE E. HARRIS aus dem Jahre 1963 sowie [3] von KRISHNA B. ATHREYA und PETER E. NEY aus dem Jahre 1972 (siehe u.a. aber auch die weiteren Referenzen auf der Seite 2).

1.1 Abgrenzung der Themenstellung und Methodik

Es soll und kann nicht annähernd derart in die Tiefe gegangen werden wie in dieser Standardliteratur. Im Vordergrund steht vielmehr, möglichst *durchgängige und vollumfängliche Beweisketten* zu konstruieren. Die Zielsetzung war dabei, besonders auf die

¹GALTON war ein Cousin des berühmteren CHARLES DARWIN – vgl. etwa [18, Acknowledgments].

²Wenngleich WATSON genau genommen irrtümlich zum Schluss kam, dass keine Stammlinie überleben kann – vgl. etwa [39, S. XVI]. Nicht zuletzt deshalb wird manchmal auch die Bezeichnung *Bienaymé-Galton-Watson-Prozesse* gebraucht, da I.J BIENAYMÉ in Frankreich vermutlich bereits zuvor eine korrekte Aussterbewahrscheinlichkeit in einem ähnlichen Zusammenhang ermittelte. Dessen Arbeit wurde aber erst 1972 'wiederentdeckt' (vgl. etwa den sehr lesenswerten historischen Abriss in [25, 1.1, S. 1 ff.] – für weitere Retrospektiven siehe u.a. z.B. [19, 1., S. 1 ff.] oder [39, S. XV ff.]).

1 Einleitung

Nachvollziehbarkeit der Argumente zu achten. Es wurde angestrebt, wesentlich mehr ins Detail gehen zu können als die erwähnten Autoren (was angesichts der dortigen Inhaltsfülle bestimmt weder machbar war noch beabsichtigt wurde).

An einigen Stellen werden auch unterschiedliche Wege dargeboten, um zum selben Resultat zu kommen. Zudem wird ein umfangreicher mathematischer Appendix mitgeliefert – teilweise mit Beweisführungen, aber durchwegs zumindest mit Literaturhinweisen. Wenn und wo nötig, können in diesem die wichtigsten mathematischen Erfordernisse eingesehen werden. Dadurch muss insgesamt deutlich weniger Grundlagenwissen vorausgesetzt werden – siehe auch die dahingehenden Vorbemerkungen im nächsten Abschnitt 1.2.

Eine der Kernannahmen des Galton-Watson-Modells ist, dass sich Individuen nicht voneinander unterscheiden. Diese Bedingung aufzuweichen und sogenannte *GW-Prozesse mit mehreren unterschiedlichen Typen* miteinzubeziehen, war ein weiteres Ziel dieser Arbeit. Darüber hinaus werden vorhandene Softwarelösungen um gewisse Instrumentarien ergänzt, welche v.a. die Analyse der theoretischen Resultate im Fokus haben. Das Vorhaben war, sozusagen eine Vorarbeit zu leisten, an welche angeknüpft werden kann.

Angesichts der beschriebenen Maßgaben werden im 2. Kapitel (nach einer zunächst informellen Beschreibung) geeignete mathematische Werkzeuge für den Galton-Watson-Prozess eingerichtet. Anschließend erfolgt die formale Definition dieses Prozesses. In Folge einer Untersuchung wichtiger stochastischer Eigenschaften kann die zentrale Frage nach der Wahrscheinlichkeit des Aussterbens beantwortet werden. Einen Einblick in asymptotische Resultate, welche sich u.a. mit dem Verhalten im Überlebensfall beschäftigen, gibt der Abschnitt 2.6.

Das 3. Kapitel widmet sich bei ähnlicher Strukturierung und etwas steigender Komplexität der Verallgemeinerung auf mehrere Typen. Teilweise kann dabei auf Vorherigem aufgebaut werden. Im Abschnitt 3.4 wird letztendlich die Hauptaussage dieser Arbeit über die diesbezügliche Aussterbewahrscheinlichkeit in geeigneten Fällen bewiesen.

Ab dem 4. Kapitel wird die Statistiksoftware `R` als Rechenunterstützung eingesetzt. Zunächst stehen erneut mathematische Zugänge zu den originalen Prozessen im Vordergrund, wobei die Theorie um manche Konzepte ergänzt wird. Eine spezielle Aufmerksamkeit wird auf Fortpflanzungsmechanismen gelegt, welche einer geometrischen Verteilung folgen. Danach wird das Hauptaugenmerk hingegen auf die Analyse der verallgemeinerten GW-Prozesse mit `R` gerichtet. Insgesamt stehen aber auch hier mathematische Aspekte im Vordergrund, weshalb bis auf wenige Ausnahmen nicht auf 'reale' Anwendungen eingegangen wird. Es sei auf weiterführende Literatur verwiesen.

Der Theorieaufbau orientiert sich hauptsächlich an den bereits genannten Büchern [19, I. und II.] und [3, I. und V.]. Wesentliche Teile sind aber auch von [25] (insbes. das 2. Kapitel) und [39] (3. Kapitel) inspiriert, welche (unter anderen) ebenso als Standardwerke einzustufen sind. Aber auch aktuellere Veröffentlichungen wurden zu Rate gezogen. Besonders herausgestrichen werden sollen [17] (worin stochastische Prozesse generell behandelt werden) oder [18] sowie [29]. Die beiden Letzteren sind von Biologen mitverfasst und spezifisch auf die Biologie ausgerichtet. Hier finden sich etwa viele einschlägige Applikationen (wie beispielsweise aus der Evolutionstheorie, Krebsforschung, Molekularbiologie oder Genetik).

1.2 Mathematische Vorbemerkungen

Es werden mathematische Grundkenntnisse – etwa in linearer Algebra, Analysis oder Wahrscheinlichkeitsrechnung bzw. -theorie – prinzipiell als bekannt vorausgesetzt. Dennoch werden gewisse benötigte Definitionen und Konventionen einleitend angeführt. Bei Bedarf können im Appendix A zudem auszugsweise einige der verwendeten mathematischen Konzepte, Begrifflichkeiten und Resultate nachgeschlagen werden. Sofern es für eine bessere Nachvollziehbarkeit hilfreich erscheint, wird auf die jeweils entsprechende Stelle in diesem Appendix referenziert. Das ist beispielsweise dann der Fall, wenn Argumente darauf beruhen oder Definitionen daraus verwendet werden.

Der Abschnitt A.1 widmet sich dabei maß- und wahrscheinlichkeitstheoretischen Grundbegriffen. Dieser beinhaltet zuerst entsprechende Vorüberlegungen aus der Mengenlehre (A.1.1) und Maßtheorie (A.1.2). Ansonsten ist er eher allgemein gehalten und soll einen groben Überblick über Wahrscheinlichkeitstheorie generell (Wahrscheinlichkeitsräume, Verteilungen, Zufallsvariablen, Unabhängigkeit und Erwartungswerte – A.1.3) und in diskreten Fällen (A.1.4) bieten.

Für die Auswahl der Auszüge in den darauf folgenden Abschnitten (A.2 wiederum aus der Wahrscheinlichkeitsrechnung bzw. -theorie sowie A.3 aus der Analysis) war hingegen mehr die spezifische Relevanz für den vorliegenden Text ausschlaggebend.

Zudem finden sich gewisse formalere, ergänzende Argumentationen bzw. Berechnungen im Appendix B unter dem Abschnitt B.1. Diese können ggf. dem besseren Verständnis dienlich sein, sind aber für den Lesefluss eher hinderlich. Sie werden als 'Nebenrechnungen' (kurz NR) bezeichnet und an den relevanten Stellen wird darauf verwiesen.

1.2.1 Grundlegende Notation, Symbole und Begrifflichkeiten

Da Notationen in der Literatur nicht immer einheitlich sind, werden zunächst einige Festlegungen getroffen.

Notation 1.1 (Allgemeine Konventionen)

- Für die Menge der natürlichen Zahlen wird die Konvention getroffen, dass die Zahl 0 bei $\mathbb{N}_0$ inkludiert und bei $\mathbb{N}$ exkludiert ist. D.h. $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$, $\mathbb{N} = \{1, 2, 3, \dots\}$ und $\mathbb{N} = \mathbb{N}_0 \setminus \{0\}$. Wenn aus dem Zusammenhang klar ist, dass eine Variable n Werte in $\mathbb{N}_0$ oder $\mathbb{N}$ annimmt, wird meist nur kurz $n \geq 0$ bzw. $n \geq 1$ (anstelle von $n \in \mathbb{N}_0$ mit $n \in \mathbb{N}$) geschrieben. Analoges gilt für entsprechende Teilmengen (z.B. $0 \leq n \leq m$ anstatt $n \in \mathbb{N}_0$ mit $n \leq m$). Für höhere Dimensionen $k \geq 1$ (d.h. $k \in \mathbb{N}$) wird $\mathbb{N}_0^k$ bzw. $\mathbb{N}^k$ geschrieben (siehe die mehrdimensionale Notation 3.1).
- Die Menge der nichtnegativen reellen Zahlen wird als $\mathbb{R}_+ := \{x \in \mathbb{R} : x \geq 0\}$ geschrieben (und für höhere Dimensionen $k \geq 1$ analog $\mathbb{R}_+^k = \{(x_i)_{i=1}^k : x_i \in \mathbb{R}_+\}$).
- Folgen, Listen, Familien und dergleichen werden mit geschwungenen Klammern geschrieben, etwa $\{a_i\}_{i \geq 1}$ für $a_1, a_2, \dots$
- Punktweise Zusammenhänge von oder zwischen Funktionen werden folgendermaßen abgekürzt:

1 Einleitung

- *Punktweise arithm. Operationen:* Für auf einer Menge X def. Funktionen f, g wird etwa nur kurz $f+g$ anstatt $(f+g)(x) = f(x)+g(x) \forall x \in X$ geschrieben.
- *Punktweise Vergleiche:* Bei einer Funktion $f : X \rightarrow Y$ und $a \in Y$ ist beispielsweise mit $f > a$ gemeint, dass $f(x) > a$ für alle $x \in X$ gilt.
- Dementsprechend beschreibt, etwa für $Y = \mathbb{R}$ oder $Y = \mathbb{R}^k$ (oder Teilmengen davon), $f \geq 0$ eine *nichtnegative Funktion bzw. Abbildung*. Im k -dimensionalen Fall ist dies komponentenweise zu verstehen (vgl. wieder die Notation 3.1; $\mathbf{f} = (f_1, \dots, f_k)^T \geq \mathbf{0}$ bedeutet also $f_i(x) \geq 0$ für jedes $1 \leq i \leq k$ und $x \in X$).
- Einseitige Grenzwerte einer Funktion f an einer Stelle x werden wie folgt abgekürzt: Der linksseitige als $f(x^-) := \lim_{s \nearrow x} f(s)$ und der rechtsseitige als $f(x^+) := \lim_{s \searrow x} f(s)$.
- Konvergiert eine Folge $\{a_i\}_{i \geq 1}$ gegen einen Grenzwert a , d.h. $\lim_{i \rightarrow \infty} a_i = a$, wird dies oft kurz als $a_i \rightarrow a$ geschrieben (vorausgesetzt die 'Laufvariable' ist klar). Gleiches gilt für bestimmte Divergenzen. So bedeutet etwa $a_i \rightarrow \infty$, dass der Grenzwert $\lim_{i \rightarrow \infty} a_i = \infty$ ist.
- Mit $a_i \nearrow a$ ist gemeint, dass etwa eine Folge $\{a_i\}_{i \geq 1}$ monoton wachsend gegen a konvergiert, d.h. $a_1 \leq a_2 \leq \dots \leq \lim_{i \rightarrow \infty} a_i = a$. Analog meint $a_i \searrow a$, dass sie monoton fallend gegen a konvergiert.
- *Leere Summe:* Einer Summe $\sum_{i \in I} a_i$ über eine leere Indexmenge $I = \emptyset$ oder einer der Art $\sum_{i=m}^n a_i$ mit $n < m$ wird der Wert 0 gegeben (vgl. z.B. [22, S. 90 f.]).
- *Leeres Produkt:* Dahingegen werden Produkte $\prod_{i \in I} b_i$ über leere Indexmengen I oder $\prod_{i=m}^n b_i$ mit $n < m$ auf 1 gesetzt (vgl. z.B. [22, S. 94]). Das ist konsistent mit der leeren Summe, denn für $b_i := s^{a_i}$ gilt allg. $\prod_{i \in I} b_i = \prod_{i \in I} s^{a_i} = s^{\sum_{i \in I} a_i}$. Mit $I = \emptyset$ folgt deshalb $1 \stackrel{\text{leeres Produkt}}{=} \prod_{i \in I} b_i = s^{\sum_{i \in I} a_i} \stackrel{\text{leere Summe}}{=} s^0$. $\triangle$

Gelegentlich wird die folgende Erweiterung der reellen Zahlen verwendet.

Definition 1.2 (Erweiterte reelle Zahlen)

Die Erweiterung der reellen Zahlen $\mathbb{R}$ um $\pm\infty$ ist definiert als $\bar{\mathbb{R}} := \mathbb{R} \cup \{+\infty, -\infty\}$.

Zudem erweist sich die folgende Notation manchmal als nützlich.

Definition 1.3 (Kronecker-Delta – vgl. z.B. [22, (16.2), S. 129])

Sei I eine beliebige Indexmenge. Für $i, j \in I$ ist das sogenannte *Kronecker-Delta* oder *Kronecker-Symbol* $\delta_{ij} : I^2 \rightarrow \{0, 1\}$ definiert als

$$\delta_{ij} := \begin{cases} 1 & \text{für } i = j, \\ 0 & \text{für } i \neq j. \end{cases} \quad \text{Die Folge } \delta_i \text{ sei definiert als } \delta_i := \{\delta_{ij}\}_{j \in I}. \quad (1.1)$$

Es ist auch möglich, das Kronecker-Delta δ_{ij} als Indikatorfunktion (siehe den letzten Punkt der Not. A.1 im Appendix) $\mathbb{1}_{\{i\}} : I \rightarrow \{0, 1\}$ zu schreiben. Es ist dann $\delta_{ij} = \mathbb{1}_{\{i\}}(j)$.

2 Einfache Galton-Watson-Verzweigungsprozesse

Zu Beginn steht jenes Modell, dessen Überlieferung auf GALTON und WATSON zurückgeht. Es soll auf den Vorschlag von HARRIS in [19, S. 2] zurückgegriffen werden, den durch das Modell beschriebenen Prozess als *den* Galton-Watson-Prozess zu bezeichnen. Dieser basiert auf besonders vereinfachenden Annahmen. Deshalb wird (etwa angelehnt an [18, 2.1, S. 13]) auch die Benennung *einfacher* Galton-Watson-Prozess verwendet. Wird also vom GW(-Verzweigungs)-Prozess – bzw. kurz nur *GWP* – gesprochen, ist dieser einfache Prozess gemeint. Ob die zu modellierenden Populationen dabei – wie beim ursprünglichen Prozess – von genealogischem Interesse sind (also Familien bzw. deren Nachnamen darstellen), es sich etwa um Zellen, Neutronen in einem Reaktor oder epidemische Ausbreitungen handelt oder aber ganz andere Dinge damit simplifiziert untersucht werden sollen, wird grundsätzlich meist nicht näher spezifiziert.

Im nächsten Abschnitt werden solche Prozesse informell vorgestellt und anschließend einige mathematische Vorüberlegungen angestellt. Danach wird die formale Definition der einfachen GW-Prozesse vorgenommen.

2.1 Informelle Beschreibung des GW-Verzweigungsprozesses

Für diese Erläuterung soll weitgehend wieder zum historischen Kontext von Familien zurückgekehrt werden. In weiterer Folge wird dies wieder allgemeiner gehalten.

GALTON und WATSON begnügten sich damit, für ihre genealogischen Untersuchungen nur die männlichen Familienmitglieder zu betrachten. Mittlerweile wäre dies längst anachronistisch, war aber damals aus traditionellen Gründen bestimmt gerechtfertigt. Dies führt zu der ersten Annahme, dass im einfachen GWP *nur eine Art von Individuen* vorkommt – im vorliegenden Setting eben Männer (und in einem weiteren Sinne können etwa die Individuen irgendeiner 'Population' oft als asexuell angesehen werden).

Nun ist für die Familienentwicklung relevant, wie viele männliche Abkömmlinge ein Mann hat. Deshalb müssen wiederum nur Söhne berücksichtigt werden, welche das zeugungsfähige Alter erreichen. Wie viele das jeweils sind, wird als zufällig angenommen. Dabei sind die (i.A. verschiedenen) *Wahrscheinlichkeiten für die möglichen Anzahlen an Nachkommen* (inkl. Null) in jeder Generation und für jedes Individuum genau *gleich*¹. D.h., jeder Mann hat die selben Chancen, einen, zwei, drei usw. oder vielleicht gar keinen

¹In der ursprünglichen Problemstellung wurden diese Wahrscheinlichkeiten als relative Häufigkeiten interpretiert. Weiters ließ die Formulierung ein Maximum von fünf Söhnen zu (vgl. [45, S. 139]). Beim GWP können aber grundsätzlich beliebig viele Anzahlen eine positive Wahrscheinlichkeit haben.

2 Einfache Galton-Watson-Verzweigungsprozesse

Sohn (der das entsprechende Alter erreicht) zu bekommen. In diesem Sinne wird also von der gänzlichen Ununterscheidbarkeit der Individuen ausgegangen.

Dabei werden keinerlei Beeinflussungen von außen oder untereinander berücksichtigt. Es wird unterstellt, dass die (zufällige) Fortpflanzung einzelner vollkommen *unabhängig voneinander* geschieht². Und zwar sowohl innerhalb einer Generation als auch generationsübergreifend (es hat also auch keine Auswirkungen auf die Anzahl der Nachkommen eines Individuums, wie viele Brüder es hat oder wer sein Vater ist).

Der Faktor Zeit wird dahingehend simplifiziert, dass nur ganze (sich nicht überlappende) Generationen betrachtet werden. Jedes Familienmitglied *lebt genau eine Generation lang*, zeugt Nachkommen und stirbt. Diese Nachkommen leben anschließend in der nächsten Generation. So wiederholt sich das ganze Prozedere (bei variierender Anzahl an Individuen pro Generation) immer wieder – bis eventuell in einer Generation gar keine Söhne gezeugt werden und dieser Familienname somit möglicherweise 'ausstirbt'.

Dazu muss auch eine Generation als Ausgangspunkt des Galton-Watson-Prozesses bestimmt werden. Diese wird im weiteren Verlauf dieses Texts als die initiale Population bezeichnet werden. Dabei spielt es angesichts der obigen Annahmen eigentlich keine Rolle, ob sie aus nur einer Familie besteht oder ob mehrere Stammlinien auf einmal betrachtet werden³. An diese Anfangspopulation wird an sich keine spezifische Anforderung gestellt. Prinzipiell kann sie selbst ebenfalls zufällig sein. Ab dem Abschnitt 2.4 wird es sich aber als nützlich erweisen, von einer festen Größe zu Beginn auszugehen. Meist wird sogar ein einzelner 'Gründungsvater' einer einzelnen Familie angenommen – was ohne Weiteres auf mehrere verallgemeinert werden kann.

²Dies stellt sicher eine drastische und in etlichen Bereichen auch unrealistische Vereinfachung dar. V.a. in physikalischen Anwendungsgebieten hat sie aber oft ihre Berechtigung (vgl. [19, S. 3]). Und etwa auch bzgl. biologischer Populationen, welche entweder hinreichend simpel oder groß sind, können u.U. vernünftige Ergebnisse erzielt werden (vgl. [39, S. XI]).

³Im klassischen Setting besteht die Ursprungspopulation aus einer gewissen Anzahl von Männern mit jeweils einem anderen Familiennamen (vgl. erneut [45, S. 139]).

2.2 Weitere mathematische Vorüberlegungen und Werkzeuge

Auf Begrifflichkeiten wie Wahrscheinlichkeitsräume (im Folgenden oft kurz als *W-Räume* bezeichnet), Wahrscheinlichkeitsmaße (kurz *W-Maße*) bzw. Verteilungen oder Zufallsvariablen (kurz *ZV*) sei ggf., wie einleitend erwähnt, erneut auf A.1.3 im Appendix verwiesen. Es kommen mehrheitlich diskrete Zufallsvariablen mit Werten in den natürlichen Zahlen $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ zur Anwendung (vgl. dazu ggf. A.1.4). Zunächst werden ein paar Vorbemerkungen bzgl. solcher $\mathbb{N}_0$ -wertiger Zufallsvariablen angeführt. Zu diesen lassen sich sogenannte *wahrscheinlichkeitserzeugende Funktionen* definieren, welche anschließend vorgestellt und untersucht werden.

2.2.1 Zufallsvariablen mit Werten in den natürlichen Zahlen

Besteht allgemein eine Ergebnismenge Ω aus den natürlichen Zahlen $\mathbb{N}_0$ und ist somit abzählbar, so kann auf die folgende Weise ein entsprechender diskreter (abzählbarer) W-Raum darauf definiert werden (vgl. ggf. die Definitionen A.39 und A.40 unter Berücksichtigung der anschließenden Bemerkungen für den Fall abzählbarer Ergebnismengen).

Definition 2.1 (Verteilung bzw. Wahrscheinlichkeitsmaß auf $\mathbb{N}_0$)

Gegeben eines Messraumes, bestehend aus der (abzählbaren) Ergebnismenge $\Omega = \mathbb{N}_0$ und dem Ereignisraum $\Sigma = \mathcal{P}(\mathbb{N}_0)$, ist $\mathbb{P} : \mathcal{P}(\mathbb{N}_0) \rightarrow [0, 1]$ ein *diskretes Wahrscheinlichkeitsmaß* bzw. eine *diskrete (Wahrscheinlichkeits-)Verteilung* (auf $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$) und somit $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0), \mathbb{P})$ ein diskreter W-Raum.

Bemerkung zur Definition 2.1. Es gilt $\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = \sum_{n \geq 0} \mathbb{P}(\{n\}) \stackrel{(A.50)}{=} 1$ und gemäß der Bemerkung zum Satz A.41 (i) ist $\mathbb{P}$ bereits durch die Folge $\{\mathbb{P}(\{n\})\}_{n \geq 0}$ eindeutig bestimmt. $\triangle$

Für alles Folgende in diesem Abschnitt sei aber $(\Omega, \Sigma, \mathbb{P})$ ein (nicht näher spezifizierter) allgemeinerer Wahrscheinlichkeitsraum. Kann eine Zufallsvariable nur Werte in $\mathbb{N}_0$ annehmen, so handelt es sich um eine diskrete Zufallsvariable (vgl. die Definition A.42 und beachte die dazugehörige Bemerkung für den Fall abzählbarer Wertebereiche). Daher bietet sich die folgende Definition an.

Definition 2.2 ($\mathbb{N}_0$ -wertige Zufallsvariable und ihre Verteilung)

Sei $X : \Omega \rightarrow \mathbb{N}_0$, also eine diskrete Zufallsvariable mit Werten in $\mathbb{N}_0$. Dann wird X als *$\mathbb{N}_0$ -wertige Zufallsvariable* bezeichnet. Das W-Maß $\mathbb{P}_X : \mathcal{P}(\mathbb{N}_0) \rightarrow [0, 1]$ ist die *Verteilung von X* auf ihrem *Bildraum* $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$. Weiters ist die Folge $p := \{p_n\}_{n \geq 0}$, wobei die einzelnen $p_n : \mathbb{N}_0 \rightarrow [0, 1]$ definiert sind als

$$p_n := p_X(n) \stackrel{\text{vgl. (A.53)}}{=} \mathbb{P}_X(\{n\}) \stackrel{\text{(A.39) Not. A.30}}{=} \mathbb{P}(X = n) \quad \text{für jedes } n \geq 0, \quad (2.1)$$

die *Wahrscheinlichkeitsfunktion* (kurz W-Funktion) von X .

2 Einfache Galton-Watson-Verzweigungsprozesse

Bemerkung zur Definition 2.2. Aufgrund der Bem. zur Def. 2.1 bzw. (2.1) lässt sich die Verteilung von X bereits durch die W-Funktion definieren (und umgekehrt). Beide Begrifflichkeiten können in diesem Sinne gleichwertig verwendet werden. Weiters gilt, analog zur soeben erwähnten Bemerkung und da das *Bildmaß* $\mathbb{P}_X$ gemäß der Def. 2.1 ein W-Maß auf $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$ ist,

$$\sum_{n \geq 0} p_n \stackrel{(2.1)}{=} \sum_{n \geq 0} \mathbb{P}(X = n) \stackrel{(2.1)}{=} \sum_{n \geq 0} \mathbb{P}_X(\{n\}) \stackrel{(A.50)}{=} 1. \quad (2.2)$$

△

Den *Erwartungswert*, kurz *EW*, einer $\mathbb{N}_0$ -wertigen ZV (bzw. den von Funktionen davon) erhält man mit dem nächsten Satz. Dieser ist eine direkte Konsequenz des Satzes A.46 über den EW von Funktionen diskreter Zufallsvariablen im Appendix (beachte für die Anwendung des Satzes, dass g auf der abzählbaren Menge $\mathbb{N}_0$ definiert und somit laut der Bemerkung zur Definition A.8 immer messbar ist).

Satz 2.3 (Erwartungswerte von (Funktionen von) $\mathbb{N}_0$ -wertigen ZV)

Sei X eine $\mathbb{N}_0$ -wertige Zufallsvariable mit W-Funktion $\{p_n\}_{n \geq 0}$. Weiters sei eine Funktion $g : \mathbb{N}_0 \rightarrow \bar{\mathbb{R}}$, die entweder nichtnegativ ist ($g \geq 0$) oder so, dass $\sum_{n \geq 0} |g(n)|p_n < \infty$ gilt, gegeben. Dann gilt für den Erwartungswert $\mathbb{E}[g(X)]$ der ZV $g \circ X = g(X)$ (wobei damit $(g \circ X)(\omega) = g(X)(\omega) := g(X(\omega))$ für jedes $\omega \in \Omega$ gemeint ist)

$$\mathbb{E}[g(X)] = \mathbb{E}g(X) \stackrel{(A.57)}{=} \sum_{n \geq 0} g(n)\mathbb{P}(X = n) \stackrel{(2.1)}{=} \sum_{n \geq 0} g(n)p_n. \quad (2.3)$$

Mit $g(n) = n$ für alle $n \geq 0$ ergibt das für den Erwartungswert von X

$$\mathbb{E}[X] = \mathbb{E}X = \sum_{n \geq 0} np_n. \quad (2.4)$$

Bemerkungen zum Satz 2.3. Im Sinne der Def. A.36 (und der Bem. dazu) besitzen die Zufallsvariablen X und $g(X)$, bei letzterer $g \geq 0$ vorausgesetzt, also immer einen Erwartungswert, er kann aber auch $+\infty$ sein. Gilt $\sum_{n \geq 0} |g(n)|p_n < \infty$, so ist der von $g(X)$ endlich, d.h., er existiert (selbiges gilt für $\mathbb{E}X$, wenn $\sum_{n \geq 0} np_n < \infty$ ist).

Manchmal wird der Satz auch als *Transformationsformel* (vgl. z.B. [20, 12.5, S. 82]) bezeichnet und gelegentlich kann es sich anbieten, anstelle von (2.4) die Formel

$$\mathbb{E}X = \sum_{n \geq 0} \mathbb{P}(X > n) \quad (2.5)$$

zu verwenden. Für die Herleitung siehe die Nebenrechnung (kurz NR) B.1.

△

2.2.2 Wahrscheinlichkeitserzeugende Funktionen

Es gibt mehrere Arten von sogenannten *erzeugenden Funktionen*. Angelehnt an [46, S. 8] kann man sagen, dass sie helfen, "eine Brücke zwischen diskreter Mathematik und stetiger Analysis zu schlagen". In der genannten Referenz, genauer unter [46, Def. 2.5], wird allgemein die *erzeugende Funktion einer Folge* $\{a_n\}_{n \geq 0}$ als die Potenzreihe (siehe die Def. A.102) $f(x) = \sum_{n \geq 0} a_n x^n$ definiert.

Wie im vorherigen Unterabschnitt beschrieben, kann die Verteilung bzw. W-Funktion einer $\mathbb{N}_0$ -wertigen Zufallsvariablen als eine Folge dargestellt werden (was allgemein für ZV mit Werten in diskreten Räumen gilt). Dadurch wird motiviert, diese ebenfalls mit einer erzeugenden Funktion – dann *wahrscheinlichkeitserzeugende Funktion* genannt – zu beschreiben. Somit stellen erzeugende Funktionen ein nützliches Analysewerkzeug für diskrete Zufallsvariablen dar.

Sei dafür im gesamten weiteren Verlauf dieses Unterabschnitts 2.2.2 immer X eine $\mathbb{N}_0$ -wertige Zufallsvariable mit der Verteilung $\mathbb{P}_X$ bzw. der Wahrscheinlichkeitsfunktion $p := \{p_k\}_{k \geq 0}$ wie in der Def. 2.2 (es soll als 'Laufvariable' k anstatt n verwendet werden).

Wahrscheinlichkeitserzeugende Funktionen und ihre Eigenschaften

Zunächst wird die wahrscheinlichkeitserzeugende Funktion der (Wahrscheinlichkeits-) Verteilung einer $\mathbb{N}_0$ -wertigen Zufallsvariablen definiert.

Definition 2.4 (Wahrscheinlichkeitserzeugende Funktion – vgl. [15, S. 110/112])

Allgemein wird für eine Verteilung $\mathbb{P}$ auf $(\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$ die *wahrscheinlichkeitserzeugende Funktion* $f_{\mathbb{P}} : [0, 1] \rightarrow [0, 1]$ als $f_{\mathbb{P}}(s) := \sum_{k \geq 0} \mathbb{P}(\{k\}) s^k$ definiert.

Die zu der $\mathbb{N}_0$ -wertigen Zufallsvariablen X mit W-Funktion $p = \{p_k\}_{k \geq 0}$ gehörige Funktion $f_X : [0, 1] \rightarrow [0, 1]$ mit

$$f_X(s) := f_{\mathbb{P}_X}(s) = \sum_{k \geq 0} \mathbb{P}_X(\{k\}) s^k \stackrel{(2.1)}{=} \sum_{k \geq 0} \mathbb{P}(X = k) s^k \stackrel{(2.1)}{=} \sum_{k \geq 0} p_k s^k \stackrel{(2.3)}{=} \mathbb{E} s^X \quad (2.6)$$

heißt *wahrscheinlichkeitserzeugende Funktion (kurz weF) von X* oder *von $\mathbb{P}_X$* . Wenn klar ist, dass die weF von X gemeint ist, wird oft auch nur f für f_X geschrieben.

Wenngleich die wahrscheinlichkeitserzeugenden Funktionen grundsätzlich auf ganz $\mathbb{R}$ (oder sogar $\mathbb{C}$ – vgl. etwa [19, I.2.2, S. 5]) definiert sind, werden im nächsten Satz grundlegende nützliche Eigenschaften einer weF auf dem Intervall $[0, 1]$ (der Def. 2.4 entsprechend) untersucht.

Satz 2.5 (Eigenschaften einer weF als Summenfunktion einer Potenzreihe)

Für die Funktion $f : [0, 1] \rightarrow [0, 1]$ mit $f(s) = \sum_{k \geq 0} p_k s^k$ der Def. 2.4 gelten auf dem Intervall $[0, 1]$ die folgenden Eigenschaften.

- (a) Der Konvergenzradius R (Def. A.102) von f (auf $\mathbb{R}$) ist mindestens 1, d.h. $R \geq 1$.
- (b) Stetige Differenzierbarkeit: f ist mindestens auf $[0, 1)$ beliebig oft stetig differenzierbar und alle Ableitungen können durch gliedweises Differenzieren erhalten werden.

2 Einfache Galton-Watson-Verzweigungsprozesse

D.h., für $s \in [0, 1]$ gilt bspw. für die ersten beiden Ableitungen $f'(s) = \sum_{k \geq 1} k p_k s^{k-1}$ bzw. $f''(s) = \sum_{k \geq 2} k(k-1) p_k s^{k-2}$. Allgemein für $l \geq 0$ gilt

$$f^{(l)}(s) = \sum_{k \geq 0} k(k-1)(k-2) \cdots (k-l+1) p_k s^{k-l} = \sum_{k \geq l} \frac{k!}{(k-l)!} p_k s^{k-l}. \quad (2.7)$$

Die Summe kann auch im mittleren Ausdruck erst bei l beginnen, denn die Summanden von $k = 0$ bis $k = l-1$ sind ohnehin 0. Der Konvergenzradius (siehe (a)) bleibt für alle Ableitungen $f^{(l)}$ gleich dem von f .

(c) Die Randpunkte von f auf $[0, 1]$ sind wohldefiniert. Es gilt

$$f(1) = 1 \quad \text{und} \quad (2.8)$$

$$f(0) = p_0. \quad (2.9)$$

(d) Auch wenn der Konvergenzradius nur 1 ist, kann die bereits in (b) festgestellte Stetigkeit von f (auf $[0, 1]$) auf den rechten Randpunkt des Konvergenzintervalls erweitert werden und f ist somit auf dem gesamten Intervall $[0, 1]$ stetig.

(e) f ist monoton wachsend (isoton) und konvex (auf $[0, 1]$). Zudem gilt:

- Falls es mindestens ein $k \geq 1$ gibt, für das $p_k > 0$ ist, ist f strikt isoton. Dies ist genau dann erfüllt, wenn $p_0 < 1$ ist.
- Falls es mindestens ein $k \geq 2$ gibt, für das $p_k > 0$ ist, ist f strikt konvex und f' strikt isoton. Dies ist genau dann erfüllt, wenn $p_0 + p_1 < 1$ ist.

Beweis. Für einige Eigenschaften kann ausgenutzt werden, dass es sich bei f um eine Potenzreihe gemäß der Definition A.102 (mit k statt n und $a_k = p_k$ bzw. $x = s$) handelt.

(a) Es soll der Kehrwert von (A.110) aus der Konvergenzradius-Def. verwendet werden:

$$\frac{1}{R} \stackrel{(A.110)}{=} \limsup_{k \rightarrow \infty} \sqrt[k]{|p_k|} \stackrel{p_k \geq 0}{=} \limsup_{k \rightarrow \infty} \underbrace{\sqrt[k]{p_k}}_{\leq 1 \quad \forall k \geq 0} \leq 1 \quad \Rightarrow R \geq 1.$$

(b) Diese Aussagen folgen aus dem Differenzierbarkeitssatz für Potenzreihen (Satz A.103 (c)). Dieser erlaubt eben das gliedweise Differenzieren, was mit den üblichen Ableitungsregeln und $f^{(l)}(s) = \sum_{k \geq 0} \left(\frac{d^l}{ds^l} p_k s^k\right)$ selbsterklärend zu (2.7) führt.

(c) Den rechten Randpunkt erhält man direkt mit $f(1) \stackrel{(2.6)}{=} \sum_{k \geq 0} p_k \overbrace{1^k}^{=1 \quad \forall k \geq 0} \stackrel{(2.2)}{=} 1$.

Mit der oft praktikablen Konvention $0^0 := 1$ kann genauso direkt der linke Randpunkt bestimmt oder wegen der rechtsseitigen Stetigkeit in 0 argumentiert werden, dass $f(0) \stackrel{\text{stetig}}{\stackrel{(b)}}{=} \sum_{k \geq 0} p_k (\lim_{s \searrow 0} s)^k = \sum_{k \geq 0} p_k \underbrace{\lim_{s \searrow 0} s^k}_{=0 \text{ für } k \geq 1} = p_0 \lim_{s \searrow 0} \underbrace{s^0}_{=1 \text{ für } s > 0} = p_0$ ist.

(d) Da laut (c) $f(1) = 1$ gilt, folgt mit dem Grenzwertsatz von Abel (Satz A.103 (d)), dass $\lim_{s \nearrow 1} f(s) = 1$ und f demnach auch im Punkt $R = 1$ linksseitig stetig ist.

2.2 Weitere mathematische Vorüberlegungen und Werkzeuge

- (e) Um die Monotonie zu überprüfen genügt (wegen (b) und der Stetigkeit (d)) die Betrachtung der ersten Ableitung auf dem Inneren $(0, 1)$ (siehe den Satz A.88). Gilt dort $f' \geq 0$, so ist f isoton. Wenn sogar $f' > 0$ gilt, ist f strikt isoton. Es gilt ohnehin, dass $p_k = \mathbb{P}(X=k) \geq 0$ ist und

$$f'(s) = \sum_{(b)} \underbrace{ks^{k-1}}_{>0 \text{ für } s \in (0,1)} \underbrace{\mathbb{P}(X=k)}_{>0 \text{ für mind. ein } k, \text{ falls } p_0 < 1 \text{ ist}} \begin{cases} \geq 0 & \Rightarrow f \text{ ist jedenfalls isoton,} \\ > 0 & \Rightarrow f \text{ ist strikt isoton, falls } p_0 < 1 \text{ ist.} \end{cases}$$

Wenn f' auf einem Intervall (strikt) isoton ist, so ist f dort auch (strikt) konvex. Analog zu vorher kann das – nun mittels der zweiten Ableitung – wieder auf $(0, 1)$ (siehe den Satz A.89) für $[0, 1]$ überprüft werden.

$$f''(s) = \sum_{(b)} \underbrace{k(k-1)s^{k-2}}_{>0 \text{ für } s \in (0,1)} \underbrace{\mathbb{P}(X=k)}_{>0 \text{ für mind. ein } k, \text{ falls } p_0 + p_1 < 1 \text{ ist}} \begin{cases} \geq 0 & \Rightarrow f \text{ ist jedenfalls konvex,} \\ > 0 & \Rightarrow f \text{ ist strikt konvex, falls } p_0 + p_1 < 1 \text{ ist.} \end{cases}$$

Bemerkungen zur Definition 2.4 und dem Satz 2.5. Die durch die weF f definierte Reihe *konvergiert* auf ihrem gesamten Definitionsbereich $[0, 1]$ *absolut*. Dies folgt für den Randpunkt 1 aus (c) und für $s \in [0, 1)$ aufgrund der Nichtnegativität der Glieder etwa aus dem Majorantenkriterium für Reihen (vgl. z.B. [22, Satz 33.4, S. 204] mit $f(1)$ als Majorante). Für $s \in (-1, 1)$ würde dies auch direkt aus (a) bzw. dem Konvergenzsatz für Potenzreihen (Satz A.103 (a)) folgen. Falls f ein Polynom ist (also eine abbrechende Potenzreihe, bei welcher nur endlich viele p_k größer als 0 sind) handelt es sich um eine endliche Summe und die Reihe würde sogar für alle Werte von $s \in \mathbb{R}$ konvergieren. Das Interesse konzentriert sich aber primär auf $s \in [0, 1]$. Dass als Zielmenge ebenso dieses Intervall formuliert wurde, begründet sich durch die Punkte (c) und (e) des Satzes. Sie ist im Fall $p_0 = 0$ exakt $[0, 1]$ und man könnte sie auf $[p_0, 1]$ spezifizieren, falls $p_0 > 0$ ist.

Demnach ist s^X für $s \in [0, 1]$ auch *integrierbar* ($\mathbb{E}s^X = f(s) \Rightarrow 0 \leq \mathbb{E}s^X \leq 1$). $\triangle$
(2.6)

Die bisher im vorherigen Satz besprochenen Eigenschaften werden also kurz gesagt daraus abgeleitet, dass die wahrscheinlichkeitserzeugende Funktion f eine Potenzreihe ist, deren Koeffizienten p_k nichtnegativ sind und in Summe 1 ergeben.

Bereits aus der Definition 2.4 bzw. (2.6) geht hervor, dass die wahrscheinlichkeitserzeugende Funktion f bzw. f_X nur von der Verteilung $\mathbb{P}_X$ der $\mathbb{N}_0$ -wertigen Zufallsvariablen X abhängt und deshalb durch das Wissen über diese (bzw. die W-Funktion $\{p_k\}_{k \geq 0}$) eindeutig bestimmt ist. Dass dies aber auch umgekehrt gilt, also dass ebenso die Verteilung in eindeutiger Weise durch die weF charakterisiert ist, zeigt der nächste Satz (für beide Richtungen vgl. z.B. auch [30, Satz 3.2 (ii), S. 80]). Dieser veranschaulicht zudem, wie die Wahrscheinlichkeiten der einzelnen möglichen Werte von X aus der weF erhalten bzw. 'erzeugt' werden können, wodurch die Namensgebung 'wahrscheinlichkeitserzeugende Funktion' gerechtfertigt ist.

2 Einfache Galton-Watson-Verzweigungsprozesse

Satz 2.6 (Erzeugung und Eindeutigkeit einer weF – vgl. z.B. [42, S. 372])

Die Wahrscheinlichkeitsverteilung bzw. -funktion der $\mathbb{N}_0$ -wertigen Zufallsvariablen X kann aus ihrer weF f wie folgt bestimmt werden: Für alle $k \geq 0$ gilt

$$p_k \stackrel{(2.1)}{=} \mathbb{P}(X = k) = \frac{1}{k!} f^{(k)}(0) \quad \text{mit } f^{(k)}(0) = \left. \frac{d^k}{ds^k} f(s) \right|_{s=0} \quad (2.10)$$

Für zwei $\mathbb{N}_0$ -wertige Zufallsvariablen X und Y mit weF f_X bzw. f_Y gilt deshalb weiters

$$X \stackrel{d}{=} Y \iff \mathbb{P}_X = \mathbb{P}_Y \iff f_X = f_Y. \quad (2.11)$$

Beweis. Für $k = 0$ erhält man $p_0 = f^{(0)}(0)$ auch sofort aus dem Satz 2.5 (c) bzw. (2.9) (beachte, dass $f^{(0)} = f$ und $0! = 1$ ist). Für ein allgemeines $k \geq 0$ bzw. (2.10) betrachte die l -te Ableitung aus (2.7). Es ist

$$f^{(l)}(s) = \sum_{k \geq l} \underbrace{\frac{k!}{(k-l)!}}_{=l! \text{ für } k=l} p_k \underbrace{s^{k-l}}_{=1 \text{ für } k=l} = l! p_l + \sum_{k \geq l+1} \frac{k!}{(k-l)!} p_k \underbrace{s^{k-l}}_{=0 \text{ für } s=0 \text{ und } k \geq l+1} \Rightarrow f^{(l)}(0) = l! p_l,$$

woraus (mit $l = k$) die Behauptung folgt. Deshalb ist p_k für jedes $k \geq 0$ und somit auch $\mathbb{P}_X$ (beachte die Bem. zur Def. 2.2) eindeutig bestimmt, was die Richtung ' $\Leftarrow$ ' der zweiten Äquivalenz in (2.11) erklärt. Wie zuvor beschrieben, ist die andere Richtung durch die Def. 2.4 klar. Die erste Äquivalenz gilt definitionsgemäß sogar für alle Zufallsvariablen (vgl. die Def. A.29 (iii)). $\square$

Da die Verteilung von X also durch ihre weF eindeutig bestimmt ist, können auch ihre Momente aus f erhalten werden. Dazu wird zunächst das folgende Lemma formuliert.

Lemma 2.7 (Hilfsformel zur Bestimmung der Momente mit einer weF – angelehnt an [35, Satz 7.2 (ii), S. 107])

Sei wieder f die weF der $\mathbb{N}_0$ -wertigen ZV X und sei $l \geq 1$. Dann gilt für den linksseitigen Grenzwert gegen 1 der l -ten Ableitung von f

$$\lim_{s \nearrow 1} f^{(l)}(s) \stackrel{\text{Not. 1.1}}{=} f^{(l)}(1^-) = \mathbb{E}[X(X-1) \cdots (X-l+1)]. \quad (2.12)$$

Beide Seiten können dabei auch $+\infty$ sein. Es gilt: $f^{(l)}(1^-) < \infty \iff \mathbb{E}X^l < \infty$. In diesem Fall ist f auch in 1 l -mal (linksseitig) stetig differenzierbar und $f^{(l)}(1^-) = f^{(l)}(1)$.

Beweis. In (2.12) sind sowohl die Koeffizienten der Potenzreihe $f^{(l)}$ als auch die ZV $X(X-1) \cdots (X-l+1)$ nichtnegativ, denn es ist $k(k-1) \cdots (k-l+1) = 0$ für alle $k < l$ bzw. ≥ 0 für alle $k \geq l$ (Δ). Laut dem Satz 2.5 (b) ist die weF f mindestens auf $[0, R)$ mit $R \geq 1$ beliebig oft stetig differenzierbar. Im Allgemeinen ist es jedoch nicht sicher, ob die l -te Ableitung $f^{(l)}$ am rechten Randpunkt des Konvergenzintervalls von f existiert. Definiere deshalb eine Funktion $h : [0, 1] \rightarrow \overline{\mathbb{R}}$, $h(s) := \sum_{k \geq 0} k(k-1) \cdots (k-l+1) p_k s^{k-l}$, welche zumindest auf $[0, R)$ mit der l -ten Ableitung übereinstimmt, d.h., es gilt

$$h(s) \stackrel{(2.7)}{=} f^{(l)}(s) \text{ für alle } s \in [0, R) \quad \text{und insbesondere } h(1^-) = f^{(l)}(1^-). \quad (\star)$$

2.2 Weitere mathematische Vorüberlegungen und Werkzeuge

Weiters ist $h \geq 0$ (siehe (Δ)). Da für $s=1$ immer $s^{k-l}=1$ ist, gilt somit mit dem Satz 2.3

$$\mathbb{E}[X(X-1)\cdots(X-l+1)] \stackrel{(2.3)}{=} \sum_{k \geq 0} k(k-1)\cdots(k-l+1)p_k = h(1). \quad (\star\star)$$

Wenn $R > 1$ ist, ist h auch im Punkt 1 stetig und es gilt $f^{(l)}(1^-) = h(1^-) = h(1)$. Aber auch bei einem Konvergenzradius $R = 1$ liefert im Fall von $h(1) < \infty$ der Satz A.103 (d) (*Grenzwertsatz von Abel*) und im Fall von $h(1) = \infty$ das Korollar dazu (Ergänzung zum Grenzwertsatz für nichtnegative Koeffizienten), dass $h(1^-) = h(1)$ ist. Es gilt also immer

$$f^{(l)}(1^-) \stackrel{(\star)}{=} h(1^-) = h(1) \stackrel{(\star\star)}{=} \mathbb{E}[X(X-1)\cdots(X-l+1)], \quad (\star\star\star)$$

womit (2.12) – auch für den Fall, dass beide Seiten $+\infty$ sind – gezeigt ist.

Für die zweite Aussage des Lemmas sollen zwei Fälle in Abhängigkeit der Existenz des Erwartungswertes $\mathbb{E}X^l$ betrachtet werden (beachte, dass die Reihe $h(1)$ nur nichtnegative Glieder hat). Für $\mathbb{E}X^l < \infty$ kann $h(1)$ von oben abgeschätzt werden. Diesfalls gilt

$$h(1) \stackrel{(\star\star)}{=} \sum_{k \geq 0} \underbrace{k(k-1)\cdots(k-l+1)}_{\leq k^l \quad \forall k \geq 0 \text{ (} l \text{ Faktoren)}} p_k \leq \sum_{k \geq 0} k^l p_k \stackrel{(2.3)}{=} \mathbb{E}X^l < \infty.$$

Sei nun $\mathbb{E}X^l = \infty$. Es gilt $\mathbb{E}X^l = \sum_{k \geq 0} k^l p_k = \underbrace{\sum_{k=0}^{2l-1} k^l p_k}_{< \infty \text{ (endl. } \Sigma)} + \sum_{k \geq 2l} k^l p_k$, woraus dann $\sum_{k \geq 2l} k^l p_k = \infty$ folgt. Eine Abschätzung von unten für $h(1)$ liefert für diesen Fall

$$h(1) \geq \sum_{k \geq 2l} \underbrace{k(k-1)\cdots(k-l+1)}_{\geq \left(\frac{k}{2}\right)^l \text{ für } k \geq 2l} p_k \geq \underbrace{\left(\frac{1}{2}\right)^l}_{\text{konst.}} \underbrace{\sum_{k \geq 2l} k^l p_k}_{=\infty} = \infty \quad \Rightarrow h(1) = \infty.$$

Es folgt also $h(1) < \infty$ aus $\mathbb{E}X^l < \infty$. Andererseits impliziert $\mathbb{E}X^l = \infty$, dass $h(1) = \infty$ ist. Mit $(\star\star\star)$ heißt das: $\mathbb{E}X^l < \infty \iff f^{(l)}(1^-) = h(1) < \infty$. Dann ist aber laut dem Satz A.103 (e) (bzw. der Bemerkung dazu) f auch in 1 l -mal (linksseitig) differenzierbar und die l -te Ableitung ist dort $f^{(l)}(1) = h(1) < \infty$. Das führt (erneut mit $(\star\star\star)$) zu $f^{(l)}(1^-) = f^{(l)}(1)$ bzw. der letzten Gleichung des Lemmas. $\square$

Bemerkung zum Lemma 2.7. Gemäß der Aussage des Lemmas ist f , wenn $\mathbb{E}X^l < \infty$ gilt, auf ihrem gesamten Definitionsbereich $[0, 1]$ l -mal stetig differenzierbar und auch an der Stelle 1 gilt (2.7) für ihre l -te Ableitung – vgl. dazu auch [42, Satz 16.1.5, S. 373]. Dort wird eine gleichwertige Aussage über den als *Binomial-Moment der Ordnung l von X* bezeichneten EW des Binomialkoeffizienten $\binom{X}{l}$ getätigt. Dieses kann (für $k, l \geq 0$ gilt $\binom{k}{l} = \frac{k!}{(k-l)!l!}$, wenn $k \geq l$ ist und $\binom{k}{l} = 0$, wenn $k < l$ ist) geschrieben werden als

$$\mathbb{E}\left[\binom{X}{l}\right] = \mathbb{E}\left[\frac{X!}{(X-l)!l!} \mathbb{1}_{\{X \geq l\}}\right] = \frac{1}{l!} \mathbb{E}[X(X-1)\cdots(X-l+1)] \stackrel{(2.12)}{=} \frac{1}{l!} f^{(l)}(1^-). \quad \Delta$$

2 Einfache Galton-Watson-Verzweigungsprozesse

Existiert der Erwartungswert von X , also das erste Moment, so könnte er bereits durch $f'(1^-) = \sum_{k \geq 1} k p_k \cdot 1^{k-1} = \mathbb{E}X$ gewonnen werden. Der *Grenzwertsatz von Abel*, welcher ja auch im Beweis des Lemmas verwendet wurde, rechtfertigt dies (analog könnte man das zweite Moment durch $f''(1^-) = \sum_{k \geq 2} k(k-1)p_k \cdot 1^{k-2} = \mathbb{E}X^2 - \mathbb{E}X$ bestimmen – vorausgesetzt es existiert).

Mit der Hilfsformel (2.12) können aber auch die höheren Momente rekursiv bestimmt werden. So ergibt sie mit $l = 1$ (auch wenn $\mathbb{E}X = +\infty$ gilt) direkt $\mathbb{E}X = f'(1^-)$. Falls $\mathbb{E}X < \infty$ ist, gilt mit $l = 2$

$$f''(1^-) \stackrel{(2.12)}{=} \mathbb{E}[X(X-1)] = \mathbb{E}X^2 - \overbrace{\mathbb{E}X}^{=f'(1^-)} \Rightarrow \mathbb{E}X^2 = f''(1^-) + f'(1^-) \quad (\star)$$

und mit $l = 3$, $\mathbb{E}X^2 < \infty$ vorausgesetzt,

$$\begin{aligned} f^{(3)}(1^-) &= \mathbb{E}[X(X-1)(X-2)] = \mathbb{E}X^3 - \overbrace{3\mathbb{E}X^2}^{=f''(1^-)+f'(1^-)} + \overbrace{2\mathbb{E}X}^{=f'(1^-)} \\ &\Rightarrow \mathbb{E}X^3 = f^{(3)}(1^-) + 3f''(1^-) + f'(1^-). \end{aligned}$$

Für die höheren Momente $\mathbb{E}X^l$ kann analog vorgegangen werden – immer $\mathbb{E}X^{l-1} < \infty$ vorausgesetzt. Falls diese Voraussetzung nicht erfüllt und $\mathbb{E}X^{l-1} = +\infty$ ist (X ist nicht-negativ), sind aber ohnehin alle $\mathbb{E}X^k$ mit $k \geq l$ ebenfalls $+\infty$ (siehe das Lemma A.38).

Zudem stellt das Lemma 2.7 anhand der Existenz der höheren Momente ein allgemeineres Kriterium auf, wann der linksseitige Grenzwert $f^{(l)}(1^-)$ verwendet werden muss und wann direkt (2.7) an der Stelle 1 ausgewertet werden darf.

Der EW und die Varianz, ausgedrückt durch ihre weF und ermittelt unter Zuhilfenahme dieses Lemmas, werden im folgenden Korollar zusammengefasst dargestellt.

Korollar 2.8 (Darstellung des Erwartungswertes und der Varianz durch weF)

Für den Erwartungswert der $\mathbb{N}_0$ -wertigen ZV X mit der weF f gilt

$$\mathbb{E}X \stackrel{(2.12)}{=} f'(1^-) \stackrel[\mathbb{E}X < \infty \text{ ist}]{\text{falls}} f'(1) \quad (2.13)$$

und für die Varianz (siehe ggf. die Def. A.53), falls $\mathbb{E}X < \infty$ ist,

$$\begin{aligned} \mathbb{V}[X] &= \mathbb{V}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{E}X^2 - \mathbb{E}X + \mathbb{E}X - (\mathbb{E}X)^2 \\ &\stackrel{(2.12)}{=} f''(1^-) + \mathbb{E}X - (\mathbb{E}X)^2 \stackrel{(2.13)}{=} f''(1^-) + f'(1) - (f'(1))^2 \\ &\stackrel[\mathbb{E}X^2 < \infty \text{ ist}]{\text{falls}} f''(1) + f'(1) - (f'(1))^2. \end{aligned} \quad (2.14)$$

Eine lineare Transformation von X mit passenden Parametern stellt wiederum eine $\mathbb{N}_0$ -wertige Zufallsvariable dar, deren weF im nächsten Lemma präsentiert wird.

Lemma 2.9 (WeF der linearen Transformation einer $\mathbb{N}_0$ -wertigen ZV)

Wird X (mit der weF f_X) durch $a, b \in \mathbb{N}_0$ linear zu der neuen $\mathbb{N}_0$ -wertigen Zufallsvariablen $Y := a + bX$ transformiert, dann gilt für die dazugehörige weF $f_Y : [0, 1] \rightarrow [0, 1]$

$$f_Y(s) = f_{a+bX}(s) = s^a f_X(s^b) \quad \text{für alle } s \in [0, 1]. \quad 'b \text{ ist ein Teiler von } (k-a)'$$

Weiters ist für alle $k \geq 0$, falls $b \neq 0$ ist, $\mathbb{P}(Y=k) = \begin{cases} \mathbb{P}(X = \frac{k-a}{b}) & \forall k \geq a : \overbrace{b|(k-a)} \\ 0 & \text{sonst.} \end{cases}$
Für $b = 0$ ist $\mathbb{P}(Y=a) = 1$.

2.2 Weitere mathematische Vorüberlegungen und Werkzeuge

Beweis. Für alle $s \in [0, 1]$ ist $f_{a+bX}(s) \stackrel{(2.6)}{=} \mathbb{E}[s^{a+bX}] = s^a \mathbb{E}[(s^b)^X] \stackrel{(2.6)}{=} s^a f_X(s^b)$. Für die zweite Aussage betrachte (mit $b \neq 0$) $\mathbb{P}(Y = k) = \mathbb{P}(a + bX = k) = \mathbb{P}(X = \frac{k-a}{b})$. $\square$

Mit $b = 1$ und $a \in \mathbb{N}_0$ kann die besprochene Lineartransformation $X + a$ auch als die Summe von X und einer deterministischen (also nichtzufälligen bzw. konstanten) 'Zufallsvariablen' $Y \equiv a$ aufgefasst werden, welche unabhängig von X ist. Für $s \in [0, 1]$ ist die weF von Y dann $f_Y(s) = \mathbb{E}s^a = s^a$. Laut dem vorherigen Lemma 2.9 gilt somit $f_{X+Y}(s) = f_{a+X}(s) = s^a f_X(s) = f_Y(s)f_X(s)$ für die weF von $X + Y$.

Dieser Zusammenhang, dass die Summation der unabhängigen, $\mathbb{N}_0$ -wertigen ZV X und Y auf der Ebene der wahrscheinlichkeits erzeugenden Funktionen in eine Multiplikation übergeht, gilt aber allgemeiner. Im Satz A.48 im Appendix A.2.1 wird gezeigt, wie die Verteilung der Summe zweier unabhängiger, $\mathbb{N}_0$ -wertiger ZV mittels der sogenannten *Faltung* der einzelnen Verteilungen (siehe die Def. im Satz bzw. auch die Def. A.98) bestimmt werden kann. Die Durchführung von Faltungen kann jedoch aufwendig sein und es ist u.U. wesentlich einfacher, mit den weF zu arbeiten. Wie die weF der Summe solcher ZV durch die Faltung ihrer Verteilungen bestimmt werden kann, wird im Korollar zum Satz A.48 demonstriert – dieses verdeutlicht auch, dass sie sich als Produkt der beiden ursprünglichen weF äußert.

Ein analoges Resultat für mehrere unabhängige, $\mathbb{N}_0$ -wertige ZV, dessen Beweis sich nicht der Reihendefinition der weF bedient, sondern mit den Erwartungswerten operiert, liefert das nächste Lemma (welches auch durch wiederholtes Anwenden des genannten Korollars bzw. (A.61) bewiesen werden könnte).

Lemma 2.10 (WeF der Summe von unabhängigen, $\mathbb{N}_0$ -wertigen ZV – vgl. z.B. [30, Satz 3.3, S. 80])

Seien $n \geq 1$ und $\{X_i\}_{i=1}^n$ eine Familie von n unabhängigen, $\mathbb{N}_0$ -wertigen ZV. Für die weF $f_{X_1+\dots+X_n} : [0, 1] \rightarrow [0, 1]$ der Summe $X_1 + \dots + X_n$ gilt

$$f_{X_1+\dots+X_n} = f_{\sum_{i=1}^n X_i} = \prod_{i=1}^n f_{X_i}. \quad (2.15)$$

Beweis. Sei $s \in [0, 1]$. Laut dem Lemma A.34 über die Funktionen unabhängiger ZV folgt aus der Unabhängigkeit der $\{X_i\}_{i=1}^n$ auch die Unabhängigkeit der Familie $\{s^{X_i}\}_{i=1}^n$ ($x \mapsto s^x$ ist messbar, da die Funktion auf der abzählbaren Menge $\mathbb{N}_0$ definiert ist – siehe die Bem. zur Def. A.8). Zudem sind, wie in den Bem. zum Satz 2.5 erläutert wurde, alle s^{X_i} integrierbar. Deshalb gilt mit dem Produktsatz für Erwartungswerte (Satz A.52)

$$f_{\sum_{i=1}^n X_i}(s) \stackrel{(2.6)}{=} \mathbb{E}[s^{\sum_{i=1}^n X_i}] = \mathbb{E}\left[\prod_{i=1}^n s^{X_i}\right] \stackrel{s^{X_i} \text{ uabh.}}{\stackrel{(A.66)}}{=} \prod_{i=1}^n \mathbb{E}s^{X_i} \stackrel{(2.6)}{=} \prod_{i=1}^n f_{X_i}(s). \quad \square$$

Etwas schwieriger wird es, wenn die Anzahl der aufzusummierenden Zufallsvariablen n ebenfalls zufällig ist. Der nun folgende Teilabschnitt (dieses Unterabschnitts) beantwortet diese Frage, wenn die X_i nicht nur unabhängig, sondern zudem *identisch verteilt* sind (also *i.i.d.* – siehe die Definition A.29 (iii)).

Summe einer zufälligen Anzahl an $\mathbb{N}_0$ -wertigen *i.i.d.*-Zufallsvariablen

Zunächst soll die wahrscheinlichkeitserzeugende Funktion der Summe einer zufälligen Anzahl an $\mathbb{N}_0$ -wertigen *i.i.d.*-Zufallsvariablen hergeleitet werden. Anhand ihr können dann die Momente dieser Summe ermittelt werden. Diese Herangehensweise ist grob angelehnt an [35, S. 110, f.]. Anschließend werden die analogen Resultate nochmals mithilfe von bedingten Erwartungswerten (vgl. ggf. A.2.3) erarbeitet und in Sätzen, welche teilweise auch etwas allgemeiner gelten, formuliert. Betrachte dafür im aktuellen Teilabschnitt stets die Zufallsvariablen, welche gleich definiert werden.

Definition 2.11. Sei $\{X_i\}_{i \geq 1}$ eine Familie unabhängiger, identisch verteilter (*i.i.d.*-) Kopien der $\mathbb{N}_0$ -wertigen Zufallsvariablen X , alle mit der weF $f_X(s)$. Es gilt also für alle $i \geq 1$ (etwa gemäß (2.11) aus dem Satz 2.6)

$$X_i \stackrel{d}{=} X \quad \text{und} \quad f_{X_i} = f_X \quad \xRightarrow{(A.46)} \quad \mathbb{E}X_i = \mathbb{E}X \quad \text{und} \quad \mathbb{V}X_i = \mathbb{V}X. \quad (2.16)$$

Weiters sei N ebenso eine $\mathbb{N}_0$ -wertige Zufallsvariable, die unabhängig von allen $\{X_i\}_{i \geq 1}$ ist und die weF $f_N(s)$ besitzt.

Definiere zunächst für die Summe der ersten $n \geq 0$ der X_i die $\mathbb{N}_0$ wertige ZV $S_n := \sum_{i=1}^n X_i$, wobei $S_0 = 0$ ist ('leere Summe' – siehe die Not. 1.1). Für ihre weF gilt

$$f_{S_n} = f_{\sum_{i=1}^n X_i} \stackrel{(2.15)}{=} \prod_{i=1}^n f_{X_i} \stackrel{(2.16)}{=} (f_X)^n \quad \left(\text{insbes. } \prod_{i=1}^0 f_{X_i} \stackrel{\text{leeres Produkt}}{=} 1 = f_{S_0} \right). \quad (2.17)$$

Dann ist auch S_N eine $\mathbb{N}_0$ -wertige Zufallsvariable (also $S_N(\omega) := S_{N(\omega)}(\omega) \forall \omega \in \Omega$), und zwar die Summe einer zufälligen Anzahl N dieser Kopien von X . D.h., dass $S_N = \sum_{i=1}^N X_i = X_1 + X_2 + \dots + X_N$ für $N > 0$ ist. Für $N = 0$ ist $S_N = 0$.

Betrachte nun für ein beliebiges $k \geq 0$ das Ereignis $\{S_N = k\}$. Für das Eintreten dieses Ereignisses müssen die Ereignisse $\{S_n = k\}$ und $\{N = n\}$ für irgendein $n \geq 0$ simultan eintreten, d.h.

$$\begin{aligned} \{S_N = k\} &= (\{S_0 = k\} \cap \{N = 0\}) \cup (\{S_1 = k\} \cap \{N = 1\}) \cup \dots \\ &= \bigcup_{n \geq 0} \{S_n = k\} \cap \{N = n\}. \end{aligned} \quad (2.18)$$

Die Ereignisse in der Vereinigung des letzten Ausdrucks sind dabei für unterschiedliche n paarweise disjunkt. Weiters folgt aus der Unabhängigkeit der $\{X_i\}_{i \geq 1}$ und N jene von S_n und N . Deshalb gilt

$$\begin{aligned} \mathbb{P}(S_N = k) &\stackrel{(2.18)}{=} \mathbb{P}\left(\bigcup_{n \geq 0} \{S_n = k\} \cap \{N = n\}\right) \\ &\stackrel{\sigma\text{-Add.}}{\stackrel{(A.27)}{=}} \sum_{n \geq 0} \mathbb{P}(S_n = k, N = n) \stackrel{N, S_n \text{ uabh.}}{\stackrel{(A.55)}{=}} \sum_{n \geq 0} \mathbb{P}(S_n = k) \mathbb{P}(N = n). \end{aligned} \quad (2.19)$$

2.2 Weitere mathematische Vorüberlegungen und Werkzeuge

Anschließend kann mit der Reihendefinition der weF operiert werden, um jene der Zufallsvariablen S_N , d.h. f_{S_N} , zu ermitteln. Für $s \in [0, 1]$ ist

$$\begin{aligned}
 f_{S_N}(s) &\stackrel{(2.6)}{=} \sum_{k \geq 0} \mathbb{P}(S_N = k) s^k \stackrel{(2.19)}{=} \sum_{k \geq 0} \left[\sum_{n \geq 0} \mathbb{P}(S_n = k) \mathbb{P}(N = n) \right] s^k \\
 &\stackrel{(\star)}{=} \sum_{n \geq 0} \mathbb{P}(N = n) \underbrace{\left[\sum_{k \geq 0} \mathbb{P}(S_n = k) s^k \right]}_{\stackrel{(2.6)}{=} f_{S_n}(s) \stackrel{(2.17)}{=} (f_X(s))^n} = \sum_{n \geq 0} \mathbb{P}(N = n) (f_X(s))^n \\
 &\stackrel{(2.6)}{=} f_N(f_X(s)).
 \end{aligned} \tag{2.20}$$

($\star$) Diese Umformung ist aufgrund der absoluten Konvergenz (siehe die Bem. zum Satz 2.5) der weF f_{S_N} , betrachtet als Reihe, erlaubt (vgl. dazu den Unterabschnitt A.3.2). Insbesondere das Vertauschen der Summen ist eine Anwendung des sogenannten *Doppelreihensatzes von Cauchy* (Satz A.101).

Eigentlich ist auch bereits die Nichtnegativität der Reihenglieder dafür hinreichend (vgl. insbesondere auch die Bem. zum Doppelreihensatz A.101 (ii)).

Nun sollen der Erwartungswert und die Varianz der Zufallsvariablen S_N bestimmt werden. Da deren weF f_{S_N} , beschrieben durch jene von X und N (also f_X bzw. f_N), mittlerweile bekannt ist, kann dafür das Lemma 2.7 bzw. das Korollar 2.8 herangezogen werden. Mit der Ketten- und der Produktregel gilt für die ersten beiden Ableitungen von $f_{S_N}(s) \stackrel{(2.20)}{=} f_N(f_X(s))$ (mindestens) auf $[0, 1]$

$$\begin{aligned}
 f'_{S_N}(s) &= f'_N(f_X(s)) f'_X(s) \quad \text{und} \\
 f''_{S_N}(s) &= f''_N(f_X(s)) (f'_X(s))^2 + f'_N(f_X(s)) f''_X(s).
 \end{aligned} \tag{2.21}$$

Zunächst werden die linksseitigen Grenzwerte gegen 1 der vorkommenden Ausdrücke auf den rechten Seiten einzeln bestimmt. Mit den üblichen Grenzwertsätzen für Funktionen können anschließend die auf den linken Seiten erhalten werden.

Es gilt $f'_X(1^-) \stackrel{(2.13)}{=} \mathbb{E}X$ und $f''_X(1^-) \stackrel{(2.12)}{=} \mathbb{E}[X^2 - X]$. Für die Bestimmung der Grenzwerte der restlichen Ausdrücke beachte zunächst, dass f_X isoton und an der Stelle 1 linksseitig stetig ist (Satz 2.5). Wenn s von unten gegen 1 strebt, strebt deswegen auch $f_X(s)$ von unten gegen $f_X(1)$. Kurz, aus $s \nearrow 1$ folgt $t := f_X(s) \nearrow f_X(1) \stackrel{(2.8)}{=} 1$ und deshalb gilt

$$\begin{aligned}
 f'_N(f_X(1^-)) &= \lim_{s \nearrow 1} f'_N(f_X(s)) = \lim_{t \nearrow 1} f'_N(t) = f'_N(1^-) \stackrel{(2.13)}{=} \mathbb{E}N \quad \text{und} \\
 f''_N(f_X(1^-)) &= \lim_{s \nearrow 1} f''_N(f_X(s)) = \lim_{t \nearrow 1} f''_N(t) = f''_N(1^-) \stackrel{(2.12)}{=} \mathbb{E}[N^2 - N].
 \end{aligned}$$

Zusammengenommen ergibt sich, wenn sowohl $\mathbb{E}N < \infty$ als auch $\mathbb{E}X < \infty$ ist,

$$f'_{S_N}(1^-) \stackrel{(2.21)}{=} \mathbb{E}N \mathbb{E}X \quad \Rightarrow \quad \mathbb{E}S_N \stackrel{(2.13)}{=} f'_{S_N}(1^-) = \mathbb{E}N \mathbb{E}X \tag{2.22}$$

2 Einfache Galton-Watson-Verzweigungsprozesse

und, wenn $\mathbb{E}N^2 < \infty$ und $\mathbb{E}X^2 < \infty$ erfüllt sind,

$$\begin{aligned}
 f''_{S_N}(1^-) &\stackrel{(2.21)}{=} (\mathbb{E}N^2 - \mathbb{E}N)(\mathbb{E}X)^2 + \mathbb{E}N(\mathbb{E}X^2 - \mathbb{E}X) & (\star) \\
 \Rightarrow \mathbb{V}S_N &\stackrel{(2.14)}{=} f''_{S_N}(1^-) + f'_{S_N}(1^-) - (f'_{S_N}(1^-))^2 = f''_{S_N}(1^-) + \mathbb{E}N\mathbb{E}X - (\mathbb{E}N)^2(\mathbb{E}X)^2 \\
 &\stackrel{(\star)}{=} \mathbb{E}N(\mathbb{E}X^2 - (\mathbb{E}X)^2) + (\mathbb{E}X)^2(\mathbb{E}N^2 - (\mathbb{E}N)^2) = \mathbb{E}N\mathbb{V}X + (\mathbb{E}X)^2\mathbb{V}N. & (2.23)
 \end{aligned}$$

Bemerkung 2.12. Für den Erwartungswert $\mathbb{E}S_N$ müsste dabei eigentlich weder die Integrierbarkeit von X noch die von N gefordert werden. Um alle Fälle abzudecken, sollte dann aber – genau genommen – zusätzlich $0 \cdot \infty = \infty \cdot 0 = 0$ gesetzt werden (vgl. [35, S. 111]). Dann wäre, etwa wenn $\mathbb{E}X = 0$ und/oder $\mathbb{E}N = 0$ ist, auch $\mathbb{E}S_N = 0$. Wenn beide ungleich 0 sind sowie mindestens einer der beiden $+\infty$ ist, so würde $\mathbb{E}S_N = +\infty$ gelten. Ähnliches würde auf die Varianz $\mathbb{V}S_N$ (in Abhängigkeit von $\mathbb{E}X^2$ und $\mathbb{E}N^2$) zutreffen. Der Einfachheit halber werden die Integrierbarkeiten aber wie oben angeführt vorausgesetzt. $\triangle$

Nun werden die selben Resultate für f_{S_N} und $\mathbb{E}S_N$ bzw. $\mathbb{V}S_N$ alternativ durch Überlegungen, die sich der bedingten Erwartung bedienen, erarbeitet. Dafür wird zunächst das folgende Hilfsresultat formuliert.

Lemma 2.13. *Sei alles wie in der Definition 2.11.*

(i) *Für den Erwartungswert und die Varianz der Summe S_n gilt*

$$\mathbb{E}S_n = n\mathbb{E}X \quad \text{bzw.} \quad \mathbb{V}S_n = n\mathbb{V}X. \quad (2.24)$$

(ii) *Seien $h : \mathbb{N}_0 \rightarrow \mathbb{R}_+$ und $n \geq 0$ so, dass $\mathbb{P}(N = n) > 0$ ist. Die bedingte Verteilung von $h(S_N)$ gegeben $N = n$ entspricht dann jener von $h(S_n)$ und es gilt somit*

$$\mathbb{E}[h(S_N)|N = n] = \mathbb{E}h(S_n). \quad (2.25)$$

Beweis.

- (i) Da $\{X_i\}_{i \geq 1}$ Kopien von X sind, ist $\mathbb{E}S_n = \mathbb{E}[\sum_{i=1}^n X_i] = \sum_{i=1}^n \overbrace{\mathbb{E}X_i}^{=\mathbb{E}X \text{ (2.16)}} = n\mathbb{E}X$. Genauso gilt (siehe wieder (2.16)) $\mathbb{V}X_i = \mathbb{V}X$ für alle $i \geq 0$. Da diese Kopien auch *i.i.d.* sind, ergibt die sogenannte *Bienaymé-Gleichung* (Satz A.55 bzw. (A.68)) aufgrund der Unabhängigkeit, dass $\mathbb{V}S_n = \mathbb{V}[\sum_{i=1}^n X_i] = \sum_{i=1}^n \mathbb{V}X_i = n\mathbb{V}X$ ist.
- (ii) Wiewohl diese Aussage auch intuitiv plausibel ist, kann man sie wie folgt formal zeigen. Beachte zuerst für ein $k \geq 0$, dass

$$\begin{aligned}
 \{S_N = k\} \cap \{N = n\} &\stackrel{(2.18)}{=} \left(\bigcup_{m \geq 0} \{S_m = k\} \cap \{N = m\} \right) \cap \{N = n\} \\
 &\stackrel{\text{Distr.-gesetz}}{=} \bigcup_{m \geq 0} \{S_m = k\} \cap \overbrace{\{N = m\} \cap \{N = n\}}^{=\emptyset \text{ für } m \neq n} = \{S_n = k\} \cap \{N = n\} & (\star)
 \end{aligned}$$

2.2 Weitere mathematische Vorüberlegungen und Werkzeuge

gilt. Wie bereits für (2.19) argumentiert, gilt aufgrund der Unabhängigkeit von S_n und N , dass $\mathbb{P}(S_n = k, N = n) = \mathbb{P}(S_n = k)\mathbb{P}(N = n)$ ist ($\star\star$). Deshalb ist $\mathbb{P}(S_N = k|N = n) \stackrel{(A.69)}{=} \frac{\mathbb{P}(S_n=k, N=n)}{\mathbb{P}(N=n)} \stackrel{(\star)}{=} \frac{\mathbb{P}(S_n=k, N=n)}{\mathbb{P}(N=n)} \stackrel{(\star\star)}{=} \mathbb{P}(S_n = k)$. Da dies für alle $k \geq 0$ gilt, sind damit die bedingte Verteilung von S_N gegeben $N = n$ (siehe die Bem. A.58) und die Verteilung von S_n (gemäß der Bem. zur Def. 2.1) eindeutig bestimmt und ident. Es gilt also $\mathbb{P}_{S_N|N=n} = \mathbb{P}_{S_n}$. Mit (A.46) folgt daraus $\mathbb{E}[h(S_N)|N=n] = \mathbb{E}h(S_n)$ (siehe dazu auch die allgemeinere NR B.2). $\square$

Es kann also ein Zusammenhang zwischen der *bedingten Erwartung von S_N gegeben des Ereignisses $\{N = n\}$* zur ('normalen') *Erwartung von S_n* hergestellt werden (mit dem Punkt (ii) des Lemmas, formuliert über die Funktion h dieser Summen). Bei ersterer handelt es sich um eine zufällige Anzahl an $\mathbb{N}_0$ -wertigen *i.i.d.*-ZV, wohingegen bei letzterer die Anzahl der aufsummierten ZV nicht mehr zufällig ist.

Wird die bedingte Erwartung von S_N oder $h(S_N)$ gegeben der $\mathbb{N}_0$ -wertigen ZV N wie im Appendix A.2.3 beschrieben formuliert (genauer gemäß der Def. A.66), gelten diverse Eigenschaften (vgl. ggf. den Punkt (i) der Bem. zur Def. A.66 bzw. dem Lemma A.67). Einige davon werden im Satz A.61 präsentiert. Insbesondere die Formel vom *totalen Erwartungswert* (Punkt (b) des Satzes) soll nun ausgenützt werden.

Satz 2.14 (WeF der Summe einer zufälligen Anzahl an $\mathbb{N}_0$ -wertigen *i.i.d.*-ZV – angelehnt an [17, 5.1 (25), S. 153])

Sei alles wie in der Def. 2.11. Für die weF der Summe $S_N = \sum_{i=1}^N X_i$ gilt

$$f_{S_N} = f_N \circ f_X, \quad \text{d.h., für } s \in [0, 1] \text{ ist } f_{S_N}(s) = f_N(f_X(s)). \quad (2.26)$$

Beweis. Sei $s \in [0, 1]$. Definiere eine Funktion $g : \mathbb{N}_0 \rightarrow \mathbb{R}_+$ als

$$g(n) := \begin{cases} \mathbb{E}[s^{S_N}|N = n] & \stackrel{(2.25)}{=} \mathbb{E}s^{S_n} & \stackrel{(2.6)}{=} f_{S_n}(s) & \stackrel{(2.17)}{=} (f_X(s))^n & \text{für } \mathbb{P}(N=n) > 0, \\ 0 & & & & \text{für } \mathbb{P}(N=n) = 0, \end{cases} \quad (\star)$$

wobei der Wert im zweiten Fall unerheblich ist, da es sich um eine $\mathbb{P}$ -Nullmenge handelt (siehe den Punkt (ii) der Bem. zur Def. A.66 bzw. dem Lemma A.67). Für die bedingte Erwartung von s^{S_N} gegeben N setze weiters $\mathbb{E}[s^{S_N}|N] : \Omega \rightarrow \mathbb{R}_+$ (da $s^{S_N} \geq 0$ ist) als

$$\mathbb{E}[s^{S_N}|N] := g(N) \stackrel{(\star)}{=} (f_X(s))^N. \quad (\star\star)$$

Für den EW dieser bedingten Erwartung gilt (wie oben beschrieben) der Satz A.61 (b) über den totalen Erwartungswert, womit man erhält, dass

$$f_{S_N}(s) \stackrel{(2.6)}{=} \mathbb{E}s^{S_N} \stackrel{\text{tot. EW}}{\stackrel{(A.74)}{=}} \mathbb{E}[\mathbb{E}[s^{S_N}|N]] \stackrel{(\star\star)}{=} \mathbb{E}[\underbrace{(f_X(s))^N}_{\in [0,1]}] \stackrel{(2.6)}{=} f_N(f_X(s)) \quad \text{ist.} \quad \square$$

Bemerkung zum Beweis des Satzes 2.14. Die Vorgehensweise im Beweis wird wiederholt in dieser Art angewendet. Deshalb wird sie zukünftig etwas kürzer beschrieben – ohne die jeweils zugrundeliegende Funktion g extra zu erwähnen bzw. zu definieren (siehe etwa bereits den Beweis des nächsten Satzes). $\triangle$

2 Einfache Galton-Watson-Verzweigungsprozesse

Satz 2.15 (Erwartungswert und Varianz der Summe einer zufälligen Anzahl an $\mathbb{N}_0$ -wertigen i.i.d.-Zufallsvariablen)

Sei wieder alles wie in der Definition 2.11. Sind X und N integrierbar, so ist auch $S_N = \sum_{i=1}^N X_i$ integrierbar und es gilt

$$\mathbb{E}S_N = \mathbb{E}N \mathbb{E}X. \quad (2.27)$$

Gilt zusätzlich $\mathbb{E}X^2 < \infty$ und $\mathbb{E}N^2 < \infty$, so ist auch S_N^2 integrierbar und es gilt

$$\mathbb{V}S_N = \mathbb{E}N \mathbb{V}X + (\mathbb{E}X)^2 \mathbb{V}N. \quad (2.28)$$

Beweis. Da für alle $n \geq 0$ mit $\mathbb{P}(N = n) > 0$

$$\mathbb{E}[S_N | N = n] \stackrel{(2.25)}{=} \mathbb{E}S_n \stackrel{(2.24)}{=} n\mathbb{E}X \quad \text{gilt, führt} \quad \mathbb{E}[S_N | N] := N \mathbb{E}X \quad \text{zu} \quad (2.29)$$

$$\mathbb{E}S_N \stackrel{\text{tot. EW}}{\stackrel{(A.74)}{=}} \mathbb{E}[\mathbb{E}[S_N | N]] \stackrel{(2.29)}{=} \mathbb{E}[N \underbrace{\mathbb{E}X}_{\text{konst.}}] = \mathbb{E}N \mathbb{E}X.$$

Mit dem Varianz-Verschiebungssatz gilt $\mathbb{V}S_n = (\mathbb{E}S_n)^2 - \mathbb{E}S_n^2$. Deswegen rechtfertigt

$$\mathbb{E}[S_N^2 | N = n] \stackrel{(2.25)}{=} \mathbb{E}S_n^2 = \mathbb{V}S_n + (\mathbb{E}S_n)^2 \stackrel{(2.24)}{=} n\mathbb{V}X + n^2(\mathbb{E}X)^2 \quad \text{die Festlegung} \\ \mathbb{E}[S_N^2 | N] := N \mathbb{V}X + N^2(\mathbb{E}X)^2. \quad (\star)$$

$$\Rightarrow \mathbb{E}S_N^2 \stackrel{\text{tot. EW}}{\stackrel{(A.74)}{=}} \mathbb{E}[\mathbb{E}[S_N^2 | N]] \stackrel{(\star)}{=} \mathbb{E}[N \underbrace{\mathbb{V}X}_{\text{konst.}} + N^2 \underbrace{(\mathbb{E}X)^2}_{\text{konst.}}] = \mathbb{E}N \mathbb{V}X + (\mathbb{E}X)^2 \mathbb{E}N^2 \quad (\star\star)$$

$$\Rightarrow \mathbb{V}S_N = \mathbb{E}S_N^2 - (\mathbb{E}S_N)^2 \stackrel{(2.27)}{\stackrel{(\star\star)}{=}} \mathbb{E}N \mathbb{V}X + (\mathbb{E}X)^2 \mathbb{E}N^2 - (\mathbb{E}N)^2 (\mathbb{E}X)^2 \\ = \mathbb{E}N \mathbb{V}X + (\mathbb{E}X)^2 [\mathbb{E}N^2 - (\mathbb{E}N)^2] = \mathbb{E}N \mathbb{V}X + (\mathbb{E}X)^2 \mathbb{V}N. \quad \square$$

Bemerkungen zum Satz 2.15. Analog zur Bem. 2.12 könnten auch hier ggf. die Integrierbarkeitsanforderungen an X und N abgeschwächt werden. Es handelt sich immer um nichtnegative und somit quasiintegrierbare ZV (siehe dazu die Bem. zur Def. A.59 (c)).

Der Satz gilt auch in einer Situation, in welcher die $\{X_i\}_{i \geq 1}$ nicht zwingend $\mathbb{N}_0$ -wertig, sondern allgemeiner reell (aber immer noch i.i.d.) sind. Als Beweisskizze beachte, dass die im Beweis zum Lemma 2.13 (ii) erwähnte Nebenrechnung B.2 auch allgemeiner funktioniert und im Beweis des Satzes die $\mathbb{N}_0$ -Wertigkeit gar nicht verwendet wurde. Zu zeigen bliebe für den EW, dass $\mathbb{E}S_N$ auch dann integrierbar ist (was mittels der Dreiecksungleichung möglich ist). Dann wird die Gleichung (2.27) für den EW auch als *Waldsche Gleichheit* oder *-Identität* bezeichnet. Die Gleichung (2.28) für die Varianz nennt man *Blackwell-Girshick-Gleichung* (vgl. etwa [30, Sätze 5.5, S. 105 und 5.10, S. 109] oder [38, Satz 10.24, S. 287 f.]). $\triangle$

Die Resultate für f_{S_N} , $\mathbb{E}S_N$ und $\mathbb{V}S_N$ der Sätze 2.14 und 2.15 sind also identisch zu jenen, welche mit der Herangehensweise zu Beginn dieses Teilabschnitts erzielt wurden (vgl. (2.20), (2.22) und (2.23)). Dort wurden die Eigenschaften der weF bzw. die ihrer Ableitungen verwendet, wenngleich etwa in (2.20) indirekt bereits die Eigenschaft des totalen Erwartungswertes für $\mathbb{E}[s^{S_N} | N]$ ausgenutzt (bzw. für diesen Fall mit gezeigt) wurde.

Fixpunkte einer wahrscheinlichkeitserzeugenden Funktion

Für diesen Teilabschnitt sei $f = f_X$ immer eine weF wie in der Definition 2.4, wobei sich die Betrachtung auf die Summenfunktion $f: [0, 1] \rightarrow [0, 1]$, $f(s) = \sum_{k \geq 0} p_k s^k$, konzentriert.

Es soll untersucht werden, unter welchen Voraussetzungen f wie viele *Fixpunkte* besitzt. Ein Fixpunkt $\psi \in [0, 1]$ von f erfüllt die Gleichung $f(\psi) = \psi$. Da f stetig und $f([0, 1]) \subseteq [0, 1]$ ist, ist generell garantiert, dass es mindestens einen Fixpunkt gibt (vgl. etwa einen allgemeinen Fixpunktsatz wie in [22, Satz 35.4, S. 223]). Das stimmt wegen (2.8) ($f(1) = 1$) für 1 tatsächlich immer.

Um eine Art Fixpunktsatz für wahrscheinlichkeitserzeugende Funktionen aufstellen zu können, werden im folgenden Lemma zunächst ein paar Beobachtungen bzw. Hilfsresultate formuliert. Es sei nochmals darauf hingewiesen, dass, sofern $f'(1^-) < \infty$ gilt, also falls die (linksseitige) erste Ableitung an der Stelle 1 endlich ist, $f'(1^-) = f'(1)$ geschrieben werden kann (siehe das Lemma 2.7).

Lemma 2.16 (Weitere Eigenschaften einer weF als Summenfunktion – tw. angelehnt an [19, Theorem I.5.1, S. 6])

- (i) Wenn entweder $f'(1^-) > 1$ ist (dafür muss $p_1 < 1$ sein) oder $p_1 < 1$ und $f'(1^-) = f'(1) = 1$ sind, gilt

$$p_0 + p_1 < 1 \xrightarrow[2.5(e)]{\text{Satz}} f \text{ und } f' \text{ sind strikt isoton und } f \text{ ist strikt konvex.}$$

- (ii) Wenn $p_0 + p_1 < 1$ ist, besitzt f maximal einen Fixpunkt $\psi < 1$, d.h. in $[0, 1)$.

- (iii) Wenn $f'(1^-) > 1$ ist, gilt $f(s) < s$ nahe bei 1. D.h., es gibt ein $t_0 \in (0, 1)$, sodass $f(s) < s$ für alle $s \in (t_0, 1)$ ist.

Beweis.

- (i) Angenommen, $p_0 + p_1 = 1$. Für jedes $s \in [0, 1]$ wäre dann $f(s) \stackrel{(2.6)}{=} p_0 + p_1 s$, somit $f'(s) = p_1$ und auch $f'(1^-) = f'(1) = p_1$. Das steht entweder im Widerspruch zu $f'(1^-) > 1$ oder, wenn $p_1 < 1$ ist, zu $f'(1) = 1$.
- (ii) Angenommen, es gäbe zwei *unterschiedliche* Fixpunkte ψ_1 und ψ_2 (für die also $f(\psi_1) = \psi_1$ und $f(\psi_2) = \psi_2$ gilt $(\star)$), sodass $0 \leq \psi_1 < \psi_2 < 1$ ist.

Da f laut dem Satz 2.5 auf ganz $[0, 1]$ stetig und zumindest auf $[0, 1)$ differenzierbar ist, müsste es dann nach dem Mittelwertsatz A.92 ein $c \in (\psi_1, \psi_2)$ und ein $d \in (\psi_2, 1)$ (also $0 \leq \psi_1 < c < \psi_2 < d < 1$) geben mit

$$f'(c) \stackrel{(A.106)}{=} \frac{f(\psi_2) - f(\psi_1)}{\psi_2 - \psi_1} \stackrel{(\star)}{=} \frac{\psi_2 - \psi_1}{\psi_2 - \psi_1} = 1, \quad f'(d) = \frac{f(1) - f(\psi_2)}{1 - \psi_2} \stackrel{(\star)}{\stackrel{(2.8)}{=}} \frac{1 - \psi_2}{1 - \psi_2} = 1.$$

Es ist aber f' strikt isoton (siehe (i)), woraus wegen $c < d$ wiederum $f'(c) < f'(d)$ folgen würde. Das stellt einen Widerspruch zu $f'(c) = f'(d) = 1$ dar.

2 Einfache Galton-Watson-Verzweigungsprozesse

- (iii) Diese Aussage wird dadurch klar, dass aufgrund der Steigung > 1 im Punkt 1 für hinreichend große $s < 1$ ebenso $f'(s) > 1$ gelten muss (f' ist stetig auf $[0, 1)$ und strikt isoton). Deshalb muss f dort unterhalb der Diagonalen, d.h. unter der ersten Mediane, verlaufen. Die NR B.3 liefert einen formalen Beweis. $\square$

Satz 2.17 (Fixpunktsatz für wahrscheinlichkeitserzeugende Funktionen)

Eine weF $f: [0, 1] \rightarrow [0, 1]$ besitzt wegen $f(1) = 1$ immer einen Fixpunkt in 1. Zudem gilt:

- (i) Wenn $p_1 = 1$ ist (dann ist $f'(1^-) = f'(1) = 1$), ist jeder weitere Punkt in $[0, 1)$ ebenfalls ein Fixpunkt und $f(s) = s$ gilt für alle $s \in [0, 1]$.
- (ii) Wenn $f'(1^-) = f'(1) \leq 1$ und $p_1 < 1$ (für $f'(1) < 1$ ohnehin erfüllt) gilt, gibt es keinen weiteren Fixpunkt in $[0, 1)$ und $\psi = 1$ ist der einzige, der $f(\psi) = \psi$ erfüllt.
- (iii) Wenn $f'(1^-) > 1$ ist (dafür muss $p_1 < 1$ sein), gibt es genau einen weiteren Fixpunkt $\psi \in [0, 1)$, der $f(\psi) = \psi$ erfüllt.

Beweis.

- (i) Aus $p_1 = 1$ folgt $p_k = 0$ für alle $k \neq 1$ und deshalb $f(s) = sp_1 = s$ ($\forall s \in [0, 1]$).
- (ii) Laut dem Satz 2.5 (d) ist f auf ganz $[0, 1]$ stetig und, wenn $f'(1) \leq 1$ gilt, zudem auch differenzierbar (siehe das Lemma 2.7 bzw. die Bemerkung dazu). Es werden zwei Fälle unterschieden. Sei für beide $s \in [0, 1)$.
 - Wenn $f'(1) < 1$ ist (wegen (i) muss dafür $p_1 < 1$ sein), ist f konvex (Satz 2.5 (e)) und somit f' isoton $\Rightarrow f'(s) \leq f'(1) < 1$.
 - Wenn $f'(1) = 1$ und $p_1 < 1$ ist, so ist f' strikt isoton (Lemma 2.16 (i)). Deshalb gilt dann, dass $f'(s) < f'(1) = 1$ ist.

Es ist also in beiden Fällen $f'(s) < 1$ ($\star$). Mit dem zuvor stellt der Mittelwertsatz A.92 für jedes $s \in [0, 1)$ sicher, dass es ein $c \in (s, 1)$ gibt, sodass

$$1 > f'(c) = \frac{f(1) - f(s)}{1 - s} \stackrel{(2.8)}{=} \frac{1 - f(s)}{1 - s} \text{ ist. Daraus folgt } 1 - f(s) < 1 - s \text{ und weiter } f(s) > s. \quad (\star)$$

- (iii) Erneut dem Lemma 2.16 (i) kann für $f'(1^-) > 1$ entnommen werden, dass f strikt isoton und $p_0 + p_1 < 1$ (und somit auch $p_1 < 1$) ist. Deshalb liefert (ii) des besagten Lemmas, dass es maximal einen Fixpunkt in $[0, 1)$ geben kann. Falls $p_0 = 0$ ist, so ist $f(0) = p_0 = 0$ und somit $\psi = 0$ dieser eine Fixpunkt und die Aussage ist gezeigt.

Sei also $p_0 > 0$. Wegen des Lemmas 2.16 (iii) gibt es dann ein $t_0 \in (0, 1)$, sodass $f(s) < s$ für alle $s \in (t_0, 1)$ gilt. Wähle nun ein beliebiges s_0 aus $(t_0, 1)$, für welches demnach $f(s_0) < s_0$ gilt ($\star\star$) und setze $g: [0, s_0] \rightarrow \mathbb{R}$ als $g(s) := f(s) - s$. Da f stetig auf $[0, 1]$ ist, ist auch g stetig. Weiters ist $g(0) = f(0) - 0 = p_0 > 0$ (wegen (2.9)) und $g(s_0) = f(s_0) - s_0 < 0$ (wegen ($\star\star$)). Der Nullstellensatz A.90 garantiert deshalb, dass g mindestens eine Nullstelle in $\psi \in (0, s_0) \subset (0, 1)$ besitzt, d.h. $g(\psi) = 0$. Daraus folgt $g(\psi) = f(\psi) - \psi = 0$ und weiters $f(\psi) = \psi$ (dabei gilt $0 < p_0 < \psi < 1$). $\square$

Iterierte wahrscheinlichkeits erzeugende Funktionen

Wahrscheinlichkeitserzeugende Funktionen können auch iteriert werden. Es wird sich zeigen, dass diese iterierten Funktionen selbst wieder eine Folge von weF darstellen. Im nächsten Abschnitt 2.3 wird ersichtlich werden, dass diese mit den wahrscheinlichkeits erzeugenden Funktionen der Populationsgrößen eines Galton-Watson-Prozesses (Def. 2.21) korrespondieren. Hier sollen solche Funktionen aber zunächst allgemein definiert werden.

Definition 2.18 (Iterierte weF – angelehnt an [19, I.2.2, S. 5])

Sei f die weF von X . Dann seien die Funktionen $f_n : [0, 1] \rightarrow [0, 1]$ – die *Iterationen der wahrscheinlichkeits erzeugenden Funktion f* – für jedes $n \geq 0$ rekursiv definiert als

$$\begin{aligned} f_0(s) &:= s, & f_1(s) &:= f(s) \text{ und} \\ f_{n+1}(s) &:= f(f_n(s)) \quad (\text{bzw. } f_{n+1} = f \circ f_n) & & \text{für } n \geq 1. \end{aligned} \quad (2.30)$$

Die n -fache Iteration f_n einer weF f nennt man auch ihre n -fache Verkettung (oder Komposition) mit sich selbst.

Korollar zur Definition 2.18 (erneut angelehnt an [19, I.2.2, S. 5])

Jede Iteration von f ist selbst wiederum eine weF. Für $m, n \geq 0$ und $s \in [0, 1]$ gilt

$$f_{m+n} = f_m \circ f_n = f_n \circ f_m \quad \text{und insbesondere} \quad (2.31)$$

$$f_{n+1} = f_n \circ f. \quad (2.32)$$

Beweis. Betrachte für $n = 0$ eine deterministische (also nichtzufällige) 'Zufallsvariable' $Y \equiv 1$. Dann ist ihre weF $f_Y(s) = \mathbb{E}s^Y = \mathbb{E}s^1 = s = f_0(s)$ für $s \in [0, 1]$. Mit $n = 1$ handelt es sich bei $f_1 = f$ definitionsgemäß um eine weF und für $n \geq 1$ soll mit vollständiger Induktion über n vorgegangen werden. Angenommen, die Behauptung trifft also für n bereits zu und f_n ist eine weF. Dann ist $f_{n+1} = f \circ f_n$ wiederum eine weF, da sie eine Verkettung zweier solcher ist – vgl. etwa (2.20) oder den Satz 2.14.

Die erste Gleichheit in (2.31) sieht man daran, dass für $s \in [0, 1]$ und $m, n \geq 1$ (für die Korrektheit der Darstellung müssten m und n genau genommen hinreichend groß sein)

$$\begin{aligned} f_n(s) &\stackrel{(2.30)}{=} f(f_{n-1}(s)) = f(f(f_{n-2}(s))) = \dots = \overbrace{f(f(\dots f(s)\dots))}^{n\text{-mal}} \text{ und somit auch} \\ f_{m+n}(s) &= \underbrace{f(f(\dots f}_{m\text{-mal}}(\underbrace{f(f(\dots f}_{n\text{-mal}}(f(\underbrace{f(f(\dots f}_{m\text{-mal}}(s)\dots))\dots))\dots))\dots))}_{m\text{-mal}}) = \underbrace{f(f(\dots f}_{m\text{-mal}}(f_n(s)\dots))\dots)}_{m\text{-mal}} = f_m(f_n(s)) \text{ gilt.} \end{aligned}$$

Für $m \geq 0$ (mit $n = 0$) gilt $f_{m+0}(s) = f_m(s) \stackrel{(2.30)}{=}_{(s=f_0(s))} f_m(f_0(s))$.

Analog dazu erhält man, da $m + n$ kommutativ ist, $f_{m+n}(s) = f_n(f_m(s))$ (es ist also auch $f_m \circ f_n$ kommutativ). Mit $m = 1$ folgt (2.32). $\square$

Eine Interpretation im Setting der Galton-Watson-Prozesse bzw. eine alternative Beweismethode für (2.31) findet sich im Abschnitt 2.4 auf der Seite 32 – anschließend zum Satz 2.28.

2 Einfache Galton-Watson-Verzweigungsprozesse

Zusätzlich zu den Eigenschaften, welche alle Iterationen f_n als weF jeweils sowieso haben, 'erben' sie auch gewisse der zugrundeliegenden weF f , wie etwa das nächste Lemma zeigt.

Lemma 2.19 (Eigenschaften der iterierten weF)

- (i) Wenn ψ ein Fixpunkt von f ist, also $f(\psi) = \psi$ gilt, so ist es auch ein Fixpunkt aller f_n . Es gilt dann also $f_n(\psi) = \psi$ für alle $n \geq 0$.
- (ii) Wenn $p_0 \stackrel{(2.9)}{=} f(0) < 1$ ist, so gilt auch $f_n(0) < 1$ für alle $n \geq 0$. Dementsprechend sind dann auch alle f_n (wie f) strikt isoton.

Beweis.

- (i) Es ist $f_0(\psi) \stackrel{(2.30)}{=} \psi$ und $\underbrace{f_1(\psi)}_{=f(\psi)} = \psi$. Induktiv folgt $f_{n+1}(\psi) \stackrel{(2.30)}{=} \underbrace{f(f_n(\psi))}_{= \psi \text{ (Ind.-Ann.)}} = \psi$ für alle $n \geq 0$.
- (ii) Die zweite Aussage darin folgt aus der ersten (Satz 2.5 (e) mit (2.9)). Diese erhält man wieder induktiv aus $f_{n+1}(0) = \underbrace{f(f_n(0))}_{<1 \text{ (Ind.-Ann.)}} < \underbrace{f(1)}_{=1 \text{ (2.8)}} = 1$, da $f_0(0) \stackrel{(2.30)}{=} 0 < 1$ und $f_1(0) = p_0 < 1$ ist. □

2.3 Definition des GW-Verzweigungsprozesses

Mathematisch gesehen ist ein (einfacher) Galton-Watson-Prozess ein *zeitdiskreter stochastischer Prozess* (vgl. etwa [38, Def. 10.1, S. 268]), wobei als Zeitindex $n \geq 0$ ($n \in \mathbb{N}_0$) verwendet wird. Dafür sei $(\Omega, \Sigma, \mathbb{P})$ ein passender, nicht näher spezifizierter Wahrscheinlichkeitsraum, auf dem der Prozess definiert werden kann.

Um den im Abschnitt 2.1 beschriebenen Grundannahmen gerecht zu werden, kann der Prozess gleich als eine sogenannte *Markov-Kette* (siehe ggf. A.2.4 im Appendix – genauer eine homogene Markov-Kette erster Ordnung) eingeführt werden. Dies wird etwa in den Standardwerken [19] oder [3] so gehandhabt. Hier soll er jedoch zunächst anhand der *Nachkommenszahl* bzw. *Nachkommensverteilung* der Individuen, welche jeweils unabhängige, identisch verteilte Kopien einer prototypischen Zufallsvariablen X sind, und den *Populationsgrößen* Z_n – für jede Generation $n \geq 0$ – formuliert werden. Anschließend wird gezeigt, dass es sich dabei tatsächlich um eine Markov-Kette handelt.

2.3.1 Grundlegende Definitionen

Definition 2.20 (Nachkommenszahl und -verteilung von Individuen)

Sei X eine $\mathbb{N}_0$ -wertige Zufallsvariable und $n \geq 0$ beschreibe die n -te Generation. Dann ist die Familie der Zufallsvariablen $\{X_{n,i}\}_{n \geq 0, i \geq 1}$, welche unabhängige, identisch verteilte (*i.i.d.*-)Kopien von X seien, die *Anzahl der Nachkommen* des jeweils i -ten Individuums der n -ten Generation (in der unmittelbar nachfolgenden Generation).

Die Nachkommenszahl X besitze eine *Nachkommensverteilung* (bzw. W-Funktion) $p = \{p_k\}_{k \geq 0}$, wobei p_k jeweils definiert ist als

$$p_k := \mathbb{P}(\text{'Individuum hat } k \text{ Nachkommen'}) = \mathbb{P}(X = k) \quad \text{für jedes } k \geq 0. \quad (2.33)$$

Da in jeder Generation $n \geq 0$ die Nachkommenszahl jedes Individuums $i \geq 1$ eine *i.i.d.*-Kopie der prototypischen Nachkommenszahl X sowie p eine Verteilung ist, gilt

$$\mathbb{P}(X_{n,i} = k) \stackrel{X_{n,i} \stackrel{d}{=} X}{\stackrel{(2.33)}{=} p_k} \quad \text{für alle } k, n \geq 0 \text{ und alle } i \geq 1 \text{ sowie } \sum_{k \geq 0} p_k \stackrel{(2.2)}{=} 1. \quad (2.34)$$

Definition 2.21 (Populationsgrößen)

Sei $\{Z_n\}_{n \geq 0}$ eine Folge $\mathbb{N}_0$ -wertiger Zufallsvariablen. Die 'nullte' bzw. *initiale Populationsgröße* Z_0 sei dabei unabhängig von der Nachkommenszahl X (und somit unabhängig von allen $\{X_{n,i}\}_{n \geq 0, i \geq 1}$). Dann definiert die Rekursion

$$Z_{n+1} := \sum_{i=1}^{Z_n} X_{n,i} \quad \text{für alle } n \geq 0 \quad (2.35)$$

jeweils die *Populationsgröße der (n+1)-ten Generation*. Ist $Z_n = 0$, so handelt es sich in (2.35) um eine leere Summe, weshalb dann $Z_{n+1} = 0$ per Konvention folgt (Not. 1.1).

2 Einfache Galton-Watson-Verzweigungsprozesse

Zusammengenommen werden $\{Z_n\}_{n \geq 0}$ und $\{X_{n,i}\}_{n \geq 0, i \geq 1}$ als **einfacher Galton-Watson-Verzweigungsprozess**, kurz **GWP**, mit der initialen Populationsgröße Z_0 und der Nachkommenszahl X bzw. der Nachkommensverteilung $p = \{p_k\}_{k \geq 0}$ bezeichnet.

Bei der Folge von Zufallsvariablen $\{Z_n\}_{n \geq 0}$ handelt es sich also um einen stochastischen Prozess, welcher gemäß der Definition 2.21 die Populationsgröße beliebiger Generationen $n \geq 0$ beschreibt. Nun ist das Verhalten dieser Populationsgrößen über die Zeit von Interesse. Hierzu sollen die Wahrscheinlichkeitsverteilung und die Momente der Z_n , insbesondere deren Erwartungswerte bzw. Varianzen, im nächsten Abschnitt 2.4 bestimmt werden. In weiterer Folge soll im Wesentlichen ermittelt werden, wie wahrscheinlich es ist, dass die Zufallssequenz Null wird, der Verzweigungsprozess also ausstirbt (Abschnitt 2.5) und wie sie sich verhält, falls er nicht ausstirbt (Abschnitt 2.6).

Zunächst soll aber – wie angekündigt – dargelegt werden, dass es sich bei einem GWP um eine Markov-Kette handelt.

2.3.2 GWP als Markov-Kette und Übergangswahrscheinlichkeiten

So wie der Prozess definiert wurde, ist es naheliegend, dass die Nachkommenszahlen der Individuen einer Generation (und auch die aller zukünftigen) nicht von der aktuellen und allen vergangenen Populationsgrößen beeinflusst werden und somit davon unabhängig sind. Diese Tatsache wird im nächsten Lemma präzisiert.

Lemma 2.22 (Unabhängigkeitsaussage)

Sei $n \geq 0$. Die Populationsgrößen $\{Z_m\}_{m \leq n}$ sind unabhängig von allen Nachkommenszahlen $\{X_{l,i}\}_{l \geq n, i \geq 1}$.

Beweis. Es soll gezeigt werden, dass Z_n für ein fixes $n \geq 0$ unabhängig von allen $\{X_{l,i}\}_{l \geq n, i \geq 1}$ ist. Daraus folgt dann die Aussage auch für jedes Z_m mit $m \leq n$.

Für $n = 0$ ist Z_0 per Definition unabhängig von allen $\{X_{l,i}\}_{l \geq 0, i \geq 1}$. Vollständige Induktion zeigt dies auch für beliebige $n \geq 0$. Gemäß (2.35) ist $Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i}$. Dabei sind die $\{X_{n,i}\}_{i \geq 1}$ unabhängig von allen $\{X_{l,i}\}_{l \neq n, i \geq 1}$ und Z_n ist laut der Induktionsannahme unabhängig von allen $\{X_{l,i}\}_{l \geq n, i \geq 1}$. Daraus folgt, dass Z_{n+1} unabhängig von allen $\{X_{l,i}\}_{l > n, i \geq 1} = \{X_{l,i}\}_{l \geq n+1, i \geq 1}$ ist. $\square$

Betrachte nun erneut die Rekursion (2.35), also $Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i}$, für ein beliebiges $n \geq 0$. Die $\{X_{n,i}\}_{i \geq 1}$ sind dabei *i.i.d.* und gemäß dem vorherigen Lemma unabhängig von Z_n . Somit liegt eine Situation wie in der Definition 2.11, welche die Summe einer zufälligen Anzahl an $\mathbb{N}_0$ -wertigen *i.i.d.*-ZV beschreibt, vor. Dies kann ausgenutzt werden, um das folgende Hilfsresultat und anschließend die Übergangswahrscheinlichkeiten zu formulieren.

Lemma 2.23. Seien $k, n \geq 0$ und $j_0, \dots, j_n \geq 0$. Für die Populationsgrößen gilt

- (i) $\mathbb{P}(Z_{n+1} = k, Z_n = j_n) = \mathbb{P}(\sum_{i=1}^{j_n} X_{n,i} = k) \mathbb{P}(Z_n = j_n)$ und
- (ii) $\mathbb{P}(Z_{n+1} = k, Z_n = j_n, \dots, Z_0 = j_0) = \mathbb{P}(\sum_{i=1}^{j_n} X_{n,i} = k) \mathbb{P}(Z_n = j_n, \dots, Z_0 = j_0)$.

2.3 Definition des GW-Verzweigungsprozesses

Beweis. Aufgrund des zuvor Beschriebenen kann analog zur Berechnung $(\star)$ im Beweis des Lemmas 2.13 (ii) vorgegangen werden, um zu erhalten, dass

$\{Z_{n+1}=k\} \cap \{Z_n=j_n\} \stackrel{(2.35)}{=} \left\{ \sum_{i=1}^{Z_n} X_{n,i}=k \right\} \cap \{Z_n=j_n\} = \left\{ \sum_{i=1}^{j_n} X_{n,i}=k \right\} \cap \{Z_n=j_n\}$ gilt. Die beiden Ereignisse im letzten Ausdruck sind hierbei laut dem Lemma 2.22 unabhängig (Z_n und $\{X_{n,i}\}_{i \geq 1}$ sind uabh.), weshalb etwa (A.55) zu (i) führt. Wie zuvor erhält man $\{Z_{n+1}=k\} \cap \{Z_n=j_n\} \cap \dots \cap \{Z_0=j_0\} = \left\{ \sum_{i=1}^{j_n} X_{n,i}=k \right\} \cap \{Z_n=j_n\} \cap \dots \cap \{Z_0=j_0\}$. Dabei sind – wieder gemäß dem Lemma 2.22 – alle $\{Z_m\}_{m \leq n}$ unabhängig von allen $\{X_{n,i}\}_{i \geq 1}$. Damit kann (ii) begründet werden. $\square$

Aus der Beziehung (i) des Lemmas 2.23 (mit j statt j_n) kann die bedingte Verteilung von Z_{n+1} gegeben $Z_n = j$ – also der Summe von j *i.i.d.*-Kopien von X – angegeben werden. Diese lässt sich mit dem Satz A.51 als j -fache Faltung von p (vgl. die Def. A.49), der W-Funktion der Nachkommenszahl X , ausdrücken. Wenn $j, n \geq 0$ und so sind, dass $\mathbb{P}(Z_n = j) > 0$ ist (dann darf durch diese Wahrscheinlichkeit – kurz *WK* – dividiert werden), erhält man für $k \geq 0$

$$\mathbb{P}(Z_{n+1} = k | Z_n = j) \stackrel{\text{Lem. 2.23 (i) und (A.69)}}{=} \mathbb{P} \left(\sum_{i=1}^j X_{n,i} = k \right) \stackrel{(A.65)}{=} p_k^{*j} = \sum_{\substack{l_1, \dots, l_j \geq 0: \\ l_1 + \dots + l_j = k}} p_{l_1} \cdots p_{l_j}. \quad (2.36)$$

Diese Überlegung führt zu der folgenden Definition der Übergangswahrscheinlichkeiten.

Definition 2.24 (Übergangswahrscheinlichkeiten)

Seien $j, k \geq 0$. Definiere die *Übergangswahrscheinlichkeit* p_{jk} für den Übergang von j , der Populationsgröße einer beliebigen Generation ($n \geq 0$), in k , jener der nächsten Generation, als

$$p_{jk} := p_k^{*j} \stackrel{(A.65)}{=} \mathbb{P} \left(\sum_{i=1}^j X_{n,i} = k \right) = \sum_{\substack{l_1, \dots, l_j \geq 0: \\ l_1 + \dots + l_j = k}} p_{l_1} \cdots p_{l_j}. \quad \text{Insbes. ist } p_{0k} \stackrel{(A.62)}{=} \delta_{0k} \quad (2.37)$$

(δ_{0k} ist das Kronecker-Delta – siehe die Def. 1.3). Für eine spezifische Generation $n \geq 0$ und $j \geq 0$, sodass $\mathbb{P}(Z_n = j) > 0$ ist, gilt

$$\mathbb{P}(Z_{n+1} = k | Z_n = j) \stackrel{(2.36)}{=} p_k^{*j} \stackrel{(2.37)}{=} p_{jk}. \quad (2.38)$$

Da $k \geq 0$ beliebig ist, wird also durch $\{p_{jk}\}_{k \geq 0}$ die bedingte Verteilung von Z_{n+1} gegeben $Z_n = j$ (auch ausgedrückt als $\mathbb{P}_{Z_{n+1}|Z_n=j}$, siehe die Bemerkung A.58) bestimmt.

Bemerkung zur Definition 2.24. Für $j \geq 0$ mit $\mathbb{P}(X_n = j) = 0$ wäre die bedingte WK in (2.36) bzw. (2.38) entweder nicht definiert oder, falls so wie in der Def. A.56 spezifiziert, kein W-Maß (siehe die Bem. (ii) zur genannten Def. und insbes. wieder die Bem. A.58). Durch die in (2.37) getroffene Festlegung der Übergangswahrscheinlichkeiten p_{jk} ist aber – da diese stets wohldefiniert sind – garantiert, dass es sich bei $\{p_{jk}\}_{k \geq 0} = \{p_k^{*j}\}_{k \geq 0} = p^{*j}$ für jedes $j \geq 0$ um eine W-Funktion handelt (vgl. die Bem. zum Satz A.51).

2 Einfache Galton-Watson-Verzweigungsprozesse

In der Literatur wird – dessen ungeachtet – dennoch oft schlicht die Formulierung $p_{jk} = \mathbb{P}(Z_{n+1} = k | Z_n = j)$ gewählt und allenfalls darauf hingewiesen, dass die rechte Seite im Fall $\mathbb{P}(Z_n = j) = 0$ streng genommen nicht als bedingte Wahrscheinlichkeit definiert ist (vgl. etwa [19, I.(2.1), S. 4] und die darauf folgende Anmerkung). $\triangle$

Mit den soeben definierten Übergangswahrscheinlichkeiten lässt sich zeigen, dass der GWP eine Markov-Kette im Sinne der Definition A.70 ist.

Satz 2.25 (GWP als Markov-Kette)

Seien Z_0 wie in der Def. 2.21 beschrieben (also insbes. unabhängig von $\{X_{n,i}\}_{n \geq 0, i \geq 1}$)

$$\text{sowie} \quad \mathbf{P} := (p_{jk})_{j,k \geq 0} \quad \text{und} \quad \lambda := \{\mathbb{P}(Z_0 = k)\}_{k \geq 0}.$$

Dann ist der GWP $\{Z_n\}_{n \geq 0}$ eine Markov-Kette mit Anfangsverteilung λ und Übergangsmatrix $\mathbf{P}$, kurz Markov($\lambda, \mathbf{P}$).

Wird die initiale Populationsgröße auf ein ganzzahliges $l \geq 1$ festgelegt (d.h. $Z_0 \equiv l$ mit $l \in \mathbb{N}$), dann ist $\{Z_n\}_{n \geq 0}$ Markov($\delta_l, \mathbf{P}$).

Beweis. Stelle zuerst fest, dass $\sum_{k \geq 0} p_{jk} \stackrel{(2.37)}{=} \sum_{k \geq 0} p_k^{*j} = 1$ für jedes $j \geq 0$ gilt (siehe erneut die Bem. zum Satz A.51). Somit ist $\mathbf{P}$ eine stochastische Matrix gemäß der Def. A.69. Weiters müssen die Punkte (i) und (ii) der Def. A.70 erfüllt sein. So wie λ angegeben ist, trifft das auf (i) zu. Für (ii) betrachte $k, n \geq 0$ und $j_0, \dots, j_n \geq 0$, sodass $\mathbb{P}(Z_n = j_n, \dots, Z_0 = j_0) > 0$ gilt ($\Rightarrow \mathbb{P}(Z_n = j_n) > 0$). Dann können die Gleichungen in den Punkten (i) und (ii) des Lemmas 2.23 jeweils durch die Wahrscheinlichkeit ganz rechts dividiert werden, um mit der Def. der bedingten WK zu erhalten, dass

$$\mathbb{P}(Z_{n+1} = k | Z_n = j_n, \dots, Z_0 = j_0) = \mathbb{P}(Z_{n+1} = k | Z_n = j_n) \stackrel{(2.38)}{=} p_{j_n k} \quad \text{gilt,} \quad (2.39)$$

womit auch (ii) erfüllt ist.

Ist $Z_0 \equiv l \geq 1$, dann gilt $\mathbb{P}(Z_0 = l) = 1$ und ansonsten $\mathbb{P}(Z_0 = k) = 0$ (für $k \neq l$). Deshalb ist laut (1.1) für jedes $k \geq 0$ die Wahrscheinlichkeit $\mathbb{P}(Z_0 = k) = \delta_{lk}$ und somit gilt $\lambda = \{\mathbb{P}(Z_0 = k)\}_{k \geq 0} = \{\delta_{lk}\}_{k \geq 0} = \delta_l$. $\square$

Wird ein GWP als Markov-Kette aufgefasst, kann jeder seiner Zustände – also die Populationsgrößen der unterschiedlichen Generationen – entweder als sogenannter *absorbierender* oder *transienter* Zustand (siehe die Definitionen A.75 und A.76) klassifiziert werden. Dies wird im nächsten Satz vorgenommen. Beachte dabei, dass jeder absorbierende auch ein sogenannter *rekurrenter* Zustand ist (siehe die Bem. A.78) und grundsätzlich alle Zustände einer Markov-Kette entweder transient oder rekurrent sind (Satz A.79).

Satz 2.26 (Klassifikation der Zustände – angelehnt an [19, Theorem I.6.2, S. 8])

Sei Z_0 wie im Satz 2.25. Die Zustände k der Markov-Kette $\{Z_n\}_{n \geq 0}$, sprich $Z_n = k$ für beliebige $n, k \geq 0$, können wie folgt klassifiziert werden.

(i) Der Zustand $k = 0$ ist immer **absorbierend** (und somit auch rekurrent).

D.h., wenn in einer Generation $n \geq 0$ die Populationsgröße $Z_n = 0$ ist, so wird dieser Zustand nie wieder verlassen und für jedes $m \geq 0$ ist dann ebenso $Z_{n+m} = 0$.

2.3 Definition des GW-Verzweigungsprozesses

(ii) Für alle anderen Zustände $k \geq 1$ gilt in Abhängigkeit von $p_1 \stackrel{(2.33)}{=} \mathbb{P}(X = 1)$:

- Falls $p_1 = 1$ ist, so ist auch jeder weitere Zustand $k \geq 1$ absorbierend.
- Falls $p_1 < 1$ ist, so ist jeder weitere Zustand $k \geq 1$ **transient**, d.h.,

$$\text{für } k \geq 1 \text{ gilt } \mathbb{P}('Z_n = k \text{ für } \infty \text{ viele } n') \stackrel{(A.6)}{=} \mathbb{P}\left(\limsup_{n \rightarrow \infty} \{Z_n = k\}\right) = 0. \quad (2.40)$$

I.a.W.: Wenn in einer Generation $n \geq 0$ die Populationsgröße $Z_n = k \geq 1$ ist, so wird die selbe Populationsgröße in zukünftigen Generationen möglicherweise (also mit positiver WK) nie mehr wieder erreicht. Für $k \geq 1$ gilt also

$$\mathbb{P}(Z_{n+m} \neq k \mid Z_n = k) > 0 \quad \text{sowie insbesondere} \quad (2.41)$$

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n = k) = 0 \quad \text{und} \quad \mathbb{P}\left(\lim_{n \rightarrow \infty} Z_n = k\right) = 0. \quad (2.42)$$

Beweis. Wird zu einem beliebigen Zeitpunkt eine Populationsgröße von 0 erreicht, so sind immer auch alle weiteren 0 – dies geht direkt aus dem letzten Satz der Def. 2.21 hervor. Ist $p_1 = \mathbb{P}(X = 1)$, so hat jedes Individuum genau einen Nachkommen und ausgehend von einer Populationsgröße $k \geq 1$ bleiben auch alle zukünftigen konstant bei k . Somit sind die Aussagen in (i) und für den Fall $p_1 = 1$ in (ii) intuitiv plausibel.

Um dies zudem formal zu zeigen, beachte, dass laut der Def. A.75 ein Zustand $k \geq 0$ genau dann absorbierend ist, wenn die Übergangs-WK $p_{kk} = 1$ beträgt. Gemäß (2.37) gilt $p_{kk} = p_k^{*k}$ für alle $k \geq 0$ und insbesondere $p_{00} = \delta_{00} = 1$ (Def. 1.3). Ist $p_1 = 1$, so ist $X = 1$ (f.s.) und $p_{kk} = p_k^{*k} \stackrel{(2.37)}{=} \mathbb{P}\left(\sum_{i=1}^k \underbrace{X_{n,i}}_{=1 \text{ f.s. } \forall n \geq 0, i \geq 1}\right) = \mathbb{P}(k = k) = 1$ für alle $k \geq 1$.

Zu zeigen ist noch der letzte Punkt. Seien dafür für den Rest des Beweises $k \geq 1$, $p_1 < 1$ und $n \geq 0$ so, dass $\mathbb{P}(Z_n = k) > 0$ ist. Laut dem Satz A.77 impliziert eine Rückkehr-WK von k nach k (nach dem ersten Sprung) kleiner 1, also $\mathbb{P}(\exists m \geq 1 : Z_{n+m} = k \mid Z_n = k) < 1$, die Transienz von k . Es werden zwei Fälle in Abhängigkeit von $p_0 = \mathbb{P}(X = 0)$ unterschieden.

- **Fall $p_0 = 0$:** Dann ist $Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i} \stackrel{\text{f.s.}}{\geq} Z_n$, da alle $X_{n,i} \stackrel{\text{f.s.}}{\geq} 1$ für $n \geq 0$ sind. Deshalb entspricht die Rückkehr-WK der Übergangs-WK von k nach k (wenn $Z_{n+1} > k$ ist, kann nie mehr nach k zurückgekehrt werden) und es gilt

$$\begin{aligned} \mathbb{P}(\exists m \geq 1 : Z_{n+m} = k \mid Z_n = k) &= p_{kk} \stackrel{(2.37)}{=} \mathbb{P}\left(\sum_{i=1}^k \widehat{X_{n,i}}^{\geq 1 \text{ f.s.}} = k\right) \\ &= \mathbb{P}(X_{n,1} = 1, \dots, X_{n,k} = 1) \stackrel{i.i.d.}{=} \stackrel{(A.55)}{=} (\mathbb{P}(X = 1))^k = p_1^k < 1 \Rightarrow k \text{ ist transient.} \end{aligned}$$

- **Fall $p_0 > 0$:** Stelle zunächst fest, dass $p_{k0} \stackrel{(2.37)}{=} \mathbb{P}\left(\sum_{i=1}^k X_{n,i} = 0\right) \stackrel{i.i.d.}{=} p_0^k > 0$ ist, woraus $1 - p_{k0} < 1$ folgt ($\star$). Generell muss, um von Z_n ausgehend irgendwann (nach $m \geq 1$ Schritten) nach k zu kommen, zumindest erfüllt sein, dass der Prozess nicht gleich nach einem Schritt auf 0 (und wegen (i) nie mehr davon weg) geht. Deshalb gilt $\{\exists m \geq 1 : Z_{n+m} = k\} \subseteq \{Z_{n+1} \neq 0\}$ und (mit dem Satz A.23 (i) und (ii))

$$\begin{aligned} \mathbb{P}(\exists m \geq 1 : Z_{n+m} = k \mid Z_n = k) &\leq \mathbb{P}(Z_{n+1} \neq 0 \mid Z_n = k) = \mathbb{P}(\{Z_{n+1} = 0\}^c \mid Z_n = k) \\ &= 1 - \mathbb{P}(Z_{n+1} = 0 \mid Z_n = k) \stackrel{(2.38)}{=} 1 - p_{k0} \stackrel{(\star)}{<} 1 \Rightarrow k \text{ ist transient.} \end{aligned}$$

2 Einfache Galton-Watson-Verzweigungsprozesse

Jeder Zustand $k \geq 1$ ist also in jedem Fall transient, was (siehe die Def. A.76 bzw. (A.91)) gleichbedeutend mit $\mathbb{P}(\limsup_{n \rightarrow \infty} \{Z_n = k\} | Z_0 = k) = 0$ und, begründet durch das Korollar zum Satz A.77 bzw. (A.95), sogar mit (2.40) ist. Das Argument für (2.41) steht in den Bem. zum Satz A.77 (ii). Aus (2.40) erhält man mit der $\limsup$ -Def. A.3 und der Stetigkeit von oben

$$0 = \mathbb{P}\left(\underbrace{\bigcap_{n \geq 0} \bigcup_{m \geq n} \{Z_m = k\}}_{\text{fallend in } n}\right) \stackrel{(A.32)}{=} \lim_{n \rightarrow \infty} \mathbb{P}\left(\underbrace{\bigcup_{m \geq n} \{Z_m = k\}}_{\supseteq \{Z_n = k\}}\right) \geq \lim_{n \rightarrow \infty} \mathbb{P}(Z_n = k),$$

woraus der erste Teil von (2.42) folgt. Für den zweiten beachte, dass (ebenfalls wegen (2.40)) die Wahrscheinlichkeit $\mathbb{P}('Z_n = k \text{ für } \infty \text{ viele } n') = 0$ ist, wodurch das Ereignis $\{\lim_{n \rightarrow \infty} Z_n = k\}$ unmöglich ist, da sowohl Z_n als auch k ganzzahlig sind. Die Nebenrechnung B.4 liefert zudem ein formales Argument dafür. $\square$

Da der GWP $\{Z_n\}_{n \geq 0}$ also eine Markov-Kette ist, spielen (von einem spezifischen Zeitpunkt aus betrachtet) die Populationsgrößen der vergangenen Generationen keine Rolle für die weitere Entwicklung in der Zukunft. Diese 'Gedächtnislosigkeit', die sogenannte *Markov-Eigenschaft* (kurz ME – vgl. den Satz A.71), gilt auch für bedingte Erwartungen. D.h., gemäß dem Satz A.72 (ii) gilt für alle $n \geq 1$ und $0 \leq k < n$

$$\mathbb{E}[Z_{n+1} | Z_0, \dots, Z_n] = \mathbb{E}[Z_{n+1} | Z_k, \dots, Z_n] \stackrel{(A.86)}{=} \mathbb{E}[Z_{n+1} | Z_n] \quad \text{f.s.} \quad (2.43)$$

Mit $m + k = n + 1$ ($\Rightarrow n = m + k - 1$) folgt daraus für $m \geq 1$ und $k \geq 0$ auch direkt

$$\mathbb{E}[Z_{m+k} | Z_k, \dots, Z_{m+k-1}] = \mathbb{E}[Z_{m+k} | Z_{m+k-1}] \quad \text{f.s.} \quad (2.44)$$

2.4 Wahrscheinlichkeitsverteilung und Momente der Populationsgrößen

In der Literatur wird der einfache Galton-Watson-Prozess häufig mit einer festen, d.h. nichtzufälligen, ganzzahligen initialen Populationsgröße $Z_0 \equiv l$ mit $l \geq 1$ eingeführt (der Fall $l = 0$ wäre uninteressant, da dann in jeder Generation keine Individuen existieren würden, also jede Populationsgröße Null wäre). Dies erscheint auch in den meisten Anwendungsfällen gerechtfertigt. Unabhängig davon, welche Szenarien modelliert werden sollen, besteht oft ein Kenntnis über die anfängliche Population.

Für viele Analysen genügt es sogar, o.B.d.A. einen GWP mit initialer Populationsgröße $Z_0 \equiv 1$ zu betrachten – dies wird im Korollar 2.31 präzisiert. Aufgrund der Unabhängigkeitsannahme der Individuen kann das Modell recht einfach auf Fälle mit $Z_0 \geq 1$ erweitert werden (so wie etwa in der Bem. zum Korollar 2.31 (ii) sowie in den Beispielen 2.34 und 2.49 illustriert). Die Entwicklung der unterschiedlichen Stränge erfolgt vollkommen unabhängig voneinander (vgl. z.B. auch [19, S. 4]).

Aus diesem Grund wird, angelehnt an [3, S. 2], die folgende Notation vereinbart.

Notation. Der GWP mit initialer Populationsgröße $Z_0 \equiv l \geq 1$ (i.a.W. $\mathbb{P}(Z_0 = l) = 1$ bzw. $Z_0 = l$ f.s.) soll zukünftig als $\{Z_n^{(l)}\}_{n \geq 0}$ geschrieben werden. Sofern nicht anders angemerkt, wird ab sofort – für den Rest des vorliegenden Kapitels – immer der Prozess

$$\{Z_n\}_{n \geq 0} := \{Z_n^{(1)}\}_{n \geq 0} \quad (\text{für alle } n \geq 0 \text{ ist also } Z_n = Z_n^{(1)} \text{ und } Z_0 \equiv 1),$$

d.h. ein GWP mit der initialen Populationsgröße $Z_0 \equiv 1$ und der Nachkommenszahl X , behandelt. Beachte, dass $\{Z_n\}_{n \geq 0}$ gemäß dem Satz 2.25 $\text{Markov}(\delta_1, \mathbf{P})$ ist. $\triangle$

Da die diskrete Wahrscheinlichkeitsverteilung eindeutig durch die im Unterabschnitt 2.2.2 eingeführten wahrscheinlichkeitserzeugenden Funktionen (weF) charakterisiert ist (Satz 2.6), sollen nun die der Populationsgrößen Z_n ermittelt werden. Anschließend können die gewonnen Erkenntnisse dazu genutzt werden, um ihre Erwartungswerte und Varianzen zu bestimmen.

2.4.1 Wahrscheinlichkeitserzeugende Funktionen der individuellen Nachkommenszahl und der Populationsgrößen

Laut der Definition 2.20 sind alle Nachkommenszahlen *i.i.d.*-Kopien von X , es gilt demnach $X_{n,i} \stackrel{d}{=} X$ für alle $n \geq 0$ und $i \geq 1$, was mit (2.11) äquivalent zu $f_{X_{n,i}} = f_X$ ist. Ihre wahrscheinlichkeitserzeugenden Funktionen sind also alle gleich.

Weiters wurde oben die initiale (oder 'nullte') Populationsgröße auf $Z_0 \equiv 1$ festgelegt, weshalb $Z_1 \stackrel{(2.35)}{=} \sum_{i=1}^{Z_0} X_{0,i} \stackrel{Z_0 \equiv 1}{=} X_{0,1}$ gilt. D.h., die Nachkommenszahl des initialen Individuums $X_{0,1}$ entspricht – wenig überraschend – der Populationsgröße der ersten Generation Z_1 . Daher gilt zudem $f_{Z_1} = f_{X_{0,1}} = f_X$ (und $\mathbb{P}(Z_1 = k) = p_k$).

2 Einfache Galton-Watson-Verzweigungsprozesse

Notation 2.27 (WeF der Nachkommenszahl X)

Für alles Weitere in diesem Kapitel ist mit f (ohne Index) die weF von X bzw. Z_1 gemeint. D.h., zusammen mit dem soeben Beschriebenen und der Def. 2.4 gilt für alle $s \in [0, 1]$, $n \geq 0$ und $i \geq 1$

$$f(s) := f_X(s) \stackrel{X \stackrel{d}{=} X_{n,i}}{\stackrel{(i.i.d.)}{=}} f_{X_{n,i}}(s) \stackrel{X_{0,1} = Z_1}{\stackrel{(2.35)}{=}} f_{Z_1}(s) \stackrel{(2.6)}{\stackrel{(2.33)}{=}} \sum_{k \geq 0} p_k s^k = \mathbb{E}s^X. \quad (2.45) \quad \triangle$$

Nun interessiert aber nicht nur die erste Generation, sondern wünschenswert wäre es, über die wahrscheinlichkeitserzeugenden Funktionen für die Populationsgrößen aller weiteren verfügen zu können. Dabei helfen die iterierten weF des Unterabschnitts 2.2.2, genauer der Definition 2.18, und das folgende Resultat.

Satz 2.28 (WeF der Populationsgrößen Z_n – angelehnt an [19, Theorem I.4.1, S. 5])
Für alle Generationen $n \geq 0$ ist die weF der Populationsgröße Z_n die n -fach iterierte von f , also die Funktion f_n laut der Def. 2.18, und somit gilt für alle $s \in [0, 1]$

$$f_{Z_n}(s) = f_n(s) \quad \left(\stackrel{(2.30)}{=} \underbrace{f(f(f(\dots f(s)\dots)))}_{n\text{-mal}} \right). \quad (2.46)$$

Beweis. Analog zum Beweis der ersten Aussage des Korollars zur Def. 2.18, erhält man mit (2.6) – da $Z_0 \equiv 1$ gesetzt wurde – $f_{Z_0}(s) = \mathbb{E}s^{Z_0} = \mathbb{E}s^1 = s = f_0(s)$ für alle $s \in [0, 1]$ sowie mit (2.45), dass $f_{Z_1} = f = f_1$ ist. Um nun induktiv vorgehen zu können, betrachte $\{X_{n,i}\}_{i \geq 1}$ für ein $n \geq 1$. Diese sind *i.i.d.*, alle mit der weF f und (laut dem Lemma 2.22) insbesondere auch unabhängig von Z_n . Aufgrund dessen und der Rekursion (2.35) ($Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i}$) kann der Satz 2.14 über die Summe einer zufälligen Anzahl an *i.i.d.* $\mathbb{N}_0$ -wertigen Zufallsvariablen angewendet werden, um zu erhalten, dass

$$f_{Z_{n+1}} \stackrel{(2.26)}{=} f_{Z_n} \circ f \stackrel{\text{Ind.}}{\stackrel{\text{Ann.}}{=}} f_n \circ f \stackrel{(2.32)}{=} f_{n+1} \quad \text{gilt.} \quad \square$$

Die Rekursion (2.35) repräsentiert den Sprung von der n -ten in die $(n+1)$ -te Generation. So aufgefasst, steht die Populationsgröße Z_{n+1} für die Summe über die Nachkommen aller Individuen der n -ten Generation (in der nächsten Generation).

Es ist aber auch eine umgekehrte Betrachtungsweise der Situation möglich. Seien dafür zunächst $m \geq 0$ und $n \geq 1$. Von der $(n+m)$ -ten Generation aus in die Vergangenheit geblickt, kann der Prozess so interpretiert werden, dass in der m -ten Generation, in welcher ja Z_m Individuen existierten, von jedem dieser Individuen ein neuer, unabhängiger Prozess (mit initialer Populationsgröße 1) startete. Der Teil der Population der $(n+m)$ -ten Generation, welcher vom i -ten Individuum der m -ten Generation stammt, soll als $Z_{n,i}$ bezeichnet werden. Für alle $n \geq 1$ sind dabei jeweils die $\{Z_{n,i}\}_{i \geq 1}$ *i.i.d.*-Kopien von Z_n , alle mit der weF f_{Z_n} . Zudem sind sie auch unabhängig von Z_m (siehe das Lemma 2.22 – beachte, dass sie nur von den Nachkommenszahlen $\{X_{l,j}\}_{l \geq m, j \geq 1}$ abhängen). Aus diesen Gründen gilt für $m, n \geq 0$ (für $n = 0$ trivialerweise, da $Z_0 \equiv 1$ ist)⁴

$$Z_{n+m} \stackrel{d}{=} \sum_{i=1}^{Z_m} Z_{n,i} \quad \stackrel{\text{Satz 2.14}}{\Rightarrow} \quad f_{Z_{n+m}} \stackrel{(2.26)}{=} f_{Z_m} \circ f_{Z_n}. \quad (2.47)$$

⁴Eingeschränkt auf $n \geq 1$ stimmt die Aussage sogar für allgemeinere initiale Populationsgrößen.

2.4 Wahrscheinlichkeitsverteilung und Momente der Populationsgrößen

Damit bietet sich ein alternativer Beweis des Satzes 2.28 an (ohne, dass (2.31) bzw. (2.32) verwendet werden müssen – stattdessen können beide Gleichungen anschließend direkt aus (2.47) erhalten werden, welche sich mit (2.46) eben liest als $f_{n+m} = f_m \circ f_n = f_n \circ f_m$ bzw. als $f_{n+1} = f_n \circ f$ mit $m = 1$): Erneut induktiv, wobei der Induktionsanfang (also $f_{Z_0} = f_0$ bzw. $f_{Z_1} = f_1$) gleich wie im Beweis des Satzes ist, erhält man mit $m = 1$ $f_{Z_{n+1}} \stackrel{(2.47)}{=} f_{Z_1} \circ f_{Z_n} = f_1 \circ f_{Z_n} \stackrel{\text{Ind.-Ann.}}{=} f_1 \circ f_n \stackrel{(2.30)}{=} f \circ f_n \stackrel{(2.30)}{=} f_{n+1}$, was ebenfalls (2.46) zeigt.

Wird der Teil der Population der $(n+1)$ -ten Generation, welcher vom i -ten Individuum der *ersten* stammt, als $Z_{n,i}$ bezeichnet, so wird die linke Seite von (2.47) (wieder mit $m=1$) zu

$$Z_{n+1} \stackrel{d}{=} \sum_{i=1}^{Z_1} Z_{n,i} \quad \text{für alle } n \geq 0, \text{ wobei } \{Z_{n,i}\}_{i \geq 1} \text{ i.i.d.-Kopien von } Z_n \text{ sind.} \quad (2.48)$$

Das kann als Sprung von der ersten in die $(n+1)$ -te Generation aufgefasst werden.

Bemerkung 2.29 (Vorwärts- und Rückwärtsgleichung – vgl. [29, 3.1, S. 37 ff.]
Jedes Individuum des GWP ab der ersten Generation gehört also zu einem Teilprozess, welcher von einem Vorfahren einer vergangenen Generation gestartet wurde – insbesondere (mit $m = 1$) auch zu einem von einem Vorfahren der ersten Generation gestarteten Prozess. I.a.W.: Jedem Individuum der $(n+1)$ -ten Generation kann ein Ahne aus der ersten Generation zugeordnet werden. Aufgrund dieses rückwärtsgerichteten Blicks wird (2.48) etwa in oben genannter Referenz als *Rückwärtsgleichung* bezeichnet. Die Rekursion (2.35) hingegen heißt dort *Vorwärtsgleichung*, da sie den Umstand, dass die Individuen der n -ten Generation jene der $(n+1)$ -ten gebären, widerspiegelt. $\triangle$

Notation 2.30 (WeF der Populationsgrößen Z_n)

Dem Satz 2.28 entsprechend soll für die weF f_{Z_n} der Populationsgrößen Z_n ab sofort der Einfachheit halber meist nur f_n geschrieben werden, d.h

$$f_n(s) \stackrel{(2.46)}{=} f_{Z_n}(s) \stackrel{(2.6)}{=} \sum_{k \geq 0} \mathbb{P}(Z_n = k) s^k = \mathbb{E} s^{Z_n} \quad \text{für } s \in [0, 1]. \quad (2.49)$$

$\triangle$

Es soll kurz zur Situation mit einer initialen Populationsgröße $l \geq 1$ zurückgekehrt werden.

Korollar 2.31 (Größere initiale Populationsgrößen)

Sei $l \geq 1$. Der GWP $\{Z_n^{(l)}\}_{n \geq 0}$ (also mit initialer Populationsgröße $Z_0^{(l)} \equiv l$) ist die Summe von l i.i.d.-Kopien eines GWP $\{Z_n\}_{n \geq 0}$ (also mit $Z_0 \equiv 1$). Für die wahrscheinlichkeitserzeugende Funktion der $Z_n^{(l)}$ mit $n \geq 0$ gilt

$$f_{Z_n^{(l)}}(s) = (f_{Z_n}(s))^l \stackrel{(2.49)}{=} (f_n(s))^l \quad \text{für alle } s \in [0, 1]. \quad (2.50)$$

Beweis. Die erste Aussage begründet sich in der Unabhängigkeit der einzelnen Stränge, welche von den unterschiedlichen Individuen der initialen Generation ausgehen. Sie ist gleichbedeutend damit, dass $Z_n^{(l)}$ für jedes $n \geq 0$ jeweils die Summe über l i.i.d.-Kopien

2 Einfache Galton-Watson-Verzweigungsprozesse

von Z_n ist. Die Überlegung wurde im Text vor und in der Bemerkung 2.29 bereits etwas genauer ausgeführt (vgl. auch [3, *Additive Property*, S. 3]). Sie kann dadurch veranschaulicht werden, dass (2.47) auch für $Z_0^{(l)} \equiv l$ gültig bleibt ($n \geq 1$ vorausgesetzt). Der erste Teil liest sich bzgl. des Prozesses $\{Z_n^{(l)}\}_{n \geq 0}$ mit $m = 0$ als

$$Z_n^{(l)} \stackrel{d}{=} \sum_{i=1}^{Z_0^{(l)}} Z_{n,i}^{(Z_0^{(l)} \equiv l)} = \sum_{i=1}^l Z_{n,i}, \quad \text{wobei } \{Z_{n,i}\}_{i=1}^l \text{ i.i.d.-Kopien von } Z_n \text{ sind,} \quad (2.51)$$

sowie der zweite als $f_{Z_n^{(l)}} = f_{Z_0^{(l)}} \circ f_{Z_n}$. Mit $f_{Z_0^{(l)}}(s) = \sum_{k \geq 0} \mathbb{P}(Z_0^{(l)} = k) s^k \stackrel{(Z_0^{(l)} \equiv l)}{=} s^l$ wird der Beweis komplettiert. Alternativ kann wegen (2.51) auch das Lemma 2.10 herangezogen werden um $f_{Z_n^{(l)}}(s) \stackrel{(2.15)}{=} \prod_{i=1}^l f_{Z_{n,i}}(s) \stackrel{(\star)}{=} (f_{Z_n}(s))^l$ zu erhalten (($\star$): $Z_{n,i} \stackrel{d}{=} Z_n$ (*i.i.d.*)). $\square$

Bemerkungen zum Korollar 2.31.

- (i) Die weF $f_{Z_n^{(l)}}$ genügt *nicht* der Def. 2.18, ist also keine Iteration von f . Es gilt zwar $f_{Z_{n+1}^{(l)}}(s) \stackrel{(2.50)}{=} (f_{n+1}(s))^l \stackrel{(2.32)}{=} (f_n(f(s)))^l \stackrel{(2.50)}{=} f_{Z_n^{(l)}}(f(s))$, also eine Entsprechung zu (2.32). Aber $f_{Z_0^{(l)}}(s) = s^l \neq s$, $f_{Z_1^{(l)}}(s) = (f(s))^l \neq f(s)$ und im Allg. ist auch $f_{Z_{n+1}^{(l)}}(s) = f_{Z_n^{(l)}}(f(s)) \neq f(f_{Z_n^{(l)}}(s))$. Letzteres würde etwa gelten, wenn $X \equiv j \geq 0$, also konstant ist, denn dann ist $f(s) = s^j$ und daher $f(f_{Z_n^{(l)}}(s)) = (f_{Z_n^{(l)}}(s))^j = ((f_n(s))^l)^j = ((f_n(s))^j)^l = (f(f_n(s)))^l = (f_{n+1}(s))^l = f_{Z_{n+1}^{(l)}}(s)$.

Daran erkennt man, dass der Satz 2.28 i.A. nur dann gilt (und die Notationsvereinbarung 2.30 nur dann gerechtfertigt ist), wenn $Z_0 \equiv 1$ ist.

- (ii) Aufgrund der Aussage des Korollars entspricht die gemeinsame Verteilung von $Z_{n_1}^{(l)}, \dots, Z_{n_m}^{(l)}$ für $1 \leq n_1 \leq \dots \leq n_m$ und $m \geq 1$ der l -fachen Faltung der gemeinsamen Verteilung von $Z_{n_1}, \dots, Z_{n_m}$ – vgl. wieder [3, *Add. Property*, S. 3]. Insbesondere sieht man anhand von (2.51) mit $n = 1$ ($Z_{1,i} \stackrel{d}{=} Z_1 \stackrel{(2.35)}{=} X_{0,1}$) für $k \geq 0$, dass

$$\mathbb{P}(Z_1^{(l)} = k) \stackrel{(2.37)}{=} p_k^{*l} = p_{lk} \text{ gilt} \quad \Rightarrow \quad f_{Z_1^{(l)}}(s) \stackrel{(2.6)}{=} \sum_{k \geq 0} p_{lk} s^k \stackrel{(2.50)}{=} (f(s))^l. \quad (2.52)$$

Der zweite Teil von (2.52) kann verwendet werden, um mittels höherer Übergangswahrscheinlichkeiten (vgl. den Satz A.74) den Induktionsschritt im Beweis zum Satz 2.28 für den GWP $\{Z_n\}_{n \geq 0}$ auf eine weitere andere Art – so wie in [3] praktiziert – durchzuführen. Siehe dazu die Nebenrechnung B.5.

- (iii) **Übergangswahrscheinlichkeiten:** Allgemein, also unabhängig von der initialen Populationsgröße, erkennt man an der letzten Gleichheit in (2.52), dass für jedes $l \geq 0$ die $\{p_{lk}\}_{k \geq 0}$ – also die Übergangswahrscheinlichkeiten der Definition 2.24 – die Koeffizienten der l -ten Potenz der Reihe $(f(s))^l$ sind. $\triangle$

2.4.2 Momente der individuellen Nachkommenszahl und der Populationsgröße

Als Nächstes sollen der Erwartungswert und die Varianz der Populationsgrößen erarbeitet werden. Es wird sich zeigen, dass diese, abgesehen von der Generation n , ausschließlich von jenen der Nachkommenszahl der Individuen abhängen.

Wie bereits im vorherigen Unterabschnitt eruiert wurde, gilt für die Populationsgröße der ersten Generation $Z_1 = X_{0,1}$ und deshalb auch $\mathbb{E}Z_1 = \mathbb{E}X$ und $\mathbb{V}Z_1 = \mathbb{V}X$. Zunächst werden Variablen für diese definiert und durch Ableitungen der wahrscheinlichkeitserzeugenden Funktion der Nachkommenszahl beschrieben.

Definition 2.32 (Momente der Nachkommenszahl eines Individuums)

Definiere μ als den EW und σ^2 als die Varianz der Nachkommenszahl X , d.h.,

$$\mu := \mathbb{E}X \stackrel{(2.13)}{=} f'(1^-) \stackrel{\text{falls } \mu < \infty \text{ ist}}{=} f'(1) \quad \text{und} \quad (= \mathbb{E}Z_1) \quad (2.53)$$

$$\sigma^2 := \mathbb{V}X \stackrel{\text{falls } \mu < \infty \text{ ist}}{=} f''(1^-) + \underbrace{f'(1)}_{=\mu} - \underbrace{(f'(1))^2}_{=\mu} \stackrel{\text{falls } \sigma^2 < \infty \text{ ist}}{=} f''(1) + \mu - \mu^2 \quad (= \mathbb{V}Z_1). \quad (2.54)$$

Die weF der Populationsgröße Z_n einer beliebigen Generation $n \geq 0$ ist in Form der n -fachen Iteration von f , der weF f_n , gegeben (Satz 2.28 bzw. Notation 2.30). Somit kann eine Vorgehensweise, welche der in 2.2.2 (genauer bei (2.21) und dem darauf folgenden Text) ähnelt, gewählt werden, um den Erwartungswert von Z_n zu bestimmen.

Mit der Kettenregel gilt für die erste Ableitung der Rekursion aus (2.30) ($f_n = f \circ f_{n-1}$) und (mindestens) auf $[0, 1)$, dass $f'_n(s) = f'(f_{n-1}(s))f'_{n-1}(s)$ ist. Für $\mu < \infty$ erhält man – wieder mit dem Stetigkeitsargument – $f'_n(1^-) = f'(1)f'_{n-1}(1^-) \stackrel{(2.53)}{=} \mu f'_{n-1}(1^-)$. Diese Beziehung führt für $n \geq 2$ iterativ zu

$$\mathbb{E}Z_n \stackrel{(2.13)}{=} f'_n(1^-) = \mu f'_{n-1}(1^-) = \mu^2 f'_{n-2}(1^-) = \dots = \mu^n \overbrace{f'_{n-n}(1^-)}^{= f'_0(1^-) = \mathbb{E}Z_0 = 1} = \mu^n. \quad (2.55)$$

Analog dazu kann, nun $\sigma^2 < \infty$ vorausgesetzt, die zweite Ableitung f''_n ermittelt werden, um anschließend mit (2.14) die Varianz von Z_n zu bestimmen – hierzu sei auf die Nebenrechnung B.6 verwiesen.

Wie aber schon im Absatz vor dem Lemma 2.23 festgestellt, formuliert die Rekursion (2.35) für jedes $n \geq 1$ die Populationsgröße der $(n+1)$ -ten Generation als die Summe einer zufälligen Anzahl an $\mathbb{N}_0$ -wertigen *i.i.d.*-ZV (nämlich als die Summe der unmittelbaren Nachkommen aller Individuen der vorherigen Generation). Dafür steht der Satz 2.15 (über den EW und die Varianz einer solchen Summe) zur Verfügung, was für das nächste Resultat ausgenützt wird.

Satz 2.33 (Momente der Populationsgrößen Z_n – vgl. z.B. [19, I.5, S. 6])

Sei $n \geq 0$. Für den Erwartungswert und die Varianz von Z_n gilt

$$\begin{aligned} & \bullet \text{ falls } \mu < \infty \text{ ist,} & \mathbb{E}Z_n &= \mu^n & \text{und} & & (2.56) \\ & \bullet \text{ falls } \sigma^2 < \infty \text{ ist,} & \mathbb{V}Z_n &= \mathbb{E}Z_n^2 - (\mathbb{E}Z_n)^2 = \begin{cases} \frac{\sigma^2 \mu^n (\mu^n - 1)}{\mu(\mu - 1)} & \text{für } \mu \neq 1, \\ n\sigma^2 & \text{für } \mu = 1. \end{cases} & (2.57) \end{aligned}$$

2 Einfache Galton-Watson-Verzweigungsprozesse

Beweis. Betrachte für $n \geq 0$ erneut die Rekursion (2.35), $Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i}$. Deren Ausgangswerte lassen sich sofort bestimmen und sie erfüllen (2.56) bzw. (2.57). Es gilt

$$Z_0 \equiv 1 \Rightarrow \mathbb{E}Z_0 = 1, \mathbb{V}Z_0 = 0 \quad \text{und} \quad \mathbb{E}Z_1 = \mathbb{E}X \stackrel{(2.53)}{=} \mu, \mathbb{V}Z_1 = \mathbb{V}X \stackrel{(2.54)}{=} \sigma^2,$$

wobei $\mathbb{E}X < \infty \iff \mu < \infty$ und $\mathbb{E}X^2 < \infty \iff \sigma^2 < \infty$ gilt.

Sei nun $\mu < \infty$. Laut dem Satz 2.15 gilt, sofern zudem $\mathbb{E}Z_n < \infty$ gegeben ist,

$$\mathbb{E}Z_{n+1} \stackrel{(2.27)}{=} \mathbb{E}Z_n \mathbb{E}X \stackrel{(2.53)}{=} \mathbb{E}Z_n \mu \quad \text{für } n \geq 0. \quad (2.58)$$

Da aber $\mathbb{E}Z_1 = \mu < \infty$ ist, ist diese Voraussetzung für $n = 1$ erfüllt und folglich wegen (2.58) auch $\mathbb{E}Z_2 < \infty$. Dies kann induktiv immer weiter geführt werden. Daraus folgt $\mathbb{E}Z_n < \infty$ für alle $n \geq 0$. Also führt eine Iteration, welche analog zu der in (2.55) ist, zu $\mathbb{E}Z_n = \mathbb{E}Z_{n-1}\mu = \mathbb{E}Z_{n-2}\mu^2 = \dots = \mathbb{E}Z_0\mu^n = \mu^n$. Somit ist (2.56) für $n \geq 0$ gezeigt.

Für (2.57) sei nun zusätzlich $\sigma^2 < \infty$. Falls zudem auch $\mathbb{E}Z_n^2 < \infty$ vorausgesetzt wird – was genau dann der Fall ist, wenn $\mathbb{V}Z_n < \infty$ ist – besagt wieder der Satz 2.15, dass

$$\mathbb{V}Z_{n+1} \stackrel{(2.28)}{=} \mathbb{E}Z_n \mathbb{V}X + (\mathbb{E}X)^2 \mathbb{V}Z_n \stackrel{\text{Def. 2.32 u. (2.56)}}{=} \mu^n \sigma^2 + \mu^2 \mathbb{V}Z_n \quad \text{für } n \geq 0 \text{ gilt.} \quad (2.59)$$

Wie zuvor folgt damit aus $\mathbb{V}Z_1 = \sigma^2 < \infty$ auch, dass $\mathbb{V}Z_2 < \infty$ und weiters – mittels Induktion – für jedes $n \geq 0$, dass $\mathbb{V}Z_n < \infty$ und somit (2.59) verifiziert ist.

Betrachte nun zunächst den Fall $\mu = 1$. Dann wird (2.59) zu $\mathbb{V}Z_{n+1} = \sigma^2 + \mathbb{V}Z_n$, woraus für $n \geq 2$ iterativ $\mathbb{V}Z_n = \sigma^2 + \mathbb{V}Z_{n-1} = 2\sigma^2 + \mathbb{V}Z_{n-2} = \dots = (n-1)\sigma^2 + \underbrace{\mathbb{V}Z_1}_{=\sigma^2} = n\sigma^2$, also (2.57) für diesen Fall, folgt.

Für den restlichen Beweis sei $\mu \neq 1$. Die rekursive Beschreibung der Varianz $\mathbb{V}Z_n$ in (2.59) stellt eine sogenannte *Differenzengleichung* – genauer eine inhomogene lineare Differenzengleichung erster Ordnung – dar. Für diese Klasse von Gleichungen gibt es Lösungsverfahren, welche beispielsweise in [10, 7.2, S. 289 ff.] beschrieben werden. Ein geeignetes Verfahren wird nun kurz skizziert und anschließend durchgeführt. Schreibe dafür die allgemeine Gleichung (2.59) (zunächst ohne Anfangswerte) vereinfacht als

$$x_{n+1} - \mu^2 x_n = \sigma^2 \mu^n \quad \text{für } n \geq 0, \text{ wobei } x_n := \mathbb{V}Z_n \text{ ist.} \quad (\star)$$

Wird der sogenannte *Störterm* auf der rechten Seite auf 0 gesetzt, erhält man das homogene Pendant zur allgemeinen Differenzengleichung, d.h. $x_{n+1} - \mu^2 x_n = 0$ ($\star\star$).

Ohne gegebene Anfangswerte gibt es eine ganze Lösungsgesamtheit für x_n , welche die inhomogene Gleichung ($\star$) erfüllt. Eine beliebige Lösung daraus wird als $x_n^{(p)}$ geschrieben und *partikulär* genannt. Wird eine solche partikuläre Lösung und zudem *die allgemeine* Lösung der homogenen Gleichung ($\star\star$), bezeichnet als $x_n^{(h)}$, gefunden, so ist die Lösungsgesamtheit der Differenzengleichung mit $x_n = x_n^{(h)} + x_n^{(p)}$ (Δ) gegeben. Dass diese die Gleichung löst, sieht man im vorliegenden Fall folgendermaßen. Da sowohl $x_{n+1}^{(p)} \stackrel{(\star)}{=} \mu^2 x_n^{(p)} + \sigma^2 \mu^n$ als auch $x_{n+1}^{(h)} \stackrel{(\star\star)}{=} \mu^2 x_n^{(h)}$ erfüllt sein müssen, gilt $x_{n+1} \stackrel{(\Delta)}{=} x_{n+1}^{(h)} + x_{n+1}^{(p)} = \mu^2(x_n^{(h)} + x_n^{(p)}) + \sigma^2 \mu^n \stackrel{(\Delta)}{=} \mu^2 x_n + \sigma^2 \mu^n$. Für einen Beweis, dass

2.4 Wahrscheinlichkeitsverteilung und Momente der Populationsgrößen

damit *alle* Lösungen gefunden werden, vgl. etwa [10, Satz 7.5, S. 290]. Anschließend kann ein bekannter Anfangswert eingesetzt werden.

Die homogene Gleichung $(\star\star)$ ist mit $x_n = 0$ für alle $n \geq 0$ trivialerweise erfüllt. Ein geeigneter nichttrivialer Lösungsansatz stellt $x_n^{(h)} = c\lambda^n$ dar, wobei $c, \lambda \in \mathbb{R} \setminus \{0\}$ sind (für passende Ansätze siehe etwa erneut die o.g. Referenz). Einsetzen des Ansatzes in die Gleichung ergibt, dass $c\lambda^{n+1} - \mu^2 c\lambda^n = 0 \iff \lambda - \mu^2 = 0 \iff \lambda = \mu^2$ ist, weshalb die allgemeine Lösung durch $x_n^{(h)} = c\lambda^n = c(\mu^2)^n = c\mu^{2n}$ mit $c \in \mathbb{R}$ angegeben werden kann.

Mit $d \in \mathbb{R}$ kann für die partikuläre Lösung der Ansatz $x_n^{(p)} = d\mu^n$ gewählt werden. Wird dieser in die inhomogene Gleichung $(\star)$ eingesetzt, lässt sich folgern, dass $d\mu^{n+1} - \mu^2 d\mu^n = \sigma^2 \mu^n \iff d\mu - d\mu^2 = \sigma^2 \iff d\mu(1 - \mu) = \sigma^2 \iff d = \frac{\sigma^2}{\mu(1 - \mu)} = -\frac{\sigma^2}{\mu(\mu - 1)}$ gilt, womit $x_n^{(p)} = d\mu^n = -\frac{\sigma^2 \mu^n}{\mu(\mu - 1)}$ eine partikuläre Lösung darstellt.

Also ist durch $x_n \stackrel{(\Delta)}{=} x_n^{(h)} + x_n^{(p)} = c\mu^{2n} - \frac{\sigma^2 \mu^n}{\mu(\mu - 1)}$ mit $c \in \mathbb{R}$ die Lösungsgesamtheit der Differenzengleichung gegeben. Der Anfangswert $x_0 = \mathbb{V}Z_0 = 0$ ist bekannt, weshalb für die gesuchte Lösung $0 = x_0 = c\mu^0 - \frac{\sigma^2 \mu^0}{\mu(\mu - 1)}$ und folglich $c = \frac{\sigma^2}{\mu(\mu - 1)}$ gelten muss. Das ergibt schließlich, wie behauptet, $\mathbb{V}Z_n \stackrel{(\star)}{=} x_n = \frac{\sigma^2}{\mu(\mu - 1)} \mu^{2n} - \frac{\sigma^2 \mu^n}{\mu(\mu - 1)} = \frac{\sigma^2 \mu^n (\mu^n - 1)}{\mu(\mu - 1)}$. $\square$

Bemerkung zum Satz 2.33. Es gibt mehrere Möglichkeiten, zu (2.56) und (2.57) zu kommen. Eine Herleitung des Erwartungswertes mittels weF wurde vor dem Satz vorgeführt. Auch (2.58) stellt eine (homogene) Differenzengleichung dar, der iterative Ansatz erscheint hier aber intuitiver als das entsprechende Lösungsverfahren. Wie bereits erwähnt, findet sich eine alternative Herleitung der Varianz über die zweite Ableitung der weF durch Iteration in der Nebenrechnung B.6 und auch hier könnte stattdessen die Differenzengleichung für $f_n''(1^-)$ gelöst werden. Umgekehrt könnte aber auch der direkte Zugang über die Varianz wie im Beweis des Satzes, also (2.59), iteriert werden.

Mit ähnlichen Ansätzen könnten auch die höheren Momente von Z_n bestimmt werden (vgl. etwa den Text vor dem Korollar 2.8 oder auch die Anmerkung in [3, I.1.2., S. 4]). $\triangle$

Beispiel 2.34 (Momente bei größeren initialen Populationsgrößen)

Betrachte für $l \geq 1$ erneut den GWP $\{Z_n^{(l)}\}_{n \geq 0}$. Mit der Aussage des Korollars 2.31 und dem Satz 2.33 können der EW und die Varianz der $Z_n^{(l)}$ für $n \geq 0$ sofort angegeben werden. Da es sich um die Summe von l i.i.d.-Kopien eines GWP $\{Z_n\}_{n \geq 0}$ handelt, muss $\mathbb{E}Z_n^{(l)} = l\mathbb{E}Z_n = l\mu^n$ und (mit der *Bienaymé-Gleichung* – Satz A.55) $\mathbb{V}Z_n^{(l)} = l\mathbb{V}Z_n$ gelten.

Diese Resultate können aber auch, bei analoger Durchführung des Beweises zum Satz 2.33 mit nur geringfügigen Anpassungen, wie folgt erhalten werden. Für den EW beachte, dass (2.58) genauso gilt und nur der Anfangswert $\mathbb{E}Z_0^{(l)} = l$ adaptiert werden muss. Die darauf folgende Iteration führt dann zu $\mathbb{E}Z_n^{(l)} = \dots = \mathbb{E}Z_0^{(l)} \mu^n = l\mu^n$.

Der Anfangswert ($\mathbb{V}Z_0^{(l)} = 0$) der Entsprechung der Differenzengleichung (2.59) für die Varianz hingegen bleibt gleich, dafür muss $\mathbb{E}Z_n^{(l)}$ zu $l\mu^n$ angepasst werden. Deshalb lautet die Gleichung dann $\mathbb{V}Z_{n+1}^{(l)} = l\mu^n \sigma^2 + \mu^2 \mathbb{V}Z_n^{(l)}$ (bzw. $x_{n+1} - \mu^2 x_n = (l\sigma^2)\mu^n$) und kann $-\sigma^2$ ersetzt durch $l\sigma^2$ – genau auf die gleiche Weise gelöst werden. Im Fall $\mu = 1$

2 Einfache Galton-Watson-Verzweigungsprozesse

führt dies zu $\mathbb{V}Z_n^{(l)} = n l \sigma^2$ sowie zu $\mathbb{V}Z_n^{(l)} = \frac{l \sigma^2 \mu^n (\mu^n - 1)}{\mu(\mu - 1)}$, falls $\mu \neq 1$ ist – also jedenfalls zu $\mathbb{V}Z_n^{(l)} = l \mathbb{V}Z_n$. $\triangle$

Abschließend zu diesem Unterabschnitt wird noch eine Normalisierung der Populationsgrößen um ihren Erwartungswert vorgenommen und gezeigt, dass es sich bei dem so neu gewonnen Prozess um ein Martingal – vgl. den Unterabschnitt A.2.7 im Appendix – handelt.

Zunächst wird aber ein Hilfsresultat bzgl. der bedingten Erwartung der Populationsgröße einer Generation – gegeben der einer der vorherigen Generationen – formuliert.

Lemma 2.35 (angelehnt an [19, I.(8.2), S. 13]). *Für alle $n \geq 0$ gilt*

$$\mathbb{E}[Z_{n+1}|Z_n] = \mu Z_n \text{ f.s.} \quad (\text{insb. } \mathbb{E}[Z_{n+1}|Z_n=j] = \mu j \quad \forall j \geq 0: \mathbb{P}(Z_n=j) > 0) \text{ und} \quad (2.60)$$

$$\mathbb{E}[Z_{n+k}|Z_n] = \mu^k Z_n \text{ f.s.} \quad \text{für } k \geq 0. \quad (2.61)$$

Beweis. Die erste Aussage (2.60) steht in anderer Form bereits im Beweis des Satzes 2.15 unter (2.29) (wie bereits praktiziert, kann der Satz auf die Rekursion (2.35) angewendet werden). Beachte für $k \geq 1$ weiters, dass $\sigma(Z_n) \subseteq \sigma(Z_n, \dots, Z_{n+k-1})$ gilt ($\star$). Zudem ist Z_n definitionsgemäß (vgl. den Satz A.11 (i) oder (A.38)) messbar bzgl. $\sigma(Z_n)$ ($\star\star$). Deshalb ergibt eine wiederholte Anwendung von (2.60) (die Gleichheiten gelten großteils f.s.)

$$\begin{aligned} \mathbb{E}[Z_{n+k}|Z_n] &\stackrel{(\star)}{=} \underbrace{\mathbb{E}[\mathbb{E}[Z_{n+k}|Z_n, \dots, Z_{n+k-1}]|Z_n]}_{=\mathbb{E}[Z_{n+k}|Z_{n+k-1}] \text{ (ME (2.44))}} \stackrel{(2.60)}{=} \mathbb{E}[\mu Z_{n+k-1}|Z_n] \\ &\stackrel{(A.75)}{=} \mu \mathbb{E}[Z_{n+k-1}|Z_n] \stackrel{(k-1)\text{-mal analog}}{=} \mu^k \mathbb{E}[Z_n|Z_n] \stackrel{(\star\star)}{=} \mu^k Z_n \quad \text{f.s.} \quad \square \end{aligned}$$

Bemerkung zum Lemma 2.35. Der EW $\mathbb{E}Z_n$ aus (2.56) ist konsistent mit (2.61), denn es gilt ($Z_0 \equiv 1$ ist nichtzufällig) $\mathbb{E}Z_n = \mathbb{E}[Z_n|Z_0] = \mathbb{E}[Z_{n+0}|Z_0] \stackrel{(2.61)}{=} \mu^n Z_0 = \mu^n$. $\triangle$

Definition 2.36 (Normalisierte Populationsgrößen W_n)

Definiere für jedes $n \geq 0$ die um den Kehrwert ihres Erwartungswertes μ^n (siehe (2.56)) normalisierte Populationsgröße W_n als

$$W_n := \frac{Z_n}{\mu^n}. \quad \text{Es gilt} \quad \mathbb{E}W_n = \frac{\mathbb{E}Z_n}{\mu^n} \stackrel{(2.56)}{=} \frac{\mu^n}{\mu^n} = 1 \quad \text{für alle } n \geq 0. \quad (2.62)$$

Satz 2.37. *Der Prozess $\{W_n\}_{n \geq 0}$ der vorherigen Definition ist ein **nichtnegatives Martingal** bezüglich seiner natürlichen Filtration, aber auch bezüglich $\{\mathcal{F}_n\}_{n \geq 0}$ mit $\mathcal{F}_n := \sigma(Z_0, \dots, Z_n)$ – also der nat. Filtration des ursprünglichen Prozesses $\{Z_n\}_{n \geq 0}$.*

Beweis. Beachte zunächst, dass durch die Multiplikationen mit den nichtzufälligen Faktoren μ^{-n} die von den entsprechenden Zufallsvariablen erzeugten σ -Algebren nicht verändert werden. I.a.W.: Es gilt für alle $n \geq 0$

$$\sigma(Z_n) = \sigma(W_n) \quad \text{bzw. auch} \quad \sigma(Z_0, \dots, Z_n) = \sigma(W_0, \dots, W_n). \quad (2.63)$$

2.4 Wahrscheinlichkeitsverteilung und Momente der Populationsgrößen

Dies begründet sich intuitiv darin, dass im Zuge der Normalisierung kein Informationsverlust (oder auch Informationszugewinn) stattfindet. In der Nebenrechnung B.8 findet sich zudem ein formales Argument dafür.

Am Ende des Unterabschnitts 2.3.2 wurde unter (2.43) bereits die Markov-Eigenschaft für die bedingten Erwartungen der Z_n formuliert. Zusammen mit der Linearität des bedingten EW (Satz A.61 (c)) und dem vorherigen Lemma 2.35 kann für $n \geq 0$

$$\begin{aligned} \mathbb{E}[W_{n+1}|W_0, \dots, W_n] &\stackrel{(2.62)}{=} \frac{1}{\mu^{n+1}} \mathbb{E}[Z_{n+1}|Z_0, \dots, Z_n] \stackrel{(2.43)}{=} \frac{1}{\mu^{n+1}} \underbrace{\mathbb{E}[Z_{n+1}|Z_n]}_{=\mu Z_n \text{ f.s. mit (2.60)}} \\ &= \frac{1}{\mu^{n+1}} \mu Z_n \stackrel{(2.62)}{=} W_n \text{ f.s.} \quad \text{gezeigt werden.} \end{aligned}$$

Damit genügt $\{W_n\}_{n \geq 0}$ der Martingaleigenschaft (A.101) der Definition A.86 – und zwar sowohl bzgl. der natürlichen Filtration $\{\sigma(W_0, \dots, W_n)\}_{n \geq 0}$ (vgl. (A.102)) als auch bzgl. $\{\sigma(Z_0, \dots, Z_n)\}_{n \geq 0}$ (wegen (2.63)). Bereits in der vorherigen Def. 2.36 wurde $\mathbb{E}W_n = 1 < \infty$ festgestellt. Aufgrund der Nichtnegativität ist damit die weitere Anforderung, welche an ein Martingal gestellt wird, ebenfalls erfüllt ($\mathbb{E}|W_n| < \infty$ für alle $n \geq 0$ – im Sinne der Bemerkung zur Definition A.86 (ii) wäre es auch bereits ausreichend, $\mathbb{E}|W_0| = 1 < \infty$ zu überprüfen). $\square$

Bemerkung 2.38 (Markov-Eigenschaft für bedingte Erwartungen und Martingaleigenschaft – größere Schritte)

Die Martingaleigenschaft gilt für $\{W_n\}_{n \geq 0}$ lt. der Bem. zur Def. A.86 (i) auch bei höheren Generationssprüngen. Deshalb können die durch die obige Beweisführung gewonnenen Erkenntnisse und das Lemma 2.35 verwendet werden, um zu sehen, dass für alle $n \geq 0$ und $k \geq 1$

$$\mathbb{E}[W_{n+k}|W_n] \stackrel{(2.62)}{=} \frac{1}{\mu^{n+k}} \overbrace{\mathbb{E}[Z_{n+k}|Z_n]}^{=\mu^k Z_n \text{ f.s. (2.61)}} = \frac{Z_n}{\mu^n} \stackrel{(2.62)}{=} W_n \stackrel{(A.103)}{=} \mathbb{E}[W_{n+k}|W_0, \dots, W_n] \text{ f.s. } (\star)$$

gilt. Damit kann bzgl. des Prozesses $\{Z_n\}_{n \geq 0}$ und wieder für $n \geq 0$ und $k \geq 1$

$$\begin{aligned} \mathbb{E}[Z_{n+k}|Z_0, \dots, Z_n] &\stackrel{(2.62)}{=} \mu^{n+k} \mathbb{E}[W_{n+k}|W_0, \dots, W_n] \\ &\stackrel{(\star)}{=} \mu^{n+k} \mathbb{E}[W_{n+k}|W_n] \stackrel{(2.62)}{=} \mathbb{E}[Z_{n+k}|Z_n] \quad \text{erhalten werden.} \end{aligned}$$

Dies veranschaulicht die Markov-Eigenschaft auf der Ebene der bedingten Erwartungen als Erweiterung von (2.43) für größere Schritte (allgemeiner für Markov-Ketten und etwa basierend auf (A.89) könnte alternativ gleich der Satz A.72 entsprechend adaptiert werden). $\triangle$

2.4.3 Basis-Annahmen bzw. triviale Fälle

Der Definition 2.20 entsprechend ist grundsätzlich jede beliebige $\mathbb{N}_0$ -wertige Zufallsvariable zur Beschreibung der Nachkommenszahl eines (und somit jedes) Individuums denkbar. Dennoch werden in der Literatur oft, etwa um weniger interessante Fälle auszuschließen oder die Dinge nicht unnötig zu verkomplizieren, gewisse grundlegende Annahmen (kurz Ann.) getroffen. In den Standardwerken wird dies unterschiedlich gehandhabt. Etwa in [19, I.3, S. 5] werden diese zu Beginn recht pauschal formuliert (und bei Abweichungen darauf hingewiesen), in [3] hingegen werden sie eher erst an den relevanten Stellen vermerkt.

Hier sollen die nun folgenden Annahmen – wie in letzterer Referenz – erstmal nicht 'automatisch' gelten, sondern es soll bei Bedarf auf sie verwiesen werden. Manche Resultate gelten allgemeiner. Der Punkt (ii), welcher intuitiv praxisnah erscheint, tritt insbesondere im Abschnitt 2.6 dauerhaft in Kraft.

Annahmen 2.39 (bzgl. der Nachkommensverteilung – angelehnt an [19, I.3, S. 5])

(i) Ausschluss trivialer Fälle:

(a) Für alle $k \geq 0$ gelte $p_k \stackrel{(2.33)}{=} \mathbb{P}(X = k) < 1$.

Konsequenzen: Damit ist garantiert, dass die individuelle Nachkommenszahl dem Zufall unterworfen ist. Ansonsten wäre sie fix und der Prozess nichtzufällig. Etwa mit $p_0 = 1$ gäbe es keine Nachkommen und es wären alle Populationsgrößen ab der ersten Generation 0 (d.h., der Prozess würde sofort aussterben – vgl. 2.5.1 im nächsten Abschnitt). Wäre $p_k = 1$, so würde jedes Individuum genau ein neues gebären und der Prozess bliebe immer konstant. Es wäre also jede Populationsgröße gleich der anfänglichen – insbesondere beim Prozess $\{Z_n\}_{n \geq 0}$, da $Z_0 \equiv 1$ ist, immer 1. Mit $p_k = 1$ für irgendein $k > 1$ würde der Prozess sicher mit einer konstanten Rate wachsen.

(b) Es gelte $p_0 + p_1 = \mathbb{P}(X = 0 \text{ oder } X = 1) < 1$.

Konsequenzen: Mathematisch gesehen ist f dann gemäß dem Satz 2.5 (e) *strikt* isoton und *strikt* konvex. Wäre $p_0 + p_1 = 1$, so hätte jedes Individuum maximal einen unmittelbaren Nachkommen. Der Prozess bliebe eine Weile lang gleich – insbesondere mit $Z_0 \equiv 1$ währenddessen durchwegs im Zustand 1 – und würde irgendwann, sofern $p_0 > 0$ ist, aussterben (vielleicht auch schon in der ersten Generation).

(ii) Endlicher Erwartungswert: Es gelte $\mu = \mathbb{E}X < \infty$. △

Insbesondere der Fall $p_1 = 1$ muss besonders oft ausgeschlossen werden, weshalb dem gegensätzlichen Fall die nächste Definition gewidmet wird.

Definition 2.40 (Nichttrivialer Fall)

Der Fall $p_1 = \mathbb{P}(X = 1) = 1$ – also wenn die Annahme 2.39 (i) (a) für $k = 1$ verletzt ist – soll als *der triviale Fall* bezeichnet werden.

Umgekehrt wird bei $p_1 < 1$ von einem *nichttrivialen Fall* gesprochen.

2.5 Aussterbe- bzw. Überlebenswahrscheinlichkeiten

Um die grundlegende Frage nach der Wahrscheinlichkeit des Aussterbens des gegenständlichen Verzweigungsprozesses (welche sich bereits GALTON und WATSON gestellt haben) zu beantworten, soll zunächst definiert werden, was das Aussterben und die korrespondierende Wahrscheinlichkeit dazu überhaupt im mathematischen Sinne bedeuten.

2.5.1 Aussterben des Prozesses

Der Prozess kann als *ausgestorben* bezeichnet werden, wenn ab irgendeiner Generation $n \geq 1$ alle Populationsgrößen Null sind, d.h., für alle $m \geq 0$ gilt $Z_{n+m} = 0$. Das führt zu der folgenden, bspw. von [19] getroffenen, grundlegenden Definition des Aussterbens.

Definition 2.41 (Aussterben des Prozesses – angelehnt an [19, Def. I.6.1, S. 7])
Aussterben des Prozesses ist das Ereignis, dass für fast alle Folgenglieder von $\{Z_n\}_{n \geq 0}$ $Z_n = 0$ gilt. I.a.W.: Alle bis auf endlich viele bzw. ab einem gewissen Index sind 0, also

$$E := \{\text{'Prozess stirbt aus'}\} = \{Z_n = 0 \text{ für fast alle } n\} \stackrel{(A.5)}{=} \liminf_{n \rightarrow \infty} \{Z_n = 0\}. \quad (2.64)$$

So wie der GWP definiert ist – insbesondere durch (2.35) – kann es ausgehend von der Populationsgröße $Z_n = 0$ in einer beliebigen Generation $n \geq 1$ keine weiteren Nachkommen mehr geben. Dieser Umstand wurde auch bereits im Satz 2.26 (i) genauer beschrieben (0 ist ein absorbierender Zustand – $\mathbb{P}(Z_{n+1} = 0 | Z_n = 0) = p_{00} = 1$). Deshalb ist bereits das Auftreten des Ereignisses $\{Z_n = 0\}$ für irgendein $n \geq 1$ äquivalent mit dem Aussterbeereignis und es impliziert alle Ereignisse $\{Z_{n+m} = 0\}$ für alle $m \geq 0$. Daraus folgt

$$\{Z_1 = 0\} \subseteq \{Z_2 = 0\} \subseteq \{Z_3 = 0\} \subseteq \dots \text{ bzw. } \{Z_n = 0\} \subseteq \{Z_{n+1} = 0\} \quad \forall n \geq 1. \quad (2.65)$$

Demnach ist $\{\{Z_n = 0\}\}_{n \geq 1}$ eine monoton wachsende Mengenfolge. In [3, S. 1] etwa wird das Aussterben auch gleich über das Erreichen des Zustandes 0 in irgendeiner Generation definiert. Die Äquivalenz zeigt sich durch (vgl. dazu ggf. auch [11, Lemma 2.1, S. 9] über die Konvergenz monotoner Mengenfolgen)

$$E \stackrel{(2.64)}{\stackrel{(A.4)}{=}} \bigcup_{n \geq 1} \underbrace{\left(\bigcap_{m \geq n} \{Z_m = 0\} \right)}_{=\{Z_n=0\} \text{ wegen (2.65)}} = \bigcup_{n \geq 1} \{Z_n = 0\} = \{\exists n \geq 1 : Z_n = 0\}. \quad (2.66)$$

= $\{Z_n=0\}$ wegen (2.65) – beachte, dass allg. aus $A \subseteq B$ auch $A \cap B = A$ folgt

2.5.2 Aussterbe- und Überlebenswahrscheinlichkeit

Nun kann die Wahrscheinlichkeit des eingeführten Aussterbeereignisses (und die des entsprechenden Gegenereignisses) definiert werden.

Definition 2.42 (Aussterbe- und Überlebenswahrscheinlichkeit)

Die *Aussterbewahrscheinlichkeit* η des GWP wird definiert als

$$\begin{aligned} \eta := \mathbb{P}(E) &\stackrel{(2.64)}{=} \mathbb{P}(\text{'Prozess stirbt aus'}) \stackrel{(2.66)}{=} \mathbb{P}\left(\bigcup_{n \geq 1} \underbrace{\{Z_n = 0\}}_{\substack{\text{wachsend} \\ \text{in } n \text{ (2.65)}}}\right) \\ &\stackrel{\text{Stetigk. v. unten (A.31)}}{=} \lim_{n \rightarrow \infty} \mathbb{P}(Z_n = 0) \stackrel{(2.9)}{=} \lim_{n \rightarrow \infty} f_n(0). \end{aligned} \quad (2.67)$$

Die Gegenwahrscheinlichkeit zur Aussterbewahrscheinlichkeit ist die *Überlebenswahrscheinlichkeit* und kann mit $\mathbb{P}(E^c) = 1 - \eta$ berechnet werden.

Bemerkung zur Definition 2.42. Die Folge $\{f_n(0)\}_{n \geq 0} \stackrel{(2.9)}{=} \{\mathbb{P}(Z_n = 0)\}_{n \geq 0}$ ist *monoton wachsend in n* , d.h., für alle $n \geq 0$ ist $f_n(0) = \mathbb{P}(Z_n = 0) \leq \mathbb{P}(Z_{n+1} = 0) = f_{n+1}(0)$. Das ist (mit dem Satz A.23 (ii) und $\{Z_0 = 0\} = \emptyset$) eine direkte Konsequenz aus (2.65). Alternativ kann mittels den weF f_n eingesehen werden, dass aufgrund ihrer Isotonie $f_n(0) \leq f_n(p_0) \stackrel{(2.9)}{=} f_n(f(0)) \stackrel{(2.32)}{=} f_{n+1}(0)$ für jedes $n \geq 0$ gilt (wobei im Fall $p_0 > 0$ die Isotonie und somit die Ungleichung bzw. das monotone Wachstum strikt sind).

Zudem ist die Folge klarerweise auch beschränkt (von unten durch 0 und von oben durch 1), womit durch das Monotonieprinzip (vgl. den Satz A.93) garantiert ist, dass sie konvergiert und der Grenzwert $\eta = \lim_{n \rightarrow \infty} f_n(0) (= \sup_{n \geq 0} f_n(0))$ in $[0, 1]$ existiert. $\triangle$

2.5.3 Bestimmung der Aussterbewahrscheinlichkeit und Kritikalität des GWP

Als Nächstes soll versucht werden, eine Methode zur Bestimmung dieser Aussterbewahrscheinlichkeit, von nun an auch nur kurz als *AWK* geschrieben, zu finden. Zudem wird eine Terminologie geboten, mit welcher in Abhängigkeit des Erwartungswertes der Nachkommenszahl μ eine Klassifizierung vorgenommen werden kann.

Zunächst wird hierzu festgestellt, dass sich mit der weF f eine Fixpunktgleichung für die Aussterbewahrscheinlichkeit aufstellen lässt.

Lemma 2.43 (Fixpunktgleichung für die Aussterbewahrscheinlichkeit)

Die Aussterbewahrscheinlichkeit η ist der kleinste Fixpunkt von f in $[0, 1]$. Demnach gilt

$$f(\eta) = \eta. \quad (2.68)$$

Beweis. In der Bem. zur Def. 2.42 wurde bereits dargelegt, dass der Grenzwert η der Folge $f_n(0)$ existiert. Da also $\lim_{n \rightarrow \infty} f_n(0) = \lim_{n \rightarrow \infty} f_{n+1}(0) \stackrel{(2.67)}{=} \eta$ gilt, folgt

$$\left. \begin{aligned} f_{n+1}(0) &\nearrow \eta \\ f_{n+1}(0) &\stackrel{(2.30)}{=} f(\underbrace{f_n(0)}_{\substack{\nearrow \eta \\ f \text{ stetig}}}) \nearrow f(\eta) \end{aligned} \right\} \Rightarrow f(\eta) = \eta.$$

D.h., η ist ein Fixpunkt von f . Um zu zeigen, dass es sich dabei auch um den kleinsten Fixpunkt in $[0, 1]$ handelt, sei – unabhängig davon, wie viele es gibt – ψ ein beliebiger Fixpunkt. Es gilt also $f(\psi) = \psi$ und $f_0(0) = 0 \leq \psi \in [0, 1]$. Für alle $n \geq 0$ folgt daraus induktiv $f_n(0) \leq \psi$, denn wegen der Isotonie von f ist $f_{n+1}(0) \stackrel{(2.30)}{=} f(\underbrace{f_n(0)}_{\leq \psi}) \leq f(\psi) = \psi$. Also ist $\lim_{n \rightarrow \infty} f_n(0) = \eta \leq \psi$ der kleinste Fixpunkt. $\square$
 $\leq \psi$ (Ind.-Ann.)

Handelt es sich um den trivialen Fall (Def. 2.40), also $p_1 = 1$ (und somit $\mu = \mathbb{E}X = 1$), lebt in jeder Generation genau ein Individuum und die AWK ist Null. Der Punkt (i) des Fixpunktsatzes 2.17 bestätigt, dass diese Beobachtung mit dem Lemma konform geht, indem er besagt, dass alle $s \in [0, 1]$ Fixpunkte und demnach $\eta = 0$ der kleinste davon ist.

Für alle anderen Fälle kann ebenfalls der erwähnte Fixpunktsatz für weF, Satz 2.17, herangezogen werden, um in Abhängigkeit der erwarteten bzw. der mittleren Nachkommenszahl μ den kleinsten Fixpunkt von f zu bestimmen oder zumindest einzugrenzen.

Satz 2.44 (Kritikalität des Prozesses – vgl. z.B. [3, S. 8 f.])

Sei $p_1 < 1$ (nichttrivialer Fall, Def. 2.40; für (i) und (iii) ohnehin vorausgesetzt). Es gilt

- (i) wenn $\mu < 1$ ist, so ist $\eta = 1$ und der Prozess heißt **subkritisch**,
- (ii) wenn $\mu = 1$ ist, so ist $\eta = 1$ und der Prozess heißt **kritisch** und
- (iii) wenn $\mu > 1$ ist, so ist $\eta < 1$ und der Prozess heißt **superkritisch**.

Beweis. Wie angekündigt, sind die Aussagen Konsequenzen des Fixpunktsatzes 2.17.

Im subkritischen (i) und kritischen (ii) Fall ist $f'(1) \stackrel{(2.13)}{=} \mathbb{E}X = \mu \leq 1$ (und $p_1 < 1$ lt. Voraussetzung – insbesondere benötigt für letzteren Fall). Deshalb kommt der Punkt (ii) des Fixpunktsatzes zum Tragen. Dieser besagt, dass 1 der einzige Fixpunkt in $[0, 1]$ ist und das vorherige Lemma 2.43 liefert dann $\eta = 1$.

Laut dem Punkt (iii) des Fixpunktsatzes gibt es im superkritischen Fall (iii), in welchem $f'(1^-) = \mu > 1$ ist, genau einen weiteren Fixpunkt in $[0, 1)$. Dieser ist der kleinste, weshalb gemäß dem Lemma 2.43 $\eta < 1$ folgt. $\square$

In Abhängigkeit von $p_0 = \mathbb{P}(X = 0) \stackrel{(2.9)}{=} f(0)$ lässt sich noch etwas mehr über die Aussterbewahrscheinlichkeit aussagen.

Korollar zum Satz 2.44.

- Die Aussterbewahrscheinlichkeit η ist genau dann 0, wenn $p_0 = 0$ ist. Andererseits bedeutet das: $\eta > 0 \iff p_0 > 0$.
- Wird $p_0 > 0$ vorausgesetzt, so gilt im superkritischen Fall $p_0 < \eta < 1$.

Beweis. Im Fall $p_0 = 0$ hat jedes Individuum mindestens einen Nachfahren, weshalb auch intuitiv klar ist, dass dann $\eta = 0$ sein muss. Außerdem ist das dann die kleinste Lösung von (2.68), d.h. $f(0) = p_0 = 0$, weshalb diese Intuition durch das Lemma 2.43 bestätigt wird. Umgekehrt ist diese Fixpunktgleichung in 0 nur erfüllbar, wenn $p_0 = 0$ ist.

Die obere Schranke im zweiten Punkt ist im superkritischen Fall durch den Satz klar. Für die untere beachte, dass wegen der ersten Aussage $\eta > 0$ ist. Da zudem f strikt isoton ist (siehe den Satz 2.5 (e); $p_0 < 1$ folgt aus $\mu > 1$), gilt deshalb $\eta = f(\eta) > f(0) = p_0$. $\square$

Die Funktionenfolge $\{f_n(s)\}_{n \geq 0}$ konvergiert aber nicht nur an der Stelle 0 – an welcher ihre Glieder $f_n(0) = \mathbb{P}(Z_n = 0)$ ja die WK dafür sind, dass der Prozess in der Generation n nicht mehr lebt – gegen die Aussterbewahrscheinlichkeit η . Es soll nun noch gezeigt werden, dass das fast auf ihrem gesamten Definitionsbereich, genauer auf $[0, 1)$, so ist. Dazu wird zunächst die folgende Hilfsüberlegung angestellt.

2 Einfache Galton-Watson-Verzweigungsprozesse

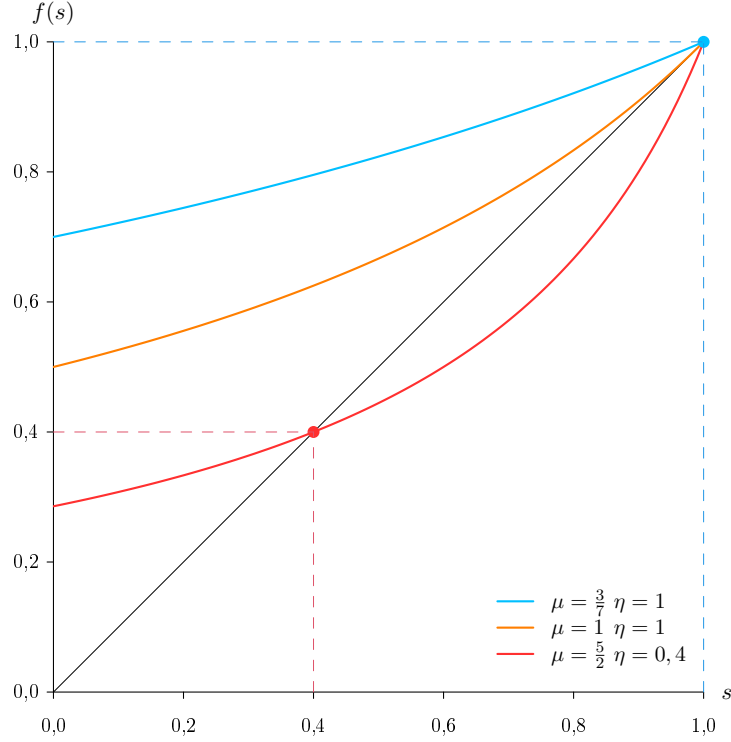


Abbildung 2.1: Darstellungen der weF für GW-Prozesse mit geometrischer Nachkommensverteilung (siehe den Abschnitt 4.1). Die oberste Kurve gehört zu einem **subkritischen** Prozess mit $\mu = \frac{3}{7}$, die mittlere zu einem **kritischen** mit $\mu = 1$. Die unterste repräsentiert die Nachkommensverteilung eines **superkritischen** geometrischen Prozesses mit $\mu = \frac{5}{2}$.

Lemma 2.45.

- (i) Falls $s \in [0, \eta)$ ist, gilt $f(s) > s$.
- (ii) Falls $s \in (\eta, 1)$ und $p_1 < 1$ ist (nichttrivialer Fall), gilt $f(s) < s$.

Beides gilt (unabhängig von μ) für alle Werte von $\eta \in [0, 1]$, wodurch die Intervalle in (i) oder (ii) u.U. leer ($\emptyset$) sein können.

Beweis.

- (i) Dieser Fall ist nur relevant, wenn $\eta > 0$ ist (da ansonsten $[0, \eta) = \emptyset$ wäre), woraus $p_0 > 0$ mit dem Korollar zum Satz 2.44 folgt. Anschaulich (siehe die Abbildung 2.1) liegt der Graph 'bis η ' über der ersten Mediane, wodurch die Aussage plausibel wird. Um sie auch analytisch zu zeigen, soll – ähnlich wie im Beweis zum Satz 2.17 (iii) – für $s \in [0, \eta]$ die Funktion $g(s) := f(s) - s$ definiert werden. Für diese gilt $g(0) = f(0) - 0 = p_0 > 0$ und $g(\eta) = f(\eta) - \eta = 0$ (2.68). Da η der kleinste Fixpunkt in $[0, 1]$ ist, muss für alle $s \in [0, \eta)$ weiters $g(s) = f(s) - s \neq 0$ sein ($\star$).

Daraus folgt $g(s) > 0$ und somit $f(s) > s$ für alle $s \in [0, \eta)$, denn: Angenommen, es gäbe ein $s_0 \in (0, \eta)$ mit $g(s_0) < 0$. Dann hätte die stetige Funktion g laut dem Nullstellensatz A.90 eine Nullstelle $t_0 \in (0, s_0)$, also mit $g(t_0) = 0$, was im Widerspruch zu $(\star)$ stünde.

- (ii) Dieser Fall ist nur relevant, wenn $\eta < 1$ ist (ansonsten wäre $(\eta, 1) = \emptyset$). Das ist wegen des Satzes 2.44 (der nichttriviale Fall $p_1 = 1$ wurde ausgeschlossen) nur möglich, wenn $f'(1^-) = \mu > 1$ ist. Laut dem Lemma 2.16 (i) ist f demnach strikt konvex. Die Aussage wurde für Werte von s , die hinreichend nahe bei 1 sind, bereits unter (iii) des besagten Lemmas gezeigt und wird anschaulich dadurch klar, dass der Graph von f 'rechts von η ' unterhalb der ersten Mediane liegt. Für einen analytischen Beweis kann die Konvexitätsdef. A.89 herangezogen werden. Wähle dafür für jedes $s \in [\eta, 1]$ ein $\lambda(s) := \frac{1-s}{1-\eta}$. Die Funktion $\lambda: [\eta, 1] \rightarrow [0, 1]$ ist strikt monoton fallend (da $\lambda' = -\frac{1}{1-\eta} < 0 \forall \eta \in [0, 1]$ ist) und es gilt $\lambda(\eta) = 1$ sowie $\lambda(1) = 0$. Daraus folgt für $s \in (\eta, 1)$, dass $\lambda(s) \in (0, 1)$ ist. Umgekehrt ist $s = 1 - (1-\eta)\lambda = \eta\lambda + 1(1-\lambda)$ $(\star\star)$. Da f strikt konvex ist, gilt nun

$$f(s) \stackrel{(\star\star)}{=} f(\eta\lambda + 1(1-\lambda)) \stackrel[\text{-def. A.89}]{\text{Konvex.}} \underbrace{f(\eta)\lambda}_{= \eta} + \underbrace{f(1)(1-\lambda)}_{= 1} \stackrel{(\star\star)}{=} s \quad \forall s \in (\eta, 1). \quad \square$$

Mithilfe des Lemmas kann die Konvergenz der f_n beschrieben werden.

Satz 2.46 (Konvergenz der Folge $\{f_n\}_{n \geq 0}$ – angelehnt an [3, Lemma I.1.2, S. 5])
 Sei wieder $p_1 < 1$ (nichttrivialer Fall). Unabhängig von μ gilt

$$\lim_{n \rightarrow \infty} f_n(s) = \eta \quad \text{für alle } s \in [0, 1). \text{ Dabei gilt}$$

- (i) immer (d.h. $\forall n \geq 0$ und auch für $\eta = 1$) $f_n(\eta) = \eta$ und somit $f_n(\eta) \rightarrow \eta$ und
- (ii) falls $s \in [0, \eta) \neq \emptyset$ ist, $f_n(s) \nearrow \eta$ (d.h., $\{f_n(s)\}_{n \geq 0}$ ist monoton wachsend) sowie
- (iii) falls $s \in (\eta, 1) \neq \emptyset$ ist, $f_n(s) \searrow \eta$ (d.h., $\{f_n(s)\}_{n \geq 0}$ ist monoton fallend).

Beweis. Der Punkt (i) ist (mit (2.68)) jedenfalls klar – vgl. auch das Lemma 2.19 (i).

Falls $p_0 = 1$ ist, folgt $\eta = 1$ und weiters, dass f sowie alle f_n mit $n \geq 1$ auf ganz $[0, 1]$ konstant 1 sind, weshalb dann alle Aussagen stimmen.

Sei also $p_0 < 1$. Dann sind laut dem Lemma 2.19 (ii) alle f_n mit $n \geq 0$ strikt isoton $(\star)$.

- (ii) Sei $s \in [0, \eta)$. Mit dem Lemma 2.45 (i) gilt, dass $f_1(s) \stackrel{(2.30)}{=} f(s) > s$ und deshalb auch für alle $n \geq 0$, dass $f_{n+1}(s) \stackrel{(2.30)}{=} f(f_n(s)) > f_n(s)$ ist. Es gilt also (für $n > 2$)

$$0 \leq s \stackrel{(2.30)}{=} f_0(s) < f_1(s) < f_2(s) < \dots < f_n(s) \stackrel{(\star)}{<} f_n(\eta) \stackrel{(i)}{=} \eta.$$

D.h., für alle $n \geq 0$ und alle $s \in [0, \eta)$ gilt $f_n(s) \in [0, \eta)$ und $\{f_n(s)\}_{n \geq 0}$ ist monoton wachsend. Gemäß dem Monotonieprinzip (Satz A.93) gibt es darum ein ψ mit

$$f_n(s) \nearrow \psi = \sup_{n \geq 0} f_n(s) \leq \sup([0, \eta)) = \eta \quad \text{für alle } s \in [0, \eta). \quad (\star\star)$$

2 Einfache Galton-Watson-Verzweigungsprozesse

- (iii) Sei $s \in (\eta, 1)$. Ähnlich wie soeben liefert nun das Lemma 2.45 (ii) $f(s) < s \Rightarrow f_{n+1}(s) = f(f_n(s)) < f_n(s) \forall n \geq 0 \Rightarrow 1 > s = f_0(s) > f_1(s) > \dots > f_n(s) \xrightarrow{(*)} f_n(\eta) = \eta$. Also gilt für alle $n \geq 0$ und $s \in (\eta, 1)$, dass $f_n(s) \in (\eta, 1)$ und $\{f_n(s)\}_{n \geq 0}$ monoton fallend ist. Wieder laut dem Monotonieprinzip folgt, dass ein ψ existiert mit

$$f_n(s) \searrow \psi = \inf_{n \geq 0} f_n(s) < \sup((\eta, 1)) = 1 \quad \text{für alle } s \in (\eta, 1). \quad (***)$$

Also gilt in jedem Fall (d.h. für alle $s \in [0, 1)$), dass es ein ψ mit $\lim_{n \rightarrow \infty} f_n(s) = \psi$ gibt. Analog zum Beweis des Lemmas 2.43 gilt aber wegen der Stetigkeit von f auch $\psi = \lim_{n \rightarrow \infty} f_{n+1}(s) = \lim_{n \rightarrow \infty} f(f_n(s)) = f(\lim_{n \rightarrow \infty} f_n(s)) = f(\psi)$, weshalb ψ ein Fixpunkt von f sein muss. Gemäß $(\star\star)$ ist im Fall (ii) $\psi \leq \eta$ und es muss, da η der kleinste Fixpunkt ist (Lemma 2.43), $\psi = \eta$ gelten. Aber auch im – nur für $\eta < 1$ relevanten – Fall (iii), in welchem $\psi < 1$ ist $(\star\star\star)$, muss $\psi = \eta$ gelten, denn lt. dem Satz 2.17 (iii) gibt es genau (also nur) einen Fixpunkt in $[0, 1)$. $\square$

Bemerkung zum Satz 2.46. Die bereits im Rahmen der Def. 2.42 gezeigte bzw. im Beweis zum Lemma 2.43 verwendete Konvergenz von $\{f_n\}_{n \geq 0}$ an der Stelle 0 gegen η stellt also einen Spezialfall des allgemeineren Satzes 2.46 dar.

An der Stelle 1 konvergiert $\{f_n\}_{n \geq 0}$ *immer* gegen 1 (vgl. etwa wieder das Lemma 2.19 (i) – oder gleich (2.8) für jedes f_n). Für $\eta < 1$ ist das also der einzige Punkt des Definitionsbereichs, an dem die Funktionenfolge *nicht* gegen die Aussterbe-WK konvergiert (ansonsten gilt das auf ganz $[0, 1]$). $\triangle$

Als direkte Konsequenz des allgemeinen Fixpunktsatzes für weF (Satz 2.17 unter Mitberücksichtigung des Lemmas 2.19 (i)), der Fixpunktgleichung des Lemmas 2.43 und des Satzes 2.44 lässt sich noch die folgende zusammenfassende Aussage formulieren.

Korollar 2.47 (Fixpunkte der weF f bzw. f_n)

Im nichttrivialen Fall ($p_1 < 1$) sind η und 1 die **einzigen Fixpunkte von f und auch von allen f_n mit $n \geq 0$.**

Im superkritischen Fall ($\mu > 1, \eta < 1$) handelt es sich dabei um genau zwei verschiedene Fixpunkte, im subkritischen und kritischen Fall ($\mu \leq 1, \eta = 1$) fallen diese zusammen und 1 ist der einzige Fixpunkt.

2.5.4 Instabilität des Prozesses

Im Unterabschnitt 2.5.1 – genauer in der Aussterbe-Definition 2.41 – wurde das Aussterben als das Ereignis, dass ab einem gewissen Index alle $Z_n = 0$ sind, beschrieben. Aufgrund der Ganzzahligkeit der Folgenglieder Z_n ist deren Konvergenz gegen Null äquivalent dazu, d.h.

$$E_{(2.64)} = \{Z_n = 0 \text{ für fast alle } n\} \stackrel{(*)}{=} \left\{ \lim_{n \rightarrow \infty} Z_n = 0 \right\} = \{Z_n \rightarrow 0\}. \quad (2.69)$$

- ($\star$) Für ein formales Argument beachte, dass die linke Seite der Limes inferior von $\{Z_n = 0\}$ ist (siehe etwa erneut (2.64) oder (A.5)). Somit kann die in der bereits

verwendeten NR B.4 durchgeführte Rechnung (B.1) (welche auch für $k = 0$ gilt) erneut herangezogen werden.

Dies führt zu der nun folgenden Aussage über das asymptotische Verhalten des GWP.

Satz 2.48 (Instabilität des Prozesses – vgl. z.B. [19, Theorem I.6.2, S. 8])

Seien $p_1 < 1$ (nichttrivialer Fall – Def. 2.40) und $\mu < \infty$ (Ann. 2.39 (ii)). Der GW-Prozess bzw. die Folge $\{Z_n\}_{n \geq 0}$ konvergiert entweder gegen 0 oder strebt gegen ∞ . Dabei gilt

$$\mathbb{P}(Z_n \rightarrow 0) = \eta \quad \text{und} \quad \mathbb{P}(Z_n \rightarrow \infty) = 1 - \eta. \quad (2.70)$$

Beweis. Der erste Teil von (2.70) ist eine direkte Konsequenz aus (2.69) und der Definition der AWK (Definition 2.42). Weiters ist aus dem Satz 2.26 (ii) für alle Zustände $k \geq 1$ ($p_1 < 1$ vorausgesetzt) bekannt, dass sowohl $\mathbb{P}(Z_n \rightarrow k) = 0$ (2.42) als auch $\mathbb{P}('Z_n = k \text{ für } \infty \text{ viele } n') = 0$ (2.40) gilt. Da also kein Zustand $k \neq 0$ unendlich oft erreicht werden kann, muss (f.s.) entweder das Ereignis, dass Z_n gegen 0 konvergiert (also $\{Z_n \rightarrow 0\}$) oder das Ereignis, dass Z_n gegen ∞ strebt (also $\{Z_n \rightarrow \infty\}$) eintreten, d.h. $\mathbb{P}(\{Z_n \rightarrow 0\} \cup \{Z_n \rightarrow \infty\}) = 1$ (was in der NR B.7 zusätzlich explizit gezeigt wird). Da sich diese beiden Ereignisse gegenseitig ausschließen, gilt deshalb auch $\mathbb{P}(Z_n \rightarrow 0) + \mathbb{P}(Z_n \rightarrow \infty) = 1$. Daraus folgt, mit dem ersten, der zweite Teil von (2.70). $\square$

Abschließend zu diesem Abschnitt soll (gewissermaßen als Fortsetzung des Korollars 2.31 und des Beispiels 2.34) kurz die Aussterbewahrscheinlichkeit bei größeren initialen Populationsgrößen, also des Prozesses $\{Z_n^{(l)}\}_{n \geq 0}$, beleuchtet werden.

Beispiel 2.49 (AWK bei größeren initialen Populationsgrößen)

Sei wieder $l \geq 1$. Um die WK für $\mathbb{E}^{(l)} := \mathbb{P}(\text{Prozess } \{Z_n^{(l)}\}_{n \geq 0} \text{ stirbt aus})$, also das diesbzgl. Aussterbeereignis, zu bestimmen, kann erneut die *Additive Property* bzw. das Korollar 2.31 ausgenützt werden.

Eine Möglichkeit ist, das Pendant zu (2.69) für den Prozess $\{Z_n^{(l)}\}$ zu verwenden:

$$\begin{aligned} \mathbb{P}(E^{(l)}) &\stackrel{(2.69)}{=} \mathbb{P}(Z_n^{(l)} \rightarrow 0) \stackrel{(2.51)}{=} \mathbb{P}\left(\sum_{i=1}^l \underbrace{Z_{n,i}}_{(*)} \rightarrow 0\right) \stackrel{Z_{n,i} \geq 0}{=} \mathbb{P}\left(\bigcap_{i=1}^l \{Z_{n,i} \rightarrow 0\}\right) \\ &\stackrel{\text{uabh.}}{\stackrel{(A.42)}{=}} \prod_{i=1}^l \mathbb{P}(\underbrace{Z_{n,i}}_{\stackrel{d}{=} Z_n \text{ (i.i.d. } (*)}) \rightarrow 0) = (\mathbb{P}(Z_n \rightarrow 0))^l \stackrel{(2.70)}{=} \eta^l. \end{aligned}$$

(*) Die $\{Z_{n,i}\}_{i=1}^l$ sind i.i.d.-Kopien von Z_n (siehe (2.51)).

Alternativ können die im erwähnten Korollar bestimmten weF herangezogen werden.

$$\mathbb{P}(E^{(l)}) \stackrel{(2.67)}{=} \lim_{n \rightarrow \infty} \mathbb{P}(Z_n^{(l)} = 0) \stackrel{(2.50)}{=} \lim_{n \rightarrow \infty} (f_n(0))^l = \left(\lim_{n \rightarrow \infty} f_n(0)\right)^l \stackrel{(2.67)}{=} \eta^l. \quad \triangle$$

2.6 Asymptotische bzw. Grenzwertresultate

Im vorherigen Unterabschnitt 2.5.4 wurde eruiert, dass nur entweder ein Aussterben (also $\{Z_n \rightarrow 0\}$ mit der WK η) oder eine 'Explosion' ($\{Z_n \rightarrow \infty\}$ mit der WK $1 - \eta$) des Prozesses möglich ist. Letzterer Fall besitzt nur mit $\mu > 1$, also bei superkritischen Prozessen, eine positive Wahrscheinlichkeit. Besonders dann lohnt es sich, den um den Kehrwert des EW μ^{-n} normalisierten Prozess $\{W_n\}_{n \geq 0}$ mit $W_n = \frac{Z_n}{\mu^n}$ zu betrachten (siehe die Def. 2.36). Dieser kann ein Gefühl für die Wachstumsgeschwindigkeit bieten.

Im vorliegenden Abschnitt gelte immer $\mu > 0$ und die Ann. 2.39 (ii), also $\mu < \infty$.

2.6.1 Konvergenz der normalisierten Populationsgrößen

Zunächst soll das asymptotische Verhalten von $\{W_n\}_{n \geq 0}$ aber unabhängig von der Kritikalität des Prozesses (d.h. für alle $0 < \mu < \infty$) beleuchtet werden. Gegen Ende des Unterabschnitts 2.4.2, im Satz 2.37, wurde bereits festgestellt, dass es sich dabei um ein nichtnegatives Martingal handelt. Für diese Klasse von Prozessen existieren diverse Konvergenzsätze. Eine passende Version davon wird unter A.2.7 im Appendix vorgestellt und nun angewendet. Für die Konvergenzarten von Zufallsvariablen siehe ggf. A.2.5.

Satz 2.50 (Konvergenzsatz für die normalisierten Populationsgrößen W_n – vgl. z.B. [3, Theorem I.6.1, S. 9])

Für jedes $n \geq 0$ sei $W_n = \frac{Z_n}{\mu^n}$, also die um den Kehrwert ihres EW normalisierte Populationsgröße einer Generation. Dann existiert eine nichtnegative Zufallsvariable W mit

$$W_n \xrightarrow{f.s.} W \quad \text{und} \quad \mathbb{E}W \leq \mathbb{E}Z_0 = 1. \quad (2.71)$$

Beweis. Gemäß dem Satz 2.37 ist $\{W_n\}_{n \geq 0}$ ein nichtnegatives Martingal (bzgl. der natürlichen Filtration). Deshalb garantiert der Martingalkonvergenzsatz A.87 (i) die Existenz einer nichtnegativen, integrierbaren Zufallsvariablen W mit $W_n \xrightarrow{f.s.} W$. Dementsprechend tritt (vgl. die entsprechende Def. A.80 (i) für die fast sichere Konvergenz) das Ereignis $\{\lim_{n \rightarrow \infty} W_n = W\}$ bzw. $\{W_n \rightarrow W\}$ fast sicher ein, kann also als $(\mathbb{P})$ -Einsmenge bezeichnet werden. Somit erhält man mithilfe des Lemmas A.17

$$\mathbb{E} \left[\liminf_{n \rightarrow \infty} W_n \right] \stackrel{(A.24)}{=} \mathbb{E} \left[\mathbb{1}_{\{W_n \rightarrow W\}} \underbrace{\liminf_{n \rightarrow \infty} W_n}_{=W \text{ auf } \{W_n \rightarrow W\}} \right] = \mathbb{E} \left[\mathbb{1}_{\{W_n \rightarrow W\}} W \right] \stackrel{(A.11)}{=} \mathbb{E}W. \quad (\star)$$

Das *Lemma von Fatou* (Satz A.16) führt schließlich entweder direkt zu

$$\mathbb{E}W \stackrel{(\star)}{=} \mathbb{E} \left[\liminf_{n \rightarrow \infty} W_n \right] \stackrel{(A.23)}{\leq} \liminf_{n \rightarrow \infty} \underbrace{\mathbb{E}W_n}_{=1 \quad \forall n \geq 0 \quad (2.62)} = 1$$

oder ausführlicher, da (siehe die Bem. zur Def. A.86 (ii)) für alle $n \geq 0$

$$\mathbb{E}W_n \stackrel{(A.104)}{=} \mathbb{E}W_0 \stackrel{(2.62)}{=} \mathbb{E} \left[\frac{Z_0}{\mu^0} \right] = \mathbb{E}Z_0 \quad \text{gilt,} \quad \text{zu} \quad \mathbb{E}W \leq \liminf_{n \rightarrow \infty} \mathbb{E}Z_0 = \mathbb{E}Z_0. \quad \square$$

2.6.2 Verteilung von W

Es konvergiert $\{W_n\}_{n \geq 0}$ also in jedem Fall fast sicher gegen eine nichtnegative Zufallsvariable W , deren Erwartungswert ≤ 1 ist. Aufgrund ihrer Nichtnegativität gibt es grundsätzlich die beiden Möglichkeiten $\{W=0\}$ oder $\{W>0\}$. Diese schließen sich gegenseitig aus, weshalb allgemein festgehalten werden kann, dass wegen

$$\mathbb{P}(\{W=0\} \cup \{W>0\}) = 1 \quad \text{auch} \quad \mathbb{P}(W=0) = 1 - \mathbb{P}(W>0) \quad (2.72)$$

gilt. Für den Fall, dass ersteres Ereignis fast sicher eintritt, also $\mathbb{P}(W=0) = 1$ ist, spricht man davon, dass W eine *degenerierte Zufallsvariable* bei 0 ist, sprich $W=0$ f.s. (oder $W \equiv 0$). Dann hat der normalisierte Prozess keine Aussagekraft – in [3, S. 9] wird etwa davon gesprochen, dass diesfalls die Normalisierungsfolge μ^n zu groß ist, sprich sie 'zu schnell wächst' (respektive die Normalisierungsfaktoren μ^{-n} zu klein sind).

Es ist also die Wahrscheinlichkeit $\mathbb{P}(W=0)$ von Interesse. Da die Z_n ganzzahlig sind (und 0 ein absorbierender Zustand ist – siehe den Satz 2.26 (i)), gilt:

$$\begin{aligned} \omega \in \{Z_n \rightarrow 0\} &\iff \lim_{n \rightarrow \infty} Z_n(\omega) = 0 \iff \exists n \geq 1 : Z_m(\omega) = 0 \quad \forall m \geq n \\ &\stackrel{(2.62)}{\iff} \exists n \geq 1 : W_m(\omega) = 0 \quad \forall m \geq n \Rightarrow \lim_{n \rightarrow \infty} W_n(\omega) = 0 \iff \omega \in \{W_n \rightarrow 0\}. \end{aligned}$$

Daraus folgt (beachte, dass im vorletzten Schritt die Implikation i.A. nur in die eine Richtung ' $\Rightarrow$ ' gilt) $\{Z_n \rightarrow 0\} \subseteq \{W_n \rightarrow 0\}$. Zudem stimmen aufgrund von (2.71), also $\mathbb{P}(W_n \rightarrow W) = 1$, die Ereignisse $\{W_n \rightarrow 0\}$ und $\{W=0\}$ nur auf einer Nullmenge nicht überein, was zusammengekommen

$$\mathbb{P}(W=0) = \mathbb{P}(W_n \rightarrow 0) \geq \mathbb{P}(Z_n \rightarrow 0) \stackrel{(2.70)}{=} \eta \quad \text{ergibt.}$$

Die Wahrscheinlichkeit $\mathbb{P}(W=0)$ ist also mindestens so groß wie die Aussterbewahrscheinlichkeit η . Um sie noch weiter eingrenzen zu können, wird zunächst ein Hilfsresultat und anschließend ein Satz für eine diesbezügliche Dichotomie formuliert.

Lemma 2.51. *Es gilt $\mathbb{P}(W=0|Z_1=k) = \mathbb{P}(W=0)^k$ für alle $k \geq 0$.*

Beweis. Bedingt auf $Z_1=0$ sind alle Glieder von $\{Z_n\}_{n \geq 1}$ – und folglich auch alle von $\{W_n\}_{n \geq 1}$ – gleich 0, weshalb $\mathbb{P}(W=0|Z_1=0) = 1 = \mathbb{P}(W=0)^0$ gilt und die Aussage somit für $k=0$ erfüllt ist.

Für $k \geq 1$ betrachte – wie etwa bereits im Unterabschnitt 2.4.1 (genauer bei der Rückwärtsgleichung (2.48)) gehandhabt – die Populationsgröße der $(n+1)$ -ten Generation als die Summe von Z_1 unabhängigen Prozessen, sprich

$$Z_{n+1} \stackrel{d}{=} \sum_{i=1}^{Z_1} Z_{n,i} \quad \text{für alle } n \geq 0, \text{ wobei } \{Z_{n,i}\}_{i \geq 1} \text{ i.i.d.-Kopien von } Z_n \text{ sind.} \quad (\star)$$

Setze $W_{n,i} := \frac{Z_{n,i}}{\mu^n}$ für jedes $n \geq 0$ und $i \geq 1$. Dann sind diese $\{W_{n,i}\}_{i \geq 1}$ i.i.d.-Kopien von W_n . Gemäß dem Satz 2.50 existiert deshalb für jedes $i \geq 1$ eine ZV $W_{(i)} \geq 0$ mit

2 Einfache Galton-Watson-Verzweigungsprozesse

$W_{n,i} \xrightarrow{\text{f.s.}} W_{(i)}$ bzw. $\lim_{n \rightarrow \infty} W_{n,i} = W_{(i)}$ f.s. ($\star\star$). Dabei sind $\{W_{(i)}\}_{i \geq 1}$ i.i.d.-Kopien von W . Weiters ist $W_{n+1} = \frac{Z_{n+1}}{\mu^{n+1}} \stackrel{d}{=} \frac{1}{\mu} \sum_{i=1}^{Z_1} \frac{Z_{n,i}}{\mu^n} = \frac{1}{\mu} \sum_{i=1}^{Z_1} W_{n,i}$ ($\star\star\star$). Daraus folgt

$$W \stackrel{(\text{f.s.})}{=} \lim_{(2.71) \ n \rightarrow \infty} W_{n+1} \stackrel{d}{=} \lim_{(\star\star\star) \ n \rightarrow \infty} \frac{1}{\mu} \sum_{i=1}^{Z_1} W_{n,i} = \frac{1}{\mu} \sum_{i=1}^{Z_1} \lim_{n \rightarrow \infty} W_{n,i} \stackrel{(\text{f.s.})}{=} \frac{1}{\mu} \sum_{i=1}^{Z_1} W_{(i)}.$$

Deshalb gilt für W , bedingt auf $Z_1 = k$, $(W|Z_1 = k) \stackrel{d}{=} \frac{1}{\mu} \sum_{i=1}^k W_{(i)}$ (Δ) und schließlich

$$\begin{aligned} \mathbb{P}(W = 0|Z_1 = k) &= \mathbb{P}_{(\Delta)} \left(\frac{1}{\mu} \sum_{i=1}^k W_{(i)} = 0 \right) \stackrel{W_{(i)} \geq 0}{=} \mathbb{P}(W_{(1)} = 0, \dots, W_{(k)} = 0) \\ &= \mathbb{P} \left(\bigcap_{i=1}^k \{W_{(i)} = 0\} \right) \stackrel{\text{uabh.}}{=} \prod_{i=1}^k \mathbb{P}(W_{(i)} = 0) \stackrel{W_{(i)} \stackrel{d}{=} W}{=} \mathbb{P}(W = 0)^k. \quad \square \end{aligned}$$

Bemerkung zum Lemma 2.51. Beachte, dass $\mathbb{P}(W = 0|Z_1 = k)$ für solche $k \geq 0$, für die $p_k = \mathbb{P}(Z_1 = k) > 0$ ist, *nicht* als die elementare bedingte WK (Def. A.56) aufgefasst werden darf. Es ist damit etwa der (stets wohldefinierte) zweite Ausdruck in der letzten Gleichung des obigen Beweises gemeint (betrachte als Beispiel einen Prozess mit $p_3 = 0$, welcher dennoch in der ersten Generation mit $Z_1 = 3$ 'gestartet' werden kann).

Aber selbst wenn man die Aussage nur für k mit $p_k > 0$ trifft, stimmt die Gleichung, in welcher sie zur Anwendung kommt (siehe den Beweis des nächsten Satzes – die betroffenen Summanden tragen nichts zu der Reihe bei, da bei diesen ohnehin $p_k = 0$ ist). $\triangle$

Satz 2.52. *Es gilt entweder $\mathbb{P}(W = 0) = \eta$ oder (bzw. ggf. und) $\mathbb{P}(W = 0) = 1$, sprich*

$$\mathbb{P}(W = 0) \in \{\eta, 1\}. \quad (2.73)$$

Beweis. Im trivialen Fall stimmt die Aussage, da einerseits $W \equiv 1$, also $\mathbb{P}(W = 0) = 0$, und andererseits $\eta = 0$ ist (für jedes $n \geq 0$ ist $Z_n \equiv 1 \Rightarrow W_n \equiv 1$). Sei also $p_1 < 1$.

Mit dem vorherigen Lemma 2.51 und dem Satz A.57 über die totale WK erhält man (beachte, dass $Z_0 = 1$, $Z_1 \stackrel{(2.35)}{=} X_{0,1}$ und somit $\mathbb{P}(Z_1 = k) \stackrel{(2.34)}{=} p_k$ ist)

$$\mathbb{P}(W = 0) \stackrel{(A.71)}{=} \sum_{k \geq 0} \underbrace{\mathbb{P}(W = 0|Z_1 = k)}_{= \mathbb{P}(W=0)^k \text{ (Lemma 2.51)}} \underbrace{\mathbb{P}(Z_1 = k)}_{= p_k} = \sum_{k \geq 0} p_k \mathbb{P}(W = 0)^k \stackrel{(2.45)}{=} f(\mathbb{P}(W = 0)).$$

Es ist also $\mathbb{P}(W = 0)$ ein Fixpunkt der weF f . Da $p_1 < 1$ ist, können dafür (etwa gemäß dem Korollar 2.47) nur η und 1 in Betracht gezogen werden. $\square$

W in Abhängigkeit von μ

Nun kann W in Abhängigkeit von μ , der erwarteten Nachkommenszahl eines Individuums, gesondert betrachtet werden. Schließe dafür den trivialen Fall ($p_1 = 1$ – siehe erneut die Def. 2.40) für den Rest dieses Kapitels aus.

2.6 Asymptotische bzw. Grenzwertresultate

In *subkritischen und kritischen Fällen*, d.h. für $\mu \leq 1$, ist die AWK $\eta = 1$ (siehe (i) und (ii) des Satzes 2.44). Es tritt demnach fast sicher das Aussterbeereignis $\{Z_n \rightarrow 0\}$ und somit ebenso $\{W_n \rightarrow 0\}$ ein. Für alle hinreichend großen n gilt dementsprechend f.s. $Z_n = 0$ bzw. $W_n = 0$ (vgl. auch erneut den Satz 2.26 (i) – 0 ist absorbierend). Weiters konvergiert W_n f.s. gegen W , weshalb es naheliegend ist, dass W in diesen Fällen die angesprochene degenerierte ZV bei 0 ist,

$$\text{d.h., wenn } \mu \leq 1 \text{ (und } p_1 < 1 \text{) ist, gilt } W = 0 \text{ f.s.} \quad (2.74)$$

Dies wird durch den Satz 2.52 ($\eta = 1 \stackrel{(2.73)}{\Rightarrow} \mathbb{P}(W = 0) = 1$) bestätigt. Aus diesem Grund ist in diesen Fällen eine weitere Analyse von W uninteressant.

Aber auch im *superkritischen Fall*, also mit $\mu > 1$ und $\eta < 1$ (Satz 2.44 (iii)), ist nicht garantiert, dass nicht die gleiche Situation vorherrscht. Unter gewissen Voraussetzungen ist dies, also eine Degeneration von W , aber nicht der Fall – beispielsweise wenn die Varianz der Nachkommenszahlen der Individuen $\sigma^2 < \infty$ ist, wie der nächste Satz zeigt.

Satz 2.53 (W -superkritische GWP mit $\sigma^2 < \infty$ – vgl. z.B. [3, Theorem I.6.2, S. 9])
Seien $1 < \mu (< \infty - \text{Ann. 2.39 (ii)})$ und $\sigma^2 < \infty$. Dann gilt

- (i) $W_n \xrightarrow{\mathcal{L}^2} W$, d.h. $\mathbb{E}(W_n - W)^2 \rightarrow 0$ (vgl. die Def. A.80 (iii)),
- (ii) $\mathbb{E}W = 1$, $\mathbb{V}W = \frac{\sigma^2}{\mu(\mu-1)}$ und
- (iii) $\mathbb{P}(W = 0) = \eta = \mathbb{P}(Z_n \rightarrow 0) < 1$, also auch
 $\mathbb{P}(W > 0) \stackrel{(2.72)}{=} 1 - \eta \stackrel{(2.70)}{=} \mathbb{P}(Z_n \rightarrow \infty) > 0$.

Beweis. Bestimme zunächst für ein beliebiges $n \geq 0$ die Varianz von $W_n \stackrel{(2.62)}{=} \frac{Z_n}{\mu^n}$:

$$\mathbb{V}W_n = \frac{1}{\mu^{2n}} \mathbb{V}Z_n \stackrel{(2.57)}{=} \frac{1}{\mu^{2n}} \frac{\sigma^2 \mu^n (\mu^n - 1)}{\mu(\mu - 1)} = \frac{\sigma^2 (1 - \mu^{-n})}{\mu(\mu - 1)}. \quad (\star)$$

Da $\{W_n\}_{n \geq 0}$ ein nichtnegatives Martingal ist, soll wieder ein Martingalkonvergenzsatz – und zwar nun der Satz A.87 (ii) (und anschließend das Korollar dazu) – angewendet werden. Dieser erfordert $\sup_{n \geq 0} \mathbb{E}W_n^2 < \infty$. Bestimme also mit dem Verschiebungssatz ($\mathbb{E}W_n = 1 \ \forall n \geq 0$ wegen (2.62)) $\mathbb{E}W_n^2 = \mathbb{V}W_n + (\mathbb{E}W_n)^2 \stackrel{(\star)}{=} \frac{\sigma^2 (1 - \mu^{-n})}{\mu(\mu - 1)} + 1 \stackrel{(\star\star)}{<} \infty$. Da $\mu > 1$ ist, fällt μ^{-n} (in n) und demzufolge ist $\{\mathbb{E}W_n^2\}_{n \geq 0}$ eine monoton wachsende, beschränkte Folge (lt. Vorauss. sind $\mu, \sigma^2 < \infty$). Daher folgt für ihr Supremum (Satz A.93) $\sup_{n \geq 0} \mathbb{E}W_n^2 = \lim_{n \rightarrow \infty} \mathbb{E}W_n^2 \stackrel{(\star\star)}{=} \frac{\sigma^2}{\mu(\mu - 1)} + 1 < \infty$. Damit ist alles für den Satz A.87 (ii) erfüllt, womit sich $W_n \xrightarrow{\mathcal{L}^2} W$, also (i), und $\mathbb{E}W^2 < \infty$ verifizieren lassen.

Das Korollar zum Satz A.87 bestätigt (ii), denn es gilt

$$\mathbb{E}W \stackrel{(A.105)}{=} \lim_{n \rightarrow \infty} \mathbb{E}W_n \stackrel{(2.62)}{=} 1, \quad \mathbb{V}W \stackrel{(A.105)}{=} \lim_{n \rightarrow \infty} \mathbb{V}W_n \stackrel{(\star)}{=} \lim_{n \rightarrow \infty} \frac{\sigma^2 (1 - \mu^{-n})}{\mu(\mu - 1)} \stackrel{\mu > 1}{=} \frac{\sigma^2}{\mu(\mu - 1)}.$$

Für (iii) beachte, dass laut dem vorherigen Satz 2.52 nur η oder 1 für $\mathbb{P}(W = 0)$ in Frage kommen (2.73). Da W nichtnegativ und $\mathbb{V}W > 0$ ist, ist $\mathbb{P}(W = 0) = 1$ ausgeschlossen (zumal bereits $\mathbb{E}W = 1 \Rightarrow \mathbb{P}(W = 0) < 1$ diesen Fall verunmöglicht). Deshalb bleibt nur die Möglichkeit $\mathbb{P}(W = 0) = \eta \stackrel{(2.70)}{=} \mathbb{P}(Z_n \rightarrow 0)$ übrig. $\square$

2 Einfache Galton-Watson-Verzweigungsprozesse

Das Kriterium $\sigma^2 < \infty$ garantiert also, dass W eine 'echte' Zufallsvariable mit einer Verteilung abseits der Punktmasse (im Ausmaß von η) bei 0 ist. Etwa gemäß [19, Theorem I.8.3, S. 16] ist diese Zufallsvariable dann auf $(0, \infty)$ sogar *absolut stetig*, wodurch der *Satz von Radon-Nikodym* sicherstellt, dass – bedingt auf das Nicht-Aussterben – eine entsprechende Dichtefunktion für W existiert (siehe dazu etwa [42, S. 255 ff.]).

Es gibt aber auch weniger starke Voraussetzungen, welche den Fall $\mathbb{P}(W = 0) = 1$ ausschließen. Eine sozusagen 'scharfes' Kriterium kann über den EW $\mathbb{E}[X \log X]$ gegeben werden. Das sogenannte *Kesten-Stigum Theorem*, erstmals publiziert in [28] (für Multitype GWP), besagt (siehe etwa [3, Theorem I.10.1, S. 24] oder [9, Theorem 1.4, S. 253 f.]), dass eine Äquivalenz zwischen der Bedingung $\mathbb{E}[X \log X] < \infty$ und jedem der Punkte (i)-(iii) des Satzes 2.53 besteht. Umgekehrt bedeutet $\mathbb{E}[X \log X] = \infty$, dass $\mathbb{E}W = 0$ (also auch $\mathbb{P}(W = 0) = 1$) und W degeneriert ist.

Im Text vor dem Lemma 2.51 wurde bereits festgestellt (alle Mengenrelationen im vorliegenden und nächsten Absatz stimmen nur auf einer Nullmenge nicht), dass allgemein, also unabhängig von μ , $\{Z_n \rightarrow 0\} \subseteq \{W = 0\}$ gilt. I.a.W. heißt das aber auch, dass $\{W > 0\} \subseteq \{Z_n \rightarrow \infty\}$ ist (denn für beliebige Mengen A, B gilt: $A \subseteq B \iff B^c \subseteq A^c$). Im subkritischen und kritischen, sowie mit $\mathbb{E}[X \log X] < \infty$ auch im superkritischen Fall, gilt ja $\mathbb{P}(Z_n \rightarrow 0) = \mathbb{P}(W = 0) = \eta$ und $\mathbb{P}(Z_n \rightarrow \infty) = \mathbb{P}(W > 0) = 1 - \eta$. Die Teilmengenrelationen gelten dann also sogar als Gleichheit.

Dahingegen handelt es sich ausschließlich im superkritischen Fall mit $\mathbb{E}[X \log X] = \infty$ um echte Teilmengen (das sieht man etwa an $\mathbb{P}(Z_n \rightarrow 0) = \eta < 1 = \mathbb{P}(W = 0)$ und $\mathbb{P}(W > 0) = 0 < 1 - \eta = \mathbb{P}(Z_n \rightarrow \infty)$). Interpretativ wird dadurch genau das angesprochene Szenario, dass auch im Fall des Nicht-Aussterbens μ^n zu groß ist und in n 'schneller' als Z_n wächst, widergespiegelt. Es ist aber auch in so einem Fall möglich, passende andere Normalisierungsfolgen $\{C_n\}_{n \geq 0}$, sogenannte *Seneta-Konstanten*, zu finden. Man kann zeigen, dass es für jeden superkritischen GWP entsprechende $\{C_n\}_{n \geq 0}$ gibt, sodass $\{\frac{Z_n}{C_n}\}_{n \geq 0}$ f.s. gegen eine nichtdegenerierte ZV konvergiert – siehe dazu etwa erneut [3, S. 24 bzw. S. 30 ff.] oder [44, Comment, S. 16]. Eine Variante wird in [25, Theorem (2.7.1), S. 31] dargestellt. Dort werden die generelle Existenz dieser Konstanten und die Möglichkeit, dass sie bei Vorliegen von $\mathbb{E}[X \log X] < \infty$ als $\{\mu^n\}_{n \geq 0}$ gewählt werden können, in einem Zug formuliert.

Generell stellt $\mathbb{E}[X \log X] < \infty$ keine besonders starke Anforderung dar. Etwa in [18, S. 116] wird – im biologischen Kontext – sogar davon gesprochen, dass Populationen mit davon abweichenden Nachkommensverteilungen unvorstellbar sind.

Laplace-Transformierte und Charakteristische Funktionen von $\{W_n\}_{n \geq 0}$ und W

Für eine weitergehende Analyse von $\{W_n\}_{n \geq 0}$ und W können deren *Laplace-Transformationen* bzw. deren *charakteristische Funktionen* von Belang sein. Diese bestimmen eindeutig die jeweilige Verteilung (vgl. ggf. die Sätze A.83 und A.84 (i) sowie eine allgemeine kurze Vorstellung solcher Funktionen im Appendix unter A.2.6). Zunächst werden beide sowohl für alle $\{W_n\}_{n \geq 0}$ als auch für W definiert und anschließend deren (punktweises) Konvergenzverhalten gezeigt. Zum Abschluss dieses Kapitels werden wichtige Funktionalgleichungen – insbesondere für die Laplace-Transformierte von W – formuliert.

Definition 2.54. Definiere für jedes $n \geq 0$ die Funktionen $\mathcal{L}_n, \mathcal{L} : \mathbb{R}_+ \rightarrow [0, 1]$ mit

$$\mathcal{L}_n(u) := \mathcal{L}_{W_n}(u) \stackrel{(A.96)}{=} \mathbb{E}[e^{-uW_n}], \quad \mathcal{L}(u) := \mathcal{L}_W(u) \stackrel{(A.96)}{=} \mathbb{E}[e^{-uW}] \quad \text{für } u \in \mathbb{R}_+ \quad (2.75)$$

als die *Laplace-Transformierten* von W_n bzw. W . Weiters seien deren *charakteristischen Funktionen* (kurz *C-Funktionen*) $\varphi_n, \varphi : \mathbb{R} \rightarrow \mathbb{C}$ für jedes $n \geq 0$ definiert durch

$$\varphi_n(t) := \varphi_{W_n}(t) \stackrel{(A.98)}{=} \mathbb{E}[e^{itW_n}], \quad \varphi(t) := \varphi_W(t) \stackrel{(A.98)}{=} \mathbb{E}[e^{itW}] \quad \text{für } t \in \mathbb{R}. \quad (2.76)$$

Für die nächsten beiden Sätze benötigt es keinerlei zusätzliche Annahmen.

Satz 2.55. *Es gilt*

$$\mathcal{L}_n(u) \xrightarrow{n \rightarrow \infty} \mathcal{L}(u) \quad \text{für alle } u \in \mathbb{R}_+ \quad \text{und} \quad \varphi_n(t) \xrightarrow{n \rightarrow \infty} \varphi(t) \quad \text{für alle } t \in \mathbb{R}. \quad (2.77)$$

Beweis. Beachte, dass mit dem Satz 2.50 allgemein $W_n \xrightarrow{\text{f.s.}} W$ gilt $(\star)$.

Sei zunächst $u \in \mathbb{R}_+$ beliebig. Die Funktion $x \mapsto e^{-ux}$ ist stetig (auf $\mathbb{R}$). Deshalb folgt aus $(\star)$ mit dem *Satz von der stetigen Abbildung* (Satz A.82 (i)) auch $e^{-uW_n} \xrightarrow{\text{f.s.}} e^{-uW}$. Da weiters $|e^{-uW_n}| \leq 1$ für alle $n \geq 0$ gilt, besagt dann der *Satz A.15 über die dominierte Konvergenz*, dass $\lim_{n \rightarrow \infty} \mathcal{L}_n(u) \stackrel{(2.75)}{=} \lim_{n \rightarrow \infty} \mathbb{E}[e^{-uW_n}] \stackrel{(A.22)}{=} \mathbb{E}[e^{-uW}] \stackrel{(2.75)}{=} \mathcal{L}(u)$ gilt.

Für die zweite Aussage in (2.77) folgt aus $(\star)$ andererseits auch (siehe den Satz A.82 (ii)), dass $\{W_n\}_{n \geq 0}$ in Verteilung gegen W konvergiert, kurz $W_n \xrightarrow{d} W$ $(\star\star)$. Anschließend kann direkt der *Stetigkeitssatz von Lévy* (Satz A.84 (ii)) für C-Funktionen verwendet werden. Gemäß ihm ist $(\star\star)$ äquivalent zu $\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t)$ für alle $t \in \mathbb{R}$. $\square$

Satz 2.56 (Funktionalgleichungen für $\{\mathcal{L}_n\}_{n \geq 0}$ und $\mathcal{L}$ – vgl. z.B. [3, I.10.(5), S. 25])

Sei $u \in \mathbb{R}_+$. Dann gilt für alle $n \geq 0$

$$\begin{aligned} \mathcal{L}_{n+1}(u) &= f\left(\mathcal{L}_n\left(\frac{u}{\mu}\right)\right) && (\text{bzw. } \mathcal{L}_{n+1}(u\mu) = f(\mathcal{L}_n(u))) \quad \text{und weiters} \\ \mathcal{L}(u) &= f\left(\mathcal{L}\left(\frac{u}{\mu}\right)\right) && (\text{bzw. } \mathcal{L}(u\mu) = f(\mathcal{L}(u))). \end{aligned} \quad (2.78)$$

Beweis. Seien $u \in \mathbb{R}_+$ und $n \geq 0$ beliebig. Es gilt

$$\begin{aligned} \mathcal{L}_n(u) &\stackrel{(2.75)}{=} \mathbb{E}[e^{-uW_n}] \stackrel{(2.62)}{=} \mathbb{E}[e^{-u\mu^{-n}Z_n}] = \mathbb{E}\left[\overbrace{(e^{-u\mu^{-n}})^{Z_n}}^{\leq 1}\right] \stackrel{(2.6)}{\stackrel{(2.46)}{=}} f_n(e^{-u\mu^{-n}}) && (2.79) \\ &\Rightarrow \mathcal{L}_{n+1}(u) \stackrel{(2.79)}{=} f_{n+1}(e^{-u\mu^{-n-1}}) \stackrel{(2.30)}{=} f\left(f_n\left(e^{-\left(\frac{u}{\mu}\right)\mu^{-n}}\right)\right) \stackrel{(2.79)}{=} f\left(\mathcal{L}_n\left(\frac{u}{\mu}\right)\right) && (\star) \\ &\Rightarrow \mathcal{L}(u) \stackrel{(2.77)}{=} \lim_{n \rightarrow \infty} \mathcal{L}_{n+1}(u) \stackrel{(\star)}{=} \lim_{n \rightarrow \infty} f\left(\mathcal{L}_n\left(\frac{u}{\mu}\right)\right) \stackrel{f}{\stackrel{\text{stetig}}{=}} f\left(\lim_{n \rightarrow \infty} \mathcal{L}_n\left(\frac{u}{\mu}\right)\right) \stackrel{(2.77)}{=} f\left(\mathcal{L}\left(\frac{u}{\mu}\right)\right). \end{aligned}$$

Die Gleichungen in Klammern folgen jeweils unmittelbar (mit $u\mu$ statt u). $\square$

Bemerkungen zum Satz 2.56.

- (i) Die Funktionalgleichung (2.78) wird etwa in der o.g. Referenz auch *Abel's equation* genannt. Wie zuvor angemerkt muss die Existenz von σ^2 für deren Gültigkeit nicht verlangt werden. Im superkritischen Fall kann man aber zeigen, dass $l(\mu u) = f(l(u))$ für $u \in \mathbb{R}_+$ unter der Voraussetzung $\mathbb{E}X^2 < \infty$ eindeutig lösbar ist, und zwar eben genau durch $l = \mathcal{L}$ (siehe etwa die Anmerkung in [18, S. 155]). Für einen Beweis der Eindeutigkeit siehe z.B. [25, Theorem 2.7.2, S. 33] (es ist dafür eigentlich wieder bereits $\mathbb{E}[X \log X] < \infty$ ausreichend – vergleichbar mit dem im letzten Teilabschnitt Besprochenen).
- (ii) Beachte, dass auch ein Pendant zu (2.78) für die C-Funktion φ gilt, nämlich

$$\varphi(t) = f\left(\varphi\left(\frac{t}{\mu}\right)\right) \quad (\text{bzw. } \varphi(u\mu) = f(\varphi(u))) \quad \text{für } t \in \mathbb{R}. \quad (2.80)$$

Ein genauer Beweis dafür findet sich etwa unter [19, Theorem I.8.2, S. 15] – formuliert für $\mathbb{E}[e^{-sW}]$, wobei für s auch komplexe Zahlen mit $\operatorname{Re}(s) \geq 0$ zugelassen sind (weshalb mit $s = -it$ die hier definierte C-Funktion φ mitberücksichtigt ist). Dafür müssen aber weF auch für komplexe Zahlen definiert sein (was in [19] der Fall ist – und zwar für $s \in \mathbb{C}$ mit $|s| \leq 1$). Dementsprechend müsste die Theorie über die weF auch im Komplexen gezeigt werden (insbesondere etwa (2.30)), was hier ausgelassen wird. Ansonsten funktioniert der Beweis aber analog zu dem des Satzes 2.56 – nur mit (2.76) und $-it$ statt u (beachte für (2.79) und $(\star)$, dass stets $|e^{it\mu^{-n}}| \stackrel{(\text{A.99})}{=} |\cos(t\mu^{-n}) + i \sin(t\mu^{-n})| = \sqrt{\cos^2(t\mu^{-n}) + \sin^2(t\mu^{-n})} = 1$ gilt, weshalb die Argumente von f , f_n und f_{n+1} jeweils im o.g. Definitionsbereich liegen.

- (iii) Eine direkte Anwendung dieser Resultate findet sich im Abschnitt 4.1 unter 4.1.2. Für eine exemplarische Analyse anhand simulierter Daten siehe das Bsp. 4.9. $\triangle$

3 Galton-Watson-Verzweigungsprozesse mit mehreren unterschiedlichen Typen

3.1 Vorbemerkungen

Ausgehend vom klassischen Modell der einfachen Galton-Watson-Prozesse gibt es mehrere Möglichkeiten der Verallgemeinerung. Eine Option ist, von der fixen Lebensdauer der Individuen (von einer Zeiteinheit) abzugehen und diese durch eine exponentialverteilte Zufallsvariable zu beschreiben (dies würde etwa zu sogenannten *Markov-Verzweigungsprozessen* in stetiger Zeit führen – vgl. z.B. [3, III.]). Als weitere denkbare Erweiterung seien u.a. die Berücksichtigung von 'Immigration' in die Population oder beliebig verteilte Lebenszeiten genannt (vgl. z.B. [3, VI.7., S. 262 ff. bzw. IV.] – in letzterem Fall handelt es sich nicht mehr um *Markov-Prozesse*).

An dieser Stelle soll aber eine Generalisierung besprochen werden, in welcher *mehrere unterschiedliche Typen an Individuen* zugelassen sind. In diesem Sinne wird von deren Ununterscheidbarkeit abgerückt. Der Aufbau dieses Kapitels ähnelt dem des vorherigen. Nach einer kurzen Konkretisierung werden zusätzliche mathematische Vorbemerkungen ergänzt. Anschließend folgt die formale Definition und Analyse.

3.1.1 Informelle Beschreibung des Multitype Galton-Watson-Prozesses

In Abgrenzung zum 2. Kapitel wird ab sofort nicht mehr von Individuen, sondern vielmehr von *Objekten* gesprochen. Bei der Erweiterung auf mehrere mögliche Typen verliert die Grundüberlegung über den genealogischen Aspekt – also Individuen eines Familienstammbaums – an Bedeutung (man könnte aber auch beispielsweise von Individuen mit unterschiedlichen Eigenschaften sprechen – vgl. etwa [18, 2.3, S. 21]).

Jedem Objekt ist ein fixer Typ zugeordnet und es kann andere Objekte unterschiedlicher Typen (re-)produzieren. Wie viele Objekte jeweils produziert werden und von welchem Typ jedes einzelne ist, ist dabei zufällig. Für jedes Objekt eines bestimmten Typs sind die Wahrscheinlichkeiten für diese Reproduktion aber identisch verteilt. D.h., jede Kombination an Anzahlen reproduzierter Objekttypen ist für alle Objekte dieses Typs gleich wahrscheinlich. Dies gilt für alle Objekttypen. Sinngemäß sind also die Objekte gleichen Typs wiederum nicht voneinander unterscheidbar. Dass diese Reproduktionswahrscheinlichkeiten aber i.A. für Objekte unterschiedlichen Typs differieren, macht gerade die verschiedenen Objekttypen aus.

Die anderen bzgl. einfacher GWP getroffenen Ann. bleiben bestehen: Die beschriebenen Reproduktionen geschehen vollständig *unabhängig* voneinander und jedes Objekt besteht (oder 'lebt') genau *eine diskrete Zeiteinheit* bzw. Generation lang (vgl. ggf. die Abb. 3.1).

Auch auf eine genauere Spezifikation der Objekte bzw. Typen wird erneut verzichtet. In [3, S. 181] etwa werden als Beispiele Genotypen in tierischen bzw. Mutationstypen in bakteriellen Populationen oder verschiedene Teilchen kosmischer Strahlung genannt.

3.1.2 Multivariate Notation und Konventionen

Ergänzend zu den allgemeinen, im Unterabschnitt 1.2.1 beschriebenen Konventionen sollen in höheren Dimensionen die folgenden Schreibweisen zur Anwendung kommen.

Notation 3.1 (Vektoren und Matrizen)

Sei für das gesamte vorliegende Kapitel $k \geq 1$ ($k \in \mathbb{N}$) beliebig. Es werden vielfach Vektoren und Matrizen verwendet. Sie selbst werden in fettgedruckten Buchstaben geschrieben, wobei für Vektoren in der Regel Klein- (z.B. $\mathbf{v}$) und für Matrizen Großbuchstaben (z.B. $\mathbf{M}$) stehen. Die Komponenten hingegen werden für gewöhnlich durch Kleinbuchstaben – nicht fettgedruckt und mit adäquaten (meist) tiefgestellten Indizes versehen – dargestellt (z.B. v_i, m_{ij}). Eine Ausnahme bilden Zufallsvariablen als Komponenten – diese erhalten auch Großbuchstaben. Meistens werden Spaltenvektoren aus Platzgründen als transponierte ($\cdot^T$) Zeilenvektoren geschrieben (und deren Elemente durch Kommas getrennt). Spaltenvektoren und Matrizen können auch durch in Klammern stehende Komponenten mit passenden Laufindizes dargestellt werden, beispielsweise

$$\mathbf{v} = (v_i)_{i=1}^k = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{pmatrix} = (v_1, v_2, \dots, v_k)^T, \quad \mathbf{M} = (m_{ij})_{i,j=1}^k = \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1k} \\ m_{21} & m_{22} & & m_{2k} \\ \vdots & & \ddots & \vdots \\ m_{k1} & m_{k2} & \cdots & m_{kk} \end{pmatrix}.$$

Hier repräsentieren also i die Zeilenindizes und j bei $\mathbf{M}$ die Spaltenindizes. Sind etwa für $1 \leq i \leq k$ Zeilenvektoren in der Art $\mathbf{v}_i = (v_{i1}, v_{i2}, \dots, v_{ik})$ gegeben, so ist mit $\mathbf{V} = (\mathbf{v}_i)_{i=1}^k$ die Matrix $\mathbf{V} = (v_{ij})_{i,j=1}^k$ gemeint.

Die Mengen, aus welchen Vektoren oder Matrizen Werte annehmen können, werden mit den Dimensionen als Hochzahlen geschrieben. Sind die Einträge von $\mathbf{v}$ bzw. $\mathbf{M}$ bspw. nichtnegativ und ganzzahlig, so wird dies als $\mathbf{v} \in \mathbb{N}_0^k$ mit $\mathbb{N}_0^k = \{(t_i)_{i=1}^k : t_i \in \mathbb{N}_0\}$ bzw. $\mathbf{M} \in \mathbb{N}_0^{k \times k}$ mit $\mathbb{N}_0^{k \times k} = \{(t_{ij})_{i,j=1}^k : t_{ij} \in \mathbb{N}_0\}$ ausgedrückt. Dass ein Vektor, wenn die Dimension klar ist, aus den entsprechenden nicht fettgedruckten Komponenten mit tiefgestellten Indizes (wie etwa $\mathbf{v}$ oben) besteht, wird manchmal stillschweigend verwendet.

Es werden des Weiteren die folgenden Konventionen getroffen (welche sich grob an [3, S. 182 f.] bzw. [19, Def. II.5.1, S. 37] orientieren). Seien dafür $\mathbf{u}$ bzw. $\mathbf{v}$ k -dimensionale Vektoren und sei $\mathbf{M}$ eine $k \times k$ -dimensionale Matrix (es kommen generell ausschließlich quadratische Matrizen zur Anwendung).

- (i) Handelt es sich bei einem Vektor $\mathbf{v}$ um den Nullvektor bzw. bei einer Matrix $\mathbf{M}$ um die Nullmatrix, d.h., alle Einträge sind 0, dann wird einfach nur $\mathbf{v} = \mathbf{0}$ bzw. $\mathbf{M} = \mathbf{0}$ geschrieben. Ebenso wird mit einem Vektor bestehend aus lauter Einsen verfahren (die Dimensionen werden dabei jeweils aus dem Kontext klar). D.h. etwa

$$\mathbf{0} = (0, \dots, 0)^T \text{ bzw. } \mathbf{1} := (1, \dots, 1)^T.$$

3.1 Vorbemerkungen

- (ii) Wird $\mathbf{u} \geq (>) \mathbf{v}$ geschrieben, so wird das komponentenweise, also $u_i \geq (>) v_i$ für alle $1 \leq i \leq k$, interpretiert. Analoges gilt für Matrizen. Dementsprechend bedeutet $\mathbf{v} < \infty$ oder $\mathbf{M} < \infty$ jeweils, dass alle Einträge $< \infty$ sind.

Ein *positiver Vektor* $\mathbf{v}$ erfüllt $\mathbf{v} > \mathbf{0}$ und eine *positive Matrix* $\mathbf{M}$ analog $\mathbf{M} > \mathbf{0}$ (alle Einträge sind strikt positiv). Selbiges gilt für *nichtnegative* $\mathbf{v} \geq \mathbf{0}$ bzw. $\mathbf{M} \geq \mathbf{0}$ (jede Komponente ist ≥ 0).

- (iii) Mit $|\mathbf{v}| := |v_1| + \dots + |v_k|$ ist die sogenannte *Summennorm* (vgl. z.B. [23, 5., S. 12] – dort als $\|\cdot\|_1$ geschrieben) gemeint. Es gelten die üblichen *Normaxiome* (vgl. z.B. [23, (N1-3), S. 11]), deshalb auch $|a\mathbf{v}| = |a||\mathbf{v}|$ für reelle a , insbesondere $|\mathbf{-v}| = |\mathbf{v}|$, sowie die *umgekehrte Dreiecksungleichung* (vgl. z.B. [23, (109.3), S. 13])

$$||\mathbf{u}| - |\mathbf{v}|| \leq |\mathbf{u} - \mathbf{v}|. \quad (3.1)$$

- (iv) Sei $1 \leq i \leq k$. Der (*kanonische*) *i-te Einheitsvektor* $\mathbf{e}_i$ ist der Spaltenvektor der Dimension k , dessen *i-te* Komponente 1 ist, während alle anderen 0 sind, d.h.

$$\mathbf{e}_i = (e_{ij})_{j=1}^k \quad \text{mit } e_{ij} := \delta_{ij} \stackrel{(1.1)}{=} \begin{cases} 1, & j = i, \\ 0, & j \neq i \end{cases} \quad \left(\mathbf{e}_i = (0, \dots, \underbrace{1}_{i\text{-te Komponente}}, \dots, 0)^T \right). \quad (3.2)$$

Die *Einheitsmatrix der Dimension* $k \times k$, also eine Matrix, deren Diagonalelemente 1 und alle anderen 0 sind, wird mit $\mathbf{I}_k$ bezeichnet. Es gilt $\mathbf{I}_k = (\mathbf{e}_j^T)_{j=1}^k$, $\mathbf{I}_k \mathbf{v} = \mathbf{v}$ bzw. $\mathbf{v}^T \mathbf{I}_k = \mathbf{v}^T$, $\mathbf{I}_k \mathbf{M} = \mathbf{M} = \mathbf{M} \mathbf{I}_k$ und $\mathbf{M}^0 = \mathbf{I}_k$.

- (v) Der k -dimensionale sogenannte *Einheitshyperwürfel* ist $[0, 1]^k = \{\mathbf{x} : \mathbf{0} \leq \mathbf{x} \leq \mathbf{1}\}$.
(vi) Es wird die folgende *Produktnotation* für Vektoren genutzt:

$$\mathbf{u}^{\mathbf{v}} = \prod_{i=1}^k u_i^{v_i} = u_1^{v_1} \cdots u_k^{v_k}. \quad (3.3)$$

△

Für die Untersuchung des Grenzverhaltens von Funktionen wird gelegentlich das sogenannte *Landau-Symbol* $\mathcal{O}$ verwendet.

Definition 3.2 (Landau-Symbol $\mathcal{O}$ – vgl. etwa [13, S. 121 ff.] für $k = 1$)

Seien $p \in \mathbb{N}$, $\mathbf{D} \subseteq \mathbb{R}^k$ und $\mathbf{s}_0 \in \mathbf{D}$ (genauer: $\mathbf{s}_0$ soll wenigstens ein *Berührungspunkt* von $\mathbf{D}$ sein, d.h., dass in jeder ε -Umgebung von $\mathbf{s}_0$ mindestens ein Punkt von $\mathbf{D}$ liegt). Weiters seien die Funktionen $\mathbf{f}, \mathbf{g} : \mathbf{D} \rightarrow \mathbb{R}^p$ gegeben. Dann schreibt man:

$$\mathbf{f}(\mathbf{s}) = \mathcal{O}(\mathbf{g}(\mathbf{s})) \quad \text{für } \mathbf{s} \rightarrow \mathbf{s}_0 \quad \Longleftrightarrow \quad \begin{aligned} &\text{für jedes } C > 0 \text{ existiert ein } \varepsilon > 0, \\ &\text{sodass } |\mathbf{f}(\mathbf{s})| < C|\mathbf{g}(\mathbf{s})| \text{ für alle } \mathbf{s} \in \mathbf{D} \text{ mit } |\mathbf{s} - \mathbf{s}_0| < \varepsilon \text{ gilt.} \end{aligned}$$

Dies bedeutet, dass $\mathbf{f}(\mathbf{s})$ bei der Annäherung von $\mathbf{s}$ zu $\mathbf{s}_0$ asymptotisch im Vergleich zu $\mathbf{g}(\mathbf{s})$ verschwindend klein bzw. 'vernachlässigbar' wird.

3.1.3 Zufallsvektoren und -matrizen

Sei für alles Weitere in diesem Abschnitt $(\Omega, \Sigma, \mathbb{P})$ ein Wahrscheinlichkeitsraum.

Zufallsvektoren, kurz *Z-Vektoren*, sind spezielle Zufallsvariablen (vgl. die Def. A.27 im Appendix) mit dem Bildraum $(\Omega', \Sigma') = (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ (oder Teilmengen davon), welche ab sofort als Spaltenvektoren aufgefasst werden sollen. Ein Z-Vektor $\mathbf{X} : \Omega \rightarrow \mathbb{R}^k$ kann also als $\mathbf{X} = (X_i)_{i=1}^k = (X_1, \dots, X_k)^T$ geschrieben werden, wobei (gemäß dem Satz A.10 bzw. der Bem. zur Def. A.27 (iii)) die einzelnen $X_i : \Omega \rightarrow \mathbb{R}$ jeweils (erweitert) reellwertige Zufallsvariablen sind.

Analoges gilt für sogenannte *Zufallsmatrizen* $\mathbf{M} : \Omega \rightarrow \mathbb{R}^{k \times k}$, $\mathbf{M} = (M_{ij})_{i,j=1}^k$ (welche hier nur implizit zur Berechnung von Kovarianzmatrizen – siehe (3.8) – benötigt werden).

Elemente aus dem *kartesischen Produkt* mehrerer Mengen werden ebenfalls als Spaltenvektoren aufgefasst. Demnach ist dann (vgl. die Notation 3.1)

$$\begin{aligned} A_1 \times \dots \times A_k &= \{ (a_1, \dots, a_k)^T : a_1 \in A_1, \dots, a_k \in A_k \} = \{ (a_i)_{i=1}^k : a_i \in A_i \} \quad \text{und} \\ A^k &= \{ (a_1, \dots, a_k)^T : a_i \in A \forall 1 \leq i \leq k \} = \{ (a_i)_{i=1}^k : a_i \in A \} \end{aligned} \quad (3.4)$$

(und mit $\mathbf{v} \in \mathbb{R}^k$ ist etwa ein k -dimensionaler reellwertiger Spaltenvektor gemeint). Für ein $\mathbf{X} = (X_1, \dots, X_k)^T$ wie oben und $A_1, \dots, A_k \subseteq \mathbb{R}$ (dann ist $A_1 \times \dots \times A_k \subseteq \mathbb{R}^k$) gilt somit

$$\mathbb{P}(\mathbf{X} \in A_1 \times \dots \times A_k) = \mathbb{P}((X_1, \dots, X_k)^T \in A_1 \times \dots \times A_k) \stackrel{(3.4)}{=} \mathbb{P}(X_1 \in A_1, \dots, X_k \in A_k) \quad (3.5)$$

$$\text{bzw. } \mathbb{P}(\mathbf{X} = (x_1, \dots, x_k)^T) = \mathbb{P}(X = x_1, \dots, X = x_k) \quad \text{für } (x_1, \dots, x_k)^T \in \mathbb{R}^k. \quad (3.6)$$

Den Definitionen A.35 und A.53 entsprechend kann die folgende Def. getroffen werden.

Definition 3.3 (EW-Vektor und Kovarianzmatrix) – vgl. z.B. [38, Def. 5.17, S. 130]) Für auf dem W-Raum $(\Omega, \Sigma, \mathbb{P})$ definierte integrierbare ZV $X_1, \dots, X_k$ ist $\mathbf{X} : \Omega \rightarrow \mathbb{R}^k$ mit $\mathbf{X} = (X_1, \dots, X_k)^T$ ein k -dimensionaler Zufallsvektor, dessen EW

$$\mathbb{E}\mathbf{X} := \begin{pmatrix} \mathbb{E}X_1 \\ \vdots \\ \mathbb{E}X_k \end{pmatrix} = (\mathbb{E}X_i)_{i=1}^k \quad (3.7)$$

auch *Erwartungswertvektor von $\mathbf{X}$* heißt. Da die Erwartungswertbildung also komponentenweise geschieht, gilt $\mathbb{E}[\mathbf{X}^T] = \mathbb{E}\mathbf{X}^T = \mathbb{E}[\mathbf{X}]^T = (\mathbb{E}\mathbf{X})^T$.

Entsprechendes gilt für die sogenannte *Erwartungswertmatrix* einer Zufallsmatrix $\mathbf{M} : \Omega \rightarrow \mathbb{R}^{k \times k}$. Sind für $1 \leq i, j \leq k$ alle Komponenten M_{ij} von $\mathbf{M}$ (das sind jeweils reellwertige Zufallsvariablen) integrierbar, so gilt $\mathbb{E}\mathbf{M} := (\mathbb{E}M_{ij})_{i,j=1}^k$.

Wenn beim Z-Vektor $\mathbf{X}$ auch alle X_i^2 mit $1 \leq i \leq k$ integrierbar sind, wird

$$\begin{aligned} \mathbb{V}\mathbf{X} &:= \text{Cov}[\mathbf{X}] := (\text{Cov}[X_i, X_j])_{i,j=1}^k \stackrel{(3.7)}{=} \mathbb{E}[(\mathbf{X} - \mathbb{E}\mathbf{X})(\mathbf{X} - \mathbb{E}\mathbf{X})^T] \\ &= (\mathbb{E}[(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)])_{i,j=1}^k \stackrel{(A.67)}{=} (\mathbb{E}[X_i X_j] - \mathbb{E}X_i \mathbb{E}X_j)_{i,j=1}^k \\ &= \mathbb{E}[\mathbf{X}\mathbf{X}^T] - \mathbb{E}\mathbf{X}\mathbb{E}\mathbf{X}^T \end{aligned} \quad (3.8)$$

als *Kovarianzmatrix von $\mathbf{X}$* bezeichnet.

Die Kovarianzmatrix *zweier* Z-Vektoren $\mathbf{X}, \mathbf{Y} : \Omega \rightarrow \mathbb{R}^k$, vorausgesetzt für jedes Tupel $1 \leq i, j \leq k$ ist $X_i \cdot Y_j$ integrierbar, lautet

$$\begin{aligned} \text{Cov}[\mathbf{X}, \mathbf{Y}] &:= (\text{Cov}[X_i, Y_j])_{i,j=1}^k = \mathbb{E}[(\mathbf{X} - \mathbb{E}\mathbf{X})(\mathbf{Y} - \mathbb{E}\mathbf{Y})^T] \\ &= \mathbb{E}[\mathbf{X}\mathbf{Y}^T] - \mathbb{E}\mathbf{X}\mathbb{E}\mathbf{Y}^T. \end{aligned} \quad (3.9)$$

Klarerweise erkennt man an (3.9), dass die Kovarianzmatrix unkorrelierter – und somit insbesondere unabhängiger – Z-Vektoren $\mathbf{0}$ ist ($\text{Cov}[X_i, Y_j] = 0 \ \forall 1 \leq i, j \leq k$ – vgl. den Satz A.54 (i)). Daraus ergibt sich der folgende Satz (welcher eine Erweiterung der *Bienaymé-Gleichung*, Satz A.55, im Fall der Unkorreliertheit für Z-Vektoren darstellt).

Satz 3.4 (Kovarianzmatrix der Summe unkorrelierter Z-Vektoren)

Seien $n \geq 1$ und $\{\mathbf{X}_i\}_{i=1}^n$ eine Familie unkorrelierter Z-Vektoren (hinreichend für die Unkorreliertheit ist ihre Unabhängigkeit). Dann gilt

$$\mathbb{V}\left[\sum_{i=1}^n \mathbf{X}_i\right] = \sum_{i=1}^n \mathbb{V}[\mathbf{X}_i]. \quad (3.10)$$

Beweis. Es genügt, den Satz für $n=2$ zu zeigen. Setze $\mathbf{X} := \mathbf{X}_1$ und $\mathbf{Y} := \mathbf{X}_2$. Dann gilt

$$\begin{aligned} \mathbb{V}[\mathbf{X} + \mathbf{Y}] &\stackrel{(3.8)}{=} \mathbb{E}[(\mathbf{X} + \mathbf{Y})(\mathbf{X} + \mathbf{Y})^T] - \mathbb{E}[\mathbf{X} + \mathbf{Y}]\mathbb{E}[\mathbf{X} + \mathbf{Y}]^T \\ &= \mathbb{E}[\mathbf{X}\mathbf{X}^T] + \mathbb{E}[\mathbf{X}\mathbf{Y}^T] + \mathbb{E}[\mathbf{Y}\mathbf{X}^T] + \mathbb{E}[\mathbf{Y}\mathbf{Y}^T] \\ &\quad - \mathbb{E}\mathbf{X}\mathbb{E}\mathbf{X}^T - \mathbb{E}\mathbf{X}\mathbb{E}\mathbf{Y}^T - \mathbb{E}\mathbf{Y}\mathbb{E}\mathbf{X}^T - \mathbb{E}\mathbf{Y}\mathbb{E}\mathbf{Y}^T \\ &\stackrel{(3.8)}{=} \mathbb{V}\mathbf{X} + \mathbb{V}\mathbf{Y} + \text{Cov}[\mathbf{X}, \mathbf{Y}] + \text{Cov}[\mathbf{Y}, \mathbf{X}], \\ &\stackrel{(3.9)}{=} \end{aligned}$$

wobei – wie vor dem Satz begründet – die beiden letzten Summanden $\mathbf{0}$ sind. $\square$

Wenn ein reeller Z-Vektor der Dimension k (oder eine $k \times k$ -Zufallsmatrix) aus *diskreten* reellen ZV gebildet wird, bleibt dieser diskret (siehe die Bemerkung A.43). Das gilt insbesondere für Zufallsvektoren, deren Komponenten aus $\mathbb{N}_0$ -wertigen Zufallsvariablen bestehen. Solchen Zufallsvektoren widmet sich der nächste Teilabschnitt.

$\mathbb{N}_0^k$ -wertige Zufallsvektoren und ihre Verteilung

Analog zur Definition 2.2 können Zufallsvektoren mit Werten in $\mathbb{N}_0^k$ aufgefasst werden.

Definition 3.5 ($\mathbb{N}_0^k$ -wertige Zufallsvektoren und ihre Verteilung)

Eine ZV $\mathbf{X} : \Omega \rightarrow \mathbb{N}_0^k$ mit $\mathbf{X} = (X_1, \dots, X_k)^T$ wird als $\mathbb{N}_0^k$ -wertiger Z-Vektor bezeichnet. Die Folge $\{p_{\mathbf{X}}(\mathbf{r})\}_{\mathbf{r} \in \mathbb{N}_0^k}$, wobei $p_{\mathbf{X}} : \mathbb{N}_0^k \rightarrow [0, 1]$ definiert ist als

$$p_{\mathbf{X}}(\mathbf{r}) := p_{\mathbf{X}}(r_1, \dots, r_k) := \mathbb{P}(\mathbf{X} = \mathbf{r}) \stackrel{(3.6)}{=} \mathbb{P}(X_1 = r_1, \dots, X_k = r_k) \quad (3.11)$$

(für $\mathbf{r} = (r_1, \dots, r_k)^T \in \mathbb{N}_0^k$), heißt *Wahrscheinlichkeitsfunktion von $\mathbf{X}$* .

3 Galton-Watson-Verzweigungsprozesse mit mehreren unterschiedlichen Typen

Bemerkung zur Def. 3.5. Die Bem. zur Def. 2.2 gilt im Wesentlichen analog für $\mathbb{N}_0^k$ -wertige Z-Vektoren. Insbesondere gilt

$$\bigcup_{\mathbf{r} \in \mathbb{N}_0^k} \{\mathbf{X} = \mathbf{r}\} \stackrel{(A.37)}{=} \bigcup_{\mathbf{r} \in \mathbb{N}_0^k} \mathbf{X}^{-1}(\mathbf{r}) \stackrel[\text{A.5 (c)}]{\text{Lemma}} \mathbf{X}^{-1} \left(\overbrace{\bigcup_{\mathbf{r} \in \mathbb{N}_0^k} \{\mathbf{r}\}}^{=\mathbb{N}_0^k} \right) \stackrel[\text{A.5(i)}]{\text{Lemma}} \Omega, \quad (3.12)$$

weshalb mit der σ -Additivität

$$\sum_{r_1, \dots, r_k \geq 0} p_{\mathbf{X}}(r_1, \dots, r_k) = \sum_{\mathbf{r} \in \mathbb{N}_0^k} p_{\mathbf{X}}(\mathbf{r}) \stackrel[\text{(A.27)}]{(3.11)} \mathbb{P} \left(\bigcup_{\mathbf{r} \in \mathbb{N}_0^k} \{\mathbf{X} = \mathbf{r}\} \right) \stackrel{(3.12)}{=} \mathbb{P}(\Omega) = 1 \quad \text{folgt.} \quad (3.13)$$

△

Die Wahrscheinlichkeitsfunktion $p_{\mathbf{X}}$ definiert also die Verteilung von $\mathbf{X}$ und somit die *gemeinsame Verteilung* aller $X_1, \dots, X_k$. Aus ihr kann aber auch die sogenannte *Marginal- oder Randverteilung* einer Komponente X_i mit $1 \leq i \leq k$ wie folgt gewonnen werden.

Lemma 3.6 (Randverteilung eines $\mathbb{N}_0^k$ -wertigen Z-Vektors)

Sei $\mathbf{X} = (X_1, \dots, X_k)^T$ wie in der Def. 3.5 und sei $1 \leq i \leq k$. Es gilt für jedes $r_i \geq 0$

$$\mathbb{P}(X_i = r_i) = \sum_{\substack{r_l \geq 0 \\ 1 \leq l \leq k; l \neq i}} \mathbb{P}(X_1 = r_1, \dots, X_k = r_k) \stackrel{(3.11)}{=} \sum_{\substack{r_l \geq 0 \\ 1 \leq l \leq k; l \neq i}} p_{\mathbf{X}}(r_1, \dots, r_k). \quad (3.14)$$

Beweis. Setze $\mathbf{r}^{(-i)} := (r_l)_{l=1; l \neq i}^k = (r_1, \dots, r_{i-1}, r_{i+1}, \dots, r_k)^T$ und def. den $\mathbb{N}_0^{k-1}$ -wertigen Z-Vektor $\mathbf{X}^{(-i)} := (X_l)_{l=1; l \neq i}^k = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_k)^T$. Dann gilt

$$\begin{aligned} \sum_{\substack{r_l \geq 0 \\ 1 \leq l \leq k; l \neq i}} \mathbb{P}(X_1 = r_1, \dots, X_k = r_k) &\stackrel{(3.6)}{=} \sum_{\mathbf{r}^{(-i)} \in \mathbb{N}_0^{k-1}} \mathbb{P}(\mathbf{X}^{(-i)} = \mathbf{r}^{(-i)}, X_i = r_i) \\ &\stackrel[\text{(A.27)}]{\sigma\text{-Add.}} \mathbb{P} \left(\bigcup_{\mathbf{r}^{(-i)} \in \mathbb{N}_0^{k-1}} (\{\mathbf{X}^{(-i)} = \mathbf{r}^{(-i)}\} \cap \{X_i = r_i\}) \right) \\ &\stackrel[\text{Distr.-gesetz}]{=} \mathbb{P} \left(\{X_i = r_i\} \cap \overbrace{\bigcup_{\mathbf{r}^{(-i)} \in \mathbb{N}_0^{k-1}} \{\mathbf{X}^{(-i)} = \mathbf{r}^{(-i)}\}}^{=\Omega \text{ ((3.12) für } \mathbf{X}^{(-i)})} \right) \stackrel{(A.33)}{=} \mathbb{P}(X_i = r_i). \quad \square \end{aligned}$$

Bemerkung zum Lemma 3.6. Durch einen sehr ähnlichen Beweis wie im Lemma erhält man für $1 \leq i, j \leq k$, $i \neq j$ und $r_i, r_j \geq 0$

$$\mathbb{P}(X_i = r_i, X_j = r_j) = \sum_{\substack{r_l \geq 0 \\ 1 \leq l \leq k; l \notin \{i, j\}}} \mathbb{P}(X_1 = r_1, \dots, X_k = r_k) = \sum_{\substack{r_l \geq 0 \\ 1 \leq l \leq k; l \notin \{i, j\}}} p_{\mathbf{X}}(r_1, \dots, r_k), \quad (3.15)$$

also die *zweidimensionale (gemeinsame) Randverteilung von X_i und X_j* (und auf eine vergleichbare Weise kann man auch höherdimensionale Randverteilungen erhalten). △

Des Weiteren soll noch ein Pendant zum Satz 2.3 (bzgl. des Erwartungswertes von Funktionen $\mathbb{N}_0^k$ -wertiger Zufallsvektoren) formuliert werden. Dieses ist ebenfalls eine direkte Anwendung des Satzes A.46 über den EW von Funktionen diskreter ZV.

Satz 3.7 (Erwartungswerte von Funktionen $\mathbb{N}_0^k$ -wertiger $\mathbf{Z}$ -Vektoren)

Sei $\mathbf{X} = (X_1, \dots, X_k)^T$ ein $\mathbb{N}_0^k$ -wertiger $\mathbf{Z}$ -Vektor mit W -Funktion $\{p_{\mathbf{X}}\}_{\mathbf{r} \in \mathbb{N}_0^k}$. Weiters sei $g : \mathbb{N}_0^k \rightarrow \bar{\mathbb{R}}$ mit $g(\mathbf{r}) := g(r_1, \dots, r_k)$ für $\mathbf{r} = (r_1, \dots, r_k)^T \in \mathbb{N}_0^k$ entweder nichtnegativ oder so, dass $\sum_{\mathbf{r} \in \mathbb{N}_0^k} |g(\mathbf{r})| p_{\mathbf{X}}(\mathbf{r}) < \infty$ gilt. Dann gilt für den EW der ZV $g(\mathbf{X})$

$$\begin{aligned}
\mathbb{E}g(\mathbf{X}) &\stackrel{(A.57)}{=} \sum_{\mathbf{r} \in \mathbb{N}_0^k} g(\mathbf{r}) p_{\mathbf{X}}(\mathbf{r}) \stackrel{(3.11)}{=} \sum_{\mathbf{r} \in \mathbb{N}_0^k} g(\mathbf{r}) \mathbb{P}(\mathbf{X} = \mathbf{r}) \\
&\stackrel{(3.6)}{=} \sum_{r_1, \dots, r_k \geq 0} g(r_1, \dots, r_k) \mathbb{P}(X_1 = r_1, \dots, X_k = r_k) \\
&\stackrel{(A.109)}{=} \sum_{(\star)} \left(\dots \left(\sum_{r_k \geq 0} g(r_1, \dots, r_k) \mathbb{P}(X_1 = r_1, \dots, X_k = r_k) \right) \dots \right).
\end{aligned} \tag{3.16}$$

- ($\star$) Es handelt sich um eine Mehrfachreihe, welche entweder absolut konvergent ist oder aus nichtnegativen Gliedern besteht. Bei solchen spielt die Summationsreihenfolge generell – im letzteren Fall auch ungeachtet des Konvergenzverhaltens – keine Rolle (siehe dazu etwa wieder die Ausführungen im Abschnitt A.3.2, insbesondere die Bemerkung zum Doppelreihensatz A.101 (iii)). Dies wird zukünftig wiederholt stillschweigend so verwendet.

3.1.4 Multivariate wahrscheinlichkeitserzeugende Funktionen

So wie im 2. Kapitel kommen auch für Multitype Galton-Watson-Verzweigungsprozesse wahrscheinlichkeitserzeugende Funktionen zum Einsatz. Diese werden aber – im Gegensatz dazu – nun verallgemeinernd multivariat definiert und beschreiben die Verteilung $\mathbb{N}_0^k$ -wertiger Zufallsvektoren.

Definition 3.8 (Multivariate wahrscheinlichkeitserzeugende Funktion)

Sei $\mathbf{X} = (X_1, \dots, X_k)^T$ ein $\mathbb{N}_0^k$ -wertiger Zufallsvektor mit W -Funktion $p_{\mathbf{X}}$ (siehe die Def. 3.5). Dann ist $f_{\mathbf{X}} : [0, 1]^k \rightarrow [0, 1]$ (also def. für alle $\mathbf{s} = (s_1, \dots, s_k)^T \in [0, 1]^k$) mit

$$\begin{aligned}
f_{\mathbf{X}}(\mathbf{s}) &:= f_{\mathbf{X}}(s_1, \dots, s_k) := \sum_{\mathbf{r} \in \mathbb{N}_0^k} p_{\mathbf{X}}(\mathbf{r}) \mathbf{s}^{\mathbf{r}} \stackrel{(3.3)}{=} \sum_{r_1, \dots, r_k \geq 0} p_{\mathbf{X}}(r_1, \dots, r_k) s_1^{r_1} \cdots s_k^{r_k} \\
&\stackrel{(3.11)}{=} \mathbb{E}[s_1^{X_1} \cdots s_k^{X_k}] = \mathbb{E} \left[\prod_{i=1}^k s_i^{X_i} \right] \stackrel{(3.3)}{=} \mathbb{E} \mathbf{s}^{\mathbf{X}}
\end{aligned} \tag{3.17}$$

die multivariate wahrscheinlichkeitserzeugende Funktion von $\mathbf{X}$, kurz weF .

Bemerkung zur Definition 3.8. Die Reihe in (3.17) konvergiert für alle $\mathbf{s} \in [0, 1]^k$, d.h., $f_{\mathbf{X}}(\mathbf{s})$ ist wohldefiniert (vgl. etwa [29, Def. A.1, Appendix A, S. 221]). $\triangle$

Notation 3.9. Es werden die folgenden – etwas saloppen – Vereinfachungen vereinbart.

- Sind partielle Ableitungen einer Funktion vorhanden und werden an einer bestimmten Stelle ausgewertet, so wird diese gleich als Argument geschrieben.

D.h., etwa bzgl. $f : \mathbb{R}^k \rightarrow \mathbb{R}$ steht $\frac{\partial f}{\partial s_i}(\tilde{s})$ anstelle von $\frac{\partial f}{\partial s_i}(s) \Big|_{s=\tilde{s}}$ ($\tilde{s} \in \mathbb{R}^k$, $1 \leq i \leq k$).

Genauso wird mit der Matrix aller partiellen Ableitungen, der sogenannten *Jacobi-Matrix* (siehe bspw. [23, Satz 164.1, S. 260]), verfahren. Diese wird beispielsweise für eine Funktion $\mathbf{f} = (f_1, \dots, f_k)^T : \mathbb{R}^k \rightarrow \mathbb{R}^k$ an der Stelle $\tilde{s} \in \mathbb{R}^k$ als

$$D\mathbf{f}(\tilde{s}) = \left(\frac{\partial f_i}{\partial s_j}(\tilde{s}) \right)_{i,j=1}^k \quad (\text{wenn } \mathbf{f} \text{ dort differenzierbar ist}) \quad \text{dargestellt.} \quad (3.18)$$

Für $f : \mathbb{R}^k \rightarrow \mathbb{R}$ ist das der transponierte Gradient, d.h. $Df(\tilde{s}) = \left(\left(\frac{\partial f}{\partial s_i}(\tilde{s}) \right)_{i=1}^k \right)^T$.

- Wenn $\mathbf{X}$ und $\mathbf{Y}$ zwei Z-Vektoren sind, ist ab sofort mit $(\mathbf{X}|\mathbf{Y}=\mathbf{y})$ der 'neue' Z-Vektor mit der *bedingten Verteilung von $\mathbf{X}$ gegeben $\mathbf{Y}=\mathbf{y}$* (vgl. die Bem. A.58) gemeint. Ist $\mathbf{X}$ $\mathbb{N}_0^k$ -wertig, so schreibe für seine weF

$$f_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(s) \stackrel{(3.17)}{=} \mathbb{E} s^{\mathbf{X}|\mathbf{Y}=\mathbf{y}} \stackrel{(\star)}{=} \mathbb{E}[s^{\mathbf{X}}|\mathbf{Y}=\mathbf{y}] \quad (3.19)$$

und bezeichne sie als *bedingte weF von $\mathbf{X}$ gegeben $\mathbf{Y}=\mathbf{y}$* (vgl. z.B. [39, S. 9 f.]).

$$(\star) \quad \mathbb{E} s^{\mathbf{X}|\mathbf{Y}=\mathbf{y}} \stackrel{(3.16)}{\stackrel{(A.73)}}{=} \sum_{\mathbf{r} \in \mathbb{N}_0^k} s^{\mathbf{r}} p_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{r}) \stackrel{(3.16)}{\stackrel{(A.73)}}{=} \mathbb{E}_{\mathbb{P}_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}} s^{\mathbf{X}} \stackrel{(A.79)}{=} \mathbb{E}[s^{\mathbf{X}}|\mathbf{Y}=\mathbf{y}].$$

Beachte, dass – streng genommen – jeweils $\mathbb{P}(\mathbf{Y}=\mathbf{y}) > 0$ sein müsste. Oft können aber auch bei Verletzung dieser Voraussetzung wieder korrekte, wohldefinierte Ausdrücke angegeben werden (bspw. bei (3.29) bzgl. einer bedingten Verteilung oder bei (3.42) bzgl. einer bedingten weF bzw. Erwartung – wird bei Letzterem der totale EW gebildet, spielen diese Nullmengen zudem ohnehin keine Rolle). Deshalb ist dies in keinem der später vorkommenden Fälle problematisch und es wird zukünftig meist nicht mehr darauf hingewiesen. $\triangle$

Nun sollen die Eigenschaften von $f_{\mathbf{X}}$ beleuchtet werden. Viele davon sind analoge Erweiterungen der Resultate im Unterabschnitt 2.2.2 und eine Auswahl davon soll der nächste Satz bieten. Insbesondere der Punkt (b) zeigt – wie im univariaten Fall (siehe den Satz 2.6) – dass $\mathbf{X}$ durch $f_{\mathbf{X}}$ eindeutig charakterisiert ist.

Satz 3.10 (Eigenschaften multivariater weF – angelehnt an [29, Theorem A.1, Appendix A, S. 221 f.]

Sei alles wie in der Def. 3.8 und $1 \leq i \leq k$ beliebig. Es gilt

- (a) *$f_{\mathbf{X}}$ ist stetig und mindestens auf $[0, 1)^k = \{\mathbf{x} : 0 \leq x_i < 1 \ \forall 1 \leq i \leq k\}$ beliebig oft stetig (total) differenzierbar. Als partielle Funktion von s_i ist $f_{\mathbf{X}}$ isoton und konvex. Falls $\mathbb{P}(X_i = 0) < 1$ gilt, ist die Isotonie sogar strikt und falls $\mathbb{P}(X_i \leq 1) < 1$ gilt, die Konvexität ebenso.*

3.1 Vorbemerkungen

(b) Die W-Funktion von $\mathbf{X}$ kann für alle $(r_1, \dots, r_k)^T \in \mathbb{N}_0^k$ erzeugt werden mit

$$p_{\mathbf{X}}(r_1, \dots, r_k) = \frac{1}{r_1! \cdots r_k!} \frac{\partial^{r_1 + \dots + r_k} f_{\mathbf{X}}}{\partial s_1^{r_1} \cdots \partial s_k^{r_k}}(\mathbf{0}).$$

(c) Es gilt $f_{\mathbf{X}}(\mathbf{0}) = p_{\mathbf{X}}(\mathbf{0})$ und $f_{\mathbf{X}}(\mathbf{1}) = 1$.

(d) Die weF f_{X_i} der Zufallsvariablen X_i kann dadurch gewonnen werden, dass die s_j für alle $1 \leq j \leq k$ mit $j \neq i$ auf 1 gesetzt werden. D.h. $f_{X_i}(s_i) = f_{\mathbf{X}}(\mathbf{s})|_{s_j=1 \forall j \neq i}$ (für $s_i \in [0, 1]$). Das ist somit die (univariate) weF der i -ten Randverteilung von $\mathbf{X}$.

(e) $\mathbb{E}X_i$ ist genau dann endlich (d.h. $\mathbb{E}X_i < \infty$), wenn die gleich folgende partielle Ableitung existiert. In diesem Fall gilt

$$\frac{\partial f_{\mathbf{X}}}{\partial s_i}(\mathbf{1}) = \mathbb{E}X_i \quad \left(\begin{array}{l} \text{falls } \mathbb{E}X_j < \infty \\ \forall 1 \leq j \leq k \text{ gilt} \end{array} \mathbf{D}f_{\mathbf{X}}(\mathbf{1}) \stackrel{(3.18)}{=} \mathbb{E}\mathbf{X}^T \right) \quad (3.20)$$

und diese erste Ableitung ist auch im Punkt $\mathbf{1}$ (zumindest linksseitig) stetig.

(f) Sei $1 \leq j \leq k$. Es ist $\mathbb{E}[X_i X_j] < \infty$ genau dann, wenn die gleich folgende partielle Ableitung zweiter Ordnung existiert und dann gilt

$$\frac{\partial^2 f_{\mathbf{X}}}{\partial s_i \partial s_j}(\mathbf{1}) = \begin{cases} \mathbb{E}[X_i X_j], & \text{wenn } i \neq j \text{ ist,} \\ \mathbb{E}[X_i(X_i - 1)] \quad \left(= \frac{\partial^2 f_{\mathbf{X}}}{\partial s_i^2}(\mathbf{1}) \right), & \text{wenn } i = j \text{ ist.} \end{cases} \quad (3.21)$$

(g) Sei $\mathbf{Y}$ ein weiterer $\mathbb{N}_0^k$ -wertiger Zufallsvektor mit der weF $f_{\mathbf{Y}}$, welcher unabhängig von $\mathbf{X}$ ist. Für die weF der Summe $\mathbf{X} + \mathbf{Y}$ gilt dann $f_{\mathbf{X}+\mathbf{Y}} = f_{\mathbf{X}} f_{\mathbf{Y}}$. Wenn $\{\mathbf{X}^{(j)}\}_{j=1}^n$ eine Familie von $n \geq 1$ unabhängigen, $\mathbb{N}_0^k$ -wertigen Z-Vektoren ist (jeder mit der weF $f_{\mathbf{X}^{(j)}}$), gilt weiters allgemein

$$f_{\mathbf{X}^{(1)} + \dots + \mathbf{X}^{(n)}} = f_{\sum_{j=1}^n \mathbf{X}^{(j)}} = \prod_{j=1}^n f_{\mathbf{X}^{(j)}}. \quad (3.22)$$

Beweis(-skizze). Die Eigenschaften sind – wie erwähnt – im Wesentlichen das multivariate Pendant zu jenen der univariaten weF. Sie werden nicht in der gleichen Ausführlichkeit wie in 2.2.2, sondern nur teilweise exakt bewiesen.

Für die Isotonie in (a) betrachte die erste partielle Ableitung nach s_i , also

$$\frac{\partial f_{\mathbf{X}}}{\partial s_i}(\mathbf{s}) = \sum_{r_1, \dots, r_k \geq 0} p_{\mathbf{X}}(r_1, \dots, r_k) s_1^{r_1} \cdots r_i s_i^{r_i-1} \cdots s_k^{r_k}, \quad (\star)$$

welche für alle $\mathbf{s} \in (0, 1)^k$ nichtnegativ ist. Die Isotonie ist nur dann *nicht* strikt, wenn diese Ableitung $(\star)$ 0 ist. Dafür muss jeder Summand 0 sein, was für solche mit $r_i = 0$ ohnedies so ist. Für alle anderen wiederum ist das genau dann der Fall, wenn

$$\sum_{r_i \geq 1} \sum_{\substack{r_l \geq 0 \\ 1 \leq l \leq k; l \neq i}} p_{\mathbf{X}}(r_1, \dots, r_k) \stackrel{(3.14)}{=} \sum_{r_i \geq 1} \mathbb{P}(X_i = r_i) \stackrel{(2.2)}{=} 1 - \mathbb{P}(X_i = 0) = 0 \quad \text{gilt.}$$

3 Galton-Watson-Verzweigungsprozesse mit mehreren unterschiedlichen Typen

Für alle $\mathbf{s}$ wie oben ist dementsprechend umgekehrt $\frac{\partial f_{\mathbf{X}}}{\partial s_i}(\mathbf{s}) > 0$, also die Isotonie bzgl. s_i strikt, sobald $1 - \mathbb{P}(X_i = 0) > 0$ (sprich $\mathbb{P}(X_i = 0) < 1$) gilt.

Für die restlichen Aussagen in (a) und (b) siehe 1. und 3. des oben genannten Theorems. Sie sind Erweiterungen der Sätze 2.5 (b), (d) und (e) sowie 2.6.

Der erste Teil von (c) ist ein Spezialfall von (b) und den zweiten erhält man mit

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{1}) &\stackrel{(3.17)}{=} \sum_{r_1, \dots, r_k \geq 0} p_{\mathbf{X}}(r_1, \dots, r_k) 1^{r_1} \dots 1^{r_k} \stackrel{(3.13)}{=} 1. & \text{Für (d) betrachte} \\ f_{\mathbf{X}}(1, \dots, 1, s_i, 1, \dots, 1) &\stackrel{(3.17)}{=} \sum_{r_1, \dots, r_k \geq 0} p_{\mathbf{X}}(r_1, \dots, r_k) \overbrace{1^{r_1} \dots 1^{r_{i-1}}}^{=1} s_i^{r_i} \overbrace{1^{r_{i+1}} \dots 1^{r_k}}^{=1} \\ &= \sum_{r_i \geq 0} \left(\sum_{\substack{r_l \geq 0 \\ 1 \leq l \leq k; l \neq i}} p_{\mathbf{X}}(r_1, \dots, r_k) \right) s_i^{r_i} = \sum_{r_i \geq 0} \mathbb{P}(X_i = r_i) s_i^{r_i} \stackrel{(2.6)}{=} f_{X_i}(s_i). \end{aligned}$$

Die Punkte (e) und (f) sind als Spezialfälle im 4. Punkt des Theorems der oben stehenden Referenz beinhaltet. Diesem können somit auch die Rechtfertigungen für die beiden Äquivalenzen entnommen werden. Außerdem sind sie die Pendants zum Lemma 2.7 (mit $l=1$ bzw. $l=2$) für den multivariaten Fall, weshalb die Argumente denen für die letzte Aussage ebenda gleichen. Um (3.20) zu zeigen, betrachte erneut die partielle Ableitung $(\star)$, nun an der Stelle $\mathbf{1}$. Ähnlich wie oben kann

$$\frac{\partial f_{\mathbf{X}}}{\partial s_i}(\mathbf{1}) = \sum_{r_i \geq 0} \left(\sum_{\substack{r_l \geq 0 \\ 1 \leq l \leq k; l \neq i}} p_{\mathbf{X}}(r_1, \dots, r_k) \right) r_i \stackrel{(2.4)}{=} \mathbb{E}X_i \quad \text{erhalten werden.}$$

Durch erneutes partielles Ableiten führt eine analoge Vorgehensweise (für (3.21)) zu

$$\begin{aligned} \frac{\partial^2 f_{\mathbf{X}}}{\partial s_i^2}(\mathbf{1}) &= \sum_{r_i \geq 0} \left(\sum_{\substack{r_l \geq 0 \\ 1 \leq l \leq k; l \neq i}} p_{\mathbf{X}}(r_1, \dots, r_k) \right) r_i(r_i - 1) \stackrel{(3.14)}{\stackrel{(2.3)}}{=} \mathbb{E}[X_i(X_i - 1)] \quad \text{für } i = j \text{ bzw.} \\ \frac{\partial^2 f_{\mathbf{X}}}{\partial s_i \partial s_j}(\mathbf{1}) &= \sum_{r_i, r_j \geq 0} \left(\sum_{\substack{r_l \geq 0 \\ 1 \leq l \leq k; l \notin \{i, j\}}} p_{\mathbf{X}}(r_1, \dots, r_k) \right) r_i r_j \stackrel{(3.16)}{=} \mathbb{E}[X_i X_j] \quad \text{für } i \neq j. \\ &\quad \stackrel{= \mathbb{P}(X_i = r_i, X_j = r_j) \text{ mit (3.15)}}{} \end{aligned}$$

Schließlich findet sich die erste Aussage in (g) als 5. Punkt im besagten Theorem, ist aber auch ein Sonderfall von (3.22), was wiederum die multivariate Version des Lemmas 2.10 ist. Beachte, dass das Unabhängigkeitsargument zu Beginn des entsprechenden Beweises genauso für die Familie $\{\mathbf{s}^{\mathbf{X}^{(j)}}\}_{j=1}^n$ gilt, da das Lemma A.34 für allgemeine Messräume formuliert wurde. Für ein beliebiges $\mathbf{s} \in [0, 1]^k$ ist dabei wegen (3.3) jedes $\mathbf{s}^{\mathbf{X}^{(j)}}$ eine (eindimensionale) ZV ($\mathbf{s}^{\mathbf{X}^{(j)}} : \Omega \rightarrow [0, 1]$). Deshalb kann der restliche Beweis völlig analog – nun mit Vektoren – durchgeführt werden. $\square$

3.2 Definition des Multitype GW-Verzweigungsprozesses

Analog zum einfachen GWP des 2. Kapitels soll der Galton-Watson-Prozess mit mehreren unterschiedlichen Typen als zeitdiskreter stochastischer Prozess definiert werden. Dafür wird wieder durchwegs $n \geq 0$ als Zeitindex verwendet und weiterhin von Generationen gesprochen. In den Vorbemerkungen unter 3.1 wurde aber bereits angekündigt, dass nun nicht mehr 'Individuen gebären' sondern 'Objekte (re-)produzieren'.

Dementsprechend sollen auch die Termini *Reproduktionszahl* und *Reproduktionsverteilung* anstelle von Nachkommenszahl bzw. -verteilung verwendet werden. Die von einem Objekt 'stammenden' bzw. ausgehenden Objekte der nächsten Generation werden schlicht als *die von diesem Objekt produzierten Objekte* betitelt. Mit *nachfolgenden Objekten* oder *Nachfolgern* sind hingegen jene aller kommenden Generationen gemeint.

Ein Objekt kann also eines Typs von k unterschiedlichen Typen sein.

3.2.1 Grundlegende Definitionen

Sei wieder $(\Omega, \Sigma, \mathbb{P})$ ein geeigneter W-Raum, auf dem nun der Multitype Galton-Watson-Prozess definiert werden kann.

Definition 3.11 (Reproduktionszahlen von Objekten)

Für jedes $1 \leq j \leq k$ sei $\mathbf{X}^{(j)} = (X_l^{(j)})_{l=0}^k = (X_1^{(j)} \cdots X_k^{(j)})^T$ ein $\mathbb{N}_0^k$ -wertiger Zufallsvektor (mit Verteilung $p_{\mathbf{X}^{(j)}}$). Er beschreibe die *Reproduktionszahlen eines typischen* (oder generischen) *Objekts vom Typ j* . Dabei ist jede Komponente $X_l^{(j)}$ eine $\mathbb{N}_0$ -wertige Zufallsvariable und steht für die Anzahl an (von diesem Objekt) produzierten Objekten vom Typ l ($1 \leq l \leq k$).

Weiters repräsentiere $n \geq 0$ die Generation. Dann stellt für jedes $1 \leq j \leq k$ die Familie der Zufallsvektoren $\{\mathbf{X}_{n,i}^{(j)}\}_{n \geq 0, i \geq 1}$ jeweils die *Anzahlen produzierter Objekte* (vom Typ $1 \leq l \leq k$) *des i -ten Typ- j -Objekts der n -ten Generation* dar. Diese

$$\{\mathbf{X}_{n,i}^{(j)}\}_{n \geq 0, i \geq 1} \text{ sind unabhängige, identisch verteilte (i.i.d.-)Kopien von } \mathbf{X}^{(j)}. \quad (3.23)$$

Definition 3.12 (Reproduktionsverteilungen der Objekte)

Schreibe für jedes $1 \leq j \leq k$ die *Reproduktionsverteilung* bzw. W-Funktion von $\mathbf{X}^{(j)}$ kürzer als $p^{(j)}$. D.h., für alle $\mathbf{r} \in \mathbb{N}_0^k$ sei

$$\begin{aligned} p^{(j)}(\mathbf{r}) &:= p^{(j)}(r_1, \dots, r_k) := p_{\mathbf{X}^{(j)}}(\mathbf{r}) \stackrel{(3.11)}{=} \mathbb{P}(\mathbf{X}^{(j)} = \mathbf{r}) \\ &= \mathbb{P}(X_1^{(j)} = r_1, \dots, X_k^{(j)} = r_k). \end{aligned} \quad \text{Es gilt } \sum_{\mathbf{r} \in \mathbb{N}_0^k} p^{(j)}(\mathbf{r}) \stackrel{(3.13)}{=} 1. \quad (3.24)$$

Sie beschreibt die WK dafür, dass ein Objekt vom Typ j r_1 Objekte vom Typ 1, r_2 Objekte vom Typ 2, usw. bis r_k Objekte vom Typ k produziert.

Definition 3.13 (Populationsgrößen des Multitype GWP)

Sei $\{\mathbf{Z}_n\}_{n \geq 0}$ eine Folge $\mathbb{N}_0^k$ -wertiger Zufallsvektoren – wobei stets $\mathbf{Z}_0$ als deterministisch, also *nichtzufällig*, angenommen wird – welche durch die Rekursion

$$\mathbf{Z}_{n+1} := \sum_{j=1}^k \sum_{i=1}^{Z_{nj}} \mathbf{X}_{n,i}^{(j)} \quad \text{für alle } n \geq 0 \quad (3.25)$$

definiert wird. Dabei beschreibt $\mathbf{Z}_n = (Z_{n1}, \dots, Z_{nk})^T$ die *Populationsgrößen der Generation* $n \geq 0$. Für jedes $1 \leq j \leq k$ ist die Komponente Z_{nj} die Gesamtanzahl an Objekten vom Typ j in der Generation n .

Ist $Z_{nj} = 0$, so ist die innere Summe in (3.25) konventionsgemäß leer, d.h. $\sum_{i=1}^0 \mathbf{X}_{n,i}^{(j)} = \mathbf{0}$. Demnach folgt aus $\mathbf{Z}_n = \mathbf{0}$ – d.h. $Z_{nj} = 0$ für alle $1 \leq j \leq k$ – auch, dass $\mathbf{Z}_{n+1} = \mathbf{0}$ ist.

Wie im einfachen Fall, werden nun die $\{\mathbf{Z}_n\}_{n \geq 0}$ zusammen mit den $\{\mathbf{X}_{n,i}^{(j)}\}_{n \geq 0, i \geq 1}$ als *Galton-Watson-Verzweigungsprozess mit k unterschiedlichen Typen* oder kurz als **Multitype GWP** bzw. nur **MGWP** bezeichnet. Beachte, dass die obigen Definitionen (mit $k = 1$) die einfachen GW-Prozesse des 2. Kapitels beinhalten. Die MGWP sind also wie beabsichtigt eine direkte Verallgemeinerung dieser.

Sie werden in der Literatur ebenfalls – etwa wieder in [19] oder [3] – oft direkt als Markov-Kette definiert. Aufgrund der zeitlich konstanten Reproduktionsverteilungen gilt aber auch mit der hier getroffenen, äquivalenten Definition erneut, dass $\{\mathbf{Z}_n\}_{n \geq 0}$ die Kriterien für eine solche Kette erfüllt. Der Beweis dafür funktioniert im Wesentlichen gleich wie im einfachen Fall und die Formulierung eines entsprechenden Satzes sowie die Definition der zeithomogenen Übergangswahrscheinlichkeiten geschieht im Unterabschnitt 3.3.4.

Zur Analyse der MGWP ist es oft hilfreich, die Spezialfälle, welche nur von einem Objekt eines gewissen Typs gestartet (bzw. initiiert) werden, gesondert zu betrachten.

Definition 3.14 (Multitype GWP initiiert von einem einzelnen Objekt)

Sei $1 \leq j \leq k$. Angenommen, $\mathbf{Z}_0$ besteht nur aus einem Objekt vom Typ j , also $\mathbf{Z}_0 \stackrel{(3.2)}{=} \mathbf{e}_j$. Dann wird der Prozess als $\{\mathbf{Z}_n^{(j)}\}_{n \geq 0}$ geschrieben, d.h., für jedes $n \geq 0$ ist

$$\mathbf{Z}_n^{(j)} = (Z_{nl}^{(j)})_{l=1}^k := \text{'}\mathbf{Z}_n \text{ gegeben } \mathbf{Z}_0 = \mathbf{e}_j\text{' } = (\mathbf{Z}_n | \mathbf{Z}_0 = \mathbf{e}_j) \quad \text{und } \mathbf{Z}_0^{(j)} = \mathbf{e}_j. \quad (3.26)$$

Bemerkungen zur Definition 3.14.

- (i) Die Rekursion (3.25) gilt klarerweise insbesondere auch für $\mathbf{Z}_n^{(j)}$.

$$\text{Deshalb gilt für } n = 1 \text{ wegen } Z_{0l}^{(j)} \stackrel{(3.2)}{\stackrel{(3.26)}}{=} \begin{cases} 1 & \text{für } l = j, \\ 0 & \text{für } l \neq j \end{cases} \quad (\star)$$

$$\mathbf{Z}_1^{(j)} \stackrel{(3.25)}{=} \sum_{l=1}^k \sum_{i=1}^{Z_{0l}^{(j)}} \mathbf{X}_{0,i}^{(l)} \stackrel{(\star)}{=} \sum_{i=1}^1 \mathbf{X}_{0,i}^{(j)} = \mathbf{X}_{0,1}^{(j)} \stackrel{(i.i.d.)}{\stackrel{d}{=}} \mathbf{X}^{(j)}. \quad (3.27)$$

Somit hat $\mathbf{Z}_1^{(j)}$ die gleiche Verteilung wie $\mathbf{X}^{(j)}$.

3.2 Definition des Multitype GW-Verzweigungsprozesses

- (ii) $\{\mathbf{Z}_n\}_{n \geq 0}$ gegeben $\mathbf{Z}_0 = \mathbf{r} \in \mathbb{N}_0^k$ kann (analog zum Korollar 2.31 bzw. (2.51)) als die Summe $r_1 + \dots + r_k$ unabhängiger Prozesse, von welchen r_1 verteilt sind wie $\{\mathbf{Z}_n^{(1)}\}_{n \geq 0}$, r_2 wie $\{\mathbf{Z}_n^{(2)}\}_{n \geq 0}$ usw. bis r_k wie $\{\mathbf{Z}_n^{(k)}\}_{n \geq 0}$, aufgefasst werden (vgl. auch [3, *Additive Property*, S. 184]). D.h., für alle $\mathbf{r} \in \mathbb{N}_0^k$ und $n \geq 0$ gilt

$$(\mathbf{Z}_n | \mathbf{Z}_0 = \mathbf{r}) \stackrel{d}{=} \sum_{j=1}^k \sum_{i=1}^{r_j} \mathbf{Z}_{n,i}^{(j)} \quad \left(\{\mathbf{Z}_{n,i}^{(j)}\}_{i \geq 1} \text{ sind i.i.d.-Kopien von } \mathbf{Z}_n^{(j)} \right). \quad (3.28)$$

Analog verhält es sich, wenn für $m \geq 0$ der Prozess $\{\mathbf{Z}_{n+m}\}_{n \geq 0}$ gegeben $\mathbf{Z}_m = \mathbf{r}$ betrachtet wird. Mit einer Überlegung, die der im einfachen Fall für (2.47) getätigten sehr ähnlich ist (vgl. auch [19, *Remark*, S. 37]), erhält man für alle $\mathbf{r} \in \mathbb{N}_0^k$ und $m, n \geq 0$

$$(\mathbf{Z}_{n+m} | \mathbf{Z}_m = \mathbf{r}) \stackrel{d}{=} \sum_{j=1}^k \sum_{i=1}^{r_j} \mathbf{Z}_{n,i}^{(j)} \quad \left(\{\mathbf{Z}_{n,i}^{(j)}\}_{i \geq 1} \text{ sind i.i.d.-Kopien von } \mathbf{Z}_n^{(j)} \right). \quad (3.29)$$

An (3.28) und (3.29) erkennt man bereits die Markov-Eigenschaft, was – wie erwähnt – im Unterabschnitt 3.3.4 präzisiert wird (siehe insbesondere die Bem. 3.44).

- (iii) Bzgl. den einfachen GWP wurde die Beziehung der Populationsgröße einer Generation zur nächsten sowohl als Vorwärts- (2.35) als auch als Rückwärtsgleichung (2.48) bezeichnet bzw. formuliert – vgl. die Bemerkung 2.29. Analoges ist auch für den Multitype GWP möglich. Fasse den Prozess $\{\mathbf{Z}_n\}_{n \geq 0}$ dafür in der $(n+1)$ -ten Generation – wie in (ii) – als die Summe $Z_{11} + \dots + Z_{1k}$ unabhängiger Prozesse, Z_{11} davon verteilt wie $\{\mathbf{Z}_n^{(1)}\}_{n \geq 0}$, Z_{12} davon verteilt wie $\{\mathbf{Z}_n^{(2)}\}_{n \geq 0}$, usw. bis Z_{1k} davon verteilt wie $\{\mathbf{Z}_n^{(k)}\}_{n \geq 0}$, auf. So lässt sich für alle $n \geq 0$ die *Rückwärtsgleichung* schreiben als

$$\mathbf{Z}_{n+1} \stackrel{d}{=} \sum_{j=1}^k \sum_{i=1}^{Z_{1j}} \mathbf{Z}_{n,i}^{(j)} \quad \left(\{\mathbf{Z}_{n,i}^{(j)}\}_{i \geq 1} \text{ sind i.i.d.-Kopien von } \mathbf{Z}_n^{(j)} \right). \quad (3.30)$$

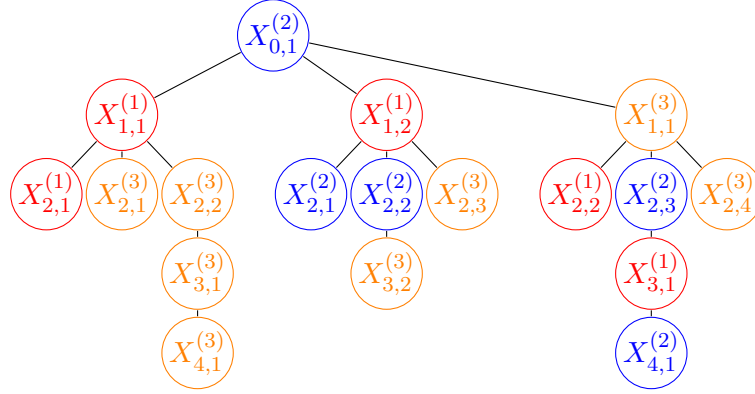
△

3.2.2 Exemplarische Realisierung

Insbesondere die Definition 3.11 der Reproduktionszahlen kann als etwas schwerfällig bzw. verklausuliert lesbar empfunden werden. Deshalb soll der definierte Prozess anhand eines konkreten Beispiels illustriert werden.

Beispiel 3.15 (Exemplarische Realisierung eines Multitype GWP)

In der Abbildung 3.1 wird eine denkbare Realisierung eines MGWP mit $k = 3$ verschiedenen Typen dargestellt. Dieser startete mit einem Objekt vom Typ zwei und ist ab der fünften Generation ausgestorben. Jede 'Zeile' steht für eine Generation und die einzelnen Objekte werden durch ihre jeweiligen Reproduktionszahlen repräsentiert.



Abbildungung 3.1: Eine mögliche Realisierung; $k = 3$.

Sei $\tilde{\omega}$ das entsprechende Ergebnis aus Ω , welches zu diesem exemplarischen Verzweigungsbaum führte. Die Realisierungen der Zufallsvektoren für die Populationsgrößen $\mathbf{Z}_n$ bis zum Aussterben, d.h. $\mathbf{z}_n := \mathbf{Z}_n(\tilde{\omega})$ mit $1 \leq n \leq 5$, sind

$$\mathbf{z}_0 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{z}_1 = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{z}_2 = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}, \quad \mathbf{z}_3 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \quad \mathbf{z}_4 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad \mathbf{z}_5 = \mathbf{0}.$$

Für jede Generation mit vorhandenen Objekten wird nun das Folgende angeführt: Die Ausdrücke in geschwungenen Klammern sind die *Reproduktionszahlen*, die für diese Entwicklung maßgeblich waren. Dabei handelt es sich nicht etwa um Matrizen, sondern um Listen, wobei in jeder j -te Zeile die Z-Vektoren $\mathbf{X}_{n,i}^{(j)}$ der Typ- j -Objekte stehen (diese beschreiben also die Anzahlen an von ihnen produzierten Objekten und alle einer Zeile besitzen definitionsgemäß jeweils die gleiche Verteilung, nämlich die von $\mathbf{X}^{(j)}$). Auch wenn die $\{\mathbf{X}_{n,i}^{(j)}\}_{n \geq 0, i \geq 1}$ grundsätzlich für alle $i \geq 1$ definiert sind, sind jene der Objekte, welche in diesem Beispiel nicht realisiert wurden, grau gedruckt. Überschreitet in der jeweiligen Generation die Spaltenanzahl die maximale Anzahl an vorhandenen Objekten pro Typ, werden die restlichen (nicht relevanten) durch $\dots$ symbolisiert.

3.2 Definition des Multitype GW-Verzweigungsprozesses

Zuvor wird jeweils die realisierte Populationsgröße $\mathbf{z}_n$ (welche alle auch bereits eingangs zusammenfassend angeführt wurden) bestimmt – bis auf die nichtzufällige initiale Populationsgröße mit der Rekursion (3.25).

An dieser Rekursion sieht man, dass sie auf den realisierten Reproduktionszahlen der vorhergehenden Generation basieren (d.h. auf $\mathbf{x}_{n,i}^{(j)} := \mathbf{X}_{n,i}^{(j)}(\tilde{\omega})$ mit $0 \leq n \leq 4$, $1 \leq j \leq 3$ und $i \geq 1$). Diese stehen – bis zur vierten Generation – jeweils abschließend dabei.

- $n = 0$: $\mathbf{z}_0 = (0, 1, 0)^T$ (initiale Populationsgröße)

$$\begin{pmatrix} \mathbf{X}_{0,1}^{(1)} & \cdots \\ \mathbf{X}_{0,1}^{(2)} & \cdots \\ \mathbf{X}_{0,1}^{(3)} & \cdots \end{pmatrix} \quad \mathbf{x}_{0,1}^{(2)} = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$$

- $n = 1$: $\mathbf{z}_1 = \sum_{j=1}^k \left(\sum_{i=1}^{z_{0j}} \mathbf{x}_{0,i}^{(j)} \right) = \mathbf{x}_{0,1}^{(2)} = (2, 0, 1)^T$

$$\begin{pmatrix} \mathbf{X}_{1,1}^{(1)} & \mathbf{X}_{1,2}^{(1)} & \cdots \\ \mathbf{X}_{1,1}^{(2)} & \mathbf{X}_{1,2}^{(2)} & \cdots \\ \mathbf{X}_{1,1}^{(3)} & \mathbf{X}_{1,2}^{(3)} & \cdots \end{pmatrix} \quad \mathbf{x}_{1,1}^{(1)} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \quad \mathbf{x}_{1,2}^{(1)} = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}, \quad \mathbf{x}_{1,1}^{(3)} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

- $n = 2$: $\mathbf{z}_2 = \sum_{j=1}^k \left(\sum_{i=1}^{z_{1j}} \mathbf{x}_{1,i}^{(j)} \right) = \left(\mathbf{x}_{1,1}^{(1)} + \mathbf{x}_{1,2}^{(1)} \right) + \mathbf{x}_{1,1}^{(3)} = (2, 3, 4)^T$

$$\begin{pmatrix} \mathbf{X}_{2,1}^{(1)} & \mathbf{X}_{2,2}^{(1)} & \mathbf{X}_{2,3}^{(1)} & \mathbf{X}_{2,4}^{(1)} & \cdots \\ \mathbf{X}_{2,1}^{(2)} & \mathbf{X}_{2,2}^{(2)} & \mathbf{X}_{2,3}^{(2)} & \mathbf{X}_{2,4}^{(2)} & \cdots \\ \mathbf{X}_{2,1}^{(3)} & \mathbf{X}_{2,2}^{(3)} & \mathbf{X}_{2,3}^{(3)} & \mathbf{X}_{2,4}^{(3)} & \cdots \end{pmatrix} \quad \mathbf{x}_{2,2}^{(2)} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{x}_{2,3}^{(2)} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{x}_{2,2}^{(3)} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\mathbf{x}_{2,1}^{(1)} = \mathbf{x}_{2,2}^{(1)} = \mathbf{x}_{2,1}^{(2)} = \mathbf{x}_{2,1}^{(3)} = \mathbf{x}_{2,3}^{(3)} = \mathbf{x}_{2,4}^{(3)} = \mathbf{0}$$

- $n = 3$: $\mathbf{z}_3 = \sum_{j=1}^k \left(\sum_{i=1}^{z_{2j}} \mathbf{x}_{2,i}^{(j)} \right) = \left(\mathbf{x}_{2,1}^{(1)} + \mathbf{x}_{2,2}^{(1)} \right) + \left(\mathbf{x}_{2,1}^{(2)} + \mathbf{x}_{2,2}^{(2)} + \mathbf{x}_{2,3}^{(2)} \right) + \left(\mathbf{x}_{2,1}^{(3)} + \mathbf{x}_{2,2}^{(3)} + \mathbf{x}_{2,3}^{(3)} + \mathbf{x}_{2,4}^{(3)} \right)$

$$= (\mathbf{0} + \mathbf{0}) + \left(\mathbf{0} + \mathbf{x}_{2,2}^{(2)} + \mathbf{x}_{2,3}^{(2)} \right) + \left(\mathbf{0} + \mathbf{x}_{2,2}^{(3)} + \mathbf{0} + \mathbf{0} \right) = (1, 0, 2)^T$$

$$\begin{pmatrix} \mathbf{X}_{3,1}^{(1)} & \mathbf{X}_{3,2}^{(1)} & \cdots \\ \mathbf{X}_{3,1}^{(2)} & \mathbf{X}_{3,2}^{(2)} & \cdots \\ \mathbf{X}_{3,1}^{(3)} & \mathbf{X}_{3,2}^{(3)} & \cdots \end{pmatrix} \quad \mathbf{x}_{3,1}^{(1)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{x}_{3,1}^{(3)} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{x}_{3,2}^{(3)} = \mathbf{0}$$

- $n = 4$: $\mathbf{z}_4 = \sum_{j=1}^k \left(\sum_{i=1}^{z_{3j}} \mathbf{x}_{3,i}^{(j)} \right) = \mathbf{x}_{3,1}^{(1)} + \left(\mathbf{x}_{3,1}^{(3)} + \mathbf{x}_{3,2}^{(3)} \right) = (0, 1, 1)^T$

$$\begin{pmatrix} \mathbf{X}_{4,1}^{(1)} & \cdots \\ \mathbf{X}_{4,1}^{(2)} & \cdots \\ \mathbf{X}_{4,1}^{(3)} & \cdots \end{pmatrix} \quad \mathbf{x}_{4,1}^{(2)} = \mathbf{x}_{4,1}^{(3)} = \mathbf{0}$$

- $n = 5$: $\mathbf{z}_5 = \sum_{j=1}^k \left(\sum_{i=1}^{z_{4j}} \mathbf{x}_{4,i}^{(j)} \right) = \mathbf{0} \Rightarrow$ Die Populationsgröße ist Null.

$\Rightarrow$ Es sind keine relevanten Realisierungen mehr vorhanden.

$\Rightarrow$ Der Prozess ist ausgestorben. △

3.3 Verteilung, Markov-Eigenschaft und Momente

Nun werden die den Prozess definierenden Zufallsvektoren eingehender untersucht.

3.3.1 Wahrscheinlichkeitserzeugende Funktionen für den Multitype GWP

Sowohl die Reproduktionszahlen der Objekte als auch die Populationsgrößen der Generationen sind also $\mathbb{N}_0^k$ -wertige $\mathbf{Z}$ -Vektoren. Deren Verteilung kann demnach durch multivariate weF eindeutig charakterisiert werden. Diese werden i.A. laut der Definition 3.8 geschrieben, etwa $f_{\mathbf{Z}_n}$ für die der Populationsgröße $\mathbf{Z}_n$.

Insbesondere für die als nichtzufällig angenommene initiale Populationsgröße $\mathbf{Z}_0$ kann die weF *allgemein* bestimmt werden. Da $f_{\mathbf{Z}_n}(\mathbf{s}) \stackrel{(3.17)}{=} \mathbb{E} \mathbf{s}^{\mathbf{Z}_n}$ ist, gilt für alle $\mathbf{s} \in [0, 1]^k$

$$f_{\mathbf{Z}_0}(\mathbf{s}) = \mathbb{E} \mathbf{s}^{\mathbf{Z}_0} \stackrel{\substack{\text{nicht} \\ \text{zufällig}}}{=} \mathbf{s}^{\mathbf{Z}_0} \stackrel{(3.3)}{=} \prod_{j=1}^k s_j^{Z_{0j}} = s_1^{Z_{01}} \cdots s_k^{Z_{0k}}. \quad (3.31)$$

Für $f_{\mathbf{X}^{(j)}}$ bzw. $f_{\mathbf{Z}_n^{(j)}}$ soll jedoch die folgende knappere Notation verwendet werden.

Definition 3.16 (WeF der Reproduktionszahlen der Objekte)

Definiere für jedes $1 \leq j \leq k$ die weF für die Reproduktionszahlen $\mathbf{X}^{(j)}$ eines Objekts vom Typ j als $f^{(j)} : \mathbb{N}_0^k \rightarrow [0, 1]$ (also für alle $\mathbf{s} = (s_1, \dots, s_k)^T \in \mathbb{N}_0^k$), d.h.

$$f^{(j)}(\mathbf{s}) := f^{(j)}(s_1, \dots, s_k) := f_{\mathbf{X}^{(j)}}(\mathbf{s}) \stackrel{(3.17)}{\stackrel{(3.24)}{=}} \sum_{\mathbf{r} \in \mathbb{N}_0^k} p^{(j)}(\mathbf{r}) \mathbf{s}^{\mathbf{r}} = \mathbb{E} \mathbf{s}^{\mathbf{X}^{(j)}}. \quad (3.32)$$

Klarerweise gilt somit für jedes $1 \leq j \leq k$ bzw. für alle $n \geq 0$, $i \geq 1$ und wegen (3.23)

$$f_{\mathbf{X}^{(j)}} = f_{\mathbf{X}_{n,i}^{(j)}} = f^{(j)}. \quad (3.33)$$

Definition 3.17 (WeF der Populationsgrößen $\mathbf{Z}_n^{(j)}$)

Betrachte den von einem einzelnen Typ- j -Objekt gestarteten Prozess. Für jedes $n \geq 0$ und $1 \leq j \leq k$ sei $f_n^{(j)} : \mathbb{N}_0^k \rightarrow [0, 1]$ die weF der Populationsgröße $\mathbf{Z}_n^{(j)}$, also

$$f_n^{(j)}(\mathbf{s}) := f_n^{(j)}(s_1, \dots, s_k) := f_{\mathbf{Z}_n^{(j)}}(\mathbf{s}) \stackrel{(3.17)}{=} \sum_{\mathbf{r} \in \mathbb{N}_0^k} \mathbb{P}(\mathbf{Z}_n^{(j)} = \mathbf{r}) \mathbf{s}^{\mathbf{r}} = \mathbb{E} \mathbf{s}^{\mathbf{Z}_n^{(j)}}. \quad (3.34)$$

$$\text{Gemäß dem Satz 3.10 (c) gilt} \quad f_n^{(j)}(\mathbf{0}) = p_{\mathbf{Z}_n^{(j)}}(\mathbf{0}) \stackrel{(3.11)}{=} \mathbb{P}(\mathbf{Z}_n^{(j)} = \mathbf{0}). \quad (3.35)$$

In der ersten sowie der initialen Generation gilt für die wahrscheinlichkeitserzeugenden Funktionen der Definition 3.17 das folgende Hilfsresultat.

Lemma 3.18. Für $1 \leq j \leq k$, $\mathbf{r} \in \mathbb{N}_0^k$ und $\mathbf{s} \in [0, 1]^k$ gilt für die weF bzw. Verteilung von

$$\bullet \mathbf{Z}_1^{(j)} : \quad f_1^{(j)}(\mathbf{s}) = f^{(j)}(\mathbf{s}) \text{ bzw.} \quad \mathbb{P}(\mathbf{Z}_1^{(j)} = \mathbf{r}) = p^{(j)}(\mathbf{r}). \quad (3.36)$$

$$\bullet \mathbf{Z}_0^{(j)} : \quad f_0^{(j)}(\mathbf{s}) = s_j \text{ bzw.} \quad \mathbb{P}(\mathbf{Z}_0^{(j)} = \mathbf{r}) = \begin{cases} 1 & \text{für } \mathbf{r} = \mathbf{e}_j, \\ 0 & \text{für } \mathbf{r} \neq \mathbf{e}_j. \end{cases} \quad (3.37)$$

Beweis. Die Bemerkung zur Def. 3.14 (i) ($\mathbf{Z}_1^{(j)} \stackrel{d}{=} \mathbf{X}^{(j)}$ (3.27)) rechtfertigt die erste Aussage – die Verteilung steht in der Def. 3.12. Für die zweite beachte, dass $\mathbf{Z}_0^{(j)} = \mathbf{e}_j$ und deshalb auch $f_0^{(j)}(\mathbf{s}) \stackrel{(3.34)}{=} f_{\mathbf{Z}_0^{(j)}}(\mathbf{s}) \stackrel{(3.26)}{=} \mathbf{s}^{\mathbf{e}_j} \stackrel{(3.31)}{=} \prod_{i=1}^k s_i^{e_{j_i}} \stackrel{(3.2)}{=} s_j$ gilt. $\square$

Als Nächstes soll eine geeignete Vektornotation eingeführt werden.

Notation 3.19 (Vektorwertige weF – angelehnt an [3, S. 183])

Schreibe alle W-Funktionen und weF der $\mathbf{X}^{(j)}$ sowie alle weF der $\mathbf{Z}_n^{(j)}$ folgendermaßen zusammen als die vektorwertigen Funktionen $\mathbf{p} : \mathbb{N}_0^k \rightarrow [0, 1]^k$ und $\mathbf{f}, \mathbf{f}_n : [0, 1]^k \rightarrow [0, 1]^k$:

$$\mathbf{p}(\mathbf{r}) := (p^{(1)}(\mathbf{r}), \dots, p^{(k)}(\mathbf{r}))^T \text{ und} \quad \mathbf{f}(\mathbf{s}) := (f^{(1)}(\mathbf{s}), \dots, f^{(k)}(\mathbf{s}))^T \text{ bzw.} \quad (3.38)$$

$$\mathbf{f}_n(\mathbf{s}) := (f_n^{(1)}(\mathbf{s}), \dots, f_n^{(k)}(\mathbf{s}))^T \quad \text{für alle } n \geq 0. \quad (3.39)$$

Damit können die weF aller $\mathbf{X}^{(j)}$ gemeinsam kompakt formuliert werden als

$$\begin{aligned} \mathbf{f}(\mathbf{s}) &\stackrel{(3.32)}{\stackrel{(3.38)}{=}} \left(\sum_{\mathbf{r} \in \mathbb{N}_0^k} p^{(1)}(\mathbf{r}) \mathbf{s}^{\mathbf{r}}, \dots, \sum_{\mathbf{r} \in \mathbb{N}_0^k} p^{(k)}(\mathbf{r}) \mathbf{s}^{\mathbf{r}} \right) \stackrel{T}{=} \sum_{\mathbf{r} \in \mathbb{N}_0^k} \underbrace{(p^{(1)}(\mathbf{r}), \dots, p^{(k)}(\mathbf{r}))^T}_{=\mathbf{p}(\mathbf{r}) \text{ mit (3.38)}} \mathbf{s}^{\mathbf{r}}, \text{ kurz} \\ \mathbf{f}(\mathbf{s}) &= \sum_{\mathbf{r} \in \mathbb{N}_0^k} \mathbf{p}(\mathbf{r}) \mathbf{s}^{\mathbf{r}} \quad \text{für alle } \mathbf{s} \in [0, 1]^k. \end{aligned}$$

Insbesondere lässt sich mit dem Lemma 3.18 für alle $\mathbf{s} \in [0, 1]^k$ schreiben, dass

$$\mathbf{f}_1(\mathbf{s}) \stackrel{(3.39)}{=} (f_1^{(1)}(\mathbf{s}), \dots, f_1^{(k)}(\mathbf{s}))^T \stackrel{(3.36)}{=} (f^{(1)}(\mathbf{s}), \dots, f^{(k)}(\mathbf{s}))^T \stackrel{(3.38)}{=} \mathbf{f}(\mathbf{s}) \quad \text{und} \quad (3.40)$$

$$\mathbf{f}_0(\mathbf{s}) \stackrel{(3.39)}{=} (f_0^{(1)}(\mathbf{s}), \dots, f_0^{(k)}(\mathbf{s}))^T \stackrel{(3.37)}{=} (s_1, \dots, s_k)^T = \mathbf{s} \quad \text{ist.} \quad (3.41)$$

$\triangle$

Mit der gewählten Schreibweise lässt sich wieder ein Zusammenhang zwischen den weF der Populationsgrößen unterschiedlicher Generationen herstellen, was im gleich folgenden Satz 3.20 geschehen soll. Betrachte für die dahingehend nötige Vorarbeit die Rekursion (3.25). Gemäß ihr ist $\mathbf{Z}_{n+1}$ gegeben $\mathbf{Z}_n = \mathbf{r} \in \mathbb{N}_0^k$ für jedes $n \geq 0$ eine (Doppel-)Summe von unabhängigen Zufallsvektoren, sprich

$$(\mathbf{Z}_{n+1} | \mathbf{Z}_n = \mathbf{r}) = \sum_{j=1}^k \sum_{i=1}^{r_j} \mathbf{X}_{n,i}^{(j)}. \quad (\star)$$

3 Galton-Watson-Verzweigungsprozesse mit mehreren unterschiedlichen Typen

Dabei haben gemäß (3.33) alle $\{\mathbf{X}_{n,i}^{(j)}\}_{i=1}^{r_j}$ jeweils die weF $f^{(j)}$. Deshalb kann der Satz 3.10 (g) angewendet werden, um die zugehörige bedingte weF zu erhalten. Es gilt

$$\begin{aligned} \mathbb{E}[\mathbf{s}^{\mathbf{Z}_{n+1}} | \mathbf{Z}_n = \mathbf{r}] &\stackrel{(3.19)}{=} f_{\mathbf{Z}_{n+1} | \mathbf{Z}_n = \mathbf{r}}(\mathbf{s}) \stackrel{(3.22)}{\underset{(\star)}{=}} \prod_{j=1}^k \prod_{i=1}^{r_j} f^{(j)}(\mathbf{s}) = \prod_{j=1}^k (f^{(j)}(\mathbf{s}))^{r_j} \\ &\stackrel{(3.3)}{=} (\mathbf{f}(\mathbf{s}))^{\mathbf{r}} \stackrel{(3.36)}{=} (\mathbf{f}_1(\mathbf{s}))^{\mathbf{r}} \quad \text{für alle } \mathbf{s} \in [0, 1]^k. \end{aligned} \quad (3.42)$$

Bemerkung. Da $f^{(j)}$ genauso die weF aller $\mathbf{Z}_1^{(j)}$ ist (siehe erneut (3.36)), kann alternativ auch gleich (3.29) mit $n = 1$ anstatt $(\star)$ für (3.42) herangezogen werden.

Völlig analog folgt aus (3.29) generell für $m, n \geq 0$, dass $\mathbf{Z}_{n+m}$ gegeben $\mathbf{Z}_m = \mathbf{r} \in \mathbb{N}_0^k$ eine Summe von $\mathbf{Z}$ -Vektoren ist, welche laut (3.34) und für $1 \leq j \leq k$ jeweils die weF $f_n^{(j)}$ besitzen. Damit erhält man (für alle $\mathbf{s} \in [0, 1]^k$)

$$\mathbb{E}[\mathbf{s}^{\mathbf{Z}_{n+m}} | \mathbf{Z}_m = \mathbf{r}] = f_{\mathbf{Z}_{n+m} | \mathbf{Z}_m = \mathbf{r}}(\mathbf{s}) \stackrel{\text{analog zu (3.42)}}{=} \prod_{j=1}^k (f_n^{(j)}(\mathbf{s}))^{r_j} \stackrel{(3.3)}{\underset{(3.39)}{=}} (\mathbf{f}_n(\mathbf{s}))^{\mathbf{r}}. \quad (3.43)$$

Bemerkung. Für $m = 1$ würde (3.43) ebenso aus der Rückwärtsgleichung (3.30) folgen.

Damit lässt sich der angekündigte nächste Satz zeigen.

Satz 3.20 (Iterierte weF – Multitype GWP vgl. z.B. [19, Theorem II.3.1, S. 36])
Für alle $n, m \geq 0$ und $\mathbf{s} \in [0, 1]^k$ gilt

$$\begin{aligned} f_{\mathbf{Z}_{n+m}}(\mathbf{s}) &= f_{\mathbf{Z}_m}(\mathbf{f}_n(\mathbf{s})) = f_{\mathbf{Z}_n}(\mathbf{f}_m(\mathbf{s})) \quad \text{und insbesondere} \\ f_{\mathbf{Z}_{n+1}}(\mathbf{s}) &= f_{\mathbf{Z}_1}(\mathbf{f}_n(\mathbf{s})) = f_{\mathbf{Z}_n}(\mathbf{f}(\mathbf{s})). \end{aligned} \quad (3.44)$$

Beweis. Es soll wieder mit dem totalen EW operiert werden. Ihm zufolge gilt

$$f_{\mathbf{Z}_{n+m}}(\mathbf{s}) \stackrel{(3.17)}{=} \mathbb{E}[\mathbf{s}^{\mathbf{Z}_{n+m}}] \stackrel{(A.74)}{=} \mathbb{E}[\mathbb{E}[\mathbf{s}^{\mathbf{Z}_{n+m}} | \mathbf{Z}_m]] \stackrel{(3.43)}{=} \mathbb{E}[(\mathbf{f}_n(\mathbf{s}))^{\mathbf{Z}_m}] \stackrel{(3.17)}{=} f_{\mathbf{Z}_m}(\mathbf{f}_n(\mathbf{s})),$$

was die erste Zeile von (3.44) zeigt (da $n+m$ kommutativ ist). Mit $m=1$ und $\mathbf{f}_1 \stackrel{(3.40)}{=} \mathbf{f}$ ergibt sich daraus auch die zweite. $\square$

Bemerkungen zum Satz 3.20.

- (i) Die zweite Zeile von (3.44) erhielte man alternativ genauso direkt mithilfe des totalen Erwartungswertes, und zwar den letzten Ausdruck über (3.42). Für den mittleren könnte das aus der Rückwärtsgleichung (3.30) gewonnene Pendant dazu, d.h. $\mathbb{E}[\mathbf{s}^{\mathbf{Z}_{n+1}} | \mathbf{Z}_1 = \mathbf{r}] = (\mathbf{f}_n(\mathbf{s}))^{\mathbf{r}}$, oder gleich (3.43) mit $m=1$ verwendet werden.
- (ii) Mit $m=0$ kann mit der ersten Zeile von (3.44) der Zusammenhang zwischen den weF der $\mathbf{Z}_n$ und jenen der $\mathbf{Z}_n^{(j)}$ gezeigt werden. Für $n \geq 0$ und $\mathbf{s} \in [0, 1]^k$ gilt

$$\begin{aligned} f_{\mathbf{Z}_n}(\mathbf{s}) &\stackrel{(3.44)}{=} f_{\mathbf{Z}_0}(\mathbf{f}_n(\mathbf{s})) \stackrel{(3.31)}{=} (\mathbf{f}_n(\mathbf{s}))^{\mathbf{Z}_0} \stackrel{(3.3)}{\underset{(3.39)}{=}} \prod_{j=1}^k (f_n^{(j)}(\mathbf{s}))^{Z_{0j}} \\ &= (f_n^{(1)}(\mathbf{s}))^{Z_{01}} \cdots (f_n^{(k)}(\mathbf{s}))^{Z_{0k}}. \end{aligned} \quad (3.45)$$

- (iii) Klarerweise gilt alles, was für $\{\mathbf{Z}_n\}_{n \geq 0}$ gilt (und somit der Satz ebenso), insbesondere auch für den Prozess $\{\mathbf{Z}_n^{(j)}\}_{n \geq 0}$. $\triangle$

Der letzte Punkt der vorherigen Bemerkungen führt unmittelbar zum folgenden Resultat, welches der Definition 2.18 im einfachen Fall bzw. dem Korollar dazu entspricht.

Korollar zum Satz 3.20 (WeF-Iterationen für die weF der $\mathbf{Z}_n^{(j)}$ – vgl. z.B. [19, II.(3.1) und II.(3.2), S. 36])

Sei $\mathbf{s} \in [0, 1]^k$ und seien $n, m \geq 0$ beliebig.

- (i) Für alle $1 \leq j \leq k$ gilt

$$\begin{aligned} f_0^{(j)}(\mathbf{s}) &\stackrel{(3.37)}{=} s_j, & f_1^{(j)}(\mathbf{s}) &\stackrel{(3.36)}{=} f^{(j)}(\mathbf{s}) && \text{und} \\ f_{n+1}^{(j)}(\mathbf{s}) &= f^{(j)}(\mathbf{f}_n(\mathbf{s})) = \mathbf{f}_n^{(j)}(\mathbf{f}(\mathbf{s})) && \text{bzw.} && (3.46) \end{aligned}$$

$$f_{n+m}^{(j)}(\mathbf{s}) = f_m^{(j)}(\mathbf{f}_n(\mathbf{s})) = \mathbf{f}_n^{(j)}(\mathbf{f}_m(\mathbf{s})). \quad (3.47)$$

- (ii) Gänzlich in Vektorform geschrieben lautet dies

$$\begin{aligned} \mathbf{f}_0(\mathbf{s}) &\stackrel{(3.41)}{=} \mathbf{s}, & \mathbf{f}_1(\mathbf{s}) &\stackrel{(3.40)}{=} \mathbf{f}(\mathbf{s}), & \mathbf{f}_{n+1}(\mathbf{s}) &= \mathbf{f}(\mathbf{f}_n(\mathbf{s})) = \mathbf{f}_n(\mathbf{f}(\mathbf{s})) && \text{bzw.} \\ \mathbf{f}_{n+m}(\mathbf{s}) &= \mathbf{f}_m(\mathbf{f}_n(\mathbf{s})) = \mathbf{f}_n(\mathbf{f}_m(\mathbf{s})). && && && (3.48) \end{aligned}$$

Beweis. Als direkte Konsequenz des Satzes folgt (siehe die Bem. zum Satz (iii)) etwa

$$f_{n+m}^{(j)}(\mathbf{s}) \stackrel{(3.34)}{=} f_{\mathbf{Z}_{n+m}^{(j)}}(\mathbf{s}) \stackrel{(3.44)}{=} f_{\mathbf{Z}_n^{(j)}}(\mathbf{f}_m(\mathbf{s})) \stackrel{(3.34)}{=} f_n^{(j)}(\mathbf{f}_m(\mathbf{s})), \quad (\star)$$

woraus sich – gleich argumentiert wie im Beweis des Satzes – (3.47) und auch (3.46) ergibt (letzte Beziehung könnte analog zu $(\star)$ ebenfalls direkt aus (3.44) abgeleitet werden).

Der Punkt (ii) entspricht genau dem Punkt (i), wobei nun alles kompakt – in vektorwertigen Funktionen gemäß der Notation 3.19 – für alle $1 \leq j \leq k$ gemeinsam geschrieben wird. Dies sieht man mit (3.39) etwa durch

$$\mathbf{f}_{n+m}(\mathbf{s}) = (f_{n+m}^{(1)}(\mathbf{s}), \dots, f_{n+m}^{(k)}(\mathbf{s}))^T \stackrel{(3.47)}{=} (f_n^{(1)}(\mathbf{f}_m(\mathbf{s})), \dots, f_n^{(k)}(\mathbf{f}_m(\mathbf{s})))^T = \mathbf{f}_n(\mathbf{f}_m(\mathbf{s})). \quad \square$$

3.3.2 Momente der Reproduktionszahlen und Populationsgrößen

Im nächsten Schritt wird eine geeignete Notation für die Erwartungswerte sowie die Kovarianzen der Reproduktionszahlen und Populationsgrößen festgelegt. Jeweils anschließend sollen die der Populationsgrößen in Abhängigkeit der Momente der Reproduktionszahlen ausgedrückt werden.

Erwartungswert(-vektoren und -matrizen)

Es ist praktikabel, die EW aller Objekttypen gemeinsam in eine Matrix zu schreiben.

Definition 3.21 (Erwartungswert(-matrix) der Reproduktionszahlen)

Definiere die $k \times k$ -Matrix $\mathbf{M}$, für jedes $1 \leq j \leq k$ den k -dimensionalen Zeilenvektor $\boldsymbol{\mu}_j$ und für jedes Tupel $1 \leq j, l \leq k$ die Zahl μ_{jl} durch

$$\mathbf{M} := (\boldsymbol{\mu}_j)_{j=1}^k := (\mu_{jl})_{j,l=1}^k := (\mathbb{E}X_l^{(j)})_{j,l=1}^k \stackrel{(3.7)}{=} (\mathbb{E}[\mathbf{X}^{(j)}]^T)_{j=1}^k, \quad \text{wobei} \quad (3.49)$$

$$\boldsymbol{\mu}_j = ((\mu_{jl})_{l=1}^k)^T = (\mu_{j1}, \dots, \mu_{jk}) \stackrel{(3.49)}{=} \mathbb{E}[(X_1^{(j)}, \dots, X_k^{(j)})] \stackrel{(3.7)}{=} \mathbb{E}[\mathbf{X}^{(j)}]^T \quad \text{und} \quad (3.50)$$

$$\mu_{jl} \stackrel{(3.49)}{=} \mathbb{E}X_l^{(j)} \stackrel{(3.51)}{\text{ist.}} \quad (3.51)$$

D.h., μ_{jl} ist die *erwartete Anzahl produzierter Typ- l -Objekte eines Objekts vom Typ j* . Somit stehen in jeder j -ten Zeile von $\mathbf{M}$ die erwarteten Reproduktionszahlen (vom Typ $1 \leq l \leq k$) eines Objektes vom Typ j . Eine solche entspricht genau dem Vektor $\boldsymbol{\mu}_j$, d.h. dem transponierten Erwartungswertvektor von $\mathbf{X}^{(j)}$.

In jeder l -ten Spalte von $\mathbf{M}$ hingegen finden sich die Erwartungswerte der Anzahlen an (von den Objekttypen $1 \leq j \leq k$) produzierten Objekten vom Typ l .

Annahme 3.22 (angelehnt an [19, *Basic assumption*, S. 37]). Seien für den Rest dieser Arbeit sämtliche Einträge von $\mathbf{M}$ endlich und nicht alle 0, d.h.

$$0 \leq \mu_{jl} < \infty \quad \text{für alle } 1 \leq j, l \leq k \quad (\Rightarrow \mathbf{M} < \infty) \quad \text{und} \quad \mathbf{M} \neq \mathbf{0}. \quad \triangle$$

Stelle für alle $1 \leq j \leq k$ sowie $n \geq 0$ und $i \geq 1$ fest, dass wegen (3.27) $(\mathbf{X}^{(j)} \stackrel{d}{=} \mathbf{Z}_1^{(j)})$ $\mathbb{E}\mathbf{Z}_1^{(j)} \stackrel{(3.26)}{=} \mathbb{E}[\mathbf{Z}_1 | \mathbf{Z}_0 = \mathbf{e}_j] \stackrel{(3.27)}{=} \mathbb{E}\mathbf{X}^{(j)} \stackrel{(3.23)}{=} \mathbb{E}\mathbf{X}_{n,i}^{(j)} \stackrel{(3.50)}{=} \boldsymbol{\mu}_j^T$ gilt. Weiters ergibt sich

$$\mu_{jl} \stackrel{(3.51)}{=} \mathbb{E}X_l^{(j)} \stackrel{(3.20)}{=} \frac{\partial f_{\mathbf{X}^{(j)}}(\mathbf{1})}{\partial s_l} \stackrel{(3.32)}{=} \frac{\partial f^{(j)}}{\partial s_l}(\mathbf{1}) \quad \text{für alle } 1 \leq l \leq k \quad (3.52)$$

mit dem Satz 3.10 (e). Direkt mit dem selben Satz oder aus (3.52) erhält man

$$\overset{\text{Ann.}}{\underset{3.22}{>}} \mathbf{M} \stackrel{(3.49)}{=} (\mu_{jl})_{j,l=1}^k = (\mathbb{E}X_l^{(j)})_{j,l=1}^k \stackrel{(3.52)}{=} \left(\frac{\partial f^{(j)}}{\partial s_l}(\mathbf{1}) \right)_{j,l=1}^k \stackrel{(3.18)}{=} \underset{(3.38)}{D} \mathbf{f}(\mathbf{1}). \quad (3.53)$$

Nun kann das gleich folgende Lemma bzgl. des weF-Vektors $\mathbf{f}_n \stackrel{(3.39)}{=} (f_n^{(1)}, \dots, f_n^{(k)})^T$ und den ersten Ableitungen der $f_{\mathbf{Z}_n}$ formuliert werden. Anschließend werden – auf eine ähnliche Weise wie in der Def. 3.21 – die EW der Populationsgrößen $\mathbf{Z}_n$ und $\mathbf{Z}_n^{(j)}$ definiert.

Lemma 3.23. Sei $n \geq 0$. Es gilt

(i) $\mathbf{f}_n(\mathbf{1}) = \mathbf{1}$ und deshalb insbesondere auch $\mathbf{f}(\mathbf{1}) \stackrel{(3.40)}{=} \mathbf{f}_1(\mathbf{1}) = \mathbf{1}$.

(ii) mindestens für $\mathbf{s} \in [0, 1]^k$ allgemein (also auch ohne die Annahme 3.22)

$$\mathbf{Df}_{\mathbf{Z}_{n+1}}(\mathbf{s}) = \mathbf{Df}_{\mathbf{Z}_n}(\mathbf{f}(\mathbf{s}))\mathbf{Df}(\mathbf{s}). \quad (\star)$$

Unter der Ann. 3.22, welche $\mathbf{Df}(\mathbf{1}) < \infty$ garantiert (siehe (3.53)), gilt zudem

$$\mathbf{Df}_{\mathbf{Z}_n}(\mathbf{1}) < \infty \quad \text{und weiter} \quad \mathbf{Df}_{\mathbf{Z}_{n+1}}(\mathbf{1}) = \mathbf{Df}_{\mathbf{Z}_n}(\mathbf{1})\mathbf{Df}(\mathbf{1}). \quad (3.54)$$

Beweis. Der erste Punkt gilt aufgrund des Satzes 3.10 (c) und (3.39).

Für (ii) betrachte das Resultat des Satzes 3.20, genauer $\mathbf{f}_{\mathbf{Z}_{n+1}}(\mathbf{s}) = \mathbf{f}_{\mathbf{Z}_n}(\mathbf{f}(\mathbf{s}))$ aus (3.44). Damit erhält man mithilfe der mehrdimensionalen Kettenregel (siehe z.B. [23, Satz 165.2, S. 267]) die Beziehung $(\star)$ mindestens auf $[0, 1]^k$ (vgl. auch den Satz 3.10 (a)). Diese bleibt, sofern die beiden Ableitungen auf der rechten Seite existieren, auch an der Stelle $\mathbf{1}$ bestehen. $\mathbf{Df}(\mathbf{1}) < \infty$ gilt unter der Ann. 3.22 ohnehin. Vorausgesetzt $\mathbf{Df}_{\mathbf{Z}_n}(\mathbf{1}) < \infty$ ist erfüllt, folgt mit der Stetigkeit (Punkt (e) des Satzes 3.10) und (i) auch, dass $\mathbf{Df}_{\mathbf{Z}_n}(\mathbf{f}(\mathbf{1})) = \mathbf{Df}_{\mathbf{Z}_n}(\mathbf{1}) < \infty$ ist.

Nun ist $\mathbf{Df}_{\mathbf{Z}_0}(\mathbf{1}) \stackrel{(3.20)}{=} \mathbb{E}[\mathbf{Z}_0]^T = \mathbf{Z}_0^T < \infty$ ($\mathbf{Z}_0$ ist nichtzufällig – Def. 3.13). Daraus folgt also auch $\mathbf{Df}_{\mathbf{Z}_1}(\mathbf{1}) < \infty$ und für alle $n \geq 0$ induktiv der erste Teil von (3.54). Deshalb gilt – wie beschrieben – auch der zweite Teil, womit (ii) gezeigt ist. $\square$

Definition 3.24 (Erwartungswert(-vektoren) der Populationsgrößen)

Definiere für jede Generation $n \geq 0$ die *erwartete Populationsgröße vom Typ l* als

$$\mu_l^{(n)} := \mathbb{E}Z_{nl} \quad \text{für } 1 \leq l \leq k \text{ und} \quad (3.55)$$

$$\boldsymbol{\mu}^{(n)} := ((\mu_l^{(n)})_{l=1}^k)^T = (\mu_1^{(n)}, \dots, \mu_k^{(n)}) \stackrel{(3.55)}{=} \mathbb{E}[(Z_{n1}, \dots, Z_{nk})] \stackrel{(3.7)}{=} \mathbb{E}\mathbf{Z}_n^T \quad (3.56)$$

als den *transponierten EW-Vektor der Populationsgröße $\mathbf{Z}_n$* . Analog werden für jedes $1 \leq j \leq k$ die folgenden Bezeichnungen für die *Erwartungswerte der $Z_{nl}^{(j)}$ bzw. $(\mathbf{Z}_n^{(j)})^T$* , also wenn der Prozess von einem Objekt vom Typ j gestartet wird, gewählt. Es seien

$$\mu_{jl}^{(n)} := \mathbb{E}Z_{nl}^{(j)} \quad \text{für } 1 \leq l \leq k. \text{ Weiters sei} \quad (3.57)$$

$$\boldsymbol{\mu}_j^{(n)} := ((\mu_{jl}^{(n)})_{l=1}^k)^T = (\mu_{j1}^{(n)}, \dots, \mu_{jk}^{(n)}) \stackrel{(3.57)}{=} \mathbb{E}[(Z_{n1}^{(j)}, \dots, Z_{nk}^{(j)})] \stackrel{(3.7)}{=} \mathbb{E}[\mathbf{Z}_n^{(j)}]^T. \quad (3.58)$$

Direkt mit dem Satz 3.10 (e) ergibt sich (wie in (3.53) für $\mathbf{Df}(\mathbf{1})$) ohne weiteren Aufwand das folgende Resultat.

Korollar zur Definition 3.24. Für jedes $n \geq 0$ gilt

$$\boldsymbol{\mu}^{(n)} \stackrel{(3.56)}{=} \mathbb{E}\mathbf{Z}_n^T \stackrel{(3.20)}{=} \left(\left(\frac{\partial \mathbf{f}_{\mathbf{Z}_n}}{\partial s_l}(\mathbf{1}) \right)_{l=1}^k \right)^T = \mathbf{Df}_{\mathbf{Z}_n}(\mathbf{1}) \quad \text{und insbesondere} \quad (3.59)$$

$$\boldsymbol{\mu}_j^{(n)} \stackrel{(3.58)}{=} \mathbb{E}[\mathbf{Z}_n^{(j)}]^T \stackrel{(3.20)}{\stackrel{(3.34)}}{=} \left(\left(\frac{\partial \mathbf{f}_n^{(j)}}{\partial s_l}(\mathbf{1}) \right)_{l=1}^k \right)^T \stackrel{(3.18)}{=} \mathbf{Df}_n^{(j)}(\mathbf{1}) \quad \text{für alle } 1 \leq j \leq k. \quad (3.60)$$

3 Galton-Watson-Verzweigungsprozesse mit mehreren unterschiedlichen Typen

Werden die Zeilenvektoren $\mu_j^{(n)}$ mit $1 \leq j \leq k$ in eine Matrix geschrieben, so erhält man

$$\underbrace{(\mu_j^{(n)})_{j=1}^k}_{\stackrel{(3.58)}{=} \mathbb{E}[\mathbf{Z}_n^{(j)}]^T} \stackrel{(3.58)}{=} \underbrace{(\mu_{jl}^{(n)})_{j,l=1}^k}_{\stackrel{(3.57)}{=} \mathbb{E}Z_{nl}^{(j)}} \stackrel{(3.60)}{=} (\mathbf{D}f_n^{(j)}(\mathbf{1}))_{j=1}^k \stackrel{(3.60)}{=} \left(\frac{\partial f_n^{(j)}}{\partial s_l}(\mathbf{1}) \right)_{j,l=1}^k \stackrel{(3.18)}{=} \stackrel{(3.39)}{=} \mathbf{D}f_n(\mathbf{1}). \quad (3.61)$$

Bemerkung. Da $\mathbf{Z}_0$ (bzw. $\mathbf{Z}_0^{(j)}$) nichtzufällig ist, gilt (für alle $1 \leq j \leq k$)

$$\begin{aligned} \mathbf{D}f_{\mathbf{Z}_0}(\mathbf{1}) &\stackrel{(3.59)}{=} \mu^{(0)} = \mathbb{E}\mathbf{Z}_0^T = \mathbf{Z}_0^T \quad \text{und} \quad \mathbf{D}f_0^{(j)}(\mathbf{1}) \stackrel{(3.60)}{=} \mu_j^{(0)} = (\mathbf{Z}_0^{(j)})^T \stackrel{(3.26)}{=} \mathbf{e}_j^T \\ &\Rightarrow \mathbf{D}f_0(\mathbf{1}) = (\mathbf{D}f_0^{(j)}(\mathbf{1}))_{j=1}^k = (\mathbf{e}_j^T)_{j=1}^k = \mathbf{I}_k. \end{aligned} \quad (3.62) \quad \triangle$$

Als Nächstes können die folgenden Zusammenhänge zwischen den Erwartungswerten der Populationsgrößen unterschiedlicher Generationen und der Matrix $\mathbf{M}$ hergestellt werden.

Satz 3.25 (EW-Vektoren der Populationsgrößen vgl. z.B. [39, 1.6, S. 13])

(i) Für alle $m, n \geq 0$ gilt

$$\mu^{(n+1)} = \mu^{(n)} \mathbf{M} \quad (\text{bzw. } \mathbb{E}\mathbf{Z}_{n+1}^T = \mathbb{E}\mathbf{Z}_n^T \mathbf{M}), \quad \text{weilers allg.} \quad (3.63)$$

$$\mu^{(n+m)} = \mu^{(m)} \mathbf{M}^n \quad (\text{bzw. } \mathbb{E}\mathbf{Z}_{n+m}^T = \mathbb{E}\mathbf{Z}_m^T \mathbf{M}^n) \quad \text{und insbes.} \quad (3.64)$$

$$\mu^{(n)} = \mu^{(0)} \mathbf{M}^n \stackrel{(3.62)}{=} \mathbf{Z}_0^T \mathbf{M}^n \quad (\text{bzw. } \mathbb{E}\mathbf{Z}_n^T = \mathbf{Z}_0^T \mathbf{M}^n). \quad (3.65)$$

(ii) Alle Beziehungen aus (i) gelten klarerweise auch für die EW der $\mathbf{Z}_n^{(j)}$ bzw. für $\mu_j^{(n)}$ (vgl. erneut die Bem. zum Satz 3.20 (iii)). D.h., es gilt (wieder für alle $m, n \geq 0$)

$$\mu_j^{(n+1)} = \mu_j^{(n)} \mathbf{M}, \quad \mu_j^{(n+m)} = \mu_j^{(m)} \mathbf{M}^n, \quad \mu_j^{(n)} = \mu_j^{(0)} \mathbf{M}^n \stackrel{(3.62)}{=} \mathbf{e}_j^T \mathbf{M}^n. \quad (3.66)$$

(iii) Dabei ist jeweils der (j, l) -te Eintrag von $\mathbf{M}^n$ der EW von $\mathbf{Z}_{nl}^{(j)}$, sprich die erwartete Anzahl an Objekten vom Typ l in der Generation n , wenn initial nur ein Objekt vom Typ j vorhanden war. D.h.

$$\mathbf{M}^n = (\mu_{jl}^{(n)})_{j,l=1}^k = (\mathbb{E}Z_{nl}^{(j)})_{j,l=1}^k \stackrel{(3.61)}{=} \mathbf{D}f_n(\mathbf{1}) \quad \text{für alle } n \geq 0. \quad (3.67)$$

Beweis. Aus (3.53) ($\mathbf{D}f(\mathbf{1}) = \mathbf{M}$), dem Korollar zur Def. 3.24 ((3.59) – $\mathbf{D}f_{\mathbf{Z}_n}(\mathbf{1}) = \mu^{(n)}$) und dem Lemma 3.23 (ii) ((3.54) – $\mathbf{D}f_{\mathbf{Z}_{n+1}}(\mathbf{1}) = \mathbf{D}f_{\mathbf{Z}_n}(\mathbf{1}) \mathbf{D}f(\mathbf{1})$) folgt (3.63) direkt. Durch iterative Anwendung kann daraus geschlossen werden, dass

$$\mu^{(m+n)} \stackrel{(3.63)}{=} \mu^{(m+n-1)} \mathbf{M} = \mu^{(m+n-2)} \mathbf{M}^2 = \dots \stackrel{\text{analog}}{=} \mu^{(m+n-n)} \mathbf{M}^n = \mu^{(m)} \mathbf{M}^n,$$

also (3.64), gilt. Mit $m = 0$ erhält man (3.65), womit (i) gezeigt ist (verwende (3.56) für die jeweils gleichwertigen Aussagen auf der rechten Seite).

Die Begründung für (ii) steht bereits in der Formulierung des Satzes. Für (iii) betrachte

$$\mu_j^{(n)} \stackrel{(3.66)}{=} \mathbf{e}_j^T \mathbf{M}^n = \text{' } j\text{-te Zeile von } \mathbf{M}^n \text{ ' } \Rightarrow \mathbf{M}^n = (\mu_j^{(n)})_{j=1}^k \stackrel{(3.58)}{=} (\mu_{jl}^{(n)})_{j,l=1}^k,$$

da Ersteres für jede Zeile von $\mathbf{M}^n$, also für alle $1 \leq j \leq k$, gilt. $\square$

Bemerkungen zum Satz 3.25.

- (i) Alternativ zum Beweis kann man $Df_{Z_{n+m}}(\mathbf{1}) = Df_{Z_n}(\mathbf{1})Df_m(\mathbf{1})$ aus (3.44) mit der Kettenregel erhalten. Falls (3.67) (also $Df_n(\mathbf{1}) = M^n$) bereits bekannt ist, ergibt sich daraus – mit (3.59) – unmittelbar $\mu^{(n+m)} = \mu^{(n)}M^m$, also (3.64) (und mit der analogen Vorgehensweise bekommt man das Pendant für $\mu_j^{(n+m)}$ aus (3.47)).
- (ii) Analog kann die (Matrizen-)Gleichung $Df_{n+m}(\mathbf{1}) = Df_n(\mathbf{1})Df_m(\mathbf{1})$ direkt aus (3.48) abgeleitet werden. Dementsprechend gilt etwa (mit $m = 1$)

$$\begin{aligned} Df_{n+1}(\mathbf{1}) &= Df_n(\mathbf{1})Df_1(\mathbf{1}) \stackrel{(3.53)}{=} Df_n(\mathbf{1})M, & \text{woraus wieder iterativ} \\ Df_n(\mathbf{1}) &= Df_{n-1}(\mathbf{1})M = Df_{n-2}(\mathbf{1})M^2 = \dots = Df_0(\mathbf{1})M^n \\ &\stackrel{(3.62)}{=} I_k M^n = M^n, & \text{also (3.67), folgt.} \end{aligned}$$

Die erste Gleichung kann dabei, unter Verwendung der zweiten, auch gelesen werden als $M^{n+1} = M^n M$, was trivialerweise erfüllt sein muss.

- (iii) Konsistent mit dem Punkt (iii) des Satzes können die Elemente $\mu_{jl}^{(n)}$ von M^n auch anders interpretiert werden. Sie sind die erwartete Anzahl an Objekten vom Typ $1 \leq l \leq k$ einer (beliebigen) Generation, welche von einem n Schritte vorher existierenden Objekt des Typs $1 \leq j \leq k$ stammen. Beachte, dass sie sich auch wie folgt durch die Regeln der Matrixmultiplikation rekursiv beschreiben lassen ($n \geq 1$):

$$\mu_{jl}^{(0)} = \delta_{jl} \ (\Rightarrow M^0 = I_k), \quad \mu_{jl}^{(1)} = \mu_{jl} \quad \text{und} \quad \mu_{jl}^{(n)} = \sum_{i=1}^k \mu_{ji}^{(n-1)} \mu_{il}. \quad (3.68)$$

Explizit ausgedrückt ist also bspw. $\mu_{jl}^{(2)} = \sum_{i=1}^k \mu_{ji} \mu_{il}$. Im ersten Schritt werden μ_{ji} Objekte vom Typ i und im zweiten Schritt, von jedem dieser, μ_{il} vom Typ l erwartet. Werden diese jeweils miteinander multipliziert und über alle $1 \leq i \leq k$ aufsummiert, ergibt sich auch intuitiv die o.g. Erwartung, also die über die Anzahl an von einem Objekt des Typs j ausgehenden Typ- l -Objekten nach zwei Schritten.

Für allgemeine n ließe sich eine analoge Interpretation formulieren. Wieder explizit hingeschrieben sind die Elemente von M^n nämlich (etwa für $n > 3$ und $1 \leq j \leq k$)

$$\mu_{jl}^{(n)} = \sum_{i_1, \dots, i_{n-1}=1}^k \mu_{ji_1} \mu_{i_1 i_2} \cdots \mu_{i_{n-2} i_{n-1}} \mu_{i_{n-1} l}. \quad \triangle$$

Mit den Ergebnissen aus dem Satz 3.25 (i) können leicht die folgenden bedingten Erwartungen – mithilfe der nun bereits öfter verwendeten Methode – formuliert werden. Etwa aus (3.63) (also $\mathbb{E}Z_{n+1}^T = \mathbb{E}Z_n^T M$) folgt $\mathbb{E}[Z_{n+1}^T | Z_n = \mathbf{r}] = \mathbf{r}^T M$ für $\mathbf{r} \in \mathbb{N}_0^k$ und somit weiter, dass $\mathbb{E}[Z_{n+1}^T | Z_n] = Z_n^T M$ f.s. ist. Dies – und die analoge aus (3.64) ableitbare Entsprechung – werden noch im folgenden Korollar zusammengefasst.

Korollar 3.26 (Bedingte EW der Populationsgrößen vgl. z.B. [19, II.(4.1), S. 37])
Für alle $n, m \geq 0$ gilt

$$\mathbb{E}[\mathbf{Z}_{n+1}^T | \mathbf{Z}_n] = \mathbf{Z}_n^T \mathbf{M} \text{ f.s.} \quad \text{und} \quad \mathbb{E}[\mathbf{Z}_{n+m}^T | \mathbf{Z}_m] = \mathbf{Z}_m^T \mathbf{M}^n \text{ f.s.}$$

Da $\mathbf{Z}_0$ nichtzufällig ist, ist $\mathbb{E}[\mathbf{Z}_n^T | \mathbf{Z}_0] = \mathbb{E} \mathbf{Z}_n^T \stackrel{(3.65)}{=} \mathbf{Z}_0^T \mathbf{M}^n$.

Bemerkung 3.27. Die bisherige Vorgehensweise für die Bestimmung der Erwartungswerte gleicht der über die weF im 2. Kapitel, etwa unter (2.22) allgemein bzw. unter (2.55) im Kontext der Populationsgrößen des GWP.

Dementsprechend bietet sich auch für den Multitype GWP eine Alternative über den totalen EW (wie in den Sätzen 2.15 bzw. 2.33 im einfachen Fall) an. Dies wird bspw. bei [19, II.(4.1), S. 37] oder in [18, S. 24 f.] so praktiziert und soll kurz skizziert werden.

Mit den nun bekannten Argumenten erhält man aus der Rekursion (3.25)

$$(\mathbf{Z}_{n+1} | \mathbf{Z}_n = \mathbf{r}) = \sum_{j=1}^k \sum_{i=1}^{r_j} \mathbf{X}_{n,i}^{(j)} \quad \text{für alle } n \geq 0, \mathbf{r} \in \mathbb{N}_0^k \quad (3.69)$$

$$\begin{aligned} \Rightarrow \mathbb{E}[\mathbf{Z}_{n+1} | \mathbf{Z}_n = \mathbf{r}] &= \sum_{j=1}^k \sum_{i=1}^{r_j} \overbrace{\mathbb{E} \mathbf{X}_{n,i}^{(j)}}^{(3.23) \mathbb{E} \mathbf{X}^{(j)} (3.50) \boldsymbol{\mu}_j^T} = \sum_{j=1}^k \boldsymbol{\mu}_j^T r_j = \boldsymbol{\mu}_1^T r_1 + \dots + \boldsymbol{\mu}_k^T r_k \\ &= ((\boldsymbol{\mu}_j)_{j=1}^k)^T \mathbf{r} \stackrel{(3.49)}{=} \mathbf{M}^T \mathbf{r} \end{aligned} \quad (3.70)$$

$$\Rightarrow \mathbb{E}[\mathbf{Z}_{n+1} | \mathbf{Z}_n] = \mathbf{M}^T \mathbf{Z}_n \text{ f.s.} \quad \text{und auch} \quad \mathbb{E}[\mathbf{Z}_{n+1}^T | \mathbf{Z}_n] = \mathbf{Z}_n^T \mathbf{M} \text{ f.s.} \quad (3.71)$$

$$\Rightarrow \mathbb{E} \mathbf{Z}_{n+1}^T \stackrel{(A.74)}{=} \mathbb{E}[\mathbb{E}[\mathbf{Z}_{n+1}^T | \mathbf{Z}_n]] \stackrel{(3.71)}{=} \mathbb{E}[\mathbf{Z}_n^T \mathbf{M}] = \mathbb{E} \mathbf{Z}_n^T \mathbf{M} \quad \text{für alle } n \geq 0.$$

Daraus können – analog wie dort – ebenfalls die anderen Beziehungen des Satzes 3.25 bzw. des Korollars 3.26 erhalten werden. $\triangle$

Kovarianzmatrizen

Für die Kovarianzmatrizen der Populationsgrößen wird nun, abweichend zu den Erwartungswerten, die Methode über den totalen Erwartungswert gewählt. Die Herangehensweise wurde soeben in der Bem. 3.27 für die EW kurz angerissen und orientiert sich an [19, II.4, S. 36 f.]. Zunächst wird eine grundlegende Annahme über die zweiten Momente der Reproduktionszahlen getroffen und eine – die Notation vereinfachende – Definition formuliert. Anschließend werden Ausdrücke für die Kovarianzmatrizen ermittelt. Teilweise werden die in der erwähnten Bemerkung gefundenen Resultate auch hierzu verwendet.

Annahme 3.28 (angelehnt an [19, Ass. II.4.1, S. 37]). Es gelte stets

$$\mathbb{E}[\mathbf{X}^{(j)} (\mathbf{X}^{(j)})^T] < \infty. \quad \triangle$$

Definition 3.29 (angelehnt an [19, Def. II.4.2, S. 37]). Definiere für jedes $1 \leq j \leq k$ die Kovarianzmatrix der Reproduktionszahlen $\mathbf{V}_j$ und für jedes $n \geq 0$ die Matrix $\mathbf{C}_n$ als

$$\mathbf{V}_j := \mathbb{V} \mathbf{X}^{(j)} = \mathbb{E}[\mathbf{X}^{(j)}(\mathbf{X}^{(j)})^T] - \mathbb{E} \mathbf{X}^{(j)} \mathbb{E}[\mathbf{X}^{(j)}]^T \stackrel{(3.50)}{=} \mathbb{E}[\mathbf{X}^{(j)}(\mathbf{X}^{(j)})^T] - \boldsymbol{\mu}_j^T \boldsymbol{\mu}_j, \quad (3.72)$$

$$\mathbf{C}_n := \mathbb{E}[\mathbf{Z}_n \mathbf{Z}_n^T] \Rightarrow \mathbf{C}_0 = \mathbb{E}[\mathbf{Z}_0 \mathbf{Z}_0^T] \stackrel{\text{nicht zufällig}}{=} \mathbf{Z}_0 \mathbf{Z}_0^T. \quad (3.73)$$

Diese Matrizen können folgendermaßen zueinander in Beziehung gesetzt werden.

Lemma 3.30 (angelehnt an [19, II.(4.2), S. 37]). Für alle $n \geq 1$ gilt

$$\mathbf{C}_{n+1} = \mathbf{M}^T \mathbf{C}_n \mathbf{M} + \sum_{j=1}^k \mu_j^{(n)} \mathbf{V}_j. \quad (3.74)$$

Beweis. Gemäß der Def. 3.3 ist für jedes $n \geq 0$ die Kovarianzmatrix der Populationsgröße $\mathbf{Z}_n$ allgemein $\mathbb{V} \mathbf{Z}_n = \mathbb{E}[\mathbf{Z}_n \mathbf{Z}_n^T] - \mathbb{E} \mathbf{Z}_n \mathbb{E} \mathbf{Z}_n^T$. Da sie also auf EW beruht, gilt für die bedingte Kovarianzmatrix von $\mathbf{Z}_{n+1}$ gegeben $\mathbf{Z}_n$ (vgl. etwa [18, A.3.3, S. 284] bzgl. der bedingten Varianz einer ZV)

$$\begin{aligned} \mathbb{V}[\mathbf{Z}_{n+1} | \mathbf{Z}_n] &= \mathbb{E}[\mathbf{Z}_{n+1} \mathbf{Z}_{n+1}^T | \mathbf{Z}_n] - \underbrace{\mathbb{E}[\mathbf{Z}_{n+1} | \mathbf{Z}_n] \mathbb{E}[\mathbf{Z}_{n+1}^T | \mathbf{Z}_n]}_{=\mathbf{M}^T \mathbf{Z}_n \text{ f.s.}} \underbrace{\mathbb{E}[\mathbf{Z}_{n+1}^T | \mathbf{Z}_n]}_{=\mathbf{Z}_n^T \mathbf{M} \text{ f.s.}} \quad (3.71) \\ &\quad \text{sowie} \\ \mathbb{V}[\mathbf{Z}_{n+1} | \mathbf{Z}_n = \mathbf{r}] &= \mathbb{E}[\mathbf{Z}_{n+1} \mathbf{Z}_{n+1}^T | \mathbf{Z}_n = \mathbf{r}] - \underbrace{\mathbb{E}[\mathbf{Z}_{n+1} | \mathbf{Z}_n = \mathbf{r}] \mathbb{E}[\mathbf{Z}_{n+1}^T | \mathbf{Z}_n = \mathbf{r}]}_{=\mathbf{M}^T \mathbf{r} \quad =\mathbf{r}^T \mathbf{M} \quad (3.70)} \end{aligned} \quad (\star)$$

für die von $\mathbf{Z}_{n+1}$ gegeben $\mathbf{Z}_n = \mathbf{r}$ (für alle $\mathbf{r} \in \mathbb{N}_0^k$).

Andererseits kann letztere bedingte Kovarianzmatrix mithilfe von (3.69) auch wie folgt als eine Summe von Kovarianzmatrizen unabhängiger Z-Vektoren beschrieben werden. Beachte dabei, dass aufgrund der Unabhängigkeit der $\mathbf{X}_{n,i}^{(j)}$ (alle sind unabhängig und identisch verteilt wie $\mathbf{X}^{(j)}$ (3.23)) der Satz 3.4 über die Kovarianzmatrizen der Summe unkorrelierter Z-Vektoren angewendet werden kann. Es gilt

$$\mathbb{V}[\mathbf{Z}_{n+1} | \mathbf{Z}_n = \mathbf{r}] \stackrel{(3.69)}{=} \mathbb{V} \left[\sum_{j=1}^k \sum_{i=1}^{r_j} \mathbf{X}_{n,i}^{(j)} \right] \stackrel{\text{uabh.}}{\stackrel{(3.10)}{=}} \sum_{j=1}^k \sum_{i=1}^{r_j} \underbrace{\mathbb{V} \mathbf{X}_{n,i}^{(j)}}_{=\mathbb{V} \mathbf{X}^{(j)}} \stackrel{(3.72)}{=} \sum_{j=1}^k r_j \mathbf{V}_j. \quad (\star\star)$$

Daraus – und mit der obigen Kovarianzmatrixformel $(\star)$ – ergibt sich

$$\begin{aligned} \mathbb{E}[\mathbf{Z}_{n+1} \mathbf{Z}_{n+1}^T | \mathbf{Z}_n = \mathbf{r}] &\stackrel{(\star)}{=} \mathbf{M}^T \mathbf{r} \mathbf{r}^T \mathbf{M} + \mathbb{V}[\mathbf{Z}_{n+1} | \mathbf{Z}_n = \mathbf{r}] \stackrel{(\star\star)}{=} \mathbf{M}^T \mathbf{r} \mathbf{r}^T \mathbf{M} + \sum_{j=1}^k r_j \mathbf{V}_j \\ &\Rightarrow \mathbb{E}[\mathbf{Z}_{n+1} \mathbf{Z}_{n+1}^T | \mathbf{Z}_n] = \mathbf{M}^T \mathbf{Z}_n \mathbf{Z}_n^T \mathbf{M} + \sum_{j=1}^k \mathbb{E} Z_{n,j} \mathbf{V}_j \text{ f.s.} \end{aligned} \quad (\star\star\star)$$

Mit dem totalen Erwartungswert (Satz A.61 (b)) erhält man abschließend

$$\begin{aligned} \mathbf{C}_{n+1} &\stackrel{(3.73)}{=} \mathbb{E}[\mathbf{Z}_{n+1} \mathbf{Z}_{n+1}^T] \stackrel{(A.74)}{=} \mathbb{E}[\mathbb{E}[\mathbf{Z}_{n+1} \mathbf{Z}_{n+1}^T | \mathbf{Z}_n]] \\ &\stackrel{(\star\star\star)}{=} \mathbf{M}^T \underbrace{\mathbb{E}[\mathbf{Z}_n \mathbf{Z}_n^T]}_{=\mathbf{C}_n} \mathbf{M} + \sum_{j=1}^k \underbrace{\mathbb{E} Z_{n,j}}_{=\mu_j^{(n)} \quad (3.55)} \mathbf{V}_j = \mathbf{M}^T \mathbf{C}_n \mathbf{M} + \sum_{j=1}^k \mu_j^{(n)} \mathbf{V}_j. \end{aligned} \quad \square$$

3 Galton-Watson-Verzweigungsprozesse mit mehreren unterschiedlichen Typen

Schließlich lässt sich der folgende Satz über die Kovarianzmatrizen der Populationsgrößen formulieren. Der Übersichtlichkeit halber werden die EW erneut dazu geschrieben.

Satz 3.31 (EW und Kovarianzmatrizen der $\mathbf{Z}_n$ – angelehnt an [19, II.(4.3), S. 37])
Für alle $n \geq 0$ gilt

$$\mathbb{E}\mathbf{Z}_n \stackrel{(3.65)}{=} (\mathbf{M}^n)^T \mathbf{Z}_0 \quad \text{und} \quad \mathbb{V}\mathbf{Z}_n = \sum_{i=1}^n (\mathbf{M}^{n-i})^T \left(\sum_{j=1}^k \mu_j^{(i-1)} \mathbf{V}_j \right) \mathbf{M}^{n-i}. \quad (3.75)$$

Beweis. Stelle zunächst – erneut mit der Definition 3.3 und dem Satz 3.25 (i) – für alle $n \geq 0$ fest, dass

$$\mathbb{V}\mathbf{Z}_n \stackrel{(3.8)}{=} \underbrace{\mathbb{E}[\mathbf{Z}_n \mathbf{Z}_n^T]}_{= \mathbf{C}_n \text{ (3.73)}} - \underbrace{\mathbb{E}\mathbf{Z}_n}_{= (\mathbf{M}^n)^T \mathbf{Z}_0 \text{ (3.65)}} \underbrace{\mathbb{E}\mathbf{Z}_n^T}_{= \mathbf{Z}_0^T \mathbf{M}^n \text{ (3.65)}} = \mathbf{C}_n - (\mathbf{M}^n)^T \mathbf{Z}_0 \mathbf{Z}_0^T \mathbf{M}^n \quad \text{gilt, und setze} \quad (\star)$$

$$\mathbf{A}_n := \sum_{j=1}^k \mu_j^{(n)} \mathbf{V}_j \quad \Rightarrow \quad \mathbf{C}_n \stackrel{(3.74)}{=} \mathbf{M}^T \mathbf{C}_{n-1} \mathbf{M} + \mathbf{A}_{n-1}. \quad (\star\star)$$

Bestimme anschließend iterativ die Matrix $\mathbf{C}_n$. Es ist (Rechnung nur für $n \geq 4$ sinnvoll)

$$\begin{aligned} \mathbf{C}_n &\stackrel{(\star\star)}{=} \mathbf{M}^T \mathbf{C}_{n-1} \mathbf{M} + \mathbf{A}_{n-1} \stackrel{(\star\star)}{=} (\mathbf{M}^T)^2 \mathbf{C}_{n-2} \mathbf{M}^2 + \mathbf{M}^T \mathbf{A}_{n-2} \mathbf{M} + \mathbf{A}_{n-1} \\ &\stackrel{(\star\star)}{=} (\mathbf{M}^T)^3 \mathbf{C}_{n-3} \mathbf{M}^3 + (\mathbf{M}^T)^2 \mathbf{A}_{n-3} \mathbf{M}^2 + \mathbf{M}^T \mathbf{A}_{n-2} \mathbf{M} + \mathbf{A}_{n-1} \\ &\stackrel{(\star\star)}{=} (\mathbf{M}^T)^4 \mathbf{C}_{n-4} \mathbf{M}^4 + (\mathbf{M}^T)^3 \mathbf{A}_{n-4} \mathbf{M}^3 + (\mathbf{M}^T)^2 \mathbf{A}_{n-3} \mathbf{M}^2 + \mathbf{M}^T \mathbf{A}_{n-2} \mathbf{M} + \mathbf{A}_{n-1} \\ &= \dots = (\mathbf{M}^T)^n \mathbf{C}_{n-n} \mathbf{M}^n + (\mathbf{M}^T)^{n-1} \mathbf{A}_{n-n} \mathbf{M}^{n-1} + \dots + \mathbf{M}^T \mathbf{A}_{n-2} \mathbf{M} + \mathbf{A}_{n-1} \\ &\stackrel{(\Delta)}{=} (\mathbf{M}^n)^T \underbrace{\mathbf{C}_0}_{= \mathbf{Z}_0 \mathbf{Z}_0^T \text{ (3.73)}} \mathbf{M}^n + \sum_{i=1}^n (\mathbf{M}^{n-i})^T \mathbf{A}_{i-1} \mathbf{M}^{n-i}. \end{aligned}$$

(Δ) Verwende, dass $\forall n \geq 0$ allgemein $(\mathbf{M}^T)^n = (\mathbf{M}^n)^T$ und $\mathbf{M}^0 = (\mathbf{M}^T)^0 = \mathbf{I}_k$ gilt.

Wird dieser Ausdruck für $\mathbf{C}_n$ in ($\star$) eingesetzt, erhält man die Aussage für $\mathbb{V}\mathbf{Z}_n$ mit dem ersten Teil von ($\star\star$). $\square$

Bemerkungen zum Satz 3.31.

- (i) Für jedes $n \geq 0$ hängt (abgesehen von n selbst) sowohl $\mathbb{E}\mathbf{Z}_n$ als auch $\mathbb{V}\mathbf{Z}_n$ in (3.75) nur von $\mathbf{Z}_0$ und der Verteilung der $\mathbf{X}^{(j)}$ mit $1 \leq j \leq k$ ab.

Wegen (3.49) ($\mathbf{M} = (\mathbb{E}[\mathbf{X}^{(j)}]^T)_{j=1}^k$) sieht man dies für den EW sofort. Bezüglich der Kovarianzmatrix beachte zudem, dass $\mathbf{V}_j \stackrel{(3.72)}{=} \mathbb{E}[\mathbf{X}^{(j)} \mathbf{X}^{(j)T}] - \mathbb{E}\mathbf{X}^{(j)} \mathbb{E}[\mathbf{X}^{(j)}]^T$ und $\mu_j^{(n)} \stackrel{(3.55)}{=} \mathbb{E}Z_{nj} = \mathbf{e}_j^T \mathbb{E}\mathbf{Z}_n \stackrel{(3.75)}{=} \mathbf{e}_j^T (\mathbf{M}^n)^T \mathbf{Z}_0$ gilt. Es lässt sich also beides, wie behauptet, nur in Abhängigkeit von $\mathbb{E}\mathbf{X}^{(j)}$ bzw. $\mathbb{E}[\mathbf{X}^{(j)} \mathbf{X}^{(j)T}]$ mit $1 \leq j \leq k$ und $\mathbf{Z}_0$ darstellen.

- (ii) Wegen (i) müssen sich $\mathbb{E}\mathbf{Z}_n$ und $\mathbb{V}\mathbf{Z}_n$ auch bereits durch die weF $\mathbf{f}$ (und $\mathbf{Z}_0$) beschreiben lassen. Für $\mathbb{E}\mathbf{Z}_n$ in (3.75) sieht man dies beispielsweise anhand von (3.53) ($\mathbf{M}^n = (\mathbf{D}\mathbf{f}(\mathbf{1}))^n$).

Bei $\mathbb{V}\mathbf{Z}_n$ 'steckt' – bis auf die $\mathbf{V}_j$ – ebenfalls alles in $\mathbf{M}^n$ (und $\mathbf{Z}_0$). Für $\mathbf{V}_j$ betrachte $\mathbf{V}_j = (v_{jil})_{i,l=1}^k$ und $v_{jil} \stackrel{(3.72)}{=} \text{Cov}[X_i^{(j)}, X_l^{(j)}] = \mathbb{E}[X_i^{(j)} X_l^{(j)}] - \underbrace{\mathbb{E}X_i^{(j)} \mathbb{E}X_l^{(j)}}_{= \frac{\partial f^{(j)}}{\partial s_i}(\mathbf{1})}$. Damit und mit dem Satz 3.10 (f) ergibt sich (für alle $1 \leq i, j, l \leq k$)

$$\text{Cov}[X_i^{(j)}, X_l^{(j)}] = \frac{\partial^2 f^{(j)}}{\partial s_i \partial s_l}(\mathbf{1}) - \frac{\partial f^{(j)}}{\partial s_i}(\mathbf{1}) \frac{\partial f^{(j)}}{\partial s_l}(\mathbf{1}) + \delta_{il} \frac{\partial f^{(j)}}{\partial s_i}(\mathbf{1}).$$

Für $i = l$ erinnert diese Form stark an (2.54) im einfachen Fall.

Etwa in [3, S. 185] ist eine Art Handlungsanleitung zu finden, wie mittels quadratischer Formen, gebildet aus den Matrizen der zweiten partiellen Ableitungen der $f_n^{(j)}$, Ausdrücke für die Kovarianzmatrizen der $\mathbf{Z}_n$ erarbeitet werden können. Dies wird hier aber nicht genauer ausgeführt (da diese bereits im Satz 3.31 formuliert wurden).

- (iii) Für $n = 0$ ist die äußere Summe in (3.75) leer und der Satz besagt (wenig überraschend und konsistent mit der Nichtzufälligkeit von $\mathbf{Z}_0$), dass $\mathbb{V}\mathbf{Z}_0 = \mathbf{0}$ ist (da $\mathbf{M}^0 = \mathbf{I}_k$ und $\mathbf{C}_0 = \mathbf{Z}_0 \mathbf{Z}_0^T$ gilt, erkennt man das auch an $(\star)$ im Beweis).

Für $n = 1$ vereinfacht sich die Kovarianzmatrix – wiederum wegen $\mathbf{M}^0 = \mathbf{I}_k$ und wegen $\mu_j^{(0)} = Z_{0j}$ – zu $\mathbb{V}\mathbf{Z}_1 = \sum_{j=1}^k Z_{0j} \mathbf{V}_j$ (was mit (3.25) auch der Satz 3.4 direkt besagen würde). Für den von nur einem Typ- l -Objekt gestarteten Prozess $\{\mathbf{Z}_n^{(l)}\}_{n \geq 0}$ vereinfacht sich diese weiter zu $\mathbb{V}\mathbf{Z}_1^{(l)} = \sum_{j=1}^k Z_{0j}^{(l)} \mathbf{V}_j = \mathbf{V}_l \stackrel{(3.72)}{=} \mathbb{V}\mathbf{X}^{(l)}$, da $Z_{0j}^{(l)} = 1$ nur für $j = l$ gilt (und 0 sonst). Dies liegt an $\mathbf{X}^{(l)} \stackrel{(3.27)}{=} \mathbf{Z}_1^{(l)}$. $\triangle$

3.3.3 Eigenschaften in Abhängigkeit von M^n

Das Hauptinteresse liegt wieder darin, wie sich der Prozess über die Zeit – also die Generationen $n \geq 0$ – verhält. Die dahingehende Beeinflussung der (im vorherigen Unterabschnitt untersuchten) ersten beiden Momente der Populationsgrößen wird mit der Bemerkung zum Satz 3.31 (i) klar. Da $\mathbf{Z}_0$ und alle $\mathbf{V}_j$ zeitunabhängig sind, spielt dafür demnach ausschließlich $\mathbf{M}$ bzw. $\mathbf{M}^n$ eine Rolle. Deshalb soll die Matrix $\mathbf{M}$ noch genauer durchleuchtet werden.

Zunächst wird jedoch ein weniger interessantes, triviales Szenario benannt und zukünftig meist nicht mitberücksichtigt.

Singulärer Prozess

Bezüglich des einfachen GWP wurde der triviale Fall, also der, in welchem jedes Individuum in der kommenden Generation genau einen Nachfahren hat (siehe die Definition 2.40), meist ausgeschlossen. Nun soll Analoges – etwas weiter gefasst – ebenfalls für die Multitype Galton-Watson-Prozesse geschehen.

Definition 3.32 (Singulärer Multitype GWP – vgl. z.B. [39, S. 11])

Ein Multitype GWP heißt *singulär*, wenn jedes Objekt genau ein neues (eines beliebigen Typs) produziert. I.a.W. heißt das: Für alle $1 \leq j \leq k$ gilt

$$\sum_{l=1}^k p^{(j)}(\mathbf{e}_l) = 1 \quad \Rightarrow \quad p^{(j)}(\mathbf{r}) = 0 \quad \text{für alle } \mathbf{r} \in \mathbb{N}_0^k \setminus \{\mathbf{e}_l : 1 \leq l \leq k\}. \quad (3.76)$$

Bemerkung zur Definition 3.32. Im Gegensatz zum trivialen Fall bei einfachen GWP ist ein singulärer Multitype GWP im Allgemeinen dem Zufall unterworfen, da der Typ dieses jeweilig produzierten Objekts nicht sicher ist (mit $k = 1$ stimmen der triviale und der singuläre Fall überein). $\triangle$

In der Literatur erfolgt die Def. häufig äquivalent wie folgt (vgl. z.B. [3, S. 184]):

$$\text{MGWP ist singulär} \iff \exists \mathbf{A} = (a_{jl})_{j,l=1}^k \in \mathbb{R}_+^{k \times k} : \mathbf{f}(\mathbf{s}) = \mathbf{A}\mathbf{s} \quad \forall \mathbf{s} \in [0, 1]^k. \quad (3.77)$$

Letzteres, also $\mathbf{f}(\mathbf{s}) = \mathbf{A}\mathbf{s}$ für alle $\mathbf{s} \in [0, 1]^k$, gilt (wegen (3.38)) genau dann, wenn

$$f^{(j)}(\mathbf{s}) = a_{j1}s_1 + \dots + a_{jk}s_k = \sum_{l=1}^k a_{jl}s_l \quad \text{für jedes } 1 \leq j \leq k \text{ ist.} \quad (3.78)$$

D.h., jedes $f^{(j)}$ ist linear in $s_1, \dots, s_k$, ohne Konstante. Um die Äquivalenz in (3.77) zu begründen, setze $R := \{\mathbf{e}_l : 1 \leq l \leq k\}$. Für jedes $\mathbf{r} \notin R$ und alle $1 \leq j \leq k$ bedingt die linke Seite, also die Singularität des Prozesses, gemäß (3.76) $p^{(j)}(\mathbf{r}) = 0$. Deshalb gilt diesfalls für alle $\mathbf{s} \in [0, 1]^k$

$$f^{(j)}(\mathbf{s}) \stackrel{(3.32)}{=} \sum_{\mathbf{r} \in \mathbb{N}_0^k} \overbrace{p^{(j)}(\mathbf{r})}^{=0 \quad \forall \mathbf{r} \notin R} \mathbf{s}^{\mathbf{r}} = \sum_{\mathbf{r} \in R} p^{(j)}(\mathbf{r}) \mathbf{s}^{\mathbf{r}} = \sum_{l=1}^k p^{(j)}(\mathbf{e}_l) \overbrace{\mathbf{s}^{\mathbf{e}_l}}^{=s_l} = \sum_{l=1}^k p^{(j)}(\mathbf{e}_l) s_l. \quad (3.79)$$

Wird dann $a_{lj} = p^{(j)}(\mathbf{e}_l)$ für jedes $1 \leq j, l \leq k$ gesetzt, ist folglich auch die rechte Seite von (3.77) erfüllt. Umgekehrt müssen, wenn die rechte Seite erfüllt ist, für jedes $\mathbf{r} \notin R$ und alle $1 \leq j \leq k$ die Koeffizienten $p^{(j)}(\mathbf{r}) = 0$ sein (was an (3.78) in Kombination mit der Reihendefinition der $f^{(j)}$ (3.32) ersichtlich ist). Daraus folgt (3.76) bzw. die Singularität.

Bei einem singulären Prozess folgt aus (3.78) zudem für jedes Tupel $1 \leq j, l \leq k$

$$\frac{\partial f^{(j)}}{\partial s_l}(\mathbf{1}) = a_{jl} \quad \Rightarrow \quad \mathbf{M}_{(3.53)} = \left(\frac{\partial f^{(j)}}{\partial s_l}(\mathbf{1}) \right)_{j,l=1}^k = (a_{jl})_{j,l=1}^k = \mathbf{A} \quad \Rightarrow \quad \mathbf{f}(\mathbf{s}) = \mathbf{M}\mathbf{s}.$$

Das kann man sich auch folgendermaßen plausibel machen: Seien $1 \leq j, l \leq k$. Aufgrund der Singularität können nur die Ereignisse $\{\mathbf{X}^{(j)} = \mathbf{e}_i\}$ mit $1 \leq i \leq k$ positive Wahrscheinlichkeiten haben. Es gilt demnach $\mathbb{P}(X_l^{(j)} = 1) = \mathbb{P}(\mathbf{X}^{(j)} = \mathbf{e}_l)$ (*) und $\mathbb{P}(X_l^{(j)} > 1) = 0$.

$$\Rightarrow \mu_{jl} \stackrel{(3.51)}{=} \mathbb{E}X_l^{(j)} = 0 \cdot \mathbb{P}(X_l^{(j)} = 0) + 1 \cdot \mathbb{P}(X_l^{(j)} = 1) \stackrel{(*)}{=} \mathbb{P}(\mathbf{X}^{(j)} = \mathbf{e}_l) \stackrel{(3.24)}{=} p^{(j)}(\mathbf{e}_l).$$

Deshalb ist $a_{jl} \stackrel{(3.78)}{=} p^{(j)}(\mathbf{e}_l) = \mu_{jl}$. Da dies für alle $1 \leq j, l \leq k$ gilt, folgt daraus ebenfalls $\mathbf{A} = \mathbf{M}$. Daran sieht man auch, dass $\mathbf{M}$ bzw. $\mathbf{A}$ des singulären Prozesses eine stochastische Matrix ist (siehe die Def. A.69), denn $\sum_{l=1}^k \mu_{jl} = \sum_{l=1}^k p^{(j)}(\mathbf{e}_l) \stackrel{(3.76)}{=} 1$.

Annahme 3.33. Für alles Folgende wird meist von der *Nichtsingularität* des MGWP ausgegangen. Der singuläre Fall wird nur anhand weniger Beispiele beleuchtet (genauer im Bsp. 3.45 sowie der Bem. zum Satz 3.46 (ii) im Kontext einer Markov-Kette). $\triangle$

Offensichtlich kann ein singulärer MGWP (wenn $\mathbf{Z}_0 \neq \mathbf{0}$ ist) nicht aussterben – siehe auch die Bemerkung zum Lemma 3.49 im Abschnitt 3.4.

Positiv regulärer Fall

Nun werden zwei weitere Eigenschaften von Multitype Galton-Watson-Prozessen eingeführt. Für die zweite und wichtigere der beiden wird anschließend gezeigt, dass sie sich unmittelbar an der Erwartungsmatrix $\mathbf{M}$, also den Reproduktionsverteilungen der $\mathbf{X}^{(j)}$ mit $1 \leq j \leq k$ festmachen lässt.

Definition 3.34 (Irreduzibler und positiv regulärer MGWP – angelehnt an [39, Definitionen 5.1 und 5.2, S. 11])

- (i) Ein MGWP heißt *irreduzibel*, wenn es für jedes Tupel $1 \leq j, l \leq k$ ein $n \geq 1$ gibt,

$$\text{sodass} \quad \mathbb{P}(Z_{nl}^{(j)} \geq 1) \stackrel{(3.26)}{=} \mathbb{P}(Z_{nl} \geq 1 | \mathbf{Z}_0 = \mathbf{e}_j) > 0 \quad \text{ist.} \quad (3.80)$$

D.h., für jedes $1 \leq j, l \leq k$ gilt $\mathbb{P}(\exists n \geq 1 : Z_{nl}^{(j)} \geq 1) > 0$ oder – i.a.W. – jedem Objekt (jeglichen Typs) kann (möglicherweise) irgendwann ein Objekt eines beliebigen Typs nachfolgen. Dies gilt wiederum für alle Typen von nachfolgenden Objekten, muss aber nicht zwingend gleichzeitig bzw. simultan zutreffen.

3 Galton-Watson-Verzweigungsprozesse mit mehreren unterschiedlichen Typen

- (ii) Ein MGWP heißt *positiv regulär*, wenn es ein $n \geq 1$ gibt, sodass (3.80) unabhängig von j und l (also für alle $1 \leq j, l \leq k$) gilt. Das kann geschrieben werden als

$$\exists n \geq 1 : \mathbb{P}(Z_{nl}^{(j)} \geq 1) > 0 \quad \text{für alle } 1 \leq j, l \leq k$$

und heißt i.a.W.: In irgendeiner der zukünftigen Generationen sind für jedes Objekt (jeglichen Typs) alle Typen als Nachfolger möglich (diese Möglichkeit muss simultan bestehen).

Klarerweise impliziert die positive Regularität (ii) die Irreduzibilität (i) eines Prozesses.

Dass ein MGWP positiv regulär ist, kann – wie oben angekündigt – äquivalent über $\mathbf{M}$ definiert werden (vgl. dazu z.B. auch [19, Def. II.5.2, S. 38] und die Bemerkung zur Definition. 3.37).

Lemma 3.35. *Es gilt:*

$$\text{Ein MGWP ist positiv regulär} \iff \exists n \geq 1 : \mathbf{M}^n > \mathbf{0}. \quad (3.81)$$

Die rechte Seite ist genau dann erfüllt, wenn es ein $n \geq 1$ gibt, sodass $\forall 1 \leq j, l \leq k$

$$\mathbb{E}Z_{nl}^{(j)} = \mu_{jl}^{(n)} > 0 \text{ gilt} \quad \left(\iff_{(3.58)} \mathbb{E}\mathbf{Z}_n^{(j)} = (\boldsymbol{\mu}_j^{(n)})^T > \mathbf{0} \right). \quad (\star)$$

Beweis. Da (für beliebige $1 \leq j, l \leq k$ und $n \geq 1$) $\mathbb{E}Z_{nl}^{(j)} \stackrel{(2.4)}{=} \sum_{i \geq 0} i \mathbb{P}(Z_{nl}^{(j)} = i)$ ist, muss die Ungleichung (3.80) erfüllt sein, damit dieser EW nicht 0 ist – und umgekehrt. Existiert ein $n \geq 1$, sodass beides für *alle* j, l gilt, so ist das gleichbedeutend mit $(\star)$ und wegen $\mathbf{M}^n \stackrel{(3.67)}{=} (\mu_{jl}^{(n)})_{j,l=1}^k$ auch mit der rechten Seite von (3.81). Andererseits entspricht dieser Fall gemäß der ersten Aussage der obigen Def. 3.34 (ii) ebenso der pos. Regularität. $\square$

Die Objekte eines irreduziblen MGWP müssen nicht selbst in der Lage sein, alle möglichen Objekttypen zu produzieren. Gemäß der Definition 3.34 ist es aber erforderlich, dass von jedem ausgehend alle Typen irgendwann – und bei positiver Regularität sogar gleichzeitig – als Nachfolger möglich sind. Abweichungen davon werden in der folgenden Bemerkung benannt.

Bemerkung 3.36.

- (i) **Reduzibler Fall** (vgl. z.B. [18, 2.3.2, S. 28]): Gibt es Objekte, denen gewisse Typen niemals – also in keiner der künftigen Generationen – nachfolgen können, so spricht man von einem *reduziblen MGWP*.
- (ii) **Periodischer Fall** (vgl. z.B. [39, 1. Remark, S. 13]): Ist ein MGWP zwar irreduzibel, aber nicht positiv regulär, ergibt sich der sogenannte *periodische Fall*.

Es kann aber jeder periodische Prozess in einen nichtperiodischen transformiert werden (vgl. dazu etwa [18, Note, S. 27] – dort ist auch ein simples Bsp. für einen periodischen Fall angeführt). Deshalb werden diese hier nicht gesondert behandelt. Es wird ggf. nur nach positiver Regularität unterschieden (bzw., wo es notwendig ist, verlangt – siehe (iii)).

- (iii) Der überwiegende Teil des vorliegenden Kapitels behält auch bei reduziblen und periodischen Fällen seine Gültigkeit. Manches hält aber nicht, wenn die positive Regularität verletzt ist. Dies betrifft insbesondere den Satz 3.46 (ii) im nächsten Unterabschnitt und die wesentlichen, finalen Resultate, welche die Frage nach der Aussterbewahrscheinlichkeit in Abhängigkeit von $\mathbf{M}$ beantworten (siehe 3.4.2). Deshalb werden dort nur positiv reguläre Fälle berücksichtigt.

Eine ausführlichere Behandlung periodischer und reduzibler Prozesse – inklusive Adaptionen der entsprechenden Theoreme – findet sich etwa unter [39, 2.2., S. 52 ff. bzw. 2.3, S. 60 ff.]. $\triangle$

Positiv reguläre Matrix und Satz von Perron-Frobenius

Dieser Teilabschnitt widmet sich Eigenschaften der Erwartungswertmatrix $\mathbf{M}$, welche generell für nichtnegative Matrizen, also nicht spezifisch für MGWP, gelten. Dafür wird zuerst die positive Regularität einer quadratischen Matrix definiert.

Definition 3.37 (Positiv reguläre Matrix – vgl. z.B. [39, Def. 6.1, S. 13])
Allgemein heißt eine nichtnegative $k \times k$ -Matrix $\mathbf{M}$ *positiv regulär*, wenn die rechte Seite von (3.81) ($\exists n \geq 1 : \mathbf{M}^n > \mathbf{0}$) und zudem $\mathbf{M} < \infty$ gilt.

Bemerkung zur Definition 3.37. Ein MGWP mit positiv regulärer Erwartungswertmatrix $\mathbf{M}$ ist also *immer* positiv regulär (siehe das Lemma 3.35). Umgekehrt hat jeder positiv reguläre MGWP unter der Annahme 3.22 ($\mathbf{M} < \infty$) eine positiv reguläre Erwartungswertmatrix – die Def. 3.34 (ii) fällt in diesem Sinne also mit der Def. 3.37 zusammen.

In [3, S. 184] wird eine Matrix $\mathbf{M}$, welche zwar $\exists n \geq 1 : \mathbf{M}^n > \mathbf{0}$, aber nicht zwingend $\mathbf{M} < \infty$ erfüllt, *strikt positiv* genannt. Dort wird die positive Regularität eines MGWP dadurch charakterisiert, dass er eine solche EW-Matrix besitzt. Das ist äquivalent zu der hier getroffenen Definition (da das Lemma 3.35 die Ann. 3.22 gar nicht erfordert).

Ohne diese Annahme müsste $\mathbf{M}$ eines positiv regulären Prozesses demnach gar nicht unbedingt positiv regulär, aber zumindest strikt positiv sein (es wären daher grundsätzlich genauso unendliche Einträge erlaubt – vgl. dazu auch [39, Remark, S. 14]). $\triangle$

Folgendes kann allgemein bzgl. nichtnegativer $k \times k$ -Matrizen beobachtet werden.

Lemma 3.38. *Sei $\mathbf{M} \geq \mathbf{0}$ eine $k \times k$ -Matrix. Es gilt:*

- (i) *Wann immer eine ganze Zeile oder Spalte von $\mathbf{M}^n$, $n \geq 1$ beliebig, nur aus Nullen besteht, gilt das auch für die entsprechende Zeile oder Spalte aller $\mathbf{M}^m$ mit $m \geq n$.*
- (ii) *Angenommen, es gibt ein $n \geq 1$ mit $\mathbf{M}^n > \mathbf{0}$. Dann ist auch $\mathbf{M}^m > \mathbf{0} \forall m \geq n$.*
- (iii) *Ist $\mathbf{M}$ positiv regulär, so kann nie – also für kein $n \geq 1$ – eine ganze Zeile oder Spalte von $\mathbf{M}^n$ nur aus Nullen bestehen.*

Beweis. Es soll die Notation $\mathbf{M}^n = (\mu_{jl}^{(n)})_{j,l=1}^k$ (wie in 3.3.2 eingeführt) verwendet werden. Für (i) bestehe für ein beliebiges $n \geq 1$ und $1 \leq j \leq k$ die j -te Zeile von $\mathbf{M}^n$ nur aus

3 Galton-Watson-Verzweigungsprozesse mit mehreren unterschiedlichen Typen

Nullen, d.h. $\mu_{ji}^{(n)} = 0 \forall 1 \leq i \leq k$. Mit der Matrixmultiplikationsformel (3.68) kann man folgern, dass somit auch $\mu_{jl}^{(n+1)} = 0 \forall 1 \leq l \leq k$ gelten muss, was bedeutet, dass auch die j -te Zeile von $\mathbf{M}^{n+1}$ ein Nullvektor ist. Induktiv lässt sich dasselbe für alle $\mathbf{M}^m$ mit $m \geq n$ schließen. Da $\mathbf{M}^{n+1} = \mathbf{M}^n \mathbf{M} = \mathbf{M} \mathbf{M}^n$ gilt, kann (3.68) auch als $\mu_{jl}^{(n+1)} = \sum_{i=1}^k \mu_{ji} \mu_{il}^{(n)}$ geschrieben und damit analog für aus Nullen bestehende Spalten argumentiert werden.

Für (ii) und (iii) sei $n \geq 1$ so, dass $\mathbf{M}^n > \mathbf{0}$ ist. Beachte, dass (i) insbesondere auch für $\mathbf{M}$ gilt. Es darf also für kein $1 \leq l \leq k$ die l -te Spalte von $\mathbf{M}$ nur aus Nullen bestehen – d.h., es gibt für jedes l mindestens ein $1 \leq i \leq k$, sodass $\mu_{il} > 0$ ist. Da für alle $1 \leq i, j \leq k$ die $\mu_{ji}^{(n)} > 0$ sind, folgt wieder mit (3.68), dass (für alle $1 \leq j, l \leq k$) jedes $\mu_{jl}^{(n+1)} > 0$ und somit auch für alle $m \geq n$ jedes $\mu_{jl}^{(m)} > 0$ ist. Damit ist der Punkt (ii) gezeigt. Dieser verunmöglicht also Nuller-Zeilen oder -Spalten in $\mathbf{M}^m$ für jedes $m \geq n$. Umgekehrt würde aber das Vorhandensein einer solchen Zeile oder Spalte in $\mathbf{M}^m$ für irgendein $m < n$ (wegen (i)) wiederum $\mathbf{M}^n > \mathbf{0}$ ausschließen, weshalb diese *nie* vorkommen dürfen. Das ist genau die Aussage in (iii). $\square$

Für positiv reguläre Matrizen der Def. 3.37 gibt es das folgende Resultat aus der linearen Algebra, welches für die Eingrenzung der Aussterbewahrscheinlichkeit positiv regulärer MGWP benötigt wird (siehe den Abschnitt 3.4). Es gilt eigentlich auch allgemeiner für strikt positive Matrizen, wie etwa den Standardwerken [19, Theorem II.5.1, S. 37 f.] oder [3, Theorem V.2.1, S. 185] zu entnehmen ist, und wird in der Literatur zu Verzweigungsprozessen meist unbewiesen angewendet. Für einen ausführlichen Beweis siehe etwa [27, Appendix 2, S. 542 ff.] (zum Eigenwertproblem sei ggf. etwa auf [34, 6, S. 121 ff.] verwiesen). Die Grundaussagen werden im nächsten Satz – ohne Beweis – und entsprechende Konsequenzen im darauffolgenden Korollar dargelegt (oft wird aber beides zusammen als ein Theorem formuliert).

Satz 3.39 (Satz von Perron-Frobenius – vgl. [39, Theorem 6.1, S. 14]; ohne Bew.)
Eine positiv reguläre $k \times k$ -Matrix $\mathbf{M}$ hat einen maximalen Eigenwert ρ , welcher positiv und einfach¹ ist und mit positiven linken bzw. rechten Eigenvektoren $\mathbf{u} = (u_1, \dots, u_k) > \mathbf{0}$ bzw. $\mathbf{v} = (v_1, \dots, v_k)^T > \mathbf{0}$ korrespondiert. D.h., für jeden entsprechenden Zeilenvektor $\mathbf{u}$ bzw. Spaltenvektor $\mathbf{v}$ gilt

$$\mathbf{u}\mathbf{M} = \mathbf{u}\rho \quad \text{und} \quad \mathbf{M}\mathbf{v} = \rho\mathbf{v}. \quad (3.82)$$

Diese Eigenvektoren können auf eindeutige Weise so normalisiert gewählt werden, dass

$$\mathbf{u} \cdot \mathbf{1} = \sum_{i=1}^k u_i = 1 \quad \text{sowie} \quad \mathbf{u}\mathbf{v} = \sum_{i=1}^k u_i v_i = 1 \quad \text{ist.} \quad (3.83)$$

Werden weiters die $k \times k$ -Matrizen $\mathbf{A}$ und $\mathbf{B}$ definiert als

$$\mathbf{A} := (a_{ij})_{i,j=1}^k := (v_i u_j)_{i,j=1}^k = \mathbf{v}\mathbf{u}, \quad \mathbf{B} := (b_{ij})_{i,j=1}^k := \mathbf{M} - \rho\mathbf{A}, \quad (3.84)$$

so existiert ein $m_0 \geq 1$ sowie ein $0 < C < \infty$ und ein $0 < \alpha < \rho$, sodass

$$b_{ij}^{(n)} \leq C\alpha^n \quad \text{für jedes } n \geq m_0 \text{ und alle } 1 \leq i, j \leq k \text{ gilt.} \quad (3.85)$$

¹Mit Vielfachheit 1 – vgl. z.B. [34, Def. 6.1.15, S. 127].

Die $b_{ij}^{(n)}$ sind dabei für jedes $n \geq 1$ die Einträge der Matrix $\mathbf{B}^n$, d.h. $\mathbf{B}^n = (b_{ij}^{(n)})_{i,j=1}^k$. Dieser maximale Eigenwert ρ wird auch **Perron-Wurzel** genannt.

Bemerkung zum Satz 3.39 von Perron-Frobenius. Die Schranke in (3.85) gilt unabhängig von i und j für jeden Eintrag von $\mathbf{B}^n$ mit $n \geq m_0$. Deshalb kann auch ein $0 < D < \infty$ gefunden werden, sodass $|\mathbf{B}^n| \leq D\alpha^n$ ist (mit $|\mathbf{B}^n| := \sum_{i,j=1}^k |b_{ij}^{(n)}|$ wird dies etwa durch $D := k^2 C$ erreicht).

Oft wird diese Beschränkung äquivalent mit dem Landau-Symbol $\mathcal{O}$ geschrieben (vgl. etwa die Referenz in der 'Klein- $\mathcal{O}$ '-Definition 3.2), d.h. z.B. $b_{ij}^{(n)} = \mathcal{O}(\alpha^n)$. $\triangle$

Korollar zum Satz 3.39 von Perron-Frobenius.

Sei alles wie im Satz und $n \geq 1$. Es gilt

- (i) $\mathbf{A}\mathbf{A} = \mathbf{A}$ bzw. $\mathbf{A}^n = \mathbf{A}$ und $u_i < 1$, $0 < v_i, a_{ij} < \infty$ für alle $1 \leq i, j \leq k$,
- (ii) $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A} = \mathbf{0}$,
- (iii) die Matrix $\mathbf{M}^n$ kann dargestellt werden als

$$\mathbf{M}^n = \rho^n \mathbf{A} + \mathbf{B}^n \quad \text{und} \quad (3.86)$$
- (iv) ρ^n korrespondiert so mit $\mathbf{M}^n$ wie ρ mit $\mathbf{M}$. D.h., ρ^n ist der Eigenwert von $\mathbf{M}^n$ mit den linken bzw. rechten Eigenvektoren $\mathbf{u}$ bzw. $\mathbf{v}$.

Beweis.

$$(i) \quad \mathbf{A}\mathbf{A} \stackrel{(3.84)}{=} \left(\sum_{l=1}^k v_l u_l v_l u_l \right)_{i,j=1}^k = \left(v_i u_j \sum_{l=1}^k u_l v_l \right)_{i,j=1}^k \stackrel{(3.83)}{=} \mathbf{A}. \quad \text{Induktiv } (\mathbf{A}^n = \mathbf{A}^{n-1} \mathbf{A}) \stackrel{= \mathbf{A} \text{ (Ind.-Ann.)}}{=} \mathbf{A}.$$

folgt die zweite Aussage. Für die dritte beachte, dass laut dem Satz $u_i, v_i > 0$ für alle $1 \leq i \leq k$ gilt. Wegen $\sum_{i=1}^k u_i \stackrel{(3.83)}{=} 1$ folgt, dass $0 < u_i < 1$ und wegen $\sum_{i=1}^k u_i v_i \stackrel{(3.83)}{=} 1$, dass $0 < v_i < \infty$ für alle $1 \leq i \leq k$ gilt. Deshalb gilt auch (vgl. (3.84)) $0 < a_{ij} < \infty$ für alle $1 \leq i, j \leq k$.

$$(ii) \quad \text{Es gilt } \mathbf{B}\mathbf{A} \stackrel{(3.84)}{=} (\mathbf{M} - \rho\mathbf{A})\mathbf{A} = \mathbf{M}\mathbf{A} - \rho\mathbf{A}\mathbf{A} \stackrel{(3.84)}{=} \mathbf{M}\mathbf{v}\mathbf{u} - \rho\mathbf{v}\mathbf{u} \stackrel{= \mathbf{M}\mathbf{v} \text{ (3.82)}}{=} \mathbf{0}. \quad \text{Analog erhält man } \mathbf{A}\mathbf{B} = \mathbf{A}(\mathbf{M} - \rho\mathbf{A}) = \mathbf{A}\mathbf{M} - \rho\mathbf{A} = \mathbf{v}\mathbf{u}\mathbf{M} - \mathbf{v}\mathbf{u}\rho \stackrel{= \mathbf{uM} \text{ (3.82)}}{=} \mathbf{0}.$$

(iii) Für $n = 1$ folgt $\mathbf{M} = \rho\mathbf{A} + \mathbf{B}$ direkt aus (3.84). Induktiv kann dann für $n \geq 2$ gezeigt werden, dass (da $\mathbf{B}\mathbf{A} = \mathbf{A}\mathbf{B} = \mathbf{0}$ ist (ii))

$$\begin{aligned} \mathbf{M}^n &= \mathbf{M}^{n-1} \mathbf{M} \stackrel{\text{Ind.-Ann.}}{=} (\mathbf{B}^{n-1} + \rho^{n-1} \mathbf{A})(\mathbf{B} + \rho\mathbf{A}) \\ &= \mathbf{B}^n + \rho \mathbf{B}^{n-2} \mathbf{B}\mathbf{A} + \rho^{n-1} \mathbf{A}\mathbf{B} + \rho^n \mathbf{A}^2 \stackrel{= \mathbf{A} \text{ (i)}}{=} \mathbf{B}^n + \rho^n \mathbf{A} \quad \text{gilt.} \end{aligned}$$

(iv) Beachte, dass $\mathbf{B}\mathbf{v} \stackrel{(3.84)}{=} (\mathbf{M} - \rho\mathbf{A})\mathbf{v} \stackrel{(3.84)}{=} \mathbf{M}\mathbf{v} - \rho\mathbf{v}\mathbf{u}\mathbf{v} \stackrel{(3.82)}{=} \mathbf{M}\mathbf{v} - \mathbf{M}\mathbf{v} = \mathbf{0}$ und analog $\mathbf{u}\mathbf{B} = \mathbf{u}\mathbf{M} - \mathbf{u}\mathbf{v}\mathbf{u}\rho = \mathbf{0}$ gilt (*).

Weiters ist (wegen $\mathbf{u}\mathbf{v} \stackrel{(3.83)}{=} \mathbf{1}$ und $\mathbf{A} \stackrel{(3.84)}{=} \mathbf{v}\mathbf{u}$) $\mathbf{A}\mathbf{v} = \mathbf{v}\mathbf{u}\mathbf{v} = \mathbf{v}$ und $\mathbf{u}\mathbf{A} = \mathbf{u}\mathbf{v}\mathbf{u} = \mathbf{u}$ (**).

$$\Rightarrow \begin{cases} \mathbf{M}^n \mathbf{v} \stackrel{(iii)}{=} \mathbf{B}^n \mathbf{v} + \rho^n \mathbf{A} \mathbf{v} \stackrel{(**)}{=} \mathbf{B}^{n-1} \mathbf{B} \mathbf{v} + \rho^n \mathbf{v} \stackrel{(*)}{=} \rho^n \mathbf{v} & \text{und analog} \\ \mathbf{u} \mathbf{M}^n \stackrel{(iii)}{=} \mathbf{u} \mathbf{B}^n + \mathbf{u} \mathbf{A} \rho^n \stackrel{(**)}{=} \mathbf{u} \mathbf{B} \mathbf{B}^{n-1} + \mathbf{u} \rho^n \stackrel{(*)}{=} \mathbf{u} \rho^n. \end{cases} \quad \square$$

3 Galton-Watson-Verzweigungsprozesse mit mehreren unterschiedlichen Typen

Um den Einfluss der *Perron-Wurzel* ρ auf positiv reguläre Multitype Galton-Watson-Prozesse zu veranschaulichen, wird kurz ein motivierendes Beispiel mit einer spezifischen initialen Populationsgröße angeführt. Im Abschnitt 3.4 wird die Bedeutung von ρ vertieft.

Beispiel 3.40 (angelehnt an [18, S. 25 f.]). Es sei ein MGWP mit positiv regulärer Erwartungswertmatrix $\mathbf{M}$ gegeben (der Prozess ist also positiv regulär), ρ sei die *Perron-Wurzel* dieser Matrix und $\mathbf{u}$ der linke Eigenvektor zu ρ , normalisiert wie im Satz 3.39.

Angenommen, es gibt eine Konstante $C \in \mathbb{R}_+$, sodass $C\mathbf{u}^T$ ganzzahlig ist. Dann gilt mit $\mathbf{Z}_0 := C\mathbf{u}^T$ für alle $n \geq 0$

$$\begin{aligned} \mathbb{E}Z_n^T &\stackrel{(3.65)}{=} \mathbf{Z}_0^T \mathbf{M}^n = C\mathbf{u}^T \mathbf{M}^n = \overbrace{C\mathbf{u}^T \mathbf{M}}^{=\rho\mathbf{u}^T (3.82)} \mathbf{M}^{n-1} = \overbrace{C\rho\mathbf{u}^T \mathbf{M}}^{=\rho\mathbf{u}^T} \mathbf{M}^{n-2} = C\rho^2\mathbf{u}^T \mathbf{M}^{n-2} \\ &= \dots = C\rho^n\mathbf{u}^T \mathbf{M}^0 = C\rho^n\mathbf{u}^T \Rightarrow \mathbb{E}Z_n = \underbrace{C\mathbf{u}^T \rho^n}_{\text{konstant}} = \mathbf{Z}_0 \rho^n. \end{aligned}$$

Es hängt also das langfristige Verhalten des Erwartungswertes der Populationsgrößen Z_n nur von ρ ab. Er fällt in n für $\rho < 1$, bleibt konstant für $\rho = 1$ und wächst für $\rho > 1$. Es wird sich aber im Unterabschnitt 3.4.2 zeigen, dass die *Perron-Wurzel* auch für beliebige $\mathbf{Z}_0$ diesen Einfluss auf den Prozess nimmt, weshalb daran – analog zu μ im einfachen Fall – die *Kritikalität* positiv regulärer MGWP festgemacht werden kann. $\triangle$

Mit dem Resultat des *Satzes von Perron-Frobenius* kann auch die folgende Aussage über die langfristigen *erwarteten relativen Häufigkeiten* einzelner Objekttypen getroffen werden (dieses Resultat ist inspiriert von [18, S. 27 f.]).

Lemma 3.41 (Erwartete relative Häufigkeiten der Populationsgrößen Z_{nl})

Betrachte einen positiv regulären MGWP mit Perron-Wurzel ρ . Die erwartete Population einzelner Objekttypen relativ zur erwarteten gesamten Population einer Generation kann asymptotisch allein am korrespondierenden linken Eigenvektor $\mathbf{u}$ festgemacht werden:

Unabhängig von der initialen Population $\mathbf{Z}_0$ gilt für alle $1 \leq l \leq k$

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}Z_{nl}}{\mathbb{E}|Z_n|} = u_l.$$

Beweis. Wähle $1 \leq l \leq k$ beliebig und $m_0 \geq 1$ so, dass (3.85) des Satzes 3.39 für $0 < \alpha < \rho$ und $0 < C < \infty$ erfüllt ist. Dann folgt für alle $n \geq m_0$ und $1 \leq i, j \leq k$

$$\frac{b_{ij}^{(n)}}{\rho^n} = \frac{\alpha^n b_{ij}^{(n)}}{\rho^n \alpha^n} \stackrel{(3.85)}{\leq} \frac{\alpha^n C \alpha^n}{\rho^n \alpha^n} = \underbrace{\left(\frac{\alpha}{\rho}\right)^n}_{\rightarrow 0} C \Rightarrow \frac{B^n}{\rho^n} = \left(\frac{b_{ij}^{(n)}}{\rho^n}\right)_{i,j=1}^{n \rightarrow \infty} \mathbf{0} \cdot C = \mathbf{0}. \quad (\star)$$

Weiters gilt für beliebige $n \geq 0$ aus $\mathbb{E}Z_n^T \stackrel{(3.65)}{=} \mathbf{Z}_0^T \mathbf{M}^n$ auch $\mathbb{E}Z_{nl} = \mathbf{Z}_0^T \mathbf{M}^n \mathbf{e}_l$ ($\star\star$) und $\mathbb{E}|Z_n| = \mathbb{E}(\sum_{l=1}^k Z_{nl}) = \sum_{l=1}^k \mathbb{E}Z_{nl} \stackrel{(\star\star)}{=} \sum_{l=1}^k \mathbf{Z}_0^T \mathbf{M}^n \mathbf{e}_l = \mathbf{Z}_0^T \mathbf{M}^n \sum_{l=1}^k \mathbf{e}_l \stackrel{(3.2)}{=} \mathbf{Z}_0^T \mathbf{M}^n \mathbf{1}$ ($\star\star\star$) für die Gesamtpopulation in der n -ten Generation. Es folgt

$$\begin{aligned} \frac{\mathbb{E}Z_{nl}}{\mathbb{E}|Z_n|} &\stackrel{(\star\star)}{=} \frac{\mathbf{Z}_0^T \mathbf{M}^n \mathbf{e}_l}{\mathbf{Z}_0^T \mathbf{M}^n \mathbf{1}} \stackrel{(3.86)}{=} \frac{\mathbf{Z}_0^T (\mathbf{A}\rho^n + \mathbf{B}^n) \mathbf{e}_l}{\mathbf{Z}_0^T (\mathbf{A}\rho^n + \mathbf{B}^n) \mathbf{1}} = \frac{\mathbf{Z}_0^T (\mathbf{A} + \frac{\mathbf{B}^n}{\rho^n}) \mathbf{e}_l}{\mathbf{Z}_0^T (\mathbf{A} + \frac{\mathbf{B}^n}{\rho^n}) \mathbf{1}} \\ &\xrightarrow{n \rightarrow \infty} \frac{\mathbf{Z}_0^T \mathbf{A} \mathbf{e}_l}{\mathbf{Z}_0^T \mathbf{A} \mathbf{1}} \stackrel{(3.84)}{=} \underbrace{\frac{\mathbf{Z}_0^T \mathbf{v}}{\mathbf{Z}_0^T \mathbf{v}}}_{\text{Skalar}} \underbrace{\frac{\mathbf{u} \mathbf{e}_l}{\mathbf{u} \mathbf{1}}}_{=1 \text{ mit (3.83)}} = \mathbf{u} \mathbf{e}_l = u_l. \end{aligned} \quad \square$$

3.3.4 Multitype GWP als Markov-Kette

Wie im Unterabschnitt 3.2.1 angekündigt, wird nun die Definition der Übergangswahrscheinlichkeiten vorgenommen und der Prozess anschließend als eine Markov-Kette formuliert. Betrachte hierzu die Gleichung (3.42), welche sich in der Reihendarstellung für alle $n \geq 0$ liest als

$$(f(s))^r \stackrel{(3.42)}{=} f_{Z_{n+1}|Z_n=r}(s) \stackrel{(3.17)}{=} \sum_{t \in \mathbb{N}_0^k} \overbrace{\mathbb{P}(Z_{n+1} = t | Z_n = r)}^{=: p(r,t)} s^t = \sum_{t \in \mathbb{N}_0^k} p(r,t) s^t. \quad (3.87)$$

Daran sieht man für alle $r, t \in \mathbb{N}_0^k$, dass die Koeffizienten dieser bedingten weF, also $p(r, t)$, unabhängig von $n \geq 0$ sind. Dies rührt daher, dass Z_{n+1} gegeben $Z_n = r$ nur durch die Z-Vektoren $\{X_{n,i}^{(j)}\}_{i=1}^{r_j}$ mit $1 \leq j \leq k$ bestimmt wird (vgl. etwa (3.69)). Deren Verteilung hängt definitionsgemäß nicht von n ab (Def. 3.11 bzw. (3.23); die Reproduktionsverteilungen der einzelnen Objekte sind – je nach Typ – in jeder Generation gleich).

Das motiviert zu der folgenden Definition.

Definition 3.42 (Übergangswahrscheinlichkeiten des Multitype GWP – vgl. z.B. [3, V.1.(3), S. 183] oder [39, (4.11), S. 10])

Die Wahrscheinlichkeit für den Übergang der Populationsgröße von r nach t in einem Generationssprung ist (unabhängig von $n \geq 0$)

$$\mathbb{P}(Z_{n+1} = t | Z_n = r) \stackrel{(3.87)}{=} p(r, t) = \text{'Koeffizient von } s^t \text{ der Reihe } (f(s))^r \text{'}. \quad (3.88)$$

Bemerkungen zur Definition 3.42. Alle Reihen in $f \stackrel{(3.38)}{=} (f^{(j)})_{j=1}^n$ und somit auch die Reihe f^r sind gemäß der Bem. zur Def. 3.8 wohldefiniert. Deshalb sind es die Übergangswahrscheinlichkeiten ebenso – auch wenn das für den ersten Ausdruck in (3.88) als elementare bedingte WK genau genommen nicht zwingend so sein muss (vergleichbar zum in der Bem. zur Def. 2.24 bzgl. des einfachen Falls Besprochenen – beachte auch erneut die Anmerkungen am Ende der Notation 3.9).

Auf das Analogon im einfachen Fall – dort betreffend den Koeffizienten der l -ten Potenz ($l \geq 0$) der Reihe $(f(s))^l$ – wurde bereits in den Bem. zum Korollar 2.31 (iii) hingewiesen. Die Koeffizienten können etwa gemäß dem Satz 3.10 (b) aus der Reihe gewonnen werden (sofern diese bekannt ist). $\triangle$

Im nächsten Satz wird nun dargelegt, dass es sich beim Multitype GWP – wie im einfachen Fall – tatsächlich um eine Markov-Kette handelt. Der Beweis ist nahezu baugleich zum entsprechenden Satz 2.25. Deshalb werden er (und insbesondere die Lemmata zuvor) nicht in aller Ausführlichkeit wiederholt. Es sei etwa auch auf die angegebene Referenz in [39] verwiesen.

Satz 3.43 (MGWP als Markov-Kette – vgl. [3, S. 183] oder [39, Theorem 4.1, S. 10])
Sei $P := (p(r, t))_{r, t \in \mathbb{N}_0^k}$. Der MGWP $\{Z_n\}_{n \geq 0}$ ist Markov(δ_{Z_0}, P).

3 Galton-Watson-Verzweigungsprozesse mit mehreren unterschiedlichen Typen

Beweis. Stelle zunächst (gemäß dem Satz 3.10 (g) oder erneut mit (3.42)) fest, dass es sich bei $(\mathbf{f}(\mathbf{s}))^{\mathbf{r}} = \prod_{j=1}^k (f^{(j)}(\mathbf{s}))^{r_j}$ für ein fixes $\mathbf{r} \in \mathbb{N}_0^k$ um eine weF handelt – nämlich um die der Summe von $r_1 + \dots + r_k$ $\mathbb{N}_0^k$ -wertigen $\mathbf{Z}$ -Vektoren. Deshalb müssen sich deren Koeffizienten auf 1 summieren, womit $\mathbf{P}$ eine stochastische Matrix ist (vgl. die Def. A.69). Da $\mathbf{Z}_0$ nichtzufällig ist, gilt weiters für alle $\mathbf{r} \in \mathbb{N}_0^k$, dass $\mathbb{P}(\mathbf{Z}_0 = \mathbf{r}) = \delta_{\mathbf{Z}_0 \mathbf{r}}^{(1,1)}$ ist. Damit ist der Punkt (i) der Markov-Ketten-Definition A.70 erfüllt.

Es ist bereits intuitiv für jedes $n \geq 0$ klar, dass die Populationsgrößen $\{\mathbf{Z}_m\}_{m \leq n}$ unabhängig von allen $\{\mathbf{X}_{l,i}^{(j)}\}_{l \geq n, i \geq 1}$ mit $1 \leq j \leq k$ sind. Das ist aber auch analog zum Beweis des Lemmas 2.22 argumentierbar.

Dementsprechend kann auch das Lemma 2.23 analog angewendet werden (nun mit vektorwertigen Variablen und basierend auf der Rekursion (3.25)). Damit erhält man für alle $n \geq 0$ und alle $\mathbf{r}, \mathbf{t}, \mathbf{i}_0, \dots, \mathbf{i}_{n-1} \in \mathbb{N}_0^k$ (mit $\mathbb{P}(\mathbf{Z}_0 = \mathbf{i}_0, \dots, \mathbf{Z}_{n-1} = \mathbf{i}_{n-1}, \mathbf{Z}_n = \mathbf{r}) > 0$)

$$\mathbb{P}(\mathbf{Z}_{n+1} = \mathbf{t} | \mathbf{Z}_0 = \mathbf{i}_0, \dots, \mathbf{Z}_{n-1} = \mathbf{i}_{n-1}, \mathbf{Z}_n = \mathbf{r}) = \mathbb{P}(\mathbf{Z}_{n+1} = \mathbf{t} | \mathbf{Z}_n = \mathbf{r}),$$

womit der Punkt (ii) der Def. A.70 ebenfalls erfüllt und der Satz gezeigt ist. $\square$

Bemerkung 3.44. Der MGWP besitzt also, genauso wie der einfache GWP, die *Markov-Eigenschaft (ME)* – siehe den Satz A.71. In der Bem. zur Def. 3.14 (ii) wurde dies bereits angemerkt und (ohne den obigen Satz 3.43 zu verwenden) argumentiert, dass

$$(\mathbf{Z}_n | \mathbf{Z}_0 = \mathbf{r}) \stackrel{d}{\underset{(3.28)/(3.29)}}{=} (\mathbf{Z}_{n+m} | \mathbf{Z}_m = \mathbf{r}) \quad \text{für alle } m, n \geq 0 \text{ und } \mathbf{r} \in \mathbb{N}_0^k \text{ gilt.} \quad (3.89)$$

$\triangle$

Beispiel 3.45 (Singulärer Prozess als endliche Markov-Kette)

Ein singulärer MGWP (siehe die Def. 3.32) – für welchen der Satz 3.43 natürlich ebenfalls gilt – ist eine Markov-Kette mit *endlichem* Zustandsraum I (vgl. z.B. auch [3, S. 184] oder [39, S. 13]). Dies liegt daran, dass jedes Objekt genau ein neues produziert.

Für $\mathbf{Z}_0 = \mathbf{r} \in \mathbb{N}_0^k$ kann der Zustandsraum des Prozesses als $I = \{\mathbf{t} \in \mathbb{N}_0^k : |\mathbf{t}| = |\mathbf{r}|\}$ angegeben werden. Klarerweise sind für potentielle $\mathbf{t} \in \mathbb{N}_0^k$ nur dann positive Übergangswahrscheinlichkeiten $p(\mathbf{r}, \mathbf{t})$ möglich, wenn $\mathbf{t} \in I$ (also $|\mathbf{t}| = |\mathbf{r}| = |\mathbf{Z}_0|$) ist. Ansonsten sind alle anderen $p(\mathbf{r}, \mathbf{t}) = 0$. $\triangle$

Unter der Annahme 3.33 (Nichtsingularität) und vorausgesetzt der Prozess ist positiv regulär (Def. 3.34 (ii)), ist abgesehen von $\mathbf{0}$ jeder Zustand der Markov-Kette transient (wie im einfachen Fall). Dies zeigt das nächste Resultat, welches dem Satz 2.26 recht ähnlich ist.

Satz 3.46 (Transienz der Zustände $\mathbf{r} \neq \mathbf{0}$ – angelehnt an [19, Theorem II.6.1, S. 39])
Betrachte einen MGWP $\{\mathbf{Z}_n\}_{n \geq 0}$ als Markov-Kette. Für die Zustände $\mathbf{r} \in \mathbb{N}_0^k$ gilt:

- (i) *Der Zustand $\mathbf{r} = \mathbf{0}$ ist immer absorbierend (also auch rekurrent).*
- (ii) *Wenn der Prozess nichtsingulär und positiv regulär ist, ist jeder andere Zustand $\mathbf{r} \in \mathbb{N}_0^k \setminus \{\mathbf{0}\}$ transient. D.h., u.a. gilt*

$$\mathbb{P}(\mathbf{Z}_n \rightarrow \mathbf{r}) = 0 \quad \text{für alle } \mathbf{r} \in \mathbb{N}_0^k \setminus \{\mathbf{0}\}. \quad (3.90)$$

3.3 Verteilung, Markov-Eigenschaft und Momente

Beweis(-skizze). Der erste Punkt (i) kann analog zum einfachen Fall begründet werden (siehe den Beweis des Satzes 2.26 – (3.88) ist konsistent mit der Def. A.75 absorbierender Zustände, denn $(\mathbf{f}(\mathbf{s}))^0 = 1$ und deshalb ist $p(\mathbf{0}, \mathbf{0}) = 1$ der Koeffizient von $\mathbf{s}^0 = 1$). Auch in der Definition 3.13 wurde bereits dargelegt, dass aus $\mathbf{Z}_n = \mathbf{0}$ ebenso $\mathbf{Z}_{n+1} = \mathbf{0}$ und somit $\mathbf{Z}_m = \mathbf{0}$ für alle $m \geq n$ folgt $(\star)$.

Für (ii) sei $\mathbf{r} \in \mathbb{N}_0^k \setminus \{\mathbf{0}\}$ beliebig. Es sollen zwei Fälle unterschieden werden (vergleichbar mit den einfachen GWP, bei welchen bereits $p_0 = 0$ und $p_0 > 0$ getrennt behandelt wurden). Für den ersten Fall sei für jeden Typ $1 \leq j \leq k$ die Wahrscheinlichkeit

$$\mathbb{P}(\mathbf{Z}_n^{(j)} = \mathbf{0}) \stackrel{(3.26)}{=} \mathbb{P}(\mathbf{Z}_n = \mathbf{0} | \mathbf{Z}_0 = \mathbf{e}_j) > 0 \quad \text{für mindestens ein } n \geq 1. \quad (\text{Fall 1})$$

D.h., wenn der Prozess von einem Objekt – egal welchen Typs – gestartet wird, besteht immer die Möglichkeit, dass er ausstirbt. Bezeichne für jedes $1 \leq j \leq k$ ein passendes $n \geq 1$, welches dem (Fall 1) genügt, als $\tilde{n}_j \geq 1$. Weiters sei $\tilde{n} := \max(\tilde{n}_1, \dots, \tilde{n}_k)$. Dann gilt für jedes j , dass $\mathbb{P}(\mathbf{Z}_{\tilde{n}}^{(j)} = \mathbf{0}) \stackrel{(\star)}{\geq} \mathbb{P}(\mathbf{Z}_{\tilde{n}_j}^{(j)} = \mathbf{0}) > 0$ ist $(\star\star)$. Unter Zuhilfenahme der Beziehung (3.28) für $\tilde{n}$ ($\{\mathbf{Z}_{\tilde{n},i}^{(j)}\}_{i \geq 1}$ sind i.i.d.-Kopien von $\mathbf{Z}_{\tilde{n}}^{(j)}$) folgt

$$\begin{aligned} \mathbb{P}(\mathbf{Z}_{\tilde{n}} = \mathbf{0} | \mathbf{Z}_0 = \mathbf{r}) &\stackrel{(3.28)}{=} \mathbb{P}\left(\sum_{j=1}^k \sum_{i=1}^{r_j} \mathbf{Z}_{\tilde{n},i}^{(j)} = \mathbf{0}\right) \stackrel{\text{nicht negativ}}{=} \mathbb{P}\left(\bigcap_{j=1}^k \left(\bigcap_{i=1}^{r_j} \{\mathbf{Z}_{\tilde{n},i}^{(j)} = \mathbf{0}\}\right)\right) \\ &\stackrel{\text{uabh.}}{=} \prod_{j=1}^k \prod_{i=1}^{r_j} \mathbb{P}(\mathbf{Z}_{\tilde{n},i}^{(j)} = \mathbf{0}) \stackrel{(A.42)}{=} \prod_{j=1}^k \underbrace{\mathbb{P}(\mathbf{Z}_{\tilde{n}}^{(j)} = \mathbf{0})}_{\stackrel{\geq \mathbb{P}(\mathbf{Z}_{\tilde{n}_j}^{(j)} = \mathbf{0}) > 0}{(\star\star)}}^{r_j} > 0 \quad (\star\star\star) \\ &\Rightarrow \mathbb{P}(\exists n \geq 1 : \mathbf{Z}_n = \mathbf{0} | \mathbf{Z}_0 = \mathbf{r}) > 0. \end{aligned}$$

Nun kann ähnlich wie im Fall $p_0 = 0$ im Beweis des Satzes 2.26 vorgegangen werden. Beachte (für $m \geq 0 : \mathbb{P}(\mathbf{Z}_m = \mathbf{r}) > 0$), dass die Rückkehrwahrscheinlichkeit (nach dem ersten Sprung) von $\mathbf{r}$ nach $\mathbf{r}$

$$\begin{aligned} \mathbb{P}(\exists n \geq 1 : \mathbf{Z}_{n+m} = \mathbf{r} | \mathbf{Z}_m = \mathbf{r}) &\stackrel{\text{ME}}{\stackrel{(3.89)}{=}} \mathbb{P}(\exists n \geq 1 : \mathbf{Z}_n = \mathbf{r} | \mathbf{Z}_0 = \mathbf{r}) \\ &\leq \mathbb{P}(\exists n \geq 1 : \mathbf{Z}_n \neq \mathbf{0} | \mathbf{Z}_0 = \mathbf{r}) = 1 - \underbrace{\mathbb{P}(\exists n \geq 1 : \mathbf{Z}_n = \mathbf{0} | \mathbf{Z}_0 = \mathbf{r})}_{>0 \ (\star\star\star)} < 1 \quad \text{ist.} \end{aligned}$$

Somit ist $\mathbf{r}$ gemäß dem Satz A.77 transient.

Im zweiten Fall gibt es Objekttypen, von welchen aus der Zustand $\mathbf{0}$ – und damit das Absorbieren – nicht mehr erreichbar ist. Dieser gestaltet sich etwas schwieriger und wird hier ausgespart. Siehe dafür [19, Beweis des Theorems II.6.1, Case 2, S. 39 f.]. In diesem Beweis wird die Positivität von $\mathbf{M}$ oder $\mathbf{M}^n$ für ein $n \geq 0$ ausgenützt, um im Wesentlichen zu zeigen, dass die Population insbesondere der betroffenen Typen immer mehr anwächst und der Prozess somit 'explodiert' (vgl. auch [3, S. 186]). Deshalb ist auch dann kein Zustand rekurrent und somit sind in beiden Fällen alle $\mathbf{r} \neq \mathbf{0}$ transient.

Die letzte Aussage (3.90) folgt aus der Transienz von $\mathbf{r}$ und kann fast gleich (mit Z-Vektoren anstatt ZV) wie der zweite Teil von (2.42) im Satz 2.26 begründet werden. $\square$

Bemerkungen zum Satz 3.46.

- (i) Für den ersten Fall im Beweis von (ii) wird die *positive Regularität* gar nicht benötigt. Eine etwas andere Beweisführung bzw. Fallunterscheidung, bei welcher sie in beiden Fällen ausgenutzt wird, findet sich unter [39, Beweis des Theorems 5.1, S. 12]. Im ersten dieser Fälle wird zumindest von gewissen Typen mit einer positiven WK gar kein Objekt (keines Typs) produziert, wohingegen dies im zweiten Fall nicht möglich ist (was wiederum auf eine 'Explosion' des Prozesses hinausläuft). Insgesamt ist dieser Beweis insgesamt etwas einfacher nachvollziehbar als der oben referenzierte in [19].
- (ii) Im *singulären* Fall (dann gibt es nur endlich viele mögliche Zustände – siehe das Bsp. 3.45) können sowohl rekurrente (absorbierende und nicht absorbierende) als auch transiente Zustände vorkommen. Dies soll anhand drei einfacher Beispiele verdeutlicht werden.
 1. Wenn $k = 1$ und $p^{(1)}(1) = 1$ ist, liegt ein einfacher singulärer GWP mit $p_1 = 1$ vor. Dieser entspricht genau dem trivialen Fall bzw. dem ersten Punkt des Satzes 2.26 (ii). Dann ist jeder Zustand absorbierend (und somit auch rekurrent).

Für die nächsten beiden kleinen Beispiele wird jeweils von singulären Prozessen mit $k = 2$ und $|\mathbf{Z}_0| = 2$ (d.h. Zustandsraum $I = \{(2, 0)^T, (1, 1)^T, (0, 2)^T\}$) ausgegangen.

2. Mit $\mathbf{M} = \begin{pmatrix} 0,5 & 0,5 \\ 0,5 & 0,5 \end{pmatrix}$ ist jeder Zustand aus I rekurrent (aber nicht absorbierend).
3. Ist $\mathbf{M} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$, so ist der Zustand $(0, 2)^T$ absorbierend (und rekurrent), $(2, 0)^T$ und $(1, 1)^T$ hingegen sind transient.

Dass Nichtsingularität alleine nicht hinreichend für die Transienz aller Zustände $\mathbf{r} \neq \mathbf{0}$ ist, zeigt etwa das Beispiel in [19, Remark 2, S. 40]. △

3.4 Aussterben und Aussterbewahrscheinlichkeit

Die Überlebenschance des Prozesses kann unter Umständen sehr stark von der initialen Population $\mathbf{Z}_0$ abhängen. Insbesondere dann, wenn die Reproduktionsverteilungen der unterschiedlichen Objekttypen deutlich differieren (etwa wenn manche mit sehr hoher und andere mit sehr niedriger Wahrscheinlichkeit gar nichts produzieren).

Deshalb konzentriert sich die Betrachtung zunächst auf Prozesse, welche von einem einzelnen Objekt initiiert werden. Die Vorgehensweise gestaltet sich sehr ähnlich wie im Abschnitt 2.5 für die einfachen GWP und wird danach kurz auf beliebige Anfangspopulationsgrößen $\mathbf{Z}_0$ verallgemeinert. Schließlich wird der Einfluss der *Perron-Wurzel* ρ auf die Aussterbewahrscheinlichkeit(en) im positiv regulären Fall untersucht.

Beachte, dass weiterhin die Annahme 3.22 ($\mathbf{M} < \infty$) und, sofern nicht anders erläutert, die Annahme 3.33 (Nichtsingularität) gelten.

3.4.1 Aussterben des Multitype Galton-Watson-Prozesses

Es soll zuerst noch einmal allgemein festgehalten werden, dass das Erreichen einer Populationsgröße von $\mathbf{0}$ in einer beliebigen Generation $n \geq 0$ dies auch für alle zukünftigen Generationen bedingt. Es gilt also für jedes $n \geq 0$ (wie schon im 2. Kapitel für (2.65) argumentiert und auch im Beweis des Satzes 3.46 unter $(\star)$ erläutert wurde)

$$\{\mathbf{Z}_n = \mathbf{0}\} \subseteq \{\mathbf{Z}_{n+1} = \mathbf{0}\} \quad \Rightarrow \quad \text{insbesondere } \{\mathbf{Z}_n^{(j)} = \mathbf{0}\} \subseteq \{\mathbf{Z}_{n+1}^{(j)} = \mathbf{0}\}. \quad (3.91)$$

Aussterben des Prozesses $\{\mathbf{Z}_n^{(j)}\}_{n \geq 0}$

Nun werden die Aussterbeereignisse der von einzelnen Objekten gestarteten Prozesse und die Wahrscheinlichkeiten dafür definiert.

Definition 3.47 (Aussterbeereignis und Aussterbewahrscheinlichkeit für den Prozess $\{\mathbf{Z}_n^{(j)}\}_{n \geq 0}$ – angelehnt an [39, 1.7, S. 15])

Sei $1 \leq j \leq k$. Definiere die Wahrscheinlichkeit dafür, dass der Prozess $\{\mathbf{Z}_n^{(j)}\}_{n \geq 0}$ in einer bestimmten Generation $n \geq 0$ ausgestorben ist, als

$$\eta_n^{(j)} := \mathbb{P}(\mathbf{Z}_n^{(j)} = \mathbf{0}) \stackrel{(3.35)}{=} f_n^{(j)}(\mathbf{0}) \stackrel{(3.91)}{\Rightarrow} \eta_n^{(j)} \leq \eta_{n+1}^{(j)} \quad \text{für alle } n \geq 0. \quad (3.92)$$

Weiters sei sein *Aussterben* allg. das Ereignis $E^{(j)} := \{\text{Prozess } \{\mathbf{Z}_n^{(j)}\}_{n \geq 0} \text{ stirbt aus}\}$.

Für die gleich definierte *Aussterbewahrscheinlichkeit* (kurz *AWK*) des Prozesses $\{\mathbf{Z}_n^{(j)}\}_{n \geq 0}$ gilt (wobei alles völlig analog zu den Argumentationen für (2.66), (2.67) und (2.69) bzgl. den einfachen Prozessen begründbar ist)

$$\eta^{(j)} := \mathbb{P}(E^{(j)}) = \mathbb{P}(\exists n \geq 1 : \mathbf{Z}_n^{(j)} = \mathbf{0}) = \mathbb{P}(\mathbf{Z}_n^{(j)} \rightarrow \mathbf{0}) \stackrel{(3.35)}{=} \lim_{n \rightarrow \infty} f_n^{(j)}(\mathbf{0}) \quad (3.93)$$

$$\stackrel{(3.92)}{\Rightarrow} \eta_n^{(j)} \rightarrow \eta^{(j)}. \quad (3.94)$$

3 Galton-Watson-Verzweigungsprozesse mit mehreren unterschiedlichen Typen

Auch diese Aussterbewahrscheinlichkeiten sollen als Vektoren geschrieben werden.

Notation 3.48 (Aussterbewahrscheinlichkeitsvektoren)

$$\text{Schreibe} \quad \boldsymbol{\eta}_n := (\eta_n^{(1)}, \dots, \eta_n^{(k)})^T \stackrel{(3.92)}{=} (f_n^{(1)}(\mathbf{0}), \dots, f_n^{(k)}(\mathbf{0}))^T \stackrel{(3.39)}{=} \mathbf{f}_n(\mathbf{0}) \quad (3.95)$$

$$\text{für } n \geq 0 \text{ und} \quad \boldsymbol{\eta} := (\eta^{(1)}, \dots, \eta^{(k)})^T. \quad (3.96)$$

Damit können der zweite Teil von (3.92) und (3.94) geschrieben werden als

$$\boldsymbol{\eta}_n \leq \boldsymbol{\eta}_{n+1} \quad \text{für alle } n \geq 0 \quad \text{und} \quad \boldsymbol{\eta}_n \rightarrow \boldsymbol{\eta}. \quad (3.97)$$

△

Als Analogon zu (2.68) aus dem Lemma 2.43 gilt für $\boldsymbol{\eta}$ das folgende Resultat.

Lemma 3.49 (Fixpunktgleichung für den AWK-Vektor)

Der AWK-Vektor $\boldsymbol{\eta}$ ist der kleinste nichtnegative Fixpunkt von $\mathbf{f}$. Es gilt also

$$\mathbf{f}(\boldsymbol{\eta}) = \boldsymbol{\eta} \quad \stackrel{(\star)}{\Rightarrow} \quad \mathbf{f}_n(\boldsymbol{\eta}) = \boldsymbol{\eta} \quad \text{für alle } n \geq 0. \quad (3.98)$$

($\star$) Es ist jeder nichtnegative Fixpunkt von $\mathbf{f}$ auch ein Fixpunkt aller $\mathbf{f}_n$ mit $n \geq 0$.

Beweis. Der Beweis funktioniert nahezu gleich wie im einfachen Fall. Für alle $n \geq 0$ gilt

$$\boldsymbol{\eta}_{n+1} \stackrel{(3.95)}{=} \mathbf{f}_{n+1}(\mathbf{0}) \stackrel{(3.48)}{=} \mathbf{f}(\mathbf{f}_n(\mathbf{0})) \stackrel{(3.95)}{=} \mathbf{f}(\boldsymbol{\eta}_n),$$

woraus die Fixpunktgleichung (3.98) mit $n \rightarrow \infty$ folgt, da $\mathbf{f}$ stetig ist. Die Aussage ($\star$) ist das Pendant zum Lemma 2.19 (i) und kann wiederum mit (3.48) genau gleich gezeigt werden. Sie wird nun auch verwendet, um noch zu zeigen, dass $\boldsymbol{\eta}$ der kleinste Fixpunkt ist. Angenommen, ein nichtnegatives $\boldsymbol{\psi}$ erfüllt ebenfalls die Fixpunktgleichung (3.98). Dann gilt für alle $n \geq 1$, ähnlich wie zuvor und da alle $\mathbf{f}$ und alle $\mathbf{f}_n$ in jedem Argument isoton sind (weil dies auch für alle $f^{(j)}$ und $f_n^{(j)}$ gilt $\Rightarrow \mathbf{f}_n(\boldsymbol{\psi}) \geq \mathbf{f}_n(\mathbf{0})$),

$$\boldsymbol{\psi} = \mathbf{f}_{n+1}(\boldsymbol{\psi}) \stackrel{(3.48)}{=} \mathbf{f}(\mathbf{f}_n(\boldsymbol{\psi})) \geq \mathbf{f}(\mathbf{f}_n(\mathbf{0})) \stackrel{(3.95)}{=} \mathbf{f}(\boldsymbol{\eta}_n) \stackrel{(3.97)}{\rightarrow} \mathbf{f}(\boldsymbol{\eta}) \stackrel{(3.98)}{=} \boldsymbol{\eta}. \quad \square$$

Bemerkung zum Lemma 3.49. Das Lemma gilt allgemein – weder die positive Regularität noch die Nichtsingularität muss vorausgesetzt werden. Bei einem *singulären Prozess* haben die $f^{(j)}$ keine Konstanten (vgl. (3.78)), weshalb $\mathbf{f}(\mathbf{0}) = \mathbf{0}$ gilt. Damit ist $\mathbf{0}$ der kleinste nichtnegative Fixpunkt von $\mathbf{f}$. Tatsächlich kann ein von einem Typ- j -Objekt gestarteter singulärer Prozess nicht aussterben und hat folglich die AWK $\eta^{(j)} = 0$ (vgl. auch [19, Remark 2, S. 41]). △

Aussterben des Prozesses $\{\mathbf{Z}_n\}_{n \geq 0}$

Bei gegebenem bzw. gefundenem $\boldsymbol{\eta}$ ist es leicht möglich, die Aussterbewahrscheinlichkeit des Prozesses $\{\mathbf{Z}_n\}_{n \geq 0}$ mit beliebigem $\mathbf{Z}_0 \in \mathbb{N}_0^k$ zu bestimmen. Wird die Aussterbewahrscheinlichkeit von $\{\mathbf{Z}_n\}_{n \geq 0}$ analog zur Definition 3.47 als

$$\mathbb{P}(\mathbf{Z}_n \rightarrow \mathbf{0}) = \lim_{n \rightarrow \infty} \mathbb{P}(\mathbf{Z}_n = \mathbf{0}) \stackrel{\text{Satz 3.10 (c)}}{=} \lim_{n \rightarrow \infty} f_{\mathbf{Z}_n}(\mathbf{0}) \quad \text{angegeben,}$$

kann die *Additive Property* des Prozesses (siehe die Bem. zur Def. 3.14 (ii)) ausgenutzt werden. Mit der genau gleichen Vorgehensweise wie in (★★) im Beweis des Satzes 3.46 (ii) über die Transienz erhält man für $\mathbf{Z}_0 = \mathbf{r} \in \mathbb{N}_0^k$

$$\begin{aligned} \mathbb{P}(\mathbf{Z}_n \rightarrow \mathbf{0} | \mathbf{Z}_0 = \mathbf{r}) &= \prod_{j=1}^k (\mathbb{P}(\mathbf{Z}_n^{(j)} \rightarrow \mathbf{0}))^{r_j} \stackrel{(3.93)}{=} \prod_{j=1}^k (\eta^{(j)})^{r_j} \stackrel{(3.3)}{\stackrel{(3.96)}}{=} \boldsymbol{\eta}^{\mathbf{r}} \\ &\Rightarrow \mathbb{P}(\mathbf{Z}_n \rightarrow \mathbf{0}) = \boldsymbol{\eta}^{\mathbf{Z}_0}. \end{aligned} \quad (3.99)$$

Alternativ kann auch die Aussage in (3.45), in der ein Zusammenhang zwischen den weF der $\mathbf{Z}_n$ und jenen der $\mathbf{Z}_n^{(j)}$ hergestellt wurde, herangezogen werden. Damit gilt ebenso

$$\mathbb{P}(\mathbf{Z}_n = \mathbf{0}) = f_{\mathbf{Z}_n}(\mathbf{0}) \stackrel{(3.45)}{=} (f_n(\mathbf{0}))^{\mathbf{Z}_0} \stackrel{(3.95)}{=} \boldsymbol{\eta}_n^{\mathbf{Z}_0} \stackrel{(3.97)}{\rightarrow} \boldsymbol{\eta}^{\mathbf{Z}_0}.$$

3.4.2 Kritikalität des MGWP – positiv regulärer Fall

Wie am Ende des Unterabschnitts 3.3.3 angekündigt und dort im Beispiel 3.40 motiviert wurde, soll nun die Kritikalität definiert bzw. untersucht werden. Erstrebt wird eine analoge Erweiterung des Satzes 2.44 für den MGWP.

In der Bemerkung 3.36 (iii) wurde bereits darauf hingewiesen, dass diesbzgl. ausschließlich positiv reguläre Prozesse behandelt werden. Dies liegt daran, dass diese sowohl für die Transienz der Zustände ungleich $\mathbf{0}$ (Satz 3.46 (ii)) als auch den *Satz von Perron-Frobenius* (Satz 3.39) vorausgesetzt werden muss und beide hier zur Anwendung kommen. Dementprechend wird nun die folgende zusätzliche Annahme formuliert.

Annahme. Ab sofort wird stets die *positiver Regularität* des MGWP vorausgesetzt. $\triangle$

Satz 3.50 (Kritikalität des MGWP – angelehnt an [19, Theorem II.7.1, S. 41 f.] und [39, Theorem 7.1, S. 16 f.]

Es sei ein nichtsingulärer MGWP mit positiv regulärer Erwartungswertmatrix $\mathbf{M}$ (der Prozess ist also positiv regulär) und mit der Perron-Wurzel ρ (wie im Satz 3.39) gegeben. Dann gilt (analog zum einfachen Fall – Satz 2.44)

- (i) *wenn $\rho \leq 1$ ist, so ist $\boldsymbol{\eta} = \mathbf{1}$. Dann heißt der Prozess **subkritisch**, falls $\rho < 1$ bzw. **kritisch**, falls $\rho = 1$ ist.*
- (ii) *wenn $\rho > 1$ ist, so ist $\boldsymbol{\eta} < \mathbf{1}$ und der Prozess wird als **superkritisch** bezeichnet.*

Bevor der Satz 3.50 bewiesen wird, werden zunächst noch drei Hilfsbeobachtungen formuliert. Diese kommen anschließend im Beweis des Satzes zur Anwendung und sind von den Referenzen, welche im Satz angegeben sind, inspiriert.

Lemma 3.51. *Es sei alles wie im Satz 3.50. Dann gilt das Folgende:*

$$(i) \quad \eta_n^{(j)} \stackrel{(3.92)}{=} f_n^{(j)}(\mathbf{0}) < 1 \quad \text{für alle } n \geq 1 \text{ und } 1 \leq j \leq k.$$

(ii) *Sei $\rho > 1$. Dann gibt es ein $n_0 \geq 1$, sodass*

$$|\mathbf{M}^{n_0} \mathbf{s}| > 2|\mathbf{s}| \quad \text{für alle } \mathbf{s} \geq \mathbf{0} \text{ mit } \mathbf{s} \neq \mathbf{0} \text{ gilt.}$$

(iii) *Sei erneut $\rho > 1$. Für jedes $n_0 \geq 1$, also insbesondere auch für das in (ii) gefundene, kann ein $\varepsilon_{n_0} > 0$ gewählt werden, sodass*

$$|\mathbf{1} - \mathbf{f}_{n_0}(\mathbf{1} - \mathbf{s})| > |\mathbf{s}| \quad \text{für alle } \mathbf{s} \geq \mathbf{0} \text{ mit } \mathbf{s} \neq \mathbf{0} \text{ und } |\mathbf{s}| < \varepsilon_{n_0} \text{ gilt.}$$

Beweis des Lemmas 3.51.

- (i) Sei $1 \leq j \leq k$ beliebig. Angenommen, es gibt ein $\tilde{n} \geq 0$, für welches $f_{\tilde{n}}^{(j)}(\mathbf{0}) \stackrel{(3.35)}{=} \mathbb{P}(\mathbf{Z}_{\tilde{n}}^{(j)} = \mathbf{0}) = 1$ ist. Mit (3.16) folgt dann $\mathbb{E}\mathbf{Z}_{\tilde{n}}^{(j)} = \mathbf{0} \stackrel{(3.58)}{\Rightarrow} \boldsymbol{\mu}_j^{(\tilde{n})} = \mathbf{0}$, also dass die j -te Zeile von $\mathbf{M}^{\tilde{n}}$ nur aus Nullen besteht. Laut dem Lemma 3.38 (iii) steht dies aber im Widerspruch zur positiven Regularität von $\mathbf{M}$.
- (ii) Da $\mathbf{M}$ positiv regulär ist, gilt der *Satz von Perron-Frobenius*, Satz 3.39. Dieser – und die dort eingeführte Notation – soll nun verwendet werden. Definiere zusätzlich $v_\star := \min(v_1, \dots, v_k)$ und $u_\star := \min(u_1, \dots, u_k)$ und stelle fest, dass

$$\begin{aligned} |\mathbf{A}\mathbf{s}| &\stackrel{(3.84)}{=} |\mathbf{v}\mathbf{u}\mathbf{s}| = |(v_i u_j)_{i,j=1}^k \mathbf{s}| = \left| \left(\sum_{j=1}^k v_i u_j s_j \right)_{i=1}^k \right| = \sum_{i=1}^k \left(\sum_{j=1}^k \overbrace{v_i u_j}^{\geq v_\star u_\star} s_j \right) \\ &\geq \left(\sum_{i=1}^k v_\star u_\star \right) \left(\sum_{j=1}^k s_j \right) = \underbrace{k v_\star u_\star}_{=: K \text{ (konstant)}} |\mathbf{s}| = K |\mathbf{s}| \quad \text{ist.} \end{aligned} \quad (\star)$$

Nun gilt mit dem Korollar zum Satz 3.39 und für jedes $n \geq 1$, dass (es ist $\mathbf{B}^n \geq \mathbf{0}$)

$$\begin{aligned} \mathbf{M}^n &\stackrel{(3.86)}{=} \rho^n \mathbf{A} + \mathbf{B}^n \geq \rho^n \mathbf{A} \quad \text{ist, woraus für jedes } \mathbf{s} \geq \mathbf{0} \text{ mit } \mathbf{s} \neq \mathbf{0} \text{ folgt, dass} \\ \mathbf{M}^n \mathbf{s} &\geq \rho^n \mathbf{A}\mathbf{s} \quad \text{ist} \quad \Rightarrow |\mathbf{M}^n \mathbf{s}| \geq \rho^n |\mathbf{A}\mathbf{s}| \geq \rho^n K |\mathbf{s}|. \end{aligned} \quad (\star)$$

Da $\rho > 1$ und K konstant ist, kann also ein $n_0 \geq 1$ gefunden werden, sodass $\rho^{n_0} K > 2$ und somit $|\mathbf{M}^{n_0} \mathbf{s}| > 2|\mathbf{s}|$ gilt.

- (iii) Betrachte $\mathbf{s} = (s_1, \dots, s_k)^T \geq \mathbf{0}$, deren Komponenten $s_i \neq 0$ und $|1 - s_i| \leq 1$ für alle $1 \leq i \leq k$ erfüllen. Für jedes $n \geq 1$ liefert eine mehrdimensionale Taylorapproximation erster Ordnung² von $\mathbf{f}$ in $\mathbf{1}$ und für $\mathbf{1} - \mathbf{s} \rightarrow \mathbf{1}$ bzw. $\underbrace{\mathbf{s} \rightarrow \mathbf{0}}_{\Rightarrow |\mathbf{s}| \rightarrow 0}$ ($\mathbf{f}_n(\mathbf{1}) = \mathbf{1}$; Lemma 3.23 (i))

$$\begin{aligned} \mathbf{f}_n(\mathbf{1} - \mathbf{s}) &= \underbrace{\mathbf{f}_n(\mathbf{1})}_{=1} + \underbrace{\mathbf{D}\mathbf{f}_n(\mathbf{1})}_{=\mathbf{M}^n \text{ (3.67)}}(-\mathbf{s}) + \underbrace{\mathcal{O}(|-\mathbf{s}|)}_{=|\mathbf{s}| \text{ (Not. 3.1 (iii))}} = \mathbf{1} - \mathbf{M}^n \mathbf{s} + \mathcal{O}(|\mathbf{s}|) \\ &\Rightarrow \mathbf{f}_n(\mathbf{1} - \mathbf{s}) - \mathbf{1} + \mathbf{M}^n \mathbf{s} = \mathcal{O}(|\mathbf{s}|) \quad \text{für } |\mathbf{s}| \rightarrow 0. \end{aligned}$$

²Vgl. z.B. [32, 2.4, S. 64 ff.].

3.4 Aussterben und Aussterbewahrscheinlichkeit

Mit der $\mathcal{O}$ -Def. 3.2 folgt weiters, dass für jedes $C > 0$ ein $\varepsilon_n > 0$ existiert, sodass

$$|\mathbf{f}_n(\mathbf{1} - \mathbf{s}) - \mathbf{1} + \mathbf{M}^n \mathbf{s}| = |\mathbf{1} - \mathbf{f}_n(\mathbf{1} - \mathbf{s}) - \mathbf{M}^n \mathbf{s}| < C|\mathbf{s}| \quad \text{für } |\mathbf{s}| < \varepsilon_n \text{ ist.}$$

Insbesondere gilt das auch für das spezifische $n_0 \geq 1$, welches im Punkt (ii) dieses Lemmas gefunden wurde. Verwende nun die umgekehrte Dreiecksungleichung, setze $C = 1$ und wähle $\varepsilon_{n_0} > 0$ so, dass für alle $\mathbf{s} \geq \mathbf{0}$ mit $\mathbf{s} \neq \mathbf{0}$ und $|\mathbf{s}| < \varepsilon_{n_0}$

$$|\mathbf{1} - \mathbf{f}_{n_0}(\mathbf{1} - \mathbf{s})| - |\mathbf{M}^{n_0} \mathbf{s}| \stackrel{(3.1)}{\leq} |\mathbf{1} - \mathbf{f}_{n_0}(\mathbf{1} - \mathbf{s}) - \mathbf{M}^{n_0} \mathbf{s}| < |\mathbf{s}| \quad \text{gilt.}$$

Für reelle Zahlen a, b gilt allgemein, dass $|a| < b \iff a < b$ und $-a < b$ ist. Letzteres impliziert

$$-|\mathbf{1} - \mathbf{f}_{n_0}(\mathbf{1} - \mathbf{s})| + \overbrace{|\mathbf{M}^{n_0} \mathbf{s}|}^{> 2|\mathbf{s}| \text{ mit (ii)}} < |\mathbf{s}| \quad \Rightarrow |\mathbf{1} - \mathbf{f}_{n_0}(\mathbf{1} - \mathbf{s})| > |\mathbf{s}| \quad \text{für } |\mathbf{s}| < \varepsilon_{n_0}. \quad \square$$

Nun soll, wie angekündigt, der Satz 3.50 bewiesen werden.

Beweis des Satzes 3.50. Beachte zuerst, dass gemäß dem Satz 3.46 (ii) alle Zustände $\mathbf{r} \in \mathbb{N}_0^k \setminus \{\mathbf{0}\}$ transient sind. Insbesondere gilt also für den Prozess $\{\mathbf{Z}_n^{(j)}\}_{n \geq 0}$, dass $\mathbb{P}(\mathbf{Z}_n^{(j)} \rightarrow \mathbf{r}) = 0$ für alle $\mathbf{r} \neq \mathbf{0}$ ist (siehe (3.90)). Analog zum Satz 2.48 im einfachen Fall argumentiert folgt deshalb die *Instabilität* dieses Prozesses, d.h.

$$\mathbb{P}(\mathbf{Z}_n^{(j)} \rightarrow \mathbf{0}) + \mathbb{P}(\mathbf{Z}_n^{(j)} \rightarrow \infty) = 1. \quad (\Delta)$$

Seien nun zunächst die *Perron-Wurzel* $\rho \leq 1$ und $n \geq 1$ beliebig. Mit dem Satz 3.39 von *Perron-Frobenius* und dem Korollar dazu (verwende die dortige Notation) gilt, dass

$$(\mathbb{E} Z_{nl}^{(j)})_{j,l=1}^k \stackrel{(3.67)}{=} (\mu_{jl}^{(n)})_{j,l=1}^k = \mathbf{M}^n \stackrel{(3.86)}{=} \rho^n \mathbf{A} + \mathbf{B}^n \quad \text{ist.}$$

Dabei gibt es ein $m_0 \geq 1$, sodass die Elemente aller $\mathbf{B}^n$ mit $n \geq m_0$ und die von $\mathbf{A}$ generell folgendermaßen für alle Tupel $1 \leq j, l \leq k$ beschränkt sind (dies gilt für erstere wegen des Satzes 3.39 bzw. (3.85) und für letztere wegen (i) des anschließenden Korollars):

$$\begin{aligned} 0 \leq b_{jl}^{(n)} &\leq C \alpha^n \quad \text{mit } 0 < \alpha < \rho \leq 1 \text{ und } 0 < C < \infty \quad \text{sowie } 0 < a_{jl} < \infty \\ \Rightarrow \mathbb{E} Z_{nl}^{(j)} &\leq \overbrace{\rho^n}^{< 1} a_{jl} + C \overbrace{\alpha^n}^{< 1} < \overbrace{a_{jl} + C}^{< \infty \text{ (konstant)}} \end{aligned} \quad \text{für jedes } n \geq m_0.$$

Die Erwartungswerte $\mathbb{E} Z_{nl}^{(j)}$ sind also für alle $1 \leq j, l \leq k$ und alle $n \geq m_0$ beschränkt, was die Ereignisse $\{\mathbf{Z}_{nl}^{(j)} \rightarrow \infty\}$ unmöglich macht (die NR B.9 zeigt dies zusätzlich etwas formaler). Deshalb sind alle $\mathbb{P}(\mathbf{Z}_{nl}^{(j)} \rightarrow \infty) = 0$ und somit ist auch $\mathbb{P}(\mathbf{Z}_n^{(j)} \rightarrow \infty) = 0$. Daraus und mit der Instabilität (Δ) folgt für jedes $1 \leq j \leq k$

$$\eta^{(j)} \stackrel{(3.93)}{=} \mathbb{P}(\mathbf{Z}_n^{(j)} \rightarrow \mathbf{0}) = 1 \quad \stackrel{(3.96)}{\Rightarrow} \boldsymbol{\eta} = \mathbf{1}.$$

3 Galton-Watson-Verzweigungsprozesse mit mehreren unterschiedlichen Typen

Um den Punkt (ii) zu zeigen, sei für den Rest des Beweises $\rho > 1$. Beachte zunächst allgemein, dass aufgrund der positiven Regularität von $\mathbf{M}$ gemäß dem Lemma 3.51 (i) $\eta_n^{(j)} = f_n^{(j)}(\mathbf{0}) < 1$ für alle $n \geq 1$ und $1 \leq j \leq k$ gilt.

Prinzipiell wäre es aber dennoch denkbar, dass entweder alle oder zumindest einige der Grenzwerte $\eta^{(j)} = 1$ sind. Um zunächst Ersteres zu widerlegen, seien $n_0 \geq 1$ und $\varepsilon_{n_0} > 0$ so, wie sie im Lemma 3.51 (ii) bzw. (iii) gefunden wurden. Für nichtnegative $\mathbf{s} \neq \mathbf{0}$ gilt also wegen (iii) des besagten Lemmas $|\mathbf{1} - \mathbf{f}_{n_0}(\mathbf{1} - \mathbf{s})| > |\mathbf{s}|$, falls $|\mathbf{s}| < \varepsilon_{n_0}$ ist $(\star)$.

Weiters gilt $\mathbf{f}_n(\mathbf{0}) \stackrel{(3.95)}{=} \boldsymbol{\eta}_n \stackrel{(3.97)}{\rightarrow} \boldsymbol{\eta}$. Angenommen, es ist $\boldsymbol{\eta} = \mathbf{1}$. Dann muss es aufgrund der Konvergenz der $\boldsymbol{\eta}_n$ gegen $\mathbf{1}$ ein $n_1 \geq 1$ mit $|\mathbf{1} - \boldsymbol{\eta}_{n_1}| = |\mathbf{1} - \mathbf{f}_{n_1}(\mathbf{0})| < \varepsilon_{n_0}$ geben

$$\Rightarrow |\mathbf{1} - \mathbf{f}_{n_0+n_1}(\mathbf{0})| \stackrel{(3.48)}{=} |\mathbf{1} - \mathbf{f}_{n_0}(\mathbf{f}_{n_1}(\mathbf{0}))| = |\mathbf{1} - \mathbf{f}_{n_0}(\mathbf{1} - (\mathbf{1} - \mathbf{f}_{n_1}(\mathbf{0})))| \stackrel{(\star)}{>} |\mathbf{1} - \mathbf{f}_{n_1}(\mathbf{0})|.$$

Umgekehrt gilt aufgrund von $\mathbf{f}_{n_0+n_1}(\mathbf{0}) \stackrel{(3.95)}{=} \boldsymbol{\eta}_{n_0+n_1} \stackrel{(3.97)}{\geq} \boldsymbol{\eta}_{n_1} \stackrel{(3.95)}{=} \mathbf{f}_{n_1}(\mathbf{0})$ aber auch

$$|\mathbf{1} - \mathbf{f}_{n_0+n_1}(\mathbf{0})| \leq |\mathbf{1} - \mathbf{f}_{n_1}(\mathbf{0})|,$$

was einen Widerspruch darstellt. Es ist also nicht möglich, dass $\boldsymbol{\eta} = \mathbf{1}$ – d.h. $\eta^{(j)} = 1$ für alle $1 \leq j \leq k$ – gilt.

Nun soll noch abschließend ausgeschlossen werden, dass auch nur einzelne $\eta^{(j)} = 1$ sein können. Nehme hierfür an, dass manche (genauer $1 \leq t < k$) gleich 1 und andere (genauer $k-t$) kleiner 1 sind. O.B.d.A. können dann die Objekttypen so angeordnet werden, dass

$$\begin{aligned} \eta^{(1)} = \dots = \eta^{(t)} = 1 \quad \text{und} \quad \eta^{(t+1)} < 1, \dots, \eta^{(k)} < 1 \quad \text{gilt} \\ \stackrel{(3.96)}{\Rightarrow} \boldsymbol{\eta} = (1, \dots, 1, \eta^{(t+1)}, \dots, \eta^{(k)})^T. \end{aligned} \quad (\star\star)$$

Aufgrund der Fixpunktgleichung (3.98) des Lemmas 3.49 gilt diesfalls für alle $n \geq 1$

$$\begin{aligned} \boldsymbol{\eta} = \mathbf{f}(\boldsymbol{\eta}) \quad \stackrel{(3.98)}{\Rightarrow} \mathbf{f}_n(\boldsymbol{\eta}) = \boldsymbol{\eta} \quad \left(\stackrel{(3.39)}{\stackrel{(3.95)}}{=} (f_n^{(1)}(\boldsymbol{\eta}), \dots, f_n^{(k)}(\boldsymbol{\eta}))^T = (\eta^{(1)}, \dots, \eta^{(k)})^T \right) \\ \stackrel{(\star\star)}{\Rightarrow} f_n^{(j)}(1, \dots, 1, \eta^{(t+1)}, \dots, \eta^{(k)}) = \eta^{(j)} = 1 \stackrel{\text{Lemma 3.23 (i)}}{=} f_n^{(j)}(1, \dots, 1) \quad \text{für alle } 1 \leq j \leq t. \quad (\star\star\star) \end{aligned}$$

Beachte nun, dass alle $f_n^{(j)}$ in jedem Argument nicht-fallend sind (Satz 3.10 (a)), was insbesondere auch bzgl. aller s_l mit $t < l \leq k$ gilt. Daher impliziert $(\star\star\star)$ für die ersten t der $f_n^{(j)}$, dass sie konstant in (bzw. unabhängig von) allen s_l mit $t < l \leq k$ sein müssen – zumindest für $s_l \geq \eta^{(l)}$. Die entsprechenden partiellen Ableitungen bei $\mathbf{1}$ sind dann also

$$\mu_{jl}^{(n)} \stackrel{(3.61)}{=} \frac{\partial f_n^{(j)}}{\partial s_l}(\mathbf{1}) = 0 \quad \text{für alle } 1 \leq j \leq t, \quad t < l \leq k \text{ und } n \geq 1,$$

womit es kein $n \geq 1$ mit $\mathbf{M}^n \stackrel{(3.67)}{=} (\mu_{jl}^{(n)})_{j,l=1}^k > \mathbf{0}$ geben kann. Gemäß (3.81) konterkariert dies aber die positive Regularität von $\mathbf{M}$.

Der Widerspruch kann alternativ direkt dadurch eingesehen werden, dass es aufgrund der Def. 3.34 (ii) der positiven Regularität ein $\tilde{n} \geq 1$ gibt, für welches $\mathbb{P}(Z_{\tilde{n}l}^{(j)} \geq 1) > 0$ – also auch $\mathbb{P}(Z_{\tilde{n}l}^{(j)} = 0) < 1$ – für jedes $1 \leq j, l \leq k$ gilt. Somit ist (wieder gemäß dem Satz 3.10 (a)) jedes $f_{\tilde{n}}^{(j)}$ strikt isoton bzgl. aller s_l – insbesondere auch für $1 \leq j \leq t$ und $t < l \leq k$. Dies widerspricht $(\star\star\star)$ für $\tilde{n}$.

Zusammengenommen muss jedes $\eta^{(j)} < 1$ und somit $\boldsymbol{\eta} < \mathbf{1}$ sein, wenn $\rho > 1$ ist. $\square$

4 Anwendungsbeispiele

Bis hierhin war die Theorie sowohl bzgl. der einfachen als auch bzgl. der Multitype GWP innerhalb der modellhaften Annahmen allgemein gehalten. An fast keiner Stelle wurden spezifische Nachkommens- oder Reproduktionsverteilungen unterstellt.

Dieses Kapitel konzentriert sich nun darauf, das Erarbeitete auf spezifische Prozesse oder Prozessklassen anzuwenden. Wiewohl es zwar etliche Einsatzbereiche aus der realen Welt gäbe, bleibt es dabei aber weiterhin auf der theoretischen Ebene. Gewisse Aspekte sollen mathematisch vertieft, andere rechnerunterstützt weiter analysiert werden. Teilweise werden hierzu Datensätze anhand konkreter Verteilungen simuliert, um gewisse Zusammenhänge zu veranschaulichen. Alle Analysen wurden durchwegs mit der Statistik-Software R (Version 4.2.2 – vgl. [55]) auf einem *Windows 10 x64*-Computer (*build 19044*) durchgeführt.

Der erste Abschnitt 4.1 widmet sich – wie das 2. Kapitel – GW-Prozessen mit nur einem Objekttyp (oder nicht unterscheidbaren Individuen). Die Betrachtung konzentriert sich dabei auf die mathematische Ebene und meist wird der sogenannte *geometrische Fall* untersucht (siehe die Def. 4.2). Teilweise ist das Erarbeitete aber auch allgemeiner gültig.

Ab dem zweiten Abschnitt 4.2 werden wieder die Multitype GW-Prozesse des 3. Kapitels behandelt. Im Fokus steht hier die softwaregestützte Analyse. Dafür werden passende R-Funktionen bereitgestellt, dokumentiert und deren Funktionsweise im Anschluss anhand eines spezifischen numerischen Beispiels vorgezeigt.

4.1 Einfache GW-Verzweigungsprozesse – geometrischer Fall

Für die exemplarische Untersuchung einfacher GWP wurden solche mit einer *geometrischen* Nachkommensverteilung auserkoren. Diese haben die praktikable Eigenschaft, dass vieles explizit berechnet werden kann (vgl. z.B. [47, 0.9, S. 10 ff.]). Insofern ist die Begründung für die Wahl dieser Prozessklasse hauptsächlich mathematischer Natur.

Direkte numerische Kalkulationen wurden mit R vollzogen. Für weitere Berechnungen und Simulationen kamen die in 4.2 vorgestellten R-Funktionen zum Einsatz, welche auch auf den Spezialfall der einfachen GWP angewendet werden können – siehe die diesbezüglichen Erläuterungen. Spezifische Anwendungshinweise finden sich unter 4.2.3.

Nach der Definition des geometrischen GWP werden seine grundlegenden Eigenschaften untersucht (4.1.1). Der Unterabschnitt 4.1.2 ist der Asymptotik *superkritischer* geometrischer Fälle gewidmet. Anschließend werden die Begriffe des *Aussterbezeitpunkts* und der *Gesamtpopulation* allgemein eingeführt (4.1.3), wobei letzterer eingehender im Fall geometrischer GWP behandelt wird (4.1.4). Der abschließend unter 4.1.5 angerissene bedingte Prozess hat wieder hauptsächlich im superkritischen Fall eine Relevanz.

4.1.1 Definition und allgemeine Analyse des geometrischen GWP

Zunächst werden die wichtigsten Eigenschaften geometrisch verteilter ZV generell zusammengefasst dargestellt sowie die diesbezüglich verwendete Notation festgelegt.

Definition 4.1 (Geometrisch Verteilte ZV – vgl. z.B. [20, 23.2, S. 180 f.]

Schreibe für eine $\mathbb{N}_0$ -wertige ZV X , welche *geometrisch verteilt* mit dem Parameter $0 < p < 1$ bzw. (äquivalent) mit den Parametern

$$0 < p < 1 \text{ und } q := 1 - p \text{ ist,} \quad \text{kurz } X \sim \text{Geom}(p) \text{ bzw. } X \sim \text{Geom}(p, q). \quad (4.1)$$

Damit ist gemeint, dass ihre Verteilung bzw. Wahrscheinlichkeitsfunktion mit

$$p_k \stackrel{(2.1)}{=} \mathbb{P}(X = k) = p(1 - p)^k \stackrel{(4.1)}{=} pq^k \quad \text{für alle } k \geq 0 \quad (4.2)$$

gegeben ist. Für ihren Erwartungswert bzw. ihre Varianz gilt

$$\mu := \mathbb{E}X = \frac{1 - p}{p} \stackrel{(4.1)}{=} \frac{q}{p} \quad \text{bzw.} \quad \sigma^2 := \mathbb{V}X = \frac{1 - p}{p^2} \stackrel{(4.1)}{=} \frac{q}{p^2}. \quad (4.3)$$

Umgekehrt können p und q auch wie folgt durch μ dargestellt werden ($p + q = 1$):

$$p = \frac{1}{\frac{1}{p}} = \frac{1}{\frac{p+q}{p}} = \frac{1}{1 + \frac{q}{p}} \stackrel{(4.3)}{=} \frac{1}{1 + \mu} \quad \Rightarrow \quad q \stackrel{(4.1)}{=} 1 - p = 1 - \frac{1}{1 + \mu} = \frac{\mu}{1 + \mu}. \quad (4.4)$$

Die weF von X kann für $s \in [0, 1]$ angegeben werden als

$$\begin{aligned} f(s) &\stackrel{(2.45)}{:=} f_X(s) \stackrel{(2.6)}{=} \sum_{k \geq 0} p_k s^k \stackrel{(4.2)}{=} \sum_{k \geq 0} pq^k s^k = p \sum_{k \geq 0} \underbrace{\left(\overbrace{q}^{\in [0,1]} \underbrace{s}_{\leq 1} \right)}_{< 1}^k \stackrel{\text{Reihe}}{\stackrel{\text{geom.}}{=}} \frac{p}{1 - qs} \\ &\stackrel{(4.1)}{\stackrel{1=p+q}{=}} f(s) = \frac{p}{p + q - qs} = \frac{1}{1 + \frac{q}{p} - \frac{q}{p}s} \stackrel{(4.3)}{=} \frac{1}{1 + \mu - \mu s}. \end{aligned} \quad (4.5)$$

$$\text{Falls } \mu = 1 \text{ ist, vereinfacht sich die weF } f \text{ zu} \quad f(s) = \frac{1}{2 - s}. \quad (4.6)$$

Bemerkung zur Definition 4.1. σ^2 lässt sich auch nur durch μ darstellen und gemäß der Definition 2.32 können μ und σ^2 genauso gut aus der weF f in (4.5) erhalten werden.

$$\begin{aligned} f'(s) &= \frac{pq}{(1 - qs)^2} \quad \Rightarrow \quad \mu \stackrel{(2.53)}{=} f'(1) = \frac{pq}{(1 - q)^2} \stackrel{(4.1)}{=} \frac{pq}{p^2} = \frac{q}{p} \quad \text{und} \quad (*) \\ f''(s) &= \frac{2pq^2}{(1 - qs)^3} \quad \Rightarrow \quad f''(1) = \frac{2pq^2}{(1 - q)^3} \stackrel{(4.1)}{=} \frac{2pq^2}{p^3} = 2 \frac{q^2}{p^2} \\ \Rightarrow \quad \sigma^2 &\stackrel{(2.54)}{=} f''(1) + \mu - \mu^2 \stackrel{(*)}{=} 2 \frac{q^2}{p^2} + \frac{q}{p} - \frac{q^2}{p^2} = \frac{q^2 + qp}{p^2} \stackrel{(4.1)}{=} \frac{q^2 + q(1 - q)}{p^2} = \frac{q}{p^2} \quad (4.7) \\ &= \left(\frac{q}{p} \right)^2 + \frac{q}{p} \stackrel{(*)}{=} \mu^2 + \mu. \quad \triangle \end{aligned}$$

Definition 4.2 (GWP mit geometrischer Nachkommensverteilung)

Sei $0 < p < 1$. Der GWP $\{Z_n\}_{n \geq 0}$ mit initialer Populationsgröße $Z_0 \equiv 1$ und Nachkommenszahl $X \sim \text{Geom}(p)$ wird als *geometrischer Galton-Watson-Verzweigungsprozess* bezeichnet. Oft wird auch nur kurz vom *geometrischen Fall* gesprochen.

Klarerweise gelten für diesen Prozess alle Beziehungen der Definition 4.1 und insbesondere folgt aus (4.3) und mit dem Satz 2.44:

$$\begin{array}{lll} \mu < 1 \iff p > \frac{1}{2}, & \mu = 1 \iff p = \frac{1}{2}, & \mu > 1 \iff p < \frac{1}{2}. \end{array} \quad (4.8)$$

subkritischer Fall kritischer Fall superkritischer Fall

Der geometrische Fall erfüllt die Annahmen 2.39 (welche somit – auch allgemein – für diesen Abschnitt immer gelten sollen). Insbesondere handelt es sich um einen nichttrivialen Fall gemäß der Definition 2.40.

Bemerkung 4.3 (Gebrochen linearer Fall – vgl. z.B. [3, 4., S.6 f.])

Es handelt sich beim geometrischen Fall um einen Spezialfall des sogenannten *gebrochen linearen Falls*. Bei jenem ist die W-Funktion der Nachkommensverteilung $\{p_k\}_{k \geq 0}$ für $0 < b, q < 1$ mit $b \leq 1 - q =: p$ ($\star$) definiert durch

$$p_k := bq^{k-1} \text{ für } k \geq 1 \text{ und} \quad p_0 := 1 - \sum_{k \geq 1} p_k. \quad (\star\star)$$

Die Namensgebung rührt daher, dass alle Glieder der Folge $\{f_n\}_{n \geq 0}$ als gebrochen lineare Funktionen dargestellt werden können. Es handelt sich sogar um die einzige nichttriviale GW-Prozessklasse, für welche bisher solche explizite Lösungen gefunden werden konnten (Stand 2015 – vgl. etwa [29, 3.1.4, S. 40]).

Da die WK keine Nachkommen zu gebären nicht auf $p = 1 - q$ festgelegt ist, wird die obige Nachkommensverteilung z.B. in [18, 2.2., S. 16] als *zero-modified*-geometrische Verteilung bezeichnet. Dort wird dieser attestiert, etwa manche biologische Gegebenheiten eher abzubilden als die geometrische Verteilung (vgl. etwa den Hinweis auf eine genealogische Applikation in [19, S. 10]). Der Einfachheit halber wird in dieser Arbeit aber nur letztere behandelt.

Aus $(\star\star)$ kann explizit $p_0 = 1 - \sum_{k \geq 1} p_k = 1 - b \sum_{k \geq 0} q^k \stackrel{\text{geom. Reihe}}{=} 1 - \frac{b}{1-q} \stackrel{(\star)}{=} \frac{p-b}{p} =: r$ mit $0 < r < 1$ ermittelt werden. Daraus folgt umgekehrt $b = p - rp = (1-r)p$, womit die

W-Funktion als $p_k \stackrel{(\star\star)}{=} \begin{cases} (1-r)pq^{k-1} & \text{für } k \geq 1, \\ r & \text{für } k = 0 \end{cases}$ geschrieben werden kann. Wird nun

$r = p = 1 - q$ gewählt, erhält man genau die geometrische Verteilung (siehe (4.2)). Daran ist erkennbar, dass der gebrochen lineare Fall eine Verallgemeinerung des geometrischen Falls darstellt. $\triangle$

Im vorliegenden Abschnitt ist mit $\{Z_n\}_{n \geq 0}$ meist der geometrische Fall gemeint. Wenn etwas allgemeiner gedacht ist, wird darauf hingewiesen (dies gilt insbesondere für die beiden Exkurse im Unterabschnitt 4.1.3 und am Anfang von 4.1.5). Aber auch dann gilt der in 2.4 auf der Seite 32 eingeführten Notation entsprechend $Z_0 \equiv 1$.

Verteilung der Populationsgrößen Z_n

Sei also ein Galton-Watson-Prozess $\{Z_n\}_{n \geq 0}$ mit geometrischer Nachkommensverteilung gegeben. Wie in der vorherigen Bemerkung 4.3 angekündigt, können explizite Ausdrücke für die iterierten wahrscheinlichkeitserzeugenden Funktionen gefunden werden.

Lemma 4.4 (WeF der Populationsgrößen im geometrischen Fall – angelehnt an [17, 5.4 (3), S. 172])

Für die weF der Populationsgrößen gilt für alle $n \geq 0$ und alle $s \in [0, 1]$

$$f_n(s) \stackrel{(2.49)}{:=} f_{Z_n}(s) = \begin{cases} \frac{\frac{n-(n-1)s}{n+1-ns}}{p(q^n - p^n - qs(q^{n-1} - p^{n-1}))} & \text{für } p = q = \frac{1}{2} \quad \text{bzw. } \mu = 1, \\ \frac{\frac{\mu^n - 1 - \mu s(\mu^{n-1} - 1)}{\mu^{n+1} - 1 - \mu s(\mu^n - 1)}}{q^{n+1} - p^{n+1} - qs(q^n - p^n)} & \text{für } p \neq q \quad \text{bzw. } \mu \neq 1. \end{cases} \quad (4.9)$$

Beweis. Es werden beide Fälle in (4.9) getrennt behandelt und jeweils mit Induktion über n vorgegangen. Die Überprüfung der ersten Zeile von (2.30) der Definition 2.18 stellt in beiden Fällen den Induktionsanfang dar und die zweite Zeile kann für den Induktionsschritt herangezogen werden.

Fall 1: $p = q = \frac{1}{2}$. Wird (4.9) für $n = 0$ bzw. $n = 1$ berechnet, ergibt sich

$$\begin{aligned} f_0(s) &\stackrel{(4.9)}{=} \frac{0 - (0-1)s}{1-0} = s & \text{und} \\ f_1(s) &\stackrel{(4.9)}{=} \frac{1-0 \cdot s}{2-s} = \frac{1}{2-s} = \frac{\frac{1}{2}}{1-\frac{1}{2}s} \stackrel{p=q=\frac{1}{2}}{=} \frac{p}{1-qs} \stackrel{(4.5)}{=} f(s), \end{aligned} \quad (\star)$$

womit die erste Zeile von (2.30) erfüllt ist. Werden mit der zweiten Zeile die weF weiterer Generationen ermittelt, etwa $f_2(s) \stackrel{(2.30)}{=} f(f(s)) \stackrel{(\star)}{=} \frac{1}{2-\frac{1}{2-s}} = \frac{2-s}{3-2s}$ und $f_3(s) \stackrel{(2.30)}{=} f_2(f(s)) \stackrel{(\star)}{=} \frac{1}{2-\frac{2-s}{3-2s}} = \frac{3-2s}{4-3s}$, liegt die oben stehende Form bereits nahe. Um dies auch zu beweisen, berechne für den Induktionsschritt mit $n \geq 1$

$$f_n(s) \stackrel{(2.30)}{=} f(f_{n-1}(s)) \stackrel{(\star)}{=} \frac{1}{2 - \frac{n-1-(n-2)s}{n-(n-1)s}} = \frac{n - (n-1)s}{\underbrace{2n - 2ns + 2s - n + 1 + ns - 2s}_{=n+1-ns}}.$$

Fall 2: $p \neq q$. Beachte zunächst, dass für alle $n \geq 0$

$$q^2 - p^2 = (q-p)(q+p) \stackrel{(4.1)}{=} (q-p)(1-p+p) = q-p \quad \text{sowie} \quad (\star\star)$$

$$q^n - p^n - \underbrace{q^n p}_{=1-q} + \underbrace{q p^n}_{=1-p} = q^n - p^n - q^n + q^{n+1} + p^n - p^{n+1} = q^{n+1} - p^{n+1} \quad (\star\star\star)$$

und (wie oben) $\underbrace{q^n p}_{=1-q} = \underbrace{q p^n}_{=1-p}$ mit (4.1)

$$\begin{aligned} f_0(s) &\stackrel{(4.9)}{=} \frac{p(1-1-qs(q^{-1}-p^{-1}))}{q-p-qs(1-1)} = \frac{-ps+qs}{q-p} = \frac{(q-p)s}{q-p} = s \quad \text{bzw.} \\ f_1(s) &\stackrel{(4.9)}{=} \frac{p(q-p-qs(1-1))}{q^2-p^2-qs(q-p)} \stackrel{(\star\star)}{=} \frac{p(q-p)}{(q-p)-qs(q-p)} = \frac{p}{1-qs} \stackrel{(4.5)}{=} f(s) \end{aligned}$$

4.1 Einfache GW-Verzweigungsprozesse – geometrischer Fall

gilt (also wieder der erste Teil von (2.30)). Für den Induktionsschritt bestimme ($n \geq 1$)

$$\begin{aligned}
 f_n(s) &\stackrel{(2.30)}{=} f(f_{n-1}(s)) \stackrel{(4.5)}{=} \frac{p}{1 - qf_{n-1}(s)} \stackrel{(4.9)}{=} \frac{p}{1 - q \frac{p(q^{n-1} - p^{n-1} - qs(q^{n-2} - p^{n-2}))}{q^n - p^n - qs(q^{n-1} - p^{n-1})}} \\
 &= \frac{p(q^n - p^n - qs(q^{n-1} - p^{n-1}))}{q^n - p^n - qs(q^{n-1} - p^{n-1}) - qp(q^{n-1} - p^{n-1} - qs(q^{n-2} - p^{n-2}))} \\
 &= \frac{p(q^n - p^n - qs(q^{n-1} - p^{n-1}))}{\underbrace{q^n - p^n - q^n p + qp^n}_{=q^{n+1} - p^{n+1}} \underbrace{- qs(q^{n-1} - p^{n-1} - q^{n-1}p + qp^{n-1})}_{=q^n - p^n}},
 \end{aligned}$$

womit der erste Ausdruck für diesen Fall in (4.9) gezeigt ist. Für den zweiten verwende (4.3), um f_n in Abhängigkeit von $\mu = \frac{q}{p}$ (und n) darzustellen. Es gilt für $p \neq q$ und alle $n \geq 0$

$$\begin{aligned}
 f_n(s) &\stackrel{(4.9)}{=} \frac{p(q^n - p^n - qs(q^{n-1} - p^{n-1}))}{q^{n+1} - p^{n+1} - qs(q^n - p^n)} \cdot \frac{1}{p^{n+1}} = \frac{\left(\frac{q}{p}\right)^n - 1 - \frac{q}{p}s\left(\left(\frac{q}{p}\right)^{n-1} - 1\right)}{\left(\frac{q}{p}\right)^{n+1} - 1 - \frac{q}{p}s\left(\left(\frac{q}{p}\right)^n - 1\right)} \\
 &= \frac{\mu^n - 1 - \mu s(\mu^{n-1} - 1)}{\mu^{n+1} - 1 - \mu s(\mu^n - 1)} \quad \text{wegen } \mu^n \stackrel{(4.3)}{=} \left(\frac{q}{p}\right)^n. \quad \square
 \end{aligned}$$

Mit den Resultaten des 2. Kapitels können die Verteilung und die Momente der Populationsgrößen bestimmt werden. Die Tabelle 4.1 zeigt dies beispielhaft mit $\mu = 1,1$ für die W-Funktion von Z_5 , also gewonnen aus f_5 (die p_k mit $0 \leq k \leq 11$ beinhalten 95,47 % der Wahrscheinlichkeitsmasse). Das Korollar 2.8 liefert $\mathbb{E}Z_5 = 1,611$ und $\mathbb{V}Z_5 = 20,648$.

p_0	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	p_{11}
0,7913	0,0271	0,0235	0,0205	0,0178	0,0155	0,0135	0,0118	0,0102	0,0089	0,0078	0,0068

Tabelle 4.1: W-Funktion $\{\mathbb{P}(Z_5 = k)\}_{k \geq 0}$ bis $k = 11$ im superkritischen geometrischen Fall mit $\mu = 1,1$ – berechnet gemäß dem Satz 2.6 bzw. (2.10).

n	1	2	3	5	10	20	30	50	100
$\mathbb{E}Z_n$ berechnet*	1,10	1,21	1,33	1,61	2,59	6,73	17,45	117,39	13 780,61
$\mathbb{E}Z_n$ simuliert**	1,09	1,20	1,32	1,64	2,65	6,86	17,74	116,78	13 683,08
$\mathbb{V}Z_n$ berechnet*	2,31	5,34	9,25	20,65	86,81	809,17	6 027,68	286 927,65	3 987 721 412,81
$\mathbb{V}Z_n$ simuliert**	2,28	5,30	9,07	21,81	91,28	829,71	6 124,28	282 438,91	3 913 938 196,20

Tabelle 4.2: Die ersten beiden Momente der Populationsgröße Z_n im superkritischen geometrischen Fall mit $\mu = 1,1$ für ausgewählte Generationen (*Berechnung mit (2.56) bzw. (4.10); **MW bzw. s^2 bei 20 000 Simulationen).

Die aus der weF gewonnen Momente können aber direkter mit dem Satz 2.33 bestimmt werden. Den Erwartungswert $\mathbb{E}Z_n = \mu^n$ erhält man unmittelbar mit (2.56). Soll die Varianz nur durch μ ausgedrückt werden, kann

$$\mathbb{V}Z_n \stackrel{(2.57)}{=} \begin{cases} \frac{\sigma^2 \mu^n (\mu^n - 1)}{\mu(\mu - 1)} \stackrel{(4.7)}{=} \frac{\mu(\mu + 1) \mu^n (\mu^n - 1)}{\mu(\mu - 1)} = \frac{\mu + 1}{\mu - 1} \mu^n (\mu^n - 1) & \text{für } \mu \neq 1, \\ n\sigma^2 \stackrel{(4.7)}{=} n(\mu^2 + \mu) & \text{für } \mu = 1 \end{cases} \quad (4.10)$$

4 Anwendungsbeispiele

geschrieben werden. Die Tabelle 4.2 zeigt die so berechneten Momente für ausgewählte Generationen (wieder für $\mu = 1,1$) und stellt ihnen die entsprechenden Mittelwerte und (korrigierten) Stichprobenvarianzen aus simulierten Daten gegenüber.

Bestimmung der Aussterbewahrscheinlichkeit

Für die Bestimmung der Aussterbewahrscheinlichkeit η im geometrischen Fall kann gemäß dem Lemma 2.43 der kleinste Fixpunkt von f in $[0, 1]$ bestimmt werden. Im Fall $\mu = 1$ wird die Fixpunktgleichung zu $\eta = f(\eta) = \frac{1}{2-\eta}$. Dies ist gleichbedeutend mit $(\eta - 1)^2 = 0$, womit $\eta = 1$ die einzige Lösung ist.

Für $\mu \neq 1$ muss $f(\eta) = \frac{1}{1+\mu-\mu\eta} = \eta \iff \eta^2 - \eta\frac{1+\mu}{\mu} + \frac{1}{\mu} = 0$ erfüllt sein. Die kleinere der beiden Lösungen ist $\eta = \frac{1+\mu}{2\mu} - \frac{1}{2\mu}\sqrt{(1+\mu)^2 - 4\mu} = \frac{1+\mu-|1-\mu|}{2\mu} \begin{cases} \frac{2\mu}{2\mu} = 1 & \text{für } \mu < 1, \\ \frac{2}{2\mu} = \frac{1}{\mu} & \text{für } \mu > 1. \end{cases}$

Zusammengenommen gilt im geometrischen Fall für die AWK

$$\eta = \begin{cases} 1 & \text{für } p \geq \frac{1}{2} \quad \text{bzw. } \mu \leq 1, \\ \frac{p}{q} = \frac{1}{\mu} & \text{für } p < \frac{1}{2} \quad \text{bzw. } \mu > 1. \end{cases} \quad (4.11)$$

Dieses Resultat deckt sich mit der Aussage des Satzes 2.44.

Da die f_n bekannt sind, bietet der Satz 2.46 eine alternative Möglichkeit zur Bestimmung der Aussterbewahrscheinlichkeit. Gemäß ihm konvergiert die Folge $\{f_n\}_{n \geq 0}$ für *alle* $s \in [0, 1)$ gegen η (für $s = 0$ wurde dies sogar bereits in der Definition 2.42 der AWK festgestellt). Tatsächlich gilt im geometrischen Fall für alle $s \in [0, 1)$

$$f_n(s) \underset{(4.9)}{=} \begin{cases} \frac{n-(n-1)s}{n+1-ns} = \frac{1-(1-\frac{1}{n})s}{1+\frac{1}{n}-s} \xrightarrow{n \rightarrow \infty} \frac{1-s}{1-s} = 1 & \text{für } \mu = 1, \\ \frac{\mu^n - 1 - \mu s(\mu^{n-1} - 1)}{\mu^{n+1} - 1 - \mu s(\mu^n - 1)} \xrightarrow{n \rightarrow \infty} \frac{-1 + \mu s}{-1 + \mu s} = 1 & \text{für } \mu < 1, \\ \frac{1 - \frac{1}{\mu^n} - s + \frac{s}{\mu^{n-1}}}{\mu - \frac{1}{\mu^n} - \mu s + \frac{s}{\mu^{n-1}}} \xrightarrow{n \rightarrow \infty} \frac{1-s}{\mu(1-s)} = \frac{1}{\mu} & \text{für } \mu > 1. \end{cases}$$

In der Abbildung 4.1 ist die Konvergenz der $\{f_n\}_{n \geq 0}$ für verschiedene μ anhand von f_0 bis f_7 sowie f_{10} , f_{13} , f_{16} und f_{20} veranschaulicht (in einem Farbspektrum von **rot** bis **blau**). Im subkritischen Fall mit $\mu = \frac{2}{3}$ und im kritischen Fall mit $\mu = 1$ konvergiert die Funktionenfolge überall monoton wachsend gegen 1. In letzterem Fall ist die Konvergenzgeschwindigkeit sehr langsam, weshalb zusätzlich f_{100} eingezeichnet ist. Dies ist auch anhand der Tabelle 4.3 nachvollziehbar. Hier sind die genauen Werte an den in der Abb. markierten Stellen ablesbar (für die selben Szenarien und einen längerfristigen Fokus).

Im superkritischen Fall mit $\mu = 2$ (unteres Bild bzw. die letzten beiden Zeilen) erkennt man gut, dass die Konvergenz gegen $\eta = \frac{1}{2}$ für $s < \eta$ monoton wachsend und für $s > \eta$ monoton fallend ist (was mit den Aussagen (ii) und (iii) des Satzes 2.46 konform geht).

Durch das Vorliegen der expliziten Formen der f_n kann auch der Punkt (i) des Lemmas 2.19 (u. tw. das Korollar 2.47) überprüft werden. Dieser besagt, dass η ein Fixpunkt aller f_n mit $n \geq 0$ sein muss. Für $\mu = 1$ sieht man dies anhand (4.9) sofort, kann aber auch sonst leicht eingesehen werden. Ist $\mu < 1$, so gilt $f_n(1) = \frac{\mu^n - 1 - \mu(\mu^{n-1} - 1)}{\mu^{n+1} - 1 - \mu(\mu^n - 1)} = \frac{\mu - 1}{\mu - 1} = 1 = \eta$ (dies entspricht auch der Aussage in (2.8)). Für $\mu > 1$ gilt $f_n(\frac{1}{\mu}) = \frac{\mu^n - \mu^{n-1}}{\mu(\mu^n - \mu^{n-1})} = \frac{1}{\mu} < 1$.

4.1 Einfache GW-Verzweigungsprozesse – geometrischer Fall

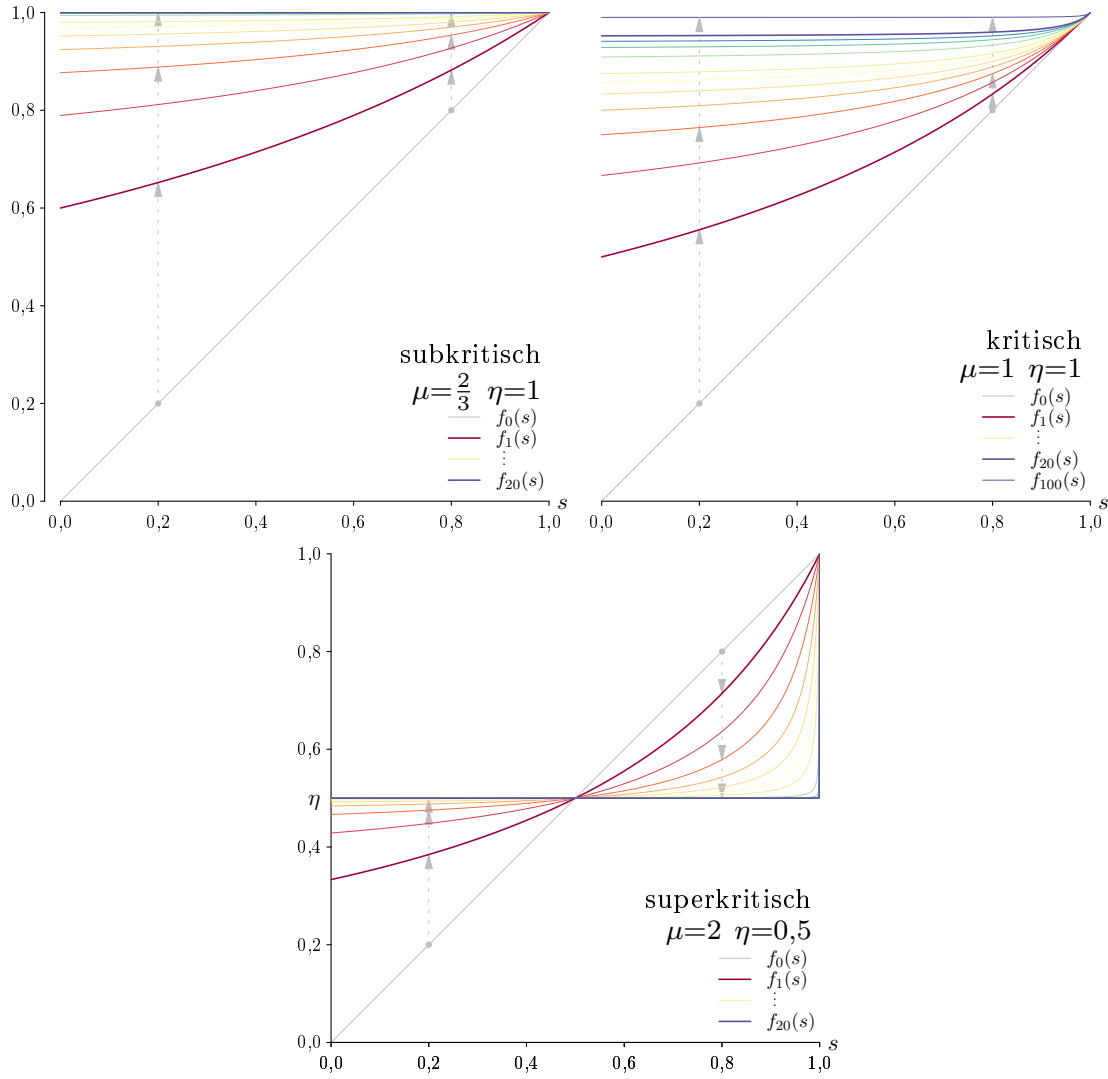


Abbildung 4.1: Konvergenz der Folge $\{f_n\}_{n \geq 0}$ für $\mu \in \{\frac{2}{3}, 1, 2\}$. Bei $s \in \{0.2; 0.8\}$ ist zusätzlich die Richtung der Konvergenz eingezeichnet (Berechnungen gemäß (4.9)).

μ	$n =$	0	1	3	5	7	10	13	20	50	100	500	1000
$\frac{2}{3}$	$f_n(0.2)$	0.2	0.652	0.889	0.956	0.981	0.995	0.998	1.000	1.000	1.000	1.000	1.000
	$f_n(0.8)$	0.8	0.882	0.954	0.980	0.991	0.998	0.999	1.000	1.000	1.000	1.000	1.000
1	$f_n(0.2)$	0.2	0.556	0.765	0.840	0.879	0.911	0.930	0.953	0.980	0.990	0.998	0.999
	$f_n(0.8)$	0.8	0.833	0.875	0.900	0.917	0.933	0.944	0.960	0.982	0.990	0.998	0.999
2	$f_n(0.2)$	0.2	0.385	0.475	0.494	0.499	0.500	0.500	0.500	0.500	0.500	0.500	0.500
	$f_n(0.8)$	0.8	0.714	0.579	0.522	0.506	0.501	0.500	0.500	0.500	0.500	0.500	0.500

Tabelle 4.3: Konvergenz der Folge $\{f_n\}_{n \geq 0}$ für $\mu \in \{\frac{2}{3}, 1, 2\}$ und ausgewählte n an den Stellen $s \in \{0.2; 0.8\}$ (Berechnungen gemäß (4.9)).

4.1.2 Asymptotisches Verhalten superkritischer geometrischer GWP

Im superkritischen Fall wachsen die Populationsgrößen bekanntermaßen mit der WK $1 - \eta$ über alle Schranken (siehe (2.70)). Insbesondere dann lohnt sich aber die Betrachtung ihrer Normalisierungen $\{W_n\}_{n \geq 0}$ (vgl. die Def. 2.36) und deren Grenzwert W (vgl. den Satz 2.50). Daraus können dann Rückschlüsse über das langfristige Wachstum der $\{Z_n\}_{n \geq 0}$ gezogen werden (was etwa im Beispiel 4.9 besprochen wird).

Im Abschnitt 2.6 wurden diverse Aussagen über die Verteilung von W getroffen. Diese kann man sich insbesondere auch für den geometrischen Fall zunutze machen.

Sei also für diesen Unterabschnitt stets $\{Z_n\}_{n \geq 0}$ ein geometrischer GWP mit $\mu > 1$ (bzw. äquivalent dazu $p < \frac{1}{2}$ – vgl. (4.8)). Da im geometrischen Fall (wegen $p > 0$ (4.1)) immer $\sigma^2 = \frac{q}{(4.3)p^2} < \infty$ gilt, besagt der Satz 2.53 (siehe auch den anschließenden Text auf der S. 52)

$$\begin{aligned} \{Z_n \rightarrow 0\} &= \{W = 0\} = \{W_n \rightarrow 0\} && \text{bzw.} \\ \Rightarrow \mathbb{P}(W = 0) &= \mathbb{P}(Z_n \rightarrow 0) = \mathbb{P}(W_n \rightarrow 0) = \eta \in (0, 1) && \text{und} \end{aligned} \quad (4.12)$$

$$\begin{aligned} \{W > 0\} &= \{Z_n \rightarrow \infty\} && \text{bzw.} \\ \Rightarrow \mathbb{P}(W > 0) &= \mathbb{P}(Z_n \rightarrow \infty) = 1 - \eta \in (0, 1). \end{aligned} \quad (4.13)$$

Betrachte nun $\{\mathcal{L}_n\}_{n \geq 0}$, $\mathcal{L}$ und $\mathcal{L}_{W|W>0}$ – also die *Laplace-Transformierten* von $\{W_n\}_{n \geq 0}$, W und $(W|W > 0)$ (vgl. die Def. 2.54 bzw. (2.75) und (A.96)). Zwischen den beiden letzteren Transformationen besteht der folgende Zusammenhang.

Lemma 4.5. *Es gilt*

$$\mathcal{L} = \eta + (1 - \eta) \mathcal{L}_{W|W>0} \quad \text{und somit auch} \quad \mathcal{L}_{W|W>0} = \frac{\mathcal{L} - \eta}{1 - \eta}. \quad (4.14)$$

Beweis. Dies ist (W ist nichtnegativ) eine direkte Anwendung der NR B.10. Es gilt

$$\mathcal{L}(u) \stackrel{(2.75)}{=} \mathbb{E}[e^{-uW}] \stackrel{(B.4)}{=} \underbrace{\mathbb{P}(W = 0)}_{=\eta \text{ (4.12)}} + \underbrace{\mathbb{P}(W > 0)}_{=1-\eta \text{ (4.13)}} \underbrace{\mathbb{E}[e^{-uW}|W > 0]}_{=\mathcal{L}_{W|W>0} \text{ (A.96)}} \quad \text{für alle } u \in \mathbb{R}_+.$$

Da $\eta < 1$ ist, folgt die zweite Aussage direkt (oder mit (B.5), wieder aus der NR B.10). $\square$

Es sollen nun explizite Formen für $\mathcal{L}$ bzw. $\mathcal{L}_{W|W>0}$ gefunden werden. Prinzipiell könnte man versuchen, die im Satz 2.56 ermittelte *Funktionalgleichung* (2.78) zu lösen. Da die $\{f_n\}_{n \geq 0}$ im geometrischen Fall aber ebenso explizit vorliegen, ist es wesentlich einfacher, mit dem Ergebnis des Satzes 2.55, also der Konvergenz der $\{\mathcal{L}_n\}_{n \geq 0}$ zu arbeiten. Dies führt zum nächsten Resultat.

Satz 4.6 (Laplace-Transformation von W – vgl. z.B. [18, (6.9), S. 155])

Für die Laplace-Transformierte von W gilt

$$\mathcal{L}(u) = \frac{u + \mu - 1}{\mu u + \mu - 1} = \frac{1 - \eta + \eta u}{1 - \eta + u} = \eta + (1 - \eta) \overbrace{\frac{1 - \eta}{1 - \eta + u}}^{=\mathcal{L}_{W|W>0} \text{ (siehe (4.14))}} \quad \text{für alle } u \in \mathbb{R}_+. \quad (4.15)$$

4.1 Einfache GW-Verzweigungsprozesse – geometrischer Fall

Beweis. Verwende für jedes $n \geq 0$, dass $\mathcal{L}_n(u) \stackrel{(2.79)}{=} f_n(e^{-u\mu^{-n}})$ $(\star)$ für alle $u \in \mathbb{R}_+$ und

$$\begin{aligned} f_n(s) &\stackrel{(4.9)}{=} \frac{\mu^n - 1 - \mu s(\mu^{n-1} - 1)}{\mu^{n+1} - 1 - \mu s(\mu^n - 1)} = \frac{\mu^n(1-s) + \mu s - 1}{\mu\mu^n(1-s) + \mu s - 1} \quad \text{für alle } s \in [0, 1] \text{ gilt} \quad (4.16) \\ \Rightarrow \mathcal{L}_n(u) &\stackrel{(\star)}{=} \frac{\mu^n(1-e^{-u\mu^{-n}}) + \mu e^{-u\mu^{-n}} - 1}{\underbrace{\mu\mu^n(1-e^{-u\mu^{-n}})}_{\rightarrow u \text{ ((B.7) aus der NR B.11)}} + \underbrace{\mu e^{-u\mu^{-n}} - 1}_{\rightarrow 1 \text{ (da } \mu > 1 \text{ ist)}} \xrightarrow{n \rightarrow \infty} \frac{u + \mu - 1}{\mu u + \mu - 1} \stackrel{(2.77)}{=} \mathcal{L}(u). \quad (\star) \end{aligned}$$

Um die zweite und anschließend gleich die dritte Gleichheit in (4.15) zu zeigen, kann nun ausgenützt werden, dass μ der Kehrwert der AWK η ist. Es gilt also $\mu \stackrel{(4.11)}{=} \frac{1}{\eta}$ und daher

$$\begin{aligned} \mathcal{L}(u) &\stackrel{(\star)}{=} \frac{u + \frac{1}{\eta} - 1}{\frac{1}{\eta}u + \frac{1}{\eta} - 1} = \frac{\eta u + 1 - \eta}{u + 1 - \eta} = \frac{1 - \eta + \eta u}{1 - \eta + u} \overbrace{-\eta + \eta}^{=0} \\ &= \frac{1 - \eta + \eta u - \eta + \eta^2 - \eta u}{1 - \eta + u} + \eta = \eta + \frac{1 - 2\eta + \eta^2}{1 - \eta + u} = \eta + (1 - \eta) \frac{1 - \eta}{1 - \eta + \mu}. \quad \square \end{aligned}$$

Nun wird kurz überprüft, ob die so gefundene Laplace-Transformierte von W tatsächlich eine Lösung der Funktionalgleichung (2.78) ist (gemäß der Bem. zum Satz 2.56 (i) ist das dann sogar die einzige). Dass dem so ist, sieht man für $u \in \mathbb{R}_+$ an

$$\begin{aligned} f\left(\mathcal{L}\left(\frac{u}{\mu}\right)\right) &\stackrel{(4.5)}{=} \frac{1}{1 + \mu - \mu\mathcal{L}\left(\frac{u}{\mu}\right)} \stackrel{(4.15)}{=} \frac{1}{1 + \mu - \frac{u + \mu^2 - \mu}{u + \mu - 1}} \\ &= \frac{u + \mu - 1}{u + \mu - 1 + u\mu + \mu^2 - \mu - u - \mu^2 + \mu} = \frac{u + \mu - 1}{u\mu + \mu - 1} \stackrel{(4.15)}{=} \mathcal{L}(u). \end{aligned}$$

Bemerkungen zum Satz 4.6.

- (i) Die Verteilung einer nichtnegativen ZV ist durch ihre Laplace-Transformierte eindeutig charakterisiert und sie kann verwendet werden, um ihre Momente zu 'erzeugen' (siehe dazu den Satz A.83). Um dies für den *EW und die Varianz von W* vorzuführen, bestimme aus $\mathcal{L}$ für $u \in \mathbb{R}_+$ (mit dem letzten Ausdruck in (4.15))

$$\begin{aligned} \mathcal{L}'(u) &= -\frac{(1 - \eta)^2}{(1 - \eta + u)^2} \quad \text{und} \quad \mathcal{L}''(u) = \frac{2(1 - \eta)^2}{(1 - \eta + u)^3} \\ \Rightarrow \mathbb{E}W &\stackrel{(A.97)}{=} -\mathcal{L}'(0^+) = \frac{(1 - \eta)^2}{\eta < 1 (1 - \eta)^2} = 1 \quad \text{und} \quad \mathbb{E}W^2 \stackrel{(A.97)}{=} \mathcal{L}''(0^+) = \frac{2}{1 - \eta} \quad (4.17) \end{aligned}$$

$$\Rightarrow \mathbb{V}W = \frac{2}{1 - \eta} - 1 = \frac{1 + \eta}{1 - \eta} \quad \left(\stackrel{(4.11)}{=} \frac{1 + \frac{1}{\mu}}{1 - \frac{1}{\mu}} \cdot \frac{\mu^2}{\mu^2} = \frac{\mu^2 + \mu}{\mu(\mu - 1)} \stackrel{(4.7)}{=} \frac{\sigma^2}{\mu(\mu - 1)} \right). \quad (4.18)$$

Beide Ergebnisse entsprechen erwartungsgemäß dem allgemeinen Satz 2.53 (ii).

- (ii) Auf die folgende Weise kann man alternativ (und genau gleichwertig) die *bedingte Laplacetransformation* von W gegeben $W > 0$ (also $\mathcal{L}_{W|W>0}$) aus der zweiten Gleichheit in (4.15) herleiten.

4 Anwendungsbeispiele

Es gilt $\mathbb{P}(W_n = 0) \stackrel{(2.62)}{=} \mathbb{P}(Z_n \mu^{-n} = 0) = \mathbb{P}(Z_n = 0) \stackrel{(2.9)}{=} f_n(0)$ für jedes $n \geq 0$. Völlig analog zum Lemma 4.5 bzw. wieder mit der NR B.10 gelangt man über $\mathcal{L}_{W_n|W_n>0} \stackrel{(2.75)}{=} \frac{\mathcal{L}_n - f_n(0)}{1 - f_n(0)} \xrightarrow{(2.67)} \frac{\mathcal{L}_n - \eta}{1 - \eta} \stackrel{(4.14)}{=} \mathcal{L}_{W|W>0}$ zu

$$\mathcal{L}_{W|W>0}(u) \stackrel{(4.15)}{=} \frac{\frac{1-\eta+\eta u}{1-\eta+u} - \eta}{1 - \eta} = \frac{(1-\eta)^2}{(1-\eta+u)(1-\eta)} = \frac{1-\eta}{1-\eta+u} \quad \text{für alle } u \in \mathbb{R}_+. \quad (4.19)$$

(iii) An (4.19) (aber auch bereits an (4.15)) erkennt man, dass $\mathcal{L}_{W|W>0}$ genau der Laplace-Transformierten einer *Exponentialverteilung mit Parameter* $\lambda = 1 - \eta$ entspricht (vgl. das Bsp. A.85 bzw. (A.100)). Demnach ist

$$\begin{aligned} (W|W > 0) &\sim \text{Exp}(1 - \eta) \stackrel{(4.11)}{=} \text{Exp}\left(1 - \frac{1}{\mu}\right) \\ &\stackrel{\text{Bsp. A.85}}{\Rightarrow} \mathbb{E}[W|W > 0] = \frac{1}{1 - \eta}, \quad \mathbb{V}[W|W > 0] = \frac{1}{(1 - \eta)^2}. \end{aligned} \quad (4.20)$$

(iv) Somit kann W – mittels (4.15) – als eine sogenannte *gemischte Zufallsvariable* (vgl. etwa [18, A.4.6, S. 289]) gesehen werden. Mit einer WK von η ist sie degeneriert und es gilt $W = 0$. Dies ist genau dann der Fall, wenn der Prozess ausstirbt. Im Überlebensfall, also mit der WK $1 - \eta$, besitzt W bzw. $(W|W > 0)$ jedoch eine stetige Dichte auf $(0, \infty)$ (im Abschnitt 2.6 wurde bereits darauf hingewiesen, dass dies zumindest für $\sigma^2 < \infty$ allgemein so sein muss). Diese entspricht im superkritischen geometrischen Fall also einer Exponentialverteilung mit Parameter $1 - \eta$. Zusammengefasst kann man das Folgende schreiben (sei $Y : \Omega \rightarrow \mathbb{R}_+$ eine ZV mit $Y \sim \text{Exp}(1 - \eta)$):

$$W \stackrel{d}{=} \begin{cases} 0 & \text{mit WK } \eta, \\ Y \sim \text{Exp}(1 - \eta) & \text{mit WK } 1 - \eta. \end{cases} \quad (4.21) \quad \triangle$$

Der letzte Punkt der obigen Bemerkung lässt auch eine etwas andere Interpretation zu: Die Realisierung von W kann als Ausgang eines *zweistufigen Experiments* (vgl. etwa [30, S. 235 oder S. 279]) aufgefasst werden. In der ersten Stufe wird zufällig bestimmt, ob $W = 0$ (mit WK η) oder größer als 0 ist. Nur in letzterem Fall gelangt man zur zweiten Stufe, in welcher (die dann positive) Realisierung von W eben exponentialverteilt ist.

Die beiden Fälle (oder Stufen) für die gemischte ZV W können auch unabhängig voneinander betrachtet werden, wie im nächsten Lemma gezeigt wird.

Lemma 4.7 (W als gemischte ZV – vgl. z.B. [18, A.4.6, S. 289])

Sei $I : \Omega \rightarrow [0, 1]$ eine *bernoulliverteilte ZV mit Parameter* $1 - \eta$, kurz $I \sim \text{Ber}(1 - \eta)$ (vgl. z.B. [30, Bsp. 1.105.(i), S. 44]). D.h. $\mathbb{P}(I = 1) = 1 - \eta$ und $\mathbb{P}(I = 0) = \eta$ ($\star$).

Sei weiters $Y : \Omega \rightarrow \mathbb{R}_+$ *unabhängig von* I und *exponentialverteilt mit Parameter* $1 - \eta$, d.h. $Y \sim \text{Exp}(1 - \eta)$ (vgl. etwa erneut das Bsp. A.85). Dann gilt

$$W \stackrel{d}{=} V := Y \cdot I \quad \left(\stackrel{(A.2)}{=} Y \mathbb{1}_{\{I=1\}} \right). \quad (4.22)$$

4.1 Einfache GW-Verzweigungsprozesse – geometrischer Fall

Beweis. Der Satz A.83 besagt: $W \stackrel{d}{=} V \iff \mathcal{L} \stackrel{(2.75)}{=} \mathcal{L}_W = \mathcal{L}_V$. Bestimme also für $u \in \mathbb{R}_+$

$$\begin{aligned} \mathcal{L}_V(u) &\stackrel{(A.96)}{=} \mathbb{E}[e^{-uV}] \stackrel{(4.22)}{=} \mathbb{E}[e^{-uYI} \mathbb{1}_\Omega] = \mathbb{E}[e^{-uYI} (\mathbb{1}_{\{I=0\}} + \mathbb{1}_{\{I=1\}})] \\ &= \mathbb{E}[e^{-uY \cdot 0} \mathbb{1}_{\{I=0\}}] + \mathbb{E}[e^{-uY \cdot 1} \mathbb{1}_{\{I=1\}}] \stackrel{Y, I \text{ uabh.}}{=} \mathbb{P}(I=0) + \mathcal{L}_Y(u) \mathbb{P}(I=1) \\ &\stackrel{(\star)}{=} \eta + (1-\eta) \frac{1-\eta}{1-\eta+u} \stackrel{(A.100)}{=} \mathcal{L}(u). \end{aligned} \quad \square$$

Damit bietet sich noch eine weitere Möglichkeit, zu den Momenten von W zu kommen – und zwar über die unabhängigen ZV I und Y (wie im vorherigen Lemma definiert). Es gilt $\mathbb{E}I = \mathbb{E}I^2 = \mathbb{E}[\mathbb{1}_{\{I=1\}}] = \mathbb{P}(I=1) = 1-\eta$ (wegen $I = I^2$). Weiters ist bekannt, dass $\mathbb{E}Y = \frac{1}{1-\eta}$ und $\mathbb{E}Y^2 = \frac{2}{(1-\eta)^2}$ gilt (da Y einer Exponentialverteilung folgt – vgl. ggf. abermals das Bsp A.85). Somit lässt sich $\mathbb{E}W \stackrel{(4.22)}{=} \mathbb{E}[YI] \stackrel{\text{uabh.}}{=} \mathbb{E}Y \mathbb{E}I = \frac{1-\eta}{1-\eta} = 1$ und

$\mathbb{V}W \stackrel{(4.22)}{=} \mathbb{E}(YI)^2 - (\mathbb{E}W)^2 \stackrel{\text{uabh.}}{=} \mathbb{E}Y^2 \mathbb{E}I^2 - 1 = \frac{2}{(1-\eta)^2} (1-\eta) - 1 = \frac{2}{1-\eta} - 1 = \frac{1+\eta}{1-\eta}$ schließen (also genau so wie in (4.17) bzw. (4.18) und wieder im Einklang mit dem Satz 2.53 (ii)).

Wie die Ergebnisse in diesem Teilabschnitt bzgl. der Verteilung von W analog über charakteristische Funktionen erhalten werden können, wird in folgender Bem. skizziert.

Bemerkung 4.8 (Charakteristische Funktionen und Verteilung von W)

Die hier mit $\{\mathcal{L}_n\}_{n \geq 0}$ bzw. $\mathcal{L}$ durchgeführten Berechnungen gestalten sich mit den C-Funktionen $\{\varphi_n\}_{n \geq 0}$ bzw. φ (von $\{W_n\}_{n \geq 0}$ bzw. W – Def. 2.54) sehr ähnlich. Etwa in der NR B.11 wird gezeigt, wie man über die Konvergenz $\varphi_n \rightarrow \varphi$ das Pendant zum Satz 4.6,

$$\begin{aligned} \varphi(t) &\stackrel{(B.8)}{=} \frac{-it + \mu - 1}{-it\mu + \mu - 1} \stackrel{(4.11)}{=} \frac{-it + \frac{1}{\eta} - 1}{-it\frac{1}{\eta} + \frac{1}{\eta} - 1} = \frac{-it\eta + 1 - \eta}{-it + 1 - \eta} \quad \text{für } t \in \mathbb{R}, \text{ erhält} \\ \Rightarrow \varphi_{W|W>0}(t) &\stackrel{(B.5)}{=} \frac{\varphi(t) - \eta}{1 - \eta} = \frac{-it\eta + 1 - \eta + it\eta - \eta + \eta^2}{(-it + 1 - \eta)(1 - \eta)} = \frac{1 - \eta}{1 - \eta - it}, \end{aligned}$$

also im Wesentlichen genau so wie in der Bem. zum Satz 4.6 (ii) bzw. in (4.19) beschrieben (die NR B.10 ist auch auf W bzw. φ anwendbar). Damit kommt man zum selben Schluss wie im Punkt (iii) der erwähnten Bem., denn die C-Funktion von W gegeben $W > 0$ ist offensichtlich die einer mit Parameter $1 - \eta$ exponentialverteilten ZV (vgl. (A.100)).

Sehr ähnlich wie für $\mathcal{L}$ überprüfbar, löst φ die entsprechende Funktionalgleichung (2.80). Auch die Momentbestimmung kann fast gleich (siehe die Bem. zum Satz 4.6 (i)) erfolgen – die Herangehensweise wird in der Bem. zum Satz A.84 erläutert. $\triangle$

Beispiel 4.9 (Exemplarische Simulation zur Untersuchung von W)

Um das längerfristige Verhalten von $\{Z_n\}_{n \geq 0}$ bzw. $\{W_n\}_{n \geq 0}$ eines superkritischen geometrischen Prozesses zu analysieren, wurde ein Datensatz von 20 000 Realisierungen mit $\mu = \frac{3}{2}$ bis zur 40. Generation simuliert (Laufzeit für die Generierung: 17 Stunden und knapp 5 Minuten).

Nach 40 Generationen waren 13 265 Pfade bereits ausgestorben, was zu einer geschätzten AWK von 66,325 % führt (zum Vergleich: $\eta = \frac{1}{\mu} = \frac{2}{3} = 0,6$ – siehe (4.11)). Die größte der 6 735 in der 40. Generation noch lebenden Populationen bestand aus 269 784 453 Individuen. Aber auch bei der kleinsten überlebenden Realisierung konnte zu diesem Zeitpunkt bereits eine Populationsgröße von 6 586 verzeichnet werden. Demnach ist davon auszugehen, dass die AWK unverzerrt geschätzt wurde (ein späteres Aussterben ist so gut wie unmöglich, denn η^{6586} ist verschwindend klein – vgl. das Bsp. 2.49 und siehe ggf. auch die Bem. 4.33 für eine genauere Behandlung dieser Thematik).

Um nun den besagten Einfluss der ZV W auf das Wachstum von $\{Z_n\}_{n \geq 0}$ zu un-

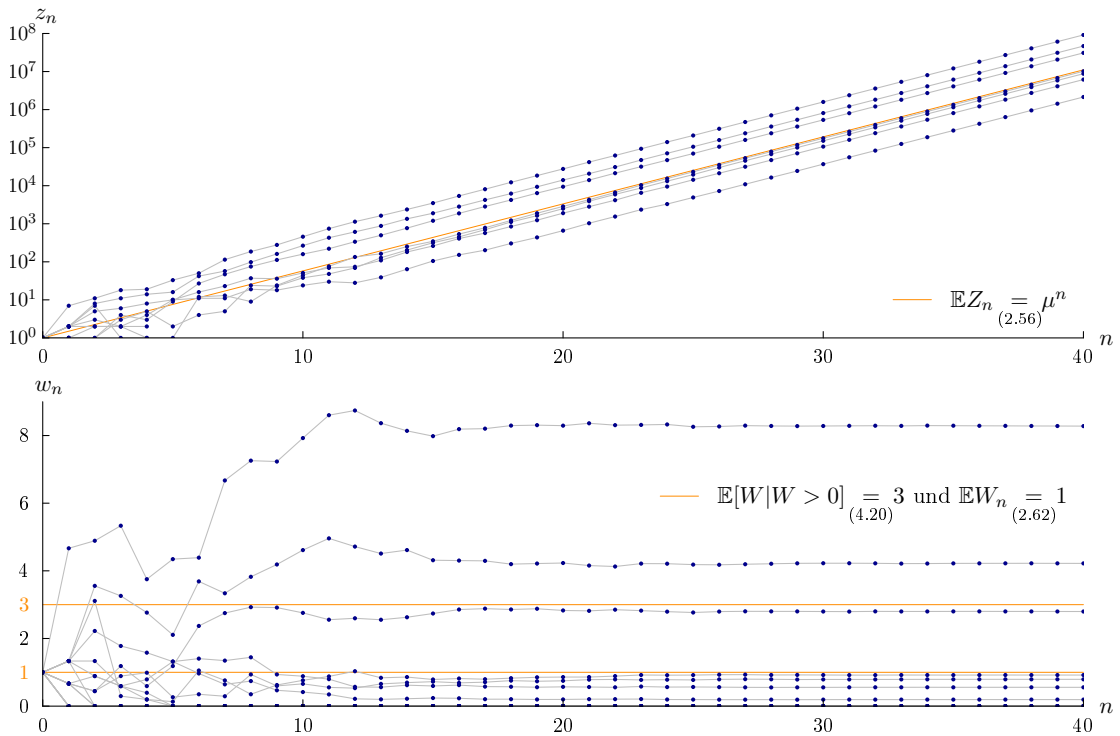


Abbildung 4.2: Beispielhafte Entwicklungⁱ 18 zufällig ausgewählter Pfade eines geometrischen GW-Prozesses mit $\mu = \frac{3}{2}$ bis zur 40. Generation (aus dem obigen Datensatz; 7 überlebende Pfade). Oben: Realisationen $\{z_n\}_{n \geq 0}$ von $\{Z_n\}_{n \geq 0}$; logarithmische y -Achse. Unten: Realisationen $\{w_n\}_{n \geq 0}$ von $\{W_n\}_{n \geq 0}$.

ⁱDie *diskreten* Punkte wurden zur besseren Nachverfolgbarkeit jeweils durch Linien verbunden (Ausnahme: Übergänge auf 0 im oberen Bild, da $\log(0) = -\infty$ ist).

4.1 Einfache GW-Verzweigungsprozesse – geometrischer Fall

tersuchen, beachte zunächst, dass $W \stackrel{(2.71)}{=} \lim_{n \rightarrow \infty} W_n \stackrel{(2.62)}{=} \lim_{n \rightarrow \infty} \frac{Z_n}{\mu^n}$ f.s. gilt $(\star)$. Auf $\{W=0\}$ ($= \{Z_n \rightarrow 0\}$) ist für hinreichend große n ohnehin $Z_n = \mu^n W_n = 0 = \mu^n W$ (siehe auch (4.12)). Wegen $(\star)$ gilt aber auch auf $\{W>0\}$ f.s. $\lim_{n \rightarrow \infty} \frac{Z_n}{\mu^n W} = 1$. I.a.W. bedeutet das, dass Z_n und $\mu^n W$ f.s. *asymptotisch gleich* sind $(\star\star)$ (vgl. etwa [13, S. 228] oder [19, S. 12] – von letzter Referenz sind auch Teile des Folgenden und insbesondere die erste Hälfte der Abb. 4.2 inspiriert). Längerfristig ist also das *exponentielle Wachstum* von $\{Z_n\}_{n \geq 0}$ durch den konstanten, nichtzufälligen Term μ^n und die ZV W bestimmt.

In der Abbildung 4.2 werden 18 zufällig aus dem obigen Datensatz ausgewählte Pfade von $\{z_n\}_{n \geq 0}$ bzw. $\{w_n\}_{n \geq 0}$ dargestellt (d.h. Realisierungen von $\{Z_n\}_{n \geq 0}$ bzw. $\{W_n\}_{n \geq 0}$). Dabei sind nur die 7 nicht-aussterbenden (38,9 %) auf längere Sicht erkennbar (6 starben sofort – also in der ersten Generation – aus und tauchen auf dem oberen Bild gar nicht auf). Für diese 7 Überlebenden ist im unteren Bild bereits die Konvergenz von W_n gegen $W > 0$ erkennbar (die anderen verbleiben zumindest ab einem gewissen Zeitpunkt für immer bei 0).

Der obere Graph für $\{z_n\}_{n \geq 0}$ ist mit einer logarithmisch skalierten Ordinate ausgestattet. Aufgrund von $(\star\star)$ sind auch $\log Z_n$ und $\log \mu^n W = \log W + n \log \mu$ im Überlebensfall f.s. asymptotisch gleich. In diesem Sinne linearisiert dann die Logarithmierung

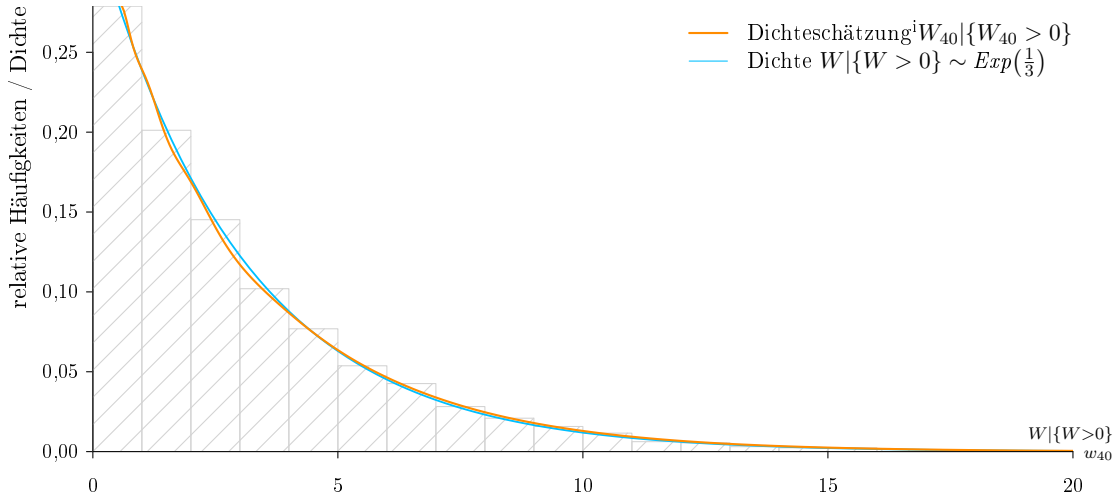


Abbildung 4.3: Histogrammⁱⁱ der in der 40. Generation noch lebenden Realisierungen w_{40} von W_{40} bei 20 000 Simulationen eines superkritischen geometrischen GWP mit $\mu = \frac{3}{2}$ (insgesamt 6 735 überlebende Pfade in der 40. Generation).

ⁱGeschätzt mittels sogenannter *Kerndichteschätzung (mit Gaußkern)* nach logarithmischer Transformation und entsprechender Rücktransformation – vgl. z.B. [14] (durch die Logarithmierung werden bei positiven ZV wesentlich bessere Ergebnisse erzielt). Die Schätzung wurde mit dem R-Paket 'logKDE' (Version 0.3.2 - vgl. [53]) durchgeführt.

ⁱⁱDargestellt werden realisierte $W_{40} \leq 20$. Im Bereich darüber befanden sich unter 0,075 % der Realisierungen (die größte lag bei 24,399). Zum Vergleich: Von der **Dichte von $W | \{W > 0\}$** wurden dadurch 0,127 % der Wahrscheinlichkeitsmasse abgeschnitten (denn $\mathbb{P}(W > 20 | W > 0) = 0,00127$).

4 Anwendungsbeispiele

den Verlauf der $\{\log z_n\}_{n \geq 0}$. Sie sorgt dafür, dass diese Pfade auf längere Sicht jeweils wahrnehmbar den Geraden $\log w + n \log \mu$ folgen (wobei w die Realisierungen von W sind). Dabei ist die Steigung dieser asymptotischen Geraden mit $\log \mu$ fix und nur die Ordinatenabschnitte $\log w$ waren zufällig.

Die überlebenden $\{z_n\}_{n \geq 0}$ und auch $\{w_n\}_{n \geq 0}$ unterlagen in den früheren Generationen stärkeren Schwankungen und stabilisierten sich später gewissermaßen – grob gesagt etwa ab der 10. bis 15. Generation. Wegen des exponentiellen Wachstums führt dies jedoch zu starken Unterschieden, welche bestehen bleiben. In [19, S. 12] wird in etwa davon gesprochen, dass die frühen Fluktuationen als Multiplikatoren für später wirken und sich deren Effekte manifestieren.

Insofern kann sozusagen an W festgemacht werden, wie 'stark' Populationen im Überlebensfall aus ihrer Anfangsperiode herauskommen. Aufgrund des bisher Besprochenen können die Realisierungen w_{40} der 20 000 vorliegenden Simulationen bereits gut als Schätzgrundlage für die Verteilung von W bzw. $(W|W > 0)$ herangezogen werden. Diese wurde für den geometrischen Fall schon theoretisch ermittelt (siehe die Bem. zum Satz 4.6). Hier bedeutet dies $(W|W > 0) \sim \text{Exp}(1-\eta) = \text{Exp}(1-\frac{1}{\mu}) = \text{Exp}(\frac{1}{3})$ (vgl. (4.20)). Die Verteilung der nicht ausgestorbenen w_{40} ist in der Abbildung 4.3 dargestellt. Nicht zuletzt, da es sich um simulierte und nicht um 'echte' Daten handelt (und auch aufgrund der großen Stichprobe) liegt die geschätzte Dichte recht nahe am theoretischen Resultat.

Auch die durch die überlebenden Realisierungen der 40. Generation geschätzten Momente von $(W|W > 0)$ liegen im Bereich der theoretischen Momente (siehe wieder (4.20) bzw. (4.21) oder das Beispiel A.85 – $\mathbb{E}[W|W > 0] = \frac{1}{1-\eta} = 3$ und $\mathbb{V}[W|W > 0] = \frac{1}{(1-\eta)^2} = 9$). So ergibt sich für die 6 735 w_{40} ein Mittelwert von 3,026 und eine (korrigierte) Stichprobenvarianz von 8,683. $\triangle$

Die wesentlichen in diesem Unterabschnitt eingesetzten theoretische Resultate haben grundsätzlich auch mit $\mu \leq 1$ Bestand (insbesondere etwa die Konvergenz $\mathcal{L}_n \rightarrow \mathcal{L}$ (2.77) sowie die Funktionalgleichung (2.78)). Für alle $u \in \mathbb{R}_+$ wird aber mit

$$\mathcal{L}_n(u) \stackrel{(2.79)}{=} f_n(e^{-u\mu^{-n}}) \stackrel{(4.9)}{=} \stackrel{(4.16)}{\begin{cases} \frac{n-ne^{-u}+e^{-u}}{n+1-ne^{-u}} = \frac{1-e^{-u}+\frac{1}{n}e^{-u}}{1+\frac{1}{n}-e^{-u}} \rightarrow \frac{1-e^{-u}}{1-e^{-u}} = 1 & \text{für } \mu = 1, \\ \frac{\mu^n(1-e^{-u\mu^{-n}})+\mu e^{-u\mu^{-n}}-1}{\underbrace{\mu}_{\rightarrow 0} \underbrace{\mu^n}_{\rightarrow 0} (1-\underbrace{e^{-u\mu^{-n}}}_{\rightarrow 0}) + \underbrace{\mu e^{-u\mu^{-n}}}_{\rightarrow 0} - 1} \rightarrow \frac{-1}{-1} = 1 & \text{für } \mu < 1 \end{cases}}$$

lediglich gezeigt, dass $\mathcal{L} = 1$ ist. Dadurch wird (2.74) bestätigt, also dass W (f.s.) degeneriert bei Null ist (beachte, dass dann $\mathcal{L}(u) = \mathbb{E}[e^{-u \cdot 0}] = 1$ für alle $u \in \mathbb{R}_+$ gilt) und in (sub-)kritischen Fällen keine Erkenntnisse von der Untersuchung dieser ZV erwartet werden können (vgl. etwa auch [3, S. 10]).

4.1.3 Exkurs: Gesamtpopulation und Aussterbezeitpunkt – allgemein

Für eine weitere Analyse können der Aussterbezeitpunkt sowie die gesamte Population eines GW-Prozesses von Interesse sein. Die entsprechenden Zufallsvariablen werden nun zunächst definiert und anschließend *allgemein* aufbereitet. Die Eigenschaften des Aussterbezeitpunkts im spezifischen geometrischen Fall werden kurz in der Bemerkung 4.14 angerissen. Einer genaueren Untersuchung der Gesamtpopulation geometrischer GWP ist der nächste Unterabschnitt 4.1.4 gewidmet.

Definitionen

Sei also $\{Z_n\}_{n \geq 0}$ ein beliebiger GWP (weiterhin mit $Z_0 \equiv 1$ und geltenden Ann. 2.39).

Definition 4.10 (Gesamtpopulation – vgl. z.B. [25, 2.11, S. 39])

Für einen GWP $\{Z_n\}_{n \geq 0}$ beschreibe allgemein die Folge $\mathbb{N}_0$ -wertiger Zufallsvariablen $\{T_n\}_{n \geq 0}$ die *Gesamtpopulationsgrößen bis zur n -ten Generation*. Bezeichne zudem jeweils die weF von T_n als $h_n : [0, 1] \rightarrow [0, 1]$. D.h., für jedes $n \geq 0$ seien

$$T_n := \sum_{j=0}^n Z_j \quad \text{und} \quad h_n(s) := f_{T_n}(s) \stackrel{(2.6)}{=} \mathbb{E}s^{T_n}. \quad (4.23)$$

Der Prozess $\{T_n\}_{n \geq 0}$ wird etwa in [29, 3.11, 8.–10., S. 69] auch als *integrierter GW-Prozess* bezeichnet.

Weiters sei die *Gesamtpopulation* des Prozesses T definiert als

$$T := \lim_{n \rightarrow \infty} T_n = \sum_{j \geq 0} Z_j = \sum_{n \geq 0} Z_n. \quad (4.24)$$

Definition 4.11 (Zeitpunkt des Aussterbens – vgl. z.B. [3, I.8.(3), S. 16])

Sei $\tau := \inf\{n \geq 0 : Z_n = 0\}$ der *Aussterbezeitpunkt* des Prozesses $\{Z_n\}_{n \geq 0}$. Es gilt:

$$\tau = n \quad \Longleftrightarrow \quad Z_{n-1} > 0 \text{ und } Z_n = 0 \quad \text{für } n \geq 1. \quad (4.25)$$

Weiters ist das Ereignis $\{\tau = \infty\}$ gleichbedeutend damit, dass es kein $n \geq 0$ mit $Z_n = 0$ gibt ($\inf \emptyset := \infty$). Dementsprechend gilt umgekehrt (da $\{Z_n \rightarrow 0\} \stackrel{(2.69)}{=} E$ gilt)

$$\{\tau < \infty\} = \{\exists n \geq 0 : Z_n = 0\} \stackrel{(2.66)}{=} \{Z_n \rightarrow 0\} \quad \Rightarrow \quad \mathbb{P}(\tau < \infty) = \mathbb{P}(Z_n \rightarrow 0) \stackrel{(2.67)}{=} \eta. \quad (4.26)$$

Bemerkung 4.12 (Wertigkeiten von T und τ)

- (i) Sowohl die Gesamtpopulation T als auch der Aussterbezeitpunkt τ können Werte in $\mathbb{N}_0$ annehmen, aber i.A. auch unendlich werden. Es handelt sich demnach um *erweitert reellwertige ZV* (vgl. die Def. A.27), für deren Bildmenge $\Omega' = \mathbb{N}_0 \cup \{\infty\}$ gesetzt werden kann ($T, \tau : \Omega \rightarrow \mathbb{N}_0 \cup \{\infty\}$ – da Ω' weiterhin abzählbar ist, handelt es sich aber dennoch um diskrete ZV, vgl. die Def. A.42 bzw. die Bem. dazu).

4 Anwendungsbeispiele

- (ii) Es gilt sowohl $\mathbb{P}(\tau < \infty) \stackrel{(4.26)}{=} \eta$ als auch $\mathbb{P}(T < \infty) \stackrel{(4.28)}{=} \eta$ (Letzteres ist leicht vorgegriffen – siehe das Lemma 4.15 (i)). Weiters ist für $\mu \leq 1$ gemäß dem Satz 2.44 $\eta = 1$. Diesfalls folgt $\mathbb{P}(T = \infty) = \mathbb{P}(\tau = \infty) = 0$ bzw. $\mathbb{P}(T \in \mathbb{N}_0) = \mathbb{P}(\tau \in \mathbb{N}_0) = 1$. Deshalb kann dann der *Träger der diskreten ZV* T (siehe erneut die Def. A.42) als $\mathbb{N}_0$ gesetzt werden. Im *subkritischen* und *kritischen Fall* können T und τ somit wie $\mathbb{N}_0$ -wertige Zufallsvariablen behandelt werden.
- (iii) Im superkritischen Fall mit $\mu > 1$ hingegen gilt $\mathbb{P}(T = \infty) = \mathbb{P}(\tau = \infty) = 1 - \eta > 0$. Für T gilt dann bspw. $\mathbb{E}[T] = \mathbb{E}[T \mathbb{1}_{\{T < \infty\}}] + \underbrace{\mathbb{E}[T \mathbb{1}_{\{T = \infty\}}]}_{=\infty}$ (und Analoges für τ), weshalb also $\mathbb{E}[T] = \mathbb{E}[\tau] = \infty$ bei $\mu > 1$ gilt. $\triangle$

Nun sollen ausgewählte Aspekte zum Aussterbezeitpunkt τ und der Gesamtpopulation T angeführt werden. Im Rahmen dieser Arbeit beschränkt sich dabei die Betrachtung von τ auf den nächsten Teilabschnitt. Dahingegen wird T im weiteren Verlauf noch etwas eingehender unter die Lupe genommen.

Resultate bezüglich des Aussterbezeitpunkts τ

In *subkritischen* Fällen kann der EW von τ folgendermaßen von oben abgeschätzt werden.

Satz 4.13 (Abschätzung des EW von τ – subkritisch – vgl. z.B. [18, 5.4, S. 115 ff.])
 Sei $\mu < 1$. Für den Zeitpunkt des Aussterbens des GWP $\{Z_n\}_{n \geq 0}$ gilt

$$\mathbb{E}[\tau] \leq \frac{1}{1 - \mu}.$$

Beweis. Der Prozess ist zu einem beliebigen Zeitpunkt $n \geq 0$ genau dann noch nicht ausgestorben, wenn die Populationsgröße dieser Generation $Z_n > 0$ ist. Dementsprechend folgt $\mathbb{P}(\tau > n) = \mathbb{P}(Z_n > 0)$ (*). Nun soll die *Markov-Ungleichung*, Satz A.37, ausgenützt werden. Diese besagt, dass $\mathbb{P}(\tau > n) = \mathbb{P}(Z_n \geq 1) \stackrel{(*)}{\leq} \frac{\mathbb{E}Z_n}{1} \stackrel{(A.47)}{=} \stackrel{(2.56)}{\mu^n}$ gilt (**)

$$\Rightarrow \mathbb{E}[\tau] \stackrel{(2.5)}{=} \sum_{n \geq 0} \mathbb{P}(\tau > n) \stackrel{(**)}{\leq} \sum_{n \geq 0} \mu^n \stackrel{\substack{\mu \leq 1 \\ \text{geom. Reihe}}}{\leq} \frac{1}{1 - \mu}. \quad \square$$

Bemerkung zum Satz 4.13. Im *superkritischen* und *kritischen* Fall gilt die Abschätzung im Beweis bis zum vorletzten Ausdruck zwar ebenso, ist aber nicht hilfreich. Da dann $\mu \geq 1$ ist, besagt sie nur $\mathbb{E}[\tau] \leq \sum_{n \geq 0} \mu^n = \infty$. $\triangle$

Bemerkung 4.14 (Verteilung von τ im geometrischen Fall)

Betrachte für den geometrischen Fall $\mathbb{P}(Z_n = 0) \stackrel{(2.9)}{=} f_n(0) \stackrel{(4.9)}{=} \begin{cases} \frac{n}{n+1} & \text{für } \mu = 1, \\ \frac{\mu^n - 1}{\mu^{n+1} - 1} & \text{für } \mu \neq 1 \end{cases}$, sowie $\{\tau = n\} = \{Z_n = 0\} \setminus \{Z_{n-1} = 0\}$ für $n \geq 1$ (siehe (4.25)). Daraus erhält man $\mathbb{P}(\tau = n) = \mathbb{P}(Z_n = 0) - \mathbb{P}(Z_{n-1} = 0)$ und für Fälle mit $\mu \neq 1$

$$\begin{aligned} \mathbb{P}(\tau = n) &= \frac{\mu^n - 1}{\mu^{n+1} - 1} - \frac{\mu^{n-1} - 1}{\mu^n - 1} = \frac{\mu^{2n} - 2\mu^n + 1 - \mu^{2n} + \mu^{n-1} + \mu^{n+1} - 1}{(\mu^{n+1} - 1)(\mu^n - 1)} \\ &= \frac{\mu^{n+1} - 2\mu^n + \mu^{n-1}}{(\mu^{n+1} - 1)(\mu^n - 1)} = \frac{\mu^{n-1}(\mu^2 - 2\mu + 1)}{(\mu^{n+1} - 1)(\mu^n - 1)} = \frac{\mu^{n-1}(\mu - 1)^2}{(\mu^{n+1} - 1)(\mu^n - 1)}. \end{aligned} \quad (4.27)$$

4.1 Einfache GW-Verzweigungsprozesse – geometrischer Fall

	1	2	3	4	5	10	20	40
$\mathbb{P}(\tau = n)$ berechnet*	0,476	0,158	0,079	0,047	0,031	$7,98 \cdot 10^{-3}$	$1,67 \cdot 10^{-3}$	$1,91 \cdot 10^{-4}$
$\mathbb{P}(\tau = n)$ simuliert**	0,479	0,157	0,076	0,047	0,032	$7,85 \cdot 10^{-3}$	$1,70 \cdot 10^{-3}$	$2,00 \cdot 10^{-4}$

Tabelle 4.4: WK für den Aussterbezeitpunkt τ im superkritischen geometrischen Fall mit $\mu = 1,1$ für ausgewählte Generationen (*Berechnung mit (4.27); **Anteil an 20 000 Simulationen, in welchen τ als n realisiert wurde).

Die Tabelle 4.4 zeigt beispielhaft einige dieser Wahrscheinlichkeiten für $\mu = 1,1$.

Mit $\mu = 1$ gilt hingegen $\mathbb{P}(\tau = n) = \frac{n}{n+1} - \frac{n-1}{n} = \frac{n^2 - (n^2 - 1)}{n(n+1)} = \frac{1}{n(n+1)}$. Dementsprechend kann der erwartete Aussterbezeitpunkt im kritischen geometrischen Fall als $\sum_{n \geq 0} \frac{1}{n+1} = \infty$ angegeben werden (harmonische Reihe – vgl. z.B. [22, S. 189]). Einerseits ist dies wegen $\eta = 1$ und $\mathbb{P}(\tau = \infty) = 0$ etwas kontraintuitiv. Andererseits passt diese Erkenntnis aber mit der – für diesen Fall unter 4.1.1 festgestellten – sehr langsamen Konvergenz von $\{f_n(0)\}_{n \geq 0}$ gegen 1 zusammen (denn $f_n(0) = \mathbb{P}(Z_n = 0)$).

Dass mit $\mu > 1$ ebenso $\mathbb{E}[\tau] = \infty$ gilt, wurde bereits in der Bem. 4.12 (iii) bestimmt. In subkritischen Fällen konnte er im vorherigen Satz 4.13 zumindest durch $\frac{1}{1-\mu}$ eingegrenzt werden. $\triangle$

Resultate bezüglich der Gesamtpopulationsgröße T

Über die Gesamtpopulationsgröße T kann die gleich folgende allgemeine Aussage getroffen werden. In diesem Lemma werden zudem die weF h_n (von T_n) aufeinanderfolgender Generationen zueinander in Beziehung gesetzt.

Lemma 4.15 (angelehnt an [25, 2.11, S. 39] bzw. [17, Problem 5.12.11, S. 207]). *Es gilt:*

- (i) *Die Gesamtpopulation T ist genau dann endlich, wenn die Populationsgrößen gegen 0 konvergieren – also wenn der Prozess ausstirbt. D.h.*

$$\begin{aligned} \{T < \infty\} &= \{Z_n \rightarrow 0\} \stackrel{(2.69)}{=} E \Rightarrow \mathbb{P}(T < \infty) = \mathbb{P}(Z_n \rightarrow 0) \stackrel{(2.64)}{=} \eta \\ &\stackrel{\text{Satz 2.44}}{\Rightarrow} \mathbb{P}(T < \infty) = \begin{cases} \eta = 1, & \text{wenn } \mu \leq 1 \text{ ist,} \\ \eta < 1, & \text{wenn } \mu > 1 \text{ ist.} \end{cases} \end{aligned} \quad (4.28)$$

- (ii) *Sei $s \in [0, 1]$. Es gilt $h_0(s) = s$ und die weF h_n von T_n erfüllen die Rekursion*

$$h_{n+1}(s) = sf(h_n(s)) \quad \text{für alle } n \geq 0. \quad \text{Weiters gilt} \quad (4.29)$$

$$h_{n+1} \leq h_n \quad \text{für alle } n \geq 0 - \text{also dass } h_n \text{ monoton fallend in } n \text{ ist.} \quad (4.30)$$

Dabei ist die Ungleichung in (4.30) auf $(0, 1)$ strikt (da $p_0 < 1$ ist – Ann. 2.39).

Beweis.

- (i) Eine Voraussetzung für $T < \infty$ – also die Konvergenz der Reihe in (4.24) – ist, dass die Glieder Z_n eine Nullfolge bilden (vgl. z.B. [22, Satz 31.3, S. 192]). Demgemäß

4 Anwendungsbeispiele

gilt $\{T < \infty\} \subseteq \{Z_n \rightarrow 0\}$. Umgekehrt gilt aber auch

$$\begin{aligned} \omega \in \{Z_n \rightarrow 0\} &\stackrel{(\star)}{\iff} \exists m \geq 1 : Z_n(\omega) = 0 \quad \forall n > m \\ &\Rightarrow \exists m \geq 1 : \sum_{n>m} Z_n(\omega) = 0 \stackrel{(\star\star)}{\iff} \omega \in \{T < \infty\}. \end{aligned}$$

($\star$) Die Begründung für diese Äquivalenz findet sich in der Rechnung (B.1) der NR B.4 (welche – wie bereits für (2.69) argumentiert – auch für $k = 0$ gilt).

($\star\star$) Für diese ω gilt $T(\omega) \stackrel{(4.24)}{=} \sum_{n \geq 0} Z_n(\omega) = \underbrace{\sum_{n=0}^m Z_n(\omega)}_{< \infty} + \underbrace{\sum_{n>m} Z_n(\omega)}_{=0}$.

Dies bedeutet, dass $\{Z_n \rightarrow 0\} \subseteq \{T < \infty\}$ ist.

Zusammengenommen gilt also $\{T < \infty\} = \{Z_n \rightarrow 0\}$.

- (ii) Aus $T_0 = Z_0 \equiv 1$ ergibt sich direkt $h_0(s) = \mathbb{E}s^{T_0} = s$. Der Beweis von (4.29) ähnelt der nun bekannten Vorgehensweise (etwa bzgl. (2.48)) und bedient sich wieder des Satzes 2.14. Sei $n \geq 0$ beliebig. Weiters sei $\{T_{n,i}\}_{i \geq 1}$ eine Familie von *i.i.d.*-Kopien von T_n (jeweils mit der weF h_n). Dann gilt $T_{n+1} \stackrel{d}{=} \underbrace{Z_0}_{=1} + \sum_{i=1}^{Z_1} T_{n,i}$ ($\star\star\star$). Für alle $s \in [0, 1]$ folgt

$$h_{n+1}(s) \stackrel{(4.23)}{=} \mathbb{E}s^{T_{n+1}} \stackrel{(\star\star\star)}{=} \mathbb{E}s^{1+\sum_{i=1}^{Z_1} T_{n,i}} = s \mathbb{E}s^{\sum_{i=1}^{Z_1} T_{n,i}} \stackrel{(2.45)}{=} s f_{Z_1} \stackrel{(2.26)}{=} s f(h_n(s)).$$

Für (4.30) betrachte beliebige $s \in [0, 1]$. Es gilt

$$h_{n+1}(s) \stackrel{(4.23)}{=} \mathbb{E}s^{T_{n+1}} \stackrel{(4.23)}{=} \mathbb{E}[s^{T_n + Z_{n+1}}] = \mathbb{E}[s^{T_n} \overbrace{s^{Z_{n+1}}}^{\leq 1}] \leq \mathbb{E}s^{T_n} \stackrel{(4.23)}{=} h_n(s).$$

Für $s \in (0, 1)$ gilt zudem $0 < s^{Z_{n+1}} < 1$ auf $\{Z_{n+1} > 0\}$. Deshalb ist (wegen $p_0 < 1 \Rightarrow \mathbb{P}(Z_{n+1} > 0) > 0$) die obige Ungleichung auf $(0, 1)$ strikt. $\square$

Nun soll der Erwartungswert der Gesamtpopulation bestimmt werden. In der Bemerkung 4.12 wurde bereits festgestellt, dass er im superkritischen Fall nicht existiert (d.h. $\mathbb{E}T = \infty$). Dies wird im nächsten Satz noch einmal bestätigt und zudem wird gezeigt, dass dies auch im kritischen Fall so ist. Bei subkritischen Prozessen hingegen existiert $\mathbb{E}T$ und es wird eine Berechnungsformel dafür angegeben.

Satz 4.16 (EW der Gesamtpopulation – angelehnt an [24, Theorem 3.5, S. 92])

Es gilt

$$\mathbb{E}T = \begin{cases} \frac{1}{1-\mu}, & \text{falls } \mu < 1 \text{ ist,} \\ \infty, & \text{falls } \mu \geq 1 \text{ ist.} \end{cases}$$

Beweis. Diese Aussage ist eine direkte Anwendung des Satzes A.14 über die monotone Konvergenz. Aufgrund der linken Seite von (4.23) und $Z_0 \equiv 1$ gilt

$$1 = T_0 \leq T_1 \leq T_2 \leq \dots \leq \lim_{n \rightarrow \infty} T_n \stackrel{(4.24)}{=} T \stackrel{(4.24)}{=} \sum_{n \geq 0} Z_n \text{ und deshalb auch}$$

$$\begin{aligned} \mathbb{E}T &\stackrel{(4.24)}{=} \mathbb{E} \left[\sum_{n \geq 0} Z_n \right] = \mathbb{E} \left[\lim_{n \rightarrow \infty} \sum_{j=0}^n Z_j \right] \stackrel{\text{MON}}{=} \lim_{n \rightarrow \infty} \mathbb{E} \left[\sum_{j=0}^n Z_j \right] \\ &\stackrel{\text{endl. Summe}}{=} \lim_{n \rightarrow \infty} \sum_{j=0}^n \mathbb{E}Z_j = \sum_{n \geq 0} \mathbb{E}Z_n \stackrel{(2.56)}{=} \sum_{n \geq 0} \mu^n. \end{aligned}$$

4.1 Einfache GW-Verzweigungsprozesse – geometrischer Fall

Der letzte Ausdruck ist dabei eine geometrische Reihe, welche für $\mu \geq 1$ divergiert und für $\mu < 1$ wie oben behauptet konvergiert. $\square$

Der Erwartungswert der Gesamtpopulation bis zur n -ten Generation T_n kann ohne Mühe direkt angegeben werden – es wird nur die n -te Partialsumme einer geometrischen Reihe benötigt.

$$\mathbb{E}T_n \stackrel{(4.23)}{=} \mathbb{E} \left[\sum_{j=0}^n Z_j \right] \stackrel{\text{endl. Summe}}{=} \sum_{j=0}^n \mathbb{E}Z_j \stackrel{(2.56)}{=} \sum_{j=0}^n \mu^j = \begin{cases} \frac{1-\mu^{n+1}}{1-\mu} & \text{für } \mu \neq 1, \\ n+1 & \text{für } \mu = 1. \end{cases}$$

In der Bem. 4.12 wurde bereits darauf hingewiesen, dass T i.A. – insbesondere im Fall $\mu > 1$ – keine $\mathbb{N}_0$ -wertige ZV ist. Um dennoch (auch in superkritischen Fällen) für weitergehende Analysen die bekannten Methoden (etwa aus 2.2.2) anwenden zu können, wird ein kleiner Umweg genommen. Im kommenden Satz wird dazu zunächst – als Behelf – eine Funktion h_T definiert, welche im Wesentlichen analog zu weF definiert ist.

Satz 4.17 (angelehnt an [25, (2.11.3), S. 39] bzw. [24, Theorem 3.2, S. 91])

Unabhängig von μ gilt für $h_T : [0, 1] \rightarrow [0, 1]$ mit $h_T(s) := \mathbb{E}s^T$ die Funktionalgleichung

$$h_T(s) = sf(h_T(s)). \quad (4.31)$$

Falls $\mu \leq 1$ ist, so ist h_T die weF der Gesamtpopulation T .

Beweis. Die letzte Aussage beruht auf der Bem. 4.12 und (2.6) (mit $\mu \leq 1$ kann T als $\mathbb{N}_0$ -wertig aufgefasst werden). Der restliche Beweis funktioniert analog zu dem für (4.29) und dem des Lemmas 4.15 (ii) – da der Satz 2.14 aber nur für weF formuliert wurde, wird nun direkt mit dem totalen EW operiert (Satz A.61 (b)). Seien $\{T_i\}_{i \geq 1}$ i.i.d.-Kopien von T (wobei jeweils $h_T(s) = \mathbb{E}s^{T_i}$ für alle $s \in [0, 1]$ gilt) $\Rightarrow T \stackrel{d}{=} 1 + \sum_{i=1}^{Z_1} T_i$ ($\star$). Dann gilt

$$\begin{aligned} \mathbb{E}[s^T | Z_1 = j] &\stackrel{(\star)}{=} \mathbb{E}[s^{1+\sum_{i=1}^j T_i}] \stackrel{(i.i.d.)}{=} s \prod_{i=1}^j \mathbb{E}s^{T_i} = s(h_T(s))^j \quad \text{für alle } j \geq 0 \quad (\star\star) \\ &\Rightarrow \mathbb{E}s^T \stackrel{(A.74)}{=} \mathbb{E}[\mathbb{E}[s^T | Z_1]] \stackrel{(\star\star)}{=} s \mathbb{E}[h_T(s)^{Z_1}] \stackrel{(2.45)}{=} sf(h_T(s)). \quad \square \end{aligned}$$

Die Funktionalgleichung (4.31) gilt grundsätzlich – wie ausgesagt – auf ganz $[0, 1]$ (vgl. etwa die erste der angegebenen Referenzen). Etwas problematisch stellt sich die Situation aber an der Stelle $s=1$ dar, wenn $\{T=\infty\}$ eine positive WK hat (1^∞ ist ein *unbestimmter Ausdruck*). Mit der im folgenden Korollar definierten Funktion h wird dieses Problem umgangen.

Korollar zum Satz 4.17. Definiere $h : [0, 1] \rightarrow [0, 1]$ als $h(s) := \mathbb{E}[s^T \mathbb{1}_{\{T < \infty\}}]$.

(i) Sei $\mu \leq 1$. Dann ist h die weF der $\mathbb{N}_0$ -wertigen ZV T und es gilt für alle $s \in [0, 1]$

$$h(s) = \mathbb{E}[s^T \mathbb{1}_{\{T < \infty\}}] = \mathbb{E}s^T \stackrel{\text{Satz 4.17}}{=} h_T(s) \stackrel{(4.31)}{\Rightarrow} h(s) = sf(h(s)). \quad (4.32)$$

(ii) Sei $\mu > 1$. Dann gilt (4.32) für alle $s \in [0, 1)$ und $\lim_{s \nearrow 1} h(s) = h(1) = \eta$.

4 Anwendungsbeispiele

Beweis. Dass h für $\mu \leq 1$ die genannte weF ist, wurde bereits im Satz argumentiert. Da dann $\{T = \infty\}$ eine Nullmenge ist ($\mathbb{P}(T < \infty) \stackrel{(4.28)}{=} \eta = 1$), komplettiert das Lemma A.17 den Beweis für (i). $=0$ für $s \in [0,1)$ und $T = \infty$

Mit $\mu > 1$ und $s \in [0,1)$ gilt $h_T(s) \stackrel{(4.31)}{=} \mathbb{E} s^T \stackrel{(A.3)}{=} \mathbb{E}[s^T \mathbb{1}_{\{T < \infty\}}] + \mathbb{E}[\overbrace{s^T}^{\text{für } T=\infty} \mathbb{1}_{\{T=\infty\}}] = h(s)$, was (4.32) bestätigt. Beachte zudem, dass $h(1) = \mathbb{E}[1 \cdot \mathbb{1}_{\{T < \infty\}}] \stackrel{(A.44)}{=} \mathbb{P}(T < \infty) = \eta < 1$ gilt und h als eine Potenzreihe dargestellt werden kann (T ist diskret). Somit garantiert (wie im Satz 2.5) der Grenzwertsatz A.103 (d) von Abel, dass $\lim_{s \nearrow 1} h(s) = h(1)$ ist. $\square$

Bemerkung zum Satz 4.17. Auch für $\mu \leq 1$ gilt klarerweise $h(1) \stackrel{(2.8)}{=} 1 \stackrel{\text{Satz 2.44}}{=} \eta$, womit $h(1) = \eta$ jedenfalls stimmt.

Für alle $s \in [0,1]$ gilt weiters $h(s) = \mathbb{E}[\overbrace{s^T}^{\leq 1} \mathbb{1}_{\{T < \infty\}}] \stackrel{(A.44)}{\leq} \mathbb{P}(T < \infty) \stackrel{(4.28)}{=} \eta$ (ebenso unabhängig von μ). Daraus folgt $h(s) \stackrel{(4.32)}{=} sf(h(s)) \leq sf(\eta) \stackrel{(2.68)}{=} s\eta$ (f ist isoton).

$$\text{Insbesondere gilt} \quad h(0) = \mathbb{E}[0^T \mathbb{1}_{\{T < \infty\}}] \stackrel{T \geq 1}{=} 0. \quad (4.33)$$

Die Funktion h liegt also für alle $s \in [0,1]$ in $[0, \eta]$. Dass diese die Funktionalgleichung in (4.32) löst, wird durch den Satz und das Korollar klar. Es lässt sich auch zeigen, dass h die einzige Lösung dieser Gleichung in $[0,1]$ ist (siehe dafür etwa [25, 2.11, S. 39 f.]). $\triangle$

In den meisten Fällen ist h genau der Grenzwert der weF h_n von T_n .

Satz 4.18 (Konvergenz der weF der n -ten Gesamtpopulationsgrößen)

(i) Sei $\mu \leq 1$. Dann konvergiert die Folge $\{h_n\}_{n \geq 0}$ gegen h , d.h. es gilt

$$h_n(s) \rightarrow h(s) \quad \text{für alle } s \in [0,1]. \quad (4.34)$$

(ii) Sei $\mu > 1$. Dann gilt (4.34) für $s \in [0,1)$.

Beweis. Beachte allg. für alle $\omega \in \Omega$, dass $\lim_{n \rightarrow \infty} s^{T_n(\omega)} = s^{\lim_{n \rightarrow \infty} T_n(\omega)} \stackrel{(4.24)}{=} s^{T(\omega)}$ gilt $(\star)$, sofern $s \in [0,1)$ ist (vgl. z.B. [7, 8.16, S. 82 f.]). Falls $\omega \in \{T < \infty\}$ ist, stimmt $(\star)$ auch für $s = 1$. Dementsprechend gilt für alle $s \in [0,1)$, und wenn $\mu \leq 1$ ist (dann ist $T < \infty$ f.s.) sogar für alle $s \in [0,1]$,

$$\begin{aligned} s &= s^{T_0} \geq s^{T_1} \geq s^{T_2} \geq \dots \geq \lim_{n \rightarrow \infty} s^{T_n} \stackrel{(\star)}{=} s^{\lim_{n \rightarrow \infty} T_n} = s^T \quad \text{f.s.} \\ \Rightarrow -s &= -s^{T_0} \leq -s^{T_1} \leq -s^{T_2} \leq \dots \leq \lim_{n \rightarrow \infty} -s^{T_n} = -s^{\lim_{n \rightarrow \infty} T_n} = -s^T \quad \text{f.s.} \\ \Rightarrow \lim_{n \rightarrow \infty} -h_n(s) &\stackrel{(4.23)}{=} \lim_{n \rightarrow \infty} \mathbb{E}[-s^{T_n}] \stackrel{\text{MON}}{\stackrel{(A.21)}}{=} \mathbb{E}\left[\lim_{n \rightarrow \infty} -s^{T_n}\right] = -\mathbb{E}[s^T] \stackrel{(4.32)}{=} -h(s). \quad \square \end{aligned}$$

Mit dem Satz 4.18 bietet sich eine alternative Möglichkeit, auf die Funktionalgleichung (4.32) zu kommen, indem der Grenzwert der Rekursion für h_n bestimmt wird. Es gilt für alle $s \in [0,1)$, und wenn $\mu \leq 1$ ist sogar für alle $s \in [0,1]$,

$$h(s) \stackrel{(4.34)}{=} \lim_{n \rightarrow \infty} h_{n+1}(s) \stackrel{(4.29)}{=} s \lim_{n \rightarrow \infty} (f(h_n(s))) \stackrel{f \text{ stetig}}{=} sf\left(\lim_{n \rightarrow \infty} h_n(s)\right) \stackrel{(4.34)}{=} sf(h(s)). \quad (4.35)$$

Im letzten Teilabschnitt des nächsten Abschnitts 4.1.4 (auf der Seite 126) wird die Konvergenz der $\{h_n\}_{n \geq 0}$ für den subkritischen geometrischen Fall zusätzlich explizit gezeigt.

Auf endliche Werte getrimmte Gesamtpopulation Y

Schließlich wird eine $\mathbb{N}_0$ -wertige ZV Y eingeführt, deren weF mit h in Verbindung gebracht werden kann (bzw. ihre Verteilung mit der von T , was das eigentliche Ziel war).

Definition 4.19 (Auf endliche Werte getrimmte Gesamtpopulation)

Definiere die $\mathbb{N}_0$ -wertige ZV Y und ihre weF $h_Y : [0, 1] \rightarrow [0, 1]$ als

$$Y := T \mathbb{1}_{\{T < \infty\}} \quad \text{und} \quad h_Y(s) := f_Y(s) \stackrel{(2.6)}{=} \mathbb{E}s^Y = \mathbb{E}s^{T \mathbb{1}_{\{T < \infty\}}}. \quad (4.36)$$

Es gilt (siehe (4.28) – wenn $\eta = 1$ gilt, ist $\{T < \infty\}$ eine Einsmenge)

$$Y = T \quad \text{f.s.,} \quad \text{falls } \mu \leq 1 \text{ ist.} \quad (4.37)$$

Lemma 4.20. *Für alle potentiellen Werte von μ gilt*

$$h_Y = h + (1 - \eta) \quad (\text{also insbesondere } h = h_Y \text{ im Fall } \mu \leq 1) \quad (4.38)$$

$$\Rightarrow h_Y^{(l)} = h^{(l)} \quad \text{mindestens auf } [0, 1) \text{ und für alle } l \geq 1 \text{ sowie} \quad (4.39)$$

$$\mathbb{P}(Y = 0) \stackrel{(2.9)}{=} h_Y(0) \stackrel{(4.33)}{=} 1 - \eta \quad \text{und allgemein} \quad (4.40)$$

$$\mathbb{P}(T = k) = \mathbb{P}(\{T \mathbb{1}_{\{T < \infty\}} = k\} \cup \overbrace{\{T \mathbb{1}_{\{T = \infty\}} = k\}}^{=\emptyset}) \stackrel{(4.36)}{=} \mathbb{P}(Y = k) \quad \text{für alle } \underbrace{k \geq 1}_{\text{(damit ist wie gewohnt } k \in \mathbb{N} \text{ gemeint, also ohne } k = \infty)}} \quad (4.41)$$

Beweis. Für alle $s \in [0, 1]$ gilt

$$\begin{aligned} s^Y \mathbb{1}_{\{T < \infty\}} &= s^{T \mathbb{1}_{\{T < \infty\}}} \mathbb{1}_{\{T < \infty\}} = s^T \mathbb{1}_{\{T < \infty\}} \quad \text{und} \\ s^Y \mathbb{1}_{\{T = \infty\}} &= \underbrace{s^{T \mathbb{1}_{\{T < \infty\}}}}_{=s^0=1 \text{ für } T=\infty} \mathbb{1}_{\{T = \infty\}} \stackrel{(A.2)}{=} \mathbb{1}_{\{T = \infty\}} \\ \Rightarrow h_Y(s) &\stackrel{(4.36)}{=} \mathbb{E}[s^Y \mathbb{1}_\Omega] \stackrel{(A.3)}{=} \mathbb{E}[s^Y (\mathbb{1}_{\{T < \infty\}} + \mathbb{1}_{\{T = \infty\}})] = \mathbb{E}[s^Y \mathbb{1}_{\{T < \infty\}}] + \mathbb{E}[s^Y \mathbb{1}_{\{T = \infty\}}] \\ &\stackrel{(\star)}{=} \mathbb{E}[s^T \mathbb{1}_{\{T < \infty\}}] + \mathbb{E}[\mathbb{1}_{\{T = \infty\}}] \stackrel{(4.32)}{=} h(s) + \mathbb{P}(T = \infty) \\ &\stackrel{(A.44)}{=} h(s) + (1 - \mathbb{P}(T < \infty)) \stackrel{(4.28)}{=} h(s) + (1 - \eta), \end{aligned} \quad (\star)$$

womit (4.38) gezeigt ist. Die anderen Beziehungen werden bereits in der Formulierung des Lemmas argumentiert. $\square$

Aufgrund von $Z_0 \equiv 1$ ist T immer mindestens 1 (vgl. (4.24)) und folglich $\mathbb{P}(T = 0) = 0$. Weiters wurde im Lemma 4.15 bereits implizit (und auch im soeben durchgeführten Beweis direkt) festgestellt, dass $\mathbb{P}(T = \infty) = 1 - \mathbb{P}(T < \infty) = 1 - \eta$ ist. Mit (4.41) kann also zusammenfassend das folgende für die Verteilung von T auf $\Omega' = \mathbb{N}_0 \cup \{\infty\}$ (siehe die Bemerkung 4.12 (i)) festgehalten werden. Es gilt

$$\mathbb{P}(T = k) = \begin{cases} 0 & \text{für } k = 0, \\ \mathbb{P}(Y = k) & \text{für } k \in \mathbb{N}, \\ 1 - \eta & \text{für } k = \infty. \end{cases} \quad (4.42)$$

4.1.4 Gesamtpopulation geometrischer GW-Prozesse

Die im vorherigen Abschnitt erarbeiteten Resultate zur Gesamtpopulation sollen nun auf den geometrischen Fall angewendet werden. Betrachte ab sofort also wieder geometrische GW-Prozesse. Zunächst wird eine analytische Form für h ermittelt. Anschließend werden explizite Ausdrücke für die Momente und die Verteilung von T gesucht.

Lösung der Funktionalgleichung für h – geometrischer Fall

Es wird versucht, die Funktionalgleichung $h(s) = sf(h(s))$ (also den zweiten Teil von (4.32)) so zu lösen, dass $h([0, 1]) \subseteq [0, 1]$ ist. Wenn $\mu \leq 1$ ist, muss h die weF von T sein (siehe das Korollar zum Satz 4.17). Aber auch wenn $\mu > 1$ ist, soll die Lösung gemäß dem Lemma 4.20 so sein, dass $h_Y = h + (1 - \eta)$ eine weF ist (vgl. (4.38)). Daher müssen die entsprechenden Eigenschaften erfüllt sein, welche anschließend auszugsweise überprüft werden. Sei zunächst $s \in [0, 1]$ beliebig. Es gilt:

$$\begin{aligned} h(s) &\stackrel{(4.32)}{=} sf(h(s)) \stackrel{(4.5)}{=} \frac{s}{1 + \mu - \mu h(s)} \iff h(s) + h(s)\mu - (h(s))^2\mu = s \\ &\iff (h(s))^2 - h(s)\frac{1+\mu}{\mu} + \frac{s}{\mu} \\ &\Rightarrow h_{1,2}(s) = \frac{1+\mu}{2\mu} \pm \sqrt{\frac{(1+\mu)^2}{4\mu^2} - \frac{s}{\mu}} = \frac{1+\mu}{2\mu} \pm \frac{1}{2\mu}\sqrt{(1+\mu)^2 - 4\mu s}. \end{aligned}$$

Für die Diskriminante $D : [0, 1] \rightarrow \mathbb{R}$ mit

$$D(s) := (1 + \mu)^2 - 4\mu s \tag{4.43}$$

gilt stets $(1-\mu)^2 \leq D(s) \leq (1+\mu)^2$. Sie ist also für alle $s \in [0, 1]$ nichtnegativ und abgesehen vom Fall $\mu = s = 1$ positiv, weshalb es bis auf diese Ausnahme immer zwei verschiedene denkbare Lösungen gibt (bezeichne h_1 als die Lösung mit '+' und h_2 als die mit '-').

Berechne nun etwa $h_1(0) = \frac{2+\mu}{2\mu} = \frac{1+\mu}{\mu} > 1$, um die erste der beiden Kandidaten auszuschließen (oder auch etwa die Tatsache, dass $h_1'(s) = -\frac{1}{\sqrt{(1+\mu)^2 - 4\mu s}} < 0$ für alle $s \in (0, 1)$ und h_1 somit monoton fallend ist, disqualifiziert diese Lösung). Die gesuchte Lösung ist demnach h_2 , d.h.

$$h(s) = \frac{1 + \mu - \sqrt{(1 + \mu)^2 - 4\mu s}}{2\mu} \stackrel{(4.43)}{=} \frac{1 + \mu - \sqrt{D(s)}}{2\mu} \quad \text{für alle } s \in [0, 1], \tag{4.44}$$

was sich im Fall $\mu = 1$ vereinfacht zu $h(s) = \frac{2 - \sqrt{4 - 4s}}{2} = 1 - \sqrt{1 - s}$.

Bemerkung 4.21 (Überprüfung der Lösung h). Wie angekündigt soll nun eine Überprüfung ausgewählter Eigenschaften von h erfolgen. Dabei werden auch Formeln bzgl. der Diskriminante D angeführt, da diese später benötigt werden. Um den Fall $\mu > 1$ ebenso abzudecken, wird gleich mit kontrolliert, ob die korrespondierende Funktion h_Y

4.1 Einfache GW-Verzweigungsprozesse – geometrischer Fall

die Eigenschaften einer weF erfüllt (mit $\mu \leq 1$ sind h und h_Y das Gleiche (4.38)).

$$D(0) \stackrel{(4.43)}{=} (1 + \mu)^2 \quad \Rightarrow h(0) \stackrel{(4.44)}{=} \frac{1}{2\mu} (1 + \mu - \sqrt{D(0)}) = 0. \quad (4.45)$$

$$D(1) \stackrel{(4.43)}{=} 1 + 2\mu + \mu^2 - 4\mu = (1 - \mu)^2 \quad \Rightarrow \sqrt{D(1)} = |1 - \mu| = \begin{cases} 1-\mu & \text{für } \mu < 1, \\ 0 & \text{für } \mu = 1, \\ \mu-1 & \text{für } \mu > 1 \end{cases} \quad (4.46)$$

$$\Rightarrow h(1) \stackrel{(4.44)}{=} \frac{1 + \mu - \sqrt{D(1)}}{2\mu} = \begin{cases} \frac{1+\mu-1+\mu}{2\mu} = 1 = \eta & \text{für } \mu < 1, \\ \frac{2}{2} = 1 = \eta & \text{für } \mu = 1, \\ \frac{1+\mu-\mu+1}{2\mu} = \frac{1}{\mu} = \eta & \text{für } \mu > 1. \end{cases} \quad (4.47)$$

$$D'(s) \stackrel{(4.43)}{=} -4\mu \quad \Rightarrow h'(s) = -\frac{1}{4\mu} (D(s))^{-\frac{1}{2}} D'(s) = \frac{1}{\sqrt{D(s)}} \quad \text{für alle } s \in [0, 1] \quad (4.48)$$

$$\Rightarrow h''(s) = -\frac{1}{2} (D(s))^{-\frac{3}{2}} D'(s) = \frac{2\mu}{(\sqrt{D(s)})^3} \quad \text{für alle } s \in [0, 1]. \quad (4.49)$$

Sofern $\mu \neq 1$ ist, gelten die Formeln für h' und h'' sogar für alle $s \in [0, 1]$. Diese Ableitungen sind aber in allen Fällen auf dem gesamten Intervall $(0, 1)$ strikt positiv, weshalb h und auch h_Y (siehe (4.39)) strikt isoton und strikt konvex sind (vgl. den Satz 2.5 (e)).

Es gilt $h_Y(0) \stackrel{(4.38)}{=} h(0) + (1 - \eta) \stackrel{(4.45)}{=} 1 - \eta \stackrel{(4.28)}{=} \mathbb{P}(T = \infty) \stackrel{(4.40)}{=} \mathbb{P}(Y = 0)$ und weiters $h_Y(1) = h(1) + (1 - \eta) \stackrel{(4.47)}{=} \eta + 1 - \eta = 1$ womit die Randpunkte korrekt sind (Satz 2.5 (c)). Somit liegen (aufgrund der Isotonie) alle Punkte in $[0, 1]$, genauer $h_Y([0, 1]) = [1 - \eta, 1] \subseteq [0, 1]$.

Abschließend wird noch nachgeprüft, ob h tatsächlich der Funktionalgleichung in (4.32) genügt, d.h. ob $h(s) = sf(h(s))$ für alle $s \in [0, 1]$ erfüllt ist (was alleine schon aufgrund des Lösungsverfahrens so sein muss). Es gilt

$$\begin{aligned} sf(h(s)) &\stackrel{(4.5)}{=} \frac{s}{1 + \mu - \mu h(s)} \stackrel{(4.44)}{=} \frac{s}{\mu + 1 - \frac{\mu + 1 - \sqrt{D(s)}}{2}} \\ &= \frac{2s}{\mu + 1 + \sqrt{D(s)}} \cdot \frac{(\mu + 1) - \sqrt{D(s)}}{(\mu + 1) - \sqrt{D(s)}} = 2s \frac{\mu + 1 - \sqrt{D(s)}}{(\mu + 1)^2 - D(s)} \\ &\stackrel{(4.43)}{=} 2s \frac{\mu + 1 - \sqrt{D(s)}}{4\mu s} = \frac{\mu + 1 - \sqrt{(1 + \mu)^2 - 4\mu s}}{2\mu} \stackrel{(4.44)}{=} h(s). \end{aligned} \quad \triangle$$

Momente der Gesamtpopulation T – geometrischer Fall

Die Funktion h kann mithilfe des Lemmas 4.20 direkt für die Bestimmung der Momente von Y herangezogen werden. Da sie nun für den geometrischen Fall bekannt ist, sollen so $\mathbb{E}Y$ und EY^2 ermittelt werden.

Für $\mu \leq 1$ kann dann durch $\mathbb{E}Y$ direkt das schon im Satz 4.16 allgemein erzielte Resultat für den Erwartungswert von T bestätigt werden (da diesfalls $Y = T$ fast sicher gilt – siehe (4.37)). Anschließend wird für die subkritischen Fälle auf die gleiche Weise zusätzlich die Varianz von T bestimmt.

Unabhängig von μ gilt mit dem erwähnten Lemma 4.20 für den EW von Y

$$\mathbb{E}Y \stackrel{(2.13)}{\stackrel{(4.39)}{=}} h'(1^-) \stackrel{(4.48)}{=} \lim_{s \nearrow 1} \frac{1}{\sqrt{D(s)}} \stackrel{\text{falls } \mu \neq 1}{\stackrel{\text{ist (4.46)}}{=}} \frac{1}{|1 - \mu|}. \quad (4.50)$$

4 Anwendungsbeispiele

Daraus folgt für die kritischen und subkritischen geometrischen Prozesse mit $\mu \leq 1$

$$\mathbb{E}T \stackrel{(4.37)}{=} \mathbb{E}Y \stackrel{(4.43)}{=} \begin{cases} \lim_{s \nearrow 1} \frac{1}{2\sqrt{1-s}} = \lim_{n \rightarrow \infty} \frac{1}{2\sqrt{1-(1-\frac{1}{n})}} = \lim_{n \rightarrow \infty} \underbrace{\frac{1}{2\sqrt{\frac{1}{n}}}}_{\rightarrow 0} = \infty & \text{für } \mu=1, \\ \frac{1}{1-\mu} & \text{für } \mu < 1. \end{cases} \quad (4.51)$$

Dies geht konform mit der Aussage des Satzes 4.16. Wiederum für $\mu \neq 1$ gilt

$$\begin{aligned} \frac{2\mu}{|1-\mu|^3} \stackrel{(4.46)}{=} h''(1^-) &\stackrel{(4.39)}{=} h_Y''(1^-) \stackrel{(2.12)}{=} \mathbb{E}Y^2 - \mathbb{E}Y \stackrel{(4.50)}{=} \mathbb{E}Y^2 - \frac{1}{|1-\mu|} \\ \Rightarrow \mathbb{E}Y^2 &= \frac{2\mu}{|1-\mu|^3} + \frac{1}{|1-\mu|} = \frac{2\mu + (1-\mu)^2}{|1-\mu|^3} = \frac{\mu^2 + 1}{|1-\mu|^3} \quad \text{für } \mu \neq 1. \end{aligned} \quad (4.52)$$

Für die Varianz in subkritischen geometrischen Prozessen erhält man somit (alternativ gleich mit (2.14)) für $\mu < 1$

$$\mathbb{V}T \stackrel{(4.37)}{=} \mathbb{E}Y^2 - (\mathbb{E}Y)^2 \stackrel{(4.50)}{=} \stackrel{(4.52)}{=} \frac{\mu^2 + 1}{(1-\mu)^3} - \frac{1}{(1-\mu)^2} = \frac{\mu^2 + 1 - (1-\mu)}{(1-\mu)^3} = \frac{\mu(1+\mu)}{(1-\mu)^3}. \quad (4.53)$$

Verteilung der Gesamtpopulation T – geometrischer Fall

Die Verteilung bzw. W-Funktion von T ist gemäß dem Lemma 4.20 bzw. laut (4.42) jedenfalls (für *alle* μ) eindeutig durch die Verteilung von Y bestimmt. Deshalb soll als Nächstes versucht werden, die W-Funktion dieser $\mathbb{N}_0$ -wertigen ZV zu eruieren. Dafür soll abermals Gebrauch von ihrer nun bekannten weF h_Y bzw. der damit verbundenen Funktion h gemacht werden. Es wird sich zeigen, dass diesbezüglich im Fall geometrischer GWP die Folge der sogenannten *Catalan-Zahlen* (siehe die Def. A.107 im Appendix oder z.B. [33]) in Erscheinung tritt.

Definition 4.22 (Verteilung von Y)

Bezeichne die W-Funktion der $\mathbb{N}_0$ -wertigen ZV $Y \stackrel{(4.36)}{=} T \mathbb{1}_{\{T < \infty\}}$ als $\{r_k\}_{k \geq 0}$, d.h.

$$r_k := \mathbb{P}(Y = k) \quad \text{für } k \geq 0 \quad \Rightarrow r_k \stackrel{(4.40)}{=} \begin{cases} 1 - \eta & \text{für } k = 0, \\ \mathbb{P}(T = k) & \text{für } k \geq 1. \end{cases} \quad (4.54)$$

Um die explizite Verteilung im geometrischen Fall zu finden, wird zunächst das folgende Hilfsresultat und anschließend ein entsprechender Satz formuliert.

Lemma 4.23. *Es gilt ($n!!$ ist dabei die Doppelfakultät – siehe die Notation A.106)*

$$h^{(k)}(0) = 2^{k-1}(2k-3)!! \mu^{k-1}(1+\mu)^{-(2k-1)} \quad \text{für alle } k \geq 1. \quad (4.55)$$

Beweis. Unter anderem mithilfe der in der Bemerkung 4.21 gefundenen Resultate soll zunächst ein Muster für die Ableitungen von h gefunden werden. Dieses wird anschließend

induktiv bekräftigt. Für die ersten vier Ableitungen gilt auf $[0, 1)$

$$\begin{aligned}
 h' &\stackrel{(4.48)}{=} D^{-\frac{1}{2}} \stackrel{=-4\mu=-2\cdot 2\cdot \mu \text{ mit (4.48)}}{=} \\
 \Rightarrow h'' &= -\frac{1}{2} D^{-\frac{3}{2}} \overbrace{D'}^{=-4\mu=-2\cdot 2\cdot \mu} = 2 \cdot 1 \cdot \mu D^{-\frac{3}{2}} \quad (\text{vgl. auch (4.49)}) \\
 \Rightarrow h''' &= 2\mu \left(-\frac{3}{2}\right) D^{-\frac{5}{2}} (-4\mu) = 2^2 \cdot 3 \cdot 1 \cdot \mu^2 D^{-\frac{5}{2}} \\
 \Rightarrow h^{(4)} &= 2^2 \cdot 3\mu^2 \left(-\frac{5}{2}\right) D^{-\frac{7}{2}} (-4\mu) = 2^3 \cdot \underbrace{5 \cdot 3 \cdot 1}_{=5!! \text{ (A.113)}} \cdot \mu^3 D^{-\frac{7}{2}}.
 \end{aligned}$$

Damit liegt die folgende Vermutung nahe ($-1!! = 1$):

$$h^{(k)}(s) = 2^{k-1} (2k-3)!! \mu^{k-1} (D(s))^{-\frac{2k-1}{2}} \quad \text{für alle } s \in [0, 1) \text{ und } k \geq 1. \quad (\star)$$

Nehme um dieses Ableitungsschema induktiv zu bestätigen an, dass es auf $h^{(k-1)}$ für ein beliebiges $k \geq 2$ bereits zutrifft. D.h. auf $[0, 1)$ gilt dann

$$\begin{aligned}
 h^{(k-1)} &= 2^{k-2} (2k-5)!! \mu^{k-2} D^{-\frac{2k-3}{2}} \\
 \Rightarrow h^{(k)} &= (h^{(k-1)})' = 2^{k-2} (2k-5)!! \mu^{k-2} \left(-\frac{2k-3}{2}\right) D^{\overbrace{\left(-\frac{2k-3}{2}-1\right)}^{=-\frac{2k+3-2}{2}=-\frac{2k-1}{2}}} \underbrace{D'}_{=-4\mu=-2\cdot 2\cdot \mu \text{ (4.48)}} \\
 &= 2^{k-1} \underbrace{(2k-3)(2k-5)!!}_{=(2k-3)!! \text{ (A.113)}} \mu^{k-1} D^{-\frac{2k-1}{2}}
 \end{aligned}$$

wie in $(\star)$ behauptet.

Ausgewertet bei $s = 0$ liefert $(\star)$ mit $D(0) \stackrel{(4.45)}{=} (1 + \mu)^2$ wie gewünscht (4.55). $\square$

Die führenden Koeffizienten in (4.55) – d.h. $2^{k-1}(2k-3)!!$ – sind

$$\{2^{k-1}(2k-3)!!\}_{k \geq 1} = 1, 2, 12, 120, 1680, 30240, 665280, 17297280, \dots$$

Zur Berechnung von r_k werden diese jeweils durch $k!$ dividiert (vgl. den Satz 2.6), was zu den Koeffizienten $\left\{\frac{2^{k-1}(2k-3)!!}{k!}\right\}_{k \geq 1} = 1, 1, 2, 5, 14, 42, 132, 429, \dots$ führt. Diese Zahlenfolge erinnert stark an die Folge der *Catalan-Zahlen*, welche somit nun ins Spiel kommt. Das führt zum nächsten Resultat.

Satz 4.24. Für $r_k \stackrel{(4.54)}{=} \mathbb{P}(Y = k) \stackrel{(4.36)}{=} \mathbb{P}(T \mathbb{1}_{\{T < \infty\}} = k)$ gilt $r_0 \stackrel{(4.40)}{=} 1 - \eta$ sowie

$$r_k = \mathbb{P}(Y = k) \stackrel{(4.54)}{=} \mathbb{P}(T = k) = \underbrace{\frac{1}{k} \binom{2(k-1)}{k-1}}_{=: c_{k-1}} p^k q^{k-1} \quad \text{für alle } k \geq 1. \quad (4.56)$$

Dabei sind die $c_k = \frac{1}{k+1} \binom{2k}{k}$ für $k \geq 0$ sogenannte *Catalan-Zahlen* (siehe die Def. A.107).

Beweis. Sei $k \geq 1$ beliebig. Beachte zuerst, dass wegen $-(2k-1) = -k - (k-1)$

$$\mu^{k-1} (1 + \mu)^{-(2k-1)} = \left(\frac{\mu}{1 + \mu}\right)^{k-1} \left(\frac{1}{1 + \mu}\right)^k \stackrel{(4.4)}{=} q^{k-1} p^k \quad \text{gilt.} \quad (\star)$$

Stelle dann mit der im Appendix unter der Not. A.106 zu findenden Beziehung (A.114)

$$(2k-3)!! = (2(k-1)-1)!! \stackrel{(A.114)}{=} \frac{(2(k-1))!}{2^{k-1}(k-1)!} = \frac{(2k-2)!}{2^{k-1}(k-1)!} \quad \text{fest.} \quad (\star\star)$$

Damit und mit dem Satz 2.6 bzw. dem Lemma 4.23 erhält man für beliebige $k \geq 1$

$$\begin{aligned} r_k &\stackrel{(4.54)}{=} \mathbb{P}(Y = k) \stackrel{\substack{(2,10) \\ (4.36)}}{=} \frac{1}{k!} h_Y^{(k)}(0) \stackrel{(4.39)}{=} \frac{1}{k!} h^{(k)}(0) \stackrel{\substack{(*) \\ (4.55)}}{=} \frac{1}{k!} 2^{k-1} (2k-3)!! q^{k-1} p^k \\ &\stackrel{(**)}{=} \frac{1}{k!} \frac{(2k-2)!}{(k-1)!} p^k q^{k-1} = \frac{1}{k} \frac{(2k-2)!}{(k-1)!(k-1)!} p^k q^{k-1} = \frac{1}{k} \binom{2k-2}{k-1} p^k q^{k-1}, \end{aligned}$$

was wegen $2k - 2 = 2(k - 1)$ der Behauptung entspricht.

Bemerkung 4.25 (Überprüfung der W-Funktion $\{r_k\}_{k \geq 0}$). Es gilt also

$$\begin{array}{llll} r_1 = 1 & \cdot p & r_2 = 1 & \cdot p^2 q \\ r_3 = 2 & \cdot p^3 q^2 & r_4 = 5 & \cdot p^4 q^3 \\ r_5 = 14 & \cdot p^5 q^4 & r_6 = 42 & \cdot p^6 q^5 \\ r_7 = 132 & \cdot p^7 q^6 & r_8 = 429 & \cdot p^8 q^7 \end{array}$$

usw. Nun soll überprüft werden, ob es sich bei den $\{r_k\}_{k \geq 0}$ der Definition 4.22 in Verbindung mit dem Ergebnis des Satzes 4.24 tatsächlich um eine W-Funktion handelt. Dafür müssen zunächst alle Glieder dieser Folge nichtnegativ sein, was zutrifft. Wenn ihre Summe über alle $k \geq 0$ zudem 1 ergibt (dann liegen auch sämtliche in $[0, 1]$), sind alle Anforderungen, die an eine W-Funktion gestellt werden, erfüllt (siehe etwa die Def. A.40).

Um dies zu zeigen, wird die Reihe $\sum_{k \geq 0} r_k$ mittels des Lemmas A.108 im Appendix auf eine sogenannte *Binomialreihe* (siehe den Satz A.105) zurückgeführt. Betrachte zunächst

$$\begin{aligned}
\sum_{k \geq 1} r_k &\stackrel{(4.56)}{=} \sum_{k \geq 1} c_{k-1} p^k q^{k-1} = \sum_{k \geq 0} c_k p^{k+1} q^k = p \sum_{k \geq 0} c_k (pq)^k \stackrel{(A.115)}{=} p \sum_{k \geq 0} \frac{1}{k+1} \binom{2k}{k} (pq)^k \\
&\stackrel{(A.116)}{=} p \sum_{k \geq 0} (-1)^k \overbrace{2^{2k+1}}^{\frac{1}{2}} \binom{\frac{1}{2}}{k+1} (pq)^k = 2p \sum_{k \geq 0} \binom{\frac{1}{2}}{k+1} (-4pq)^k \\
&= -\frac{2p}{4pq} \sum_{k \geq 0} \binom{\frac{1}{2}}{k+1} (-4pq)^{k+1} = -\frac{1}{2q} \sum_{k \geq 1} \underbrace{\binom{\frac{1}{2}}{k}}_{=1 \text{ für } k=0} (-4pq)^k \\
&= -\frac{1}{2q} \left(\sum_{k \geq 0} \left(\binom{\frac{1}{2}}{k} (-4pq)^k \right) - 1 \right) = \frac{1}{2q} - \frac{1}{2q} \sum_{k \geq 0} \binom{\frac{1}{2}}{k} \overbrace{(-4pq)^k}^{\geq -4 \cdot \frac{1}{2} \cdot \frac{1}{2} = -1} \\
&\stackrel{\text{Binomialreihe}}{=} \frac{1}{2q} - \frac{1}{2q} (1 - 4pq)^{\frac{1}{2}} \stackrel{(4.1)}{=} \frac{1 - \sqrt{1 - 4p(1-p)}}{2q} = \frac{1 - \sqrt{(1-2p)^2}}{2q} \\
&\stackrel{((A.112) \text{ mit } \alpha > 0)}{=} \begin{cases} \frac{1 - (2p-1)}{2q} \stackrel{(4.1)}{=} \frac{2-2p}{2(1-p)} = 1 \stackrel{(4.11)}{=} \eta, & \text{falls } p \geq \frac{1}{2} \text{ (} \stackrel{(4.8)}{\iff} \mu \leq 1 \text{) ist,} \\ \frac{1 - (1-2p)}{2q} = \frac{2p}{2q} = \frac{p}{q} \stackrel{(4.3)}{=} \frac{1}{\mu} \stackrel{(4.11)}{=} \eta, & \text{falls } p < \frac{1}{2} \text{ (} \stackrel{(4.8)}{\iff} \mu > 1 \text{) ist.} \end{cases}
\end{aligned}$$

Unabhängig von μ gilt also jedenfalls $\sum_{k>1} r_k = \eta$ ($\star$), womit

$$\sum_{k>0} \mathbb{P}(Y = k) \stackrel{(4.54)}{=} \sum_{k>0} r_k = r_0 + \sum_{k>1} r_k \stackrel{(4.54)}{\stackrel{(\star)}{=}} (1 - \eta) + \eta = 1 \quad \text{folgt.}$$

4.1 Einfache GW-Verzweigungsprozesse – geometrischer Fall

Dabei befindet sich bei subkritischen und kritischen Prozessen wegen $\eta = 1$ die gesamte Wahrscheinlichkeitsmasse von Y und somit auch jene von T auf $\mathbb{N}$. Im superkritischen Fall hingegen liegen sowohl Y also auch T nur mit der AWK η in $\mathbb{N}$ und im Überlebensfall (also mit einer WK von $1 - \eta$) gilt $Y = 0$ bzw. $T = \infty$ (was erwartungsgemäß mit den Aussagen des Lemmas 4.20 bzw. mit jener von (4.42) übereinstimmt). $\triangle$

Mit dem Satz 4.24 kann also – in Verbindung mit (4.42) – die Verteilung bzw. W-Funktion von T für jeden geometrischen Fall berechnet werden. Diese Resultate wurden anhand verschiedener Simulationen einer Praxisprüfung unterzogen. Für die Tabelle 4.5 wurde die Berechnung etwa für subkritische (mit $\mu = \frac{2}{3}$), kritische (also $\mu = 1$) und superkritische (mit $\mu = \frac{3}{2}$) geometrische Prozesse durchgeführt und jeweils entsprechenden Simulationsergebnissen (bis zur 40. Generation) gegenübergestellt.

μ		r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9
$\frac{2}{3}$	berechnet*	0,600	0,144	0,069	0,041	0,028	0,020	$1,51 \cdot 10^{-2}$	$1,18 \cdot 10^{-2}$	$9,44 \cdot 10^{-3}$
	simuliert**	0,606	0,142	0,067	0,038	0,028	0,020	$1,48 \cdot 10^{-2}$	$1,17 \cdot 10^{-2}$	$1,00 \cdot 10^{-2}$
1	berechnet*	0,500	0,125	0,062	0,039	0,027	0,021	$1,61 \cdot 10^{-2}$	$1,31 \cdot 10^{-2}$	$1,09 \cdot 10^{-2}$
	simuliert**	0,501	0,128	0,059	0,038	0,028	0,020	$1,47 \cdot 10^{-2}$	$1,38 \cdot 10^{-2}$	$1,11 \cdot 10^{-2}$
$\frac{3}{2}$	berechnet*	0,400	0,096	0,046	0,028	0,019	0,013	$1,01 \cdot 10^{-2}$	$7,87 \cdot 10^{-3}$	$6,30 \cdot 10^{-3}$
	simuliert**	0,396	0,098	0,046	0,027	0,016	0,014	$1,02 \cdot 10^{-2}$	$7,90 \cdot 10^{-3}$	$6,90 \cdot 10^{-3}$

Tabelle 4.5: Auszüge der Verteilung bzw. W-Funktion $\{\mathbb{P}(T = k)\}_{k \geq 0} \stackrel{(4.42)}{=} \{r_k\}_{k \geq 0}$ für $1 \leq k \leq 9$ – geometrische GWP mit $\mu \in \{\frac{2}{3}, 1, \frac{3}{2}\}$ (*Berechnung mit (4.56); **Anteil an jeweils 20 000 Simulationen, in welchen T als k realisiert wurde).

Die Tabelle 4.6 vergleicht die theoretische (kumulierte) Verteilungsfunktion von T (siehe etwa die Def. A.29 (iv)) mit den beobachteten Ergebnissen. Im ersten Simulationsdatensatz mit $\mu = \frac{2}{3}$ waren alle Pfade nach spätestens 22 Generationen ausgestorben (überhaupt nur einer erreichte eine Gesamtpopulation von über 100 – genauer: 116). Auch deshalb liegen der MW (2,983) und die Stichprobenvarianz (28,696) bereits zu diesem Zeitpunkt in der Größenordnung der in diesem Fall endlichen theoretischen Momente ($\mathbb{E}T = 3$ und $\mathbb{V}T = 30$ mit $\mu = \frac{2}{3}$ – siehe (4.51) bzw. (4.53)).

μ		$\mathbb{P}(T \leq 1)$	$\mathbb{P}(T \leq 5)$	$\mathbb{P}(T \leq 10)$	$\mathbb{P}(T \leq 15)$	$\mathbb{P}(T \leq 25)$	$\mathbb{P}(T \leq 40)$	$\mathbb{P}(T \leq 100)$	$\mathbb{P}(T \leq 200)$	$\mathbb{P}(T > 200)$	$\mathbb{P}(T = \infty)$
$\frac{2}{3}$	berechnet*	0,600	0,882	0,947	0,970	0,988	0,996	1,000	1,000	$7,28 \cdot 10^{-7}$	0
	simuliert**	0,606	0,882	0,947	0,970	0,988	0,997	1,000	1,000	$0,00 \cdot 10^0$	-
1	berechnet*	0,500	0,754	0,824	0,856	0,888	0,911	0,944	0,960	$3,99 \cdot 10^{-2}$	0
	simuliert**	0,501	0,755	0,825	0,855	0,888	0,912	0,945	0,961	$3,89 \cdot 10^{-2}$	-
$\frac{3}{2}$	berechnet*	0,400	0,588	0,631	0,647	0,659	0,664	0,667	0,667	$3,33 \cdot 10^{-1}$	0,333
	simuliert**	0,396	0,583	0,626	0,642	0,654	0,661	0,663	0,663	$3,37 \cdot 10^{-1}$	-

Tabelle 4.6: Ausgewählte Werte der Verteilungsfunktion von T bzw. kumulierte relative Häufigkeiten entsprechender Realisierungen – geometrische GWP mit $\mu \in \{\frac{2}{3}, 1, \frac{3}{2}\}$ (*Berechnung mit (4.56) bzw. letzte Spalte mit (4.11); **jeweils 20 000 Simulationen).

4 Anwendungsbeispiele

Für $\mu = 1$ und $\mu = \frac{3}{2}$ ist anhand dieser Daten eine ähnliche Schätzung der diesfalls ohnehin nicht existierenden Momente nicht möglich (letzterer Datensatz ist jener, der auch für das Bsp. 4.9 verwendet wurde – für die mittlere Gesamtpopulation seiner bereits ausgestorbener Pfade siehe 4.1.5, S. 129). Im kritischen Fall waren nach 40 Generationen zwar bereits 97,55 % der Pfade ausgestorben, es ist aber dennoch eine mittlere Gesamtpopulation von über 40 und eine maximale von 6 477 zu beobachten. Beide Werte stiegen bei Versuchen mit einer immer höheren Anzahl an simulierten Generationen unaufhörlich – genau so, wie aufgrund der theoretischen Resultate zu erwarten war.

Konvergenz von h_n im Fall subkritischer geometrischer GWP

Als Ergänzung wird die Konvergenz der Folge $\{h_n\}_{n \geq 0}$ im Folgenden explizit für subkritische geometrische Prozesse auf eine andere Art gezeigt. Sie wurde bereits im Satz 4.18 allgemein bewiesen, hier kommt man aber ohne den Satz der monotonen Konvergenz aus. Diese Vorgehensweise ist inspiriert von [29, 3.11, 8.–10., S. 69].

Sei dafür also $\mu < 1$. Für $n \geq 0$ gilt für die weF h_n (der Gesamtpopulationsgröße bis zur n -ten Generation T_n) gemäß dem Lemma 4.15 (ii) bzw. (4.29) die Rekursion $h_{n+1}(s) = sf(h_n(s))$ für alle $s \in [0, 1]$. In der Nebenrechnung B.12 wird (allgemein, für beliebige μ) gezeigt, dass für alle $s \in [0, 1]$

$$|h_{n+1}(s) - h_n(s)| \leq \mu |h_n(s) - h_{n-1}(s)| \quad \text{für alle } n \geq 1 \text{ gilt.} \quad (4.57)$$

Sei nun $n \geq 0$ beliebig. Es gilt für alle $s \in [0, 1]$ (Darstellung nur für $n \geq 2$ sinnvoll)

$$\begin{aligned} |h_{n+1} - h_n| &\stackrel{(4.57)}{\leq} \mu |h_n - h_{n-1}| \stackrel{(4.57)}{\leq} \mu^2 |h_{n-1} - h_{n-2}| \leq \dots \leq \mu^n |h_1 - h_0| \Rightarrow \\ |h_{n+1}(s) - h_n(s)| &\leq \mu^n |h_1(s) - h_0(s)| \stackrel{(4.29)}{=} \mu^n |sf(s) - s| = \mu^n \underbrace{|s(f(s) - 1)|}_{\leq 1 \text{ für } s \in [0, 1]} \leq \mu^n. \end{aligned} \quad (4.58)$$

Weiters erhält man mit der Dreiecksungleichung

$$\begin{aligned} |h_{n+2} - h_n| &= |h_{n+2} - h_{n+1} + h_{n+1} - h_n| \leq |h_{n+2} - h_{n+1}| + |h_{n+1} - h_n| \\ &\stackrel{(4.57)}{\leq} \mu |h_{n+1} - h_n| + |h_{n+1} - h_n| = (\mu + 1) |h_{n+1} - h_n|. \end{aligned}$$

Analog ergibt sich $|h_{n+3} - h_n| \leq (\mu^2 + \mu + 1) |h_{n+1} - h_n|$ und allgemein für $m > n$

$$|h_m - h_n| \leq \sum_{i=0}^{m-n-1} \mu^i |h_{n+1} - h_n| \stackrel{\substack{\text{geom.} \\ \text{Reihe} \\ (\mu < 1)}}{\leq} \frac{1}{1 - \mu} |h_{n+1} - h_n| \stackrel{(4.58)}{\leq} \frac{\mu^n}{1 - \mu}.$$

Da $\mu < 1$ ist, kann der letzte (nichtnegative) Ausdruck und folglich auch $|h_m - h_n|$ beliebig klein gemacht werden, wenn nur n hinreichend groß gewählt wird. Somit handelt es sich um eine *Cauchy-Folge* und $\{h_n\}_{n \geq 0}$ konvergiert – vgl. den Satz A.94.

Diese Konvergenz gilt für jeden Punkt $s \in [0, 1]$, d.h. $l(s) := \lim_{n \rightarrow \infty} h_n(s)$ existiert (und muss ebenso in $[0, 1]$ liegen). Es kann genau so wie in (4.35) argumentiert werden, dass diese Grenzwerte jeweils der Funktionalgleichung $l(s) = s(f \circ l)(s)$ genügen müssen. Nun hat diese Gleichung gemäß der Bem. zum Satz 4.17 nur eine Lösung in $[0, 1]$, weshalb $l = h$ gelten muss. Alternativ besagt dies eben auch der Satz 4.18 direkt.

4.1.5 Der auf das Aussterben bedingte Prozess

Nun soll noch kurz eine 'neuer' Prozess untersucht werden, welcher durch das *Bedingen auf das Ereignis des Aussterbens* kreiert werden kann. Dies ist gewissermaßen ähnlich zu Teilen des Unterabschnitts 4.1.2, wo umgekehrt das Verhalten von $\{W_n\}_{n \geq 0}$ (bzw. W) u.a. bedingt auf das *Nicht-Aussterben* im Fokus stand. Wie dort gilt das nun Folgende prinzipiell für alle Werte von μ , hat aber eigentlich nur in superkritischen Fällen eine Relevanz (siehe die Bemerkung 4.30).

Exkurs: Der auf das Aussterben bedingte Prozess – allgemein

Dieser Teilabschnitt bezieht sich nicht nur auf geometrische GWP (wie der Exkurs 4.1.3).

Definition 4.26 (Der auf das Aussterben bedingte Prozess)

Definiere für jedes $n \geq 0$

$$\tilde{Z}_n := (Z_n | Z_n \rightarrow 0) \stackrel{(4.28)}{=} (Z_n | T < \infty) \quad \text{sowie} \quad \tilde{T} := (T | T < \infty). \quad (4.59)$$

Der Prozess $\{\tilde{Z}_n\}_{n \geq 0}$ wird als *auf das Aussterben bedingter Prozess* bezeichnet.

Nach dem gleich folgenden Hilfsresultat (bzgl. $\{Z_n\}_{n \geq 0}$) wird im Lemma 4.28 die Verteilung und die weF der Populationsgrößen von $\{\tilde{Z}_n\}_{n \geq 0}$ bestimmt.

Lemma 4.27. Für alle $k, n \geq 0$ gilt $\mathbb{P}(Z_n = k, T < \infty) = \mathbb{P}(Z_n = k)\eta^k$.

Beweis. Die Aussage ist intuitiv verständlich, da aufgrund der Unabhängigkeit der Stränge von $Z_n = k$ ausgehend sozusagen k unabhängige Prozesse aussterben müssen. Deshalb potenziert sich die AWK entsprechend. Für ein formales Argument seien $k, n \geq 0$ beliebig. Es gilt

$$\begin{aligned} \mathbb{P}(Z_n = k, T < \infty) &\stackrel{(4.28)}{=} \mathbb{P}\left(Z_n = k, Z_{n+m} \xrightarrow{(m \rightarrow \infty)} 0\right) \stackrel{(*)}{=} \mathbb{P}\left(Z_n = k, \sum_{i=1}^k Z_{m,i} \rightarrow 0\right) \\ &\stackrel{\substack{Z_n, \{Z_{m,i}\}_{i \geq 0} \\ \text{uabh.}}}{=} \mathbb{P}(Z_n = k) \mathbb{P}\left(\bigcap_{i=1}^k \{Z_{m,i} \rightarrow 0\}\right) \stackrel{(i.i.d.)}{=} \mathbb{P}(Z_n = k) \prod_{i=1}^k \underbrace{\mathbb{P}(Z_{m,i} \rightarrow 0)}_{= \eta \text{ ((2.67) u. (2.69))}}. \end{aligned}$$

(*) Betrachte für ein beliebiges $m \geq 0$ wie in (2.47) $Z_{n+m} \stackrel{d}{=} \sum_{i=1}^{Z_n} Z_{m,i}$, wobei die $\{Z_{m,i}\}_{i \geq 1}$ i.i.d.-Kopien von Z_m seien. $\square$

Lemma 4.28 (Verteilung / weF der $\tilde{Z}_n$ – angelehnt an [17, Exercise 6.7.2, S. 246])
Es gilt für alle $k, n \geq 0$

$$\mathbb{P}(\tilde{Z}_n = k) = \mathbb{P}(Z_n = k)\eta^{k-1} \quad \text{und deshalb insbesondere} \quad (4.60)$$

$$\tilde{p}_k := \mathbb{P}(\tilde{Z}_1 = k) = \mathbb{P}(Z_1 = k)\eta^{k-1} \stackrel{\substack{Z_1 \stackrel{d}{=} X \\ (2.33)}}{=} p_k \eta^{k-1}. \quad \text{Weiters gilt} \quad (4.61)$$

$$\tilde{f}_n(s) := f_{\tilde{Z}_n}(s) = \frac{1}{\eta} f_n(s\eta) \quad \text{für alle } n \geq 0 \text{ und } s \in [0, 1]. \quad (4.62)$$

4 Anwendungsbeispiele

Beweis. Seien $k, n \geq 0$ beliebig. Gemäß dem Lemma 4.27 gilt

$$\mathbb{P}(\tilde{Z}_n = k) \stackrel{(4.59)}{=} \mathbb{P}(Z_n = k | T < \infty) = \frac{\mathbb{P}(Z_n = k, T < \infty)}{\mathbb{P}(T < \infty)} \stackrel{\text{Lemma 4.27}}{\stackrel{(4.28)}}{=} \frac{\mathbb{P}(Z_n = k)\eta^k}{\eta},$$

womit (4.60) folgt. Daraus ergibt sich mit $n = 1$ direkt (4.61). Für (4.62) berechne

$$f_{\tilde{Z}_n}(s) \stackrel{(2.6)}{=} \sum_{k \geq 0} \mathbb{P}(\tilde{Z}_n = k) s^k \stackrel{(4.60)}{=} \frac{1}{\eta} \sum_{k \geq 0} \mathbb{P}(Z_n = k) (s\eta)^k \stackrel{(2.6)}{\stackrel{(2.46)}}{=} \frac{1}{\eta} f_n(s\eta). \quad \square$$

Beim Prozess $\{\tilde{Z}_n\}_{n \geq 0}$ handelt es sich wiederum um einen GWP.

Satz 4.29 ($\{\tilde{Z}_n\}_{n \geq 0}$ ist ein GW-Prozess – vgl. z.B. [3, Theorem I.8.3, S. 52 f.])

Der Prozess $\{\tilde{Z}_n\}_{n \geq 0}$ ist ein GWP mit Nachkommensverteilung $\{\tilde{p}_k\}_{k \geq 0}$ und weF der Populationsgrößen $\{\tilde{f}_n\}_{n \geq 0}$ wie im Lemma 4.28 definiert. Insbesondere gilt für die korrespondierende Nachkommenszahl $\tilde{X}$ und deren weF $\tilde{f}$ (für alle $k \geq 0$ bzw. $s \in [0, 1]$)

$$\mathbb{P}(\tilde{X} = k) \stackrel{(2.33)}{=} \tilde{p}_k \stackrel{(4.61)}{=} p_k \eta^{k-1}, \quad \tilde{f}(s) = \frac{1}{\eta} f(s\eta) \quad \text{und} \quad \tilde{\mu} := \mathbb{E}[\tilde{X}] = f'(\eta). \quad (4.63)$$

Dabei beziehen sich alle Benennungen mit ' $\sim$ ' auf den 'neuen', den auf das Aussterben bedingten Prozess und solche ohne ' $\sim$ ' auf den ursprünglichen Prozess $\{Z_n\}_{n \geq 0}$. Ist letzter superkritisch ($\mu > 0$), so ist $\{\tilde{Z}_n\}_{n \geq 0}$ subkritisch, d.h. $\tilde{\mu} < 1$.

Beweis. Wenn $\mu \leq 1$ ist, gilt $\mathbb{P}(T < \infty) \stackrel{(4.28)}{=} 1 \Rightarrow \{\tilde{Z}_n\}_{n \geq 0} \stackrel{(4.59)}{=} \{Z_n | T < \infty\}_{n \geq 0} \stackrel{d}{=} \{Z_n\}_{n \geq 0}$. Deshalb sind dann beide Prozesse äquivalent und die Aussage und alle Beziehungen in (4.63) gelten trivialerweise (wegen $\eta = 1$ – siehe auch die Bem. 4.30).

Sei also $\mu > 1$, d.h. der ursprüngliche Prozess superkritisch. Diesfalls findet sich ein Beweis dafür, dass der 'neue' Prozess ein GWP ist (und auch für die ersten beiden Gleichungen in (4.63)) in der NR B.13. Beachte zudem, dass (f' ist strikt isoton – Satz 2.5 (e))

$$\tilde{\mu} \stackrel{(2.53)}{=} \tilde{f}'(1^-) \stackrel{(4.63)}{=} \lim_{s \nearrow 1} \left(\frac{1}{\eta} f(s\eta) \right)' = \frac{1}{\eta} \lim_{s \nearrow 1} (f(s\eta))' \eta = f'(\eta) \stackrel{\eta < 1}{<} 1 \quad \text{gilt.} \quad \square$$

Nun können $\mathbb{E}\tilde{T}$ und $\mathbb{V}\tilde{T}$ wie folgt aus $\mathbb{E}Y$ und $\mathbb{E}Y^2$ gewonnen werden:

$$\mathbb{E}\tilde{T} \stackrel{(4.59)}{=} \mathbb{E}[T | T < \infty] \stackrel{(A.79)}{=} \frac{\mathbb{E}[T \mathbf{1}_{\{T < \infty\}}]}{\mathbb{P}(T < \infty)} \stackrel{(4.36)}{\stackrel{(4.28)}}{=} \frac{\mathbb{E}Y}{\eta} \quad \text{und analog} \quad (4.64)$$

$$\mathbb{V}\tilde{T} = \mathbb{E}[T^2 | T < \infty] - (\mathbb{E}[T | T < \infty])^2 = \frac{\mathbb{E}Y^2}{\eta} - \frac{(\mathbb{E}Y)^2}{\eta^2}. \quad (4.65)$$

Bemerkung 4.30. Wie bereits zu Beginn des obigen Beweises festgestellt, entsteht in subkritischen und kritischen Fällen mit $\{\tilde{Z}_n\}_{n \geq 0}$ eigentlich gar kein 'neuer' Prozess. Es stimmen also (wegen $\eta = 1$) auch die Aussagen des Lemmas 4.28 sowie (4.64) und (4.65) trivialerweise. Die Betrachtung von $\{\tilde{Z}_n\}_{n \geq 0}$ ist also nur in superkritischen Fällen sinnvoll. $\triangle$

Anwendung im superkritischen geometrischen Fall

Betrachte nun wieder einen *geometrischen* GWP $\{Z_n\}_{n \geq 0}$, welcher superkritisch sein soll. Folglich sei $\mu > 1$. Der daraus lt. der Def. 4.26 gebildete, auf das Aussterben bedingte Prozess $\{\tilde{Z}_n\}_{n \geq 0}$ muss also gemäß dem Satz 4.29 ein subkritischer GWP sein. Basierend auf $f'(s) \stackrel{(4.5)}{=} \frac{\mu}{(1+\mu-\mu s)^2}$ kann zunächst für alle $s \in [0, 1]$ bestimmt werden, dass

$$\tilde{\mu} \stackrel{(4.63)}{=} f'(\eta) \stackrel{(4.11)}{=} f'\left(\frac{1}{\mu}\right) = \frac{\mu}{(1+\mu-1)^2} = \frac{1}{\mu} \quad \text{gilt} \quad \Rightarrow \tilde{\mu} < 1. \quad (4.66)$$

Weiters erhält man für die weF von $\{\tilde{Z}_n\}_{n \geq 0}$ für alle $s \in [0, 1]$

$$\tilde{f}(s) \stackrel{(4.11)}{\stackrel{(4.63)}{=}} \mu f\left(s \frac{1}{\mu}\right) \stackrel{(4.5)}{=} \mu \frac{1}{1+\mu-s} = \frac{1}{\frac{1}{\mu}+1-\frac{s}{\mu}} \stackrel{(4.66)}{=} \frac{1}{1+\tilde{\mu}-\tilde{\mu}s} \quad (\text{vgl. (4.5)}).$$

Der Prozess lässt sich natürlich vollkommen durch die weF seiner Nachkommensverteilung $\tilde{f}$ beschreiben. Deshalb kann man daraus direkt schließen, dass es sich bei $\{\tilde{Z}_n\}_{n \geq 0}$ um einen *subkritischen geometrischen GWP* mit $\tilde{\mu} = \frac{1}{\mu} < 1$ handelt. Seine Nachkommenszahl ist geometrisch verteilt mit dem Parameter $\tilde{p}$, d.h. $\tilde{X} \sim \text{Geom}(\tilde{p})$, wobei $\tilde{p} \stackrel{(4.4)}{=} \frac{1}{1+\tilde{\mu}} \stackrel{(4.66)}{=} \frac{1}{1+\frac{1}{\mu}} = \frac{\mu}{\mu+1} \stackrel{(4.4)}{=} q$ ist.

Möchte man etwa die Momente der korrespondierenden Gesamtpopulation $\tilde{T}$ erhalten, können direkt (4.51) bzw. (4.53) aus dem vorherigen Unterabschnitt angewendet werden. Alternativ gelangt man mit den vorher allgemeiner bestimmten Beziehungen zu

$$\begin{aligned} \mathbb{E}\tilde{T} &\stackrel{(4.11)}{\stackrel{(4.64)}{=}} \mu \mathbb{E}Y \stackrel{(4.50)}{\stackrel{\mu > 1}{=}} \frac{\mu}{\mu-1} = \frac{1}{1-\frac{1}{\mu}} \stackrel{(4.66)}{=} \frac{1}{1-\tilde{\mu}} \quad (\text{vgl. (4.51) für } \mu < 1), \\ \mathbb{V}\tilde{T} &\stackrel{(4.11)}{\stackrel{(4.65)}{=}} \mu \mathbb{E}Y^2 - \mu^2 (\mathbb{E}Y)^2 \stackrel{(4.50)}{\stackrel{(4.52)}{=}} \frac{\mu^3 + \mu}{(\mu-1)^3} - \frac{\mu^2}{(\mu-1)^2} = \frac{\mu^3 + \mu - \mu^2(\mu-1)}{(\mu-1)^3} \\ &= \frac{\mu^2 + \mu}{(\mu-1)^3} = \frac{\mu(\mu+1)}{(\mu-1)^3} \cdot \frac{\frac{1}{\mu^3}}{\frac{1}{\mu^3}} = \frac{\frac{1}{\mu}(1+\frac{1}{\mu})}{(1-\frac{1}{\mu})^3} \stackrel{(4.66)}{=} \frac{\tilde{\mu}(1+\tilde{\mu})}{(1-\tilde{\mu})^3} \quad (\text{vgl. (4.53)}). \end{aligned}$$

Die letzten beiden Resultate können etwa mit dem schon im Bsp. 4.9 (und bei den Tabellen 4.5 und 4.6) erörterten Datensatz abgeglichen werden ($\mu = \frac{3}{2}$; 20 000 Simulationen über 40 Generationen).

Bei jenem waren 6 735 der Realisierungen in der 40. Generation bereits ausgestorben (ein späteres Aussterben der noch lebenden ist – wie im Bsp. besprochen – nicht mehr zu erwarten). Unter diesen konnte eine mittlere Population von 3,056 bei einer Stichprobenvarianz von 31,479 verzeichnet werden ($\mathbb{E}\tilde{T} = 3$ und $\mathbb{V}\tilde{T} = 30$) – also vergleichbar mit den auf der S. 125 berichteten Ergebnissen bzgl. des subkritischen Pendants mit $\mu = \frac{2}{3}$.

4.2 Multitype GW-Verzweigungsprozesse mit R

Es sind diverse Pakete für die softwaregestützte Analyse unterschiedlicher Verzweigungsprozesse mit R online abrufbar. In Bezug auf die in dieser Arbeit besprochenen Prozesse ist im Besonderen das R-Paket '**Branching**' (Version 0.9.4 – siehe [57]) zu nennen. Dieses ist auf Multitype GWP¹ maßgeschneidert, ermöglicht diesbezügliche Parameterschätzungen anhand von Daten und stellt eine recht universale Simulationsfunktion bereit. Für Simulationen ist die Eingabe der Reproduktionsverteilungen erforderlich.

Das an dieser Stelle verfolgte Ziel ist aber nicht die Arbeit mit 'echten' Daten – vielmehr sollen die theoretischen Resultate der vorherigen Kapitel weiter analysiert werden können. Lediglich für die Erzeugung exemplarischer Realisationen wurde dabei auf die letztgenannte Funktion in '**Branching**' zurückgegriffen.

Um weitere Ergebnisse im Notationsrahmen des 3. Kapitels nachvollziehen zu können, wurde ein kleines Bündel eigener R-Funktionen erstellt. Dabei sollte u.a. auch eine Möglichkeit geschaffen werden, die weF der Reproduktionsverteilungen als Ausgangspunkt nehmen zu können. Die angefertigten Routinen werden sogleich vorgestellt und deren Funktionsweise anschließend unter 4.2.2 an einem konkreten Beispiel vorgeführt.

4.2.1 R-Funktionen zur Analyse eines Multitype GWP

Nun sollen die erwähnten R-Funktionen beschrieben werden. Die folgende Dokumentation orientiert sich zwar an üblichen Konventionen, erhebt aber nicht den Anspruch, einer standardmäßigen R-Paketdokumentation zu genügen.

Sie ist wie folgt gegliedert: Zunächst werden Funktionen aufgelistet, die ausgehend vom weF-Vektor $\mathbf{f}$ eines MGWP weitere Dinge wie die iterierten weF $\{\mathbf{f}_n\}_{n \geq 0}$, die W-Funktion $\mathbf{p}$, die Erwartungswertmatrix $\mathbf{M}$ oder den AWK-Vektor $\boldsymbol{\eta}$ berechnen können. Dafür ist die Vorgabe einer initialen Population nicht notwendig. Routinen für die Berechnung der Momente der Populationsgrößen bei spezifischen Anfangspopulationen werden anschließend vorgestellt. Zuletzt sind solche angeführt, mit welchen es möglich sein soll, die Simulationsfunktion aus '**Branching**' zu starten und anschließend die Simulationsergebnisse zu analysieren.

Für eine Übersicht siehe die Liste dokumentierter Funktionen im Appendix B.

Bemerkung 4.31 (Einbindung der dokumentierten R-Funktionen) Prinzipiell besteht die Möglichkeit, diese Funktionen zukünftig in ein eigenes R-Paket einzugliedern. Sie sind u.a. auch als Grundlage bzw. Vorarbeit für einen etwaigen Ausbau zu einem vollwertigen Paket gedacht.

Einstweilen werden sie aber über eine Quelldatei in die globale Umgebung (das *Environment* `R_GlobalEnv`) geladen. Es sind auch weitere – nicht im Detail dokumentierte – sogenannte 'Helferfunktionen' notwendig. Eine kurze Handlungsanleitung für die Einbindung aller Funktionen und anderer Abhängigkeiten findet sich unter B.2.2. Dort kann auch der Quellcode der Haupt- und Hilfsfunktionen nachgeschlagen werden. $\triangle$

¹In '**Branching**' werden MGWP als Multitype *Bienaymé-Galton-Watson*-Prozesse bezeichnet – kurz BGWM (vgl. die Fußnote ² auf der Seite 1).

R-Funktionen für die Analyse der weF eines MGWP

<code>wef.create</code>	Umwandlung einer (multivariaten) weF und Bereitstellung diverser Funktionen (der Klassen <code>function</code> bzw. <code>character</code>).
-------------------------	---

Verwendung

```
wef.create(f.str, save = FALSE)
```

Argumente

<code>f.str</code>	WeF aller Reproduktionszahlen $(f^{(1)}, \dots, f^{(k)})^T \stackrel{(3.38)}{=} \mathbf{f}$. Eingabe als Zeichenketten (<i>strings</i>) in einem <code>character</code> -Vektor der Länge k , wobei das j -te Element $f^{(j)}(\mathbf{s})$ repräsentiert – siehe (3.38). Die Komponenten von $\mathbf{s} = (s_1, \dots, s_k)^T$ müssen als <code>s1</code> , ..., <code>sk</code> geschrieben werden (vgl. z.B. die R-Ausgabe 4.1).
<code>save</code>	<code>logical</code> -Variable. Wenn sie <code>TRUE</code> ist, werden die unten angegebenen Ausgabevariablen direkt in dem <i>Environment</i> gespeichert, in welchem <code>wef.create</code> aufgerufen wird.

Ausgabewert

Falls `save = FALSE` ist, ein Listenobjekt (Klasse `list`) mit den folgenden Komponenten:

<code>fj.list</code>	<code>list</code> -Objekt der skalarwertigen Funktionen $f^{(j)}$, Inputs $s_1, \dots, s_k$.
<code>fj</code>	Skalarwertige Funktion $f^{(j)}$ einzeln, Inputs $s_1, \dots, s_k$ und j .
<code>f</code>	k -dimensionale vektorwertige Funktion f , Inputs $s_1, \dots, s_k$ und j .
<code>fj.vec.list</code>	Analog zu den obigen drei Objekten, allerdings mit Funktionsinput $\mathbf{s}$
<code>fj.vec</code> , <code>f.vec</code>	(also einem k -dimensionalen Vektor).
<code>f.str.vec</code>	<code>character</code> -Vektor wie <code>f.str</code> , nur mit <code>s[j]</code> anstelle von <code>sj</code> .
<code>k</code>	Anzahl der unterschiedlichen Typen (Länge von <code>f.str</code>).

Details

Im Zuge der Umwandlungen wird auch überprüft, ob die $f^{(j)}$ in `f.str` überhaupt als weF in Frage kommen, d.h. ob $f^{(j)}(\mathbf{1}) = 1$ für alle $1 \leq j \leq k$ erfüllt ist (Satz 3.10 (c)). $\triangle$

<code>wfkt.calc</code>	W-Funktion $\mathbf{p}$ aus den weF $f^{(j)}$ berechnen (mit dem Satz 3.10 (b)).
------------------------	--

Verwendung

```
wfkt.calc(f.str, max.order = 10, red = FALSE, time = FALSE,
          time.break = 20, loop = TRUE, tol.loop = 1e-04)
```

Argumente

<code>f.str</code>	WeF $\mathbf{f} \stackrel{(3.38)}{=} (f^{(1)}, \dots, f^{(k)})^T$ als <code>character</code> -Vektor (siehe <code>wef.create</code>).
<code>max.order</code>	Nichtnegativer <code>integer</code> – maximale Ordnung für die Bestimmung partieller Ableitungen bzgl. einzelner Komponenten (von $\mathbf{s}$).
<code>red</code>	<code>logical</code> – wenn <code>TRUE</code> , werden nur $\mathbf{p}(\mathbf{r}) > 0$ ausgegeben.
<code>time</code>	<code>logical</code> – wenn <code>TRUE</code> , wird die benötigte Zeit berichtet.

4 Anwendungsbeispiele

<code>time.break</code>	Positiver Wert in Sekunden – Berechnung bricht ab, wenn die Dauer nach einem Durchlauf (siehe Details) überschritten ist.
<code>loop</code>	<code>logical</code> – Durchläufe wenn <code>TRUE</code> über <code>for</code> -Schleife, sonst mit <code>sapply</code> .
<code>tol.loop</code>	Positiver Wert in $[0, 1]$. Berechnung wird vorzeitig beendet, wenn $\sum_{\mathbf{r} \in \mathbb{N}_0^k} \mathbf{p}(\mathbf{r}) > 1 - \text{tol.loop}$ ist (inaktiv bei <code>tol.loop = 0</code>).

Ausgabewert

Ein Listenobjekt (Klasse `list`) mit den folgenden Komponenten:

<code>pr</code>	Matrix mit den Reproduktionsverteilungen $\mathbf{p}^{(j)}(\mathbf{r})$ (eine Spalte je Typ).
<code>r.matrix</code>	Matrix gleicher Dimension mit den entsprechenden Vektoren $\mathbf{r}$.

Details / Limitationen

Es wird grundsätzlich eher die Variante mittels Schleife (`loop = TRUE`) empfohlen. Dann wird die Berechnung beim Erreichen von $\sum_{\mathbf{r} \in \mathbb{N}_0^k} \mathbf{p}(\mathbf{r}) = \mathbf{1}$ vorzeitig abgeschlossen (oder ev. bereits früher, wenn `tol.loop > 0` ist - siehe oben). Deshalb wird in der Regel weniger Zeit benötigt (vgl. auch die R-Ausgabe 4.1). Insbesondere bei *nicht* abbrechenden Potenzreihen kann die Berechnung sehr aufwendig werden bzw. ist ab einer gewissen Ordnung nicht mehr durchführbar (dann werden `NA`-Werte produziert und es wird eine Warnung ausgegeben – es kann aber dennoch mit der `W`-Funktion weitergerechnet werden).

Mit einem Durchlauf ist die Berechnung von $\mathbf{p}(\mathbf{r})$ für ein $\mathbf{r} \in \mathbb{N}_0^k$ mit $r_j \leq \text{max.order}$ für alle $1 \leq j \leq k$ gemeint. $\triangle$

<code>M.calc</code>	Erwartungswertmatrix der Reproduktionszahlen $\mathbf{M}$ und <i>Perron-Wurzel</i> ρ aus den weF $\mathbf{f}^{(j)}$ berechnen (lt. (3.53) bzw. dem Satz 3.39).
---------------------	---

Verwendung

```
M.calc(f.str, check.PR = FALSE, n.check.PR = 10)
```

Argumente

<code>f.str</code>	WeF $\mathbf{f}$ als <code>character</code> -Vektor (siehe <code>wef.create</code>).
<code>check.PR</code>	<code>logical</code> – wenn <code>TRUE</code> , wird bis zur Generation <code>n.check.PR</code> überprüft, ob über $M^n > \mathbf{0}$ die positive Regularität nachgewiesen werden kann (siehe das Lemma 3.35 bzw. (3.81)).
<code>n.check.PR</code>	Positiver <code>integer</code> – siehe oben.

Ausgabewert

Ein Listenobjekt (Klasse `list`) mit den folgenden Komponenten:

<code>M</code>	Erwartungswertmatrix $\mathbf{M}$ der Reproduktionszahlen.
<code>rho</code>	Perron-Wurzel ρ (siehe den <i>Satz 3.39 von Perron-Frobenius</i>).
<code>u, v</code>	Korrespondierender linker ($\mathbf{u}$) und rechter ($\mathbf{v}$) Eigenvektor.
<code>Klasse</code>	<code>character</code> -Wert, entweder <code>"subkritisch"</code> , <code>"kritisch"</code> oder <code>"superkritisch"</code> (gemäß dem Satz 3.50). $\triangle$

<code>fn.sym.calc</code> ,	WeF der Populationsgrößen $\mathbf{f}_n \stackrel{(3.39)}{=} (f_n^{(1)}, \dots, f_n^{(k)})^T$ aus den weF $f^{(j)}$
<code>fns.calc</code>	berechnen – (optionale) Auswertung an einer Stelle $\mathbf{s}$ (siehe (3.48)).

Verwendung

```
fn.sym.calc(f.str, n = 1, eval = FALSE, s = rep(0, length(f.str)),
            all = FALSE, include.f0 = FALSE, time = FALSE)
fns.calc(f.str, n = 1, s = rep(0, length(f.str)),
         all = FALSE, include.f0 = FALSE, time = FALSE)
```

Argumente

<code>f.str</code>	WeF $\mathbf{f}$ als <code>character</code> -Vektor (siehe <code>wef.create</code>).
<code>n</code>	Positiver <code>integer</code> – Generation n .
<code>eval</code>	<code>logical</code> – wenn <code>FALSE</code> , explizite Ausgabe von $\mathbf{f}_n$ als <code>yac_symbol</code> -Objekt (siehe Details – nur relevant für <code>fn.sym.calc</code>).
<code>s</code>	Numerischer Vektor der Länge k – Auswertungsstelle $\mathbf{s} \in [0, 1]^k$.
<code>all</code>	<code>logical</code> – wenn <code>TRUE</code> , werden $\mathbf{f}_1$ bis $\mathbf{f}_n$ ausgegeben, ansonsten nur $\mathbf{f}_n$ – siehe Ausgabewert.
<code>include.f0</code>	<code>logical</code> – wenn <code>TRUE</code> , wird $\mathbf{f}_0$ mit ausgegeben – nur mit <code>all</code> (und bei <code>fn.sym.calc</code> zusätzlich mit <code>eval</code>) relevant.
<code>time</code>	<code>logical</code> – wenn <code>TRUE</code> , wird die benötigte Zeit berichtet.

Ausgabewert

Je nach gewählter Funktion bzw. Konfiguration eines der folgenden Objekte:

Funktion `fn.sym.calc` mit `eval = FALSE`:

- Falls `all = FALSE`: `yac_symbol`-Vektor der Länge k .
- Falls `all = TRUE`: Liste mit `yac_symbol`-Vektoren der Länge k .

Funktion `fns.sym` sowie Funktion `fn.sym.calc` mit `eval = TRUE`:

- Falls `all = FALSE`: Numerischer Vektor der Länge k .
- Falls `all = TRUE`: Matrix mit n (oder $n + 1$) Zeilen und k Spalten.

Details / Limitationen

Es handelt sich eigentlich jeweils um einen sogenannten *Wrapper* um eine passende rekursive Funktion, welche sich weitere $(n-1)$ -mal selbst aufruft (gemäß (3.48)). Die Komposition erfolgt bei `fn.sym.calc` über *YACAS* (kurz für *Yet Another Computer Algebra System* - vgl. [52]), welches in R über das Paket 'Ryacas' (Version 1.1.5 - vgl. [49] und [1]) eingebunden wird. Da diese Berechnung schnell an Kapazitätsgrenzen gelangt, sollte sie nur für wenige Generationen angewendet werden (siehe als exemplarische Anwendung die NR B.16 im Setting des nächsten Unterabschnitts). Ist man an Werten von $\mathbf{f}_n(\mathbf{s})$ für explizite $\mathbf{s} \in [0, 1]^k$ interessiert, wird die wesentlich effizientere Routine `fns.calc` empfohlen (`eval` für `fn.sym.calc` liefert ebenso konkrete Werte, wird aber nur der Vollständigkeit halber mit angeboten – vgl. etwa die R-Ausgabe 4.3). $\triangle$

<code>eta.calc</code>	Berechnung des AWK-Vektors $\boldsymbol{\eta}$ (lt. dem Lemma 3.49) und der AWK (gemäß (3.99)).
-----------------------	---

Verwendung

```
eta.calc(f.str = NULL, Z0 = NULL, f.vec = wef.create(f.str)$f.vec,
        k = length(f.str))
```

Argumente

<code>f.str</code>	WeF $\boldsymbol{f}$ als <code>character</code> -Vektor (siehe <code>wef.create</code>).
<code>Z0</code>	Optional (nur für die Berechnung der AWK): Nichtnegativer <code>integer</code> -Vektor der Länge k – initiale Populationsgröße $\boldsymbol{Z}_0$.
<code>f.vec, k</code>	Alternative zu <code>f.str</code> : Eingabe von $\boldsymbol{f}$ als Vektorfunktion (vorzugsweise <code>\$f.vec</code> aus <code>wef.create</code>). Diesfalls muss k mitangegeben werden.

Ausgabewert

Wenn `Z0 = NULL` ist, nur `eta` (siehe unten), ansonsten `list` mit den folgenden Einträgen:

<code>eta</code>	AWK-vektor $\boldsymbol{\eta} \stackrel{(3.96)}{=} (\eta^{(1)}, \dots, \eta^{(k)})^T$.
<code>AWK</code>	Aussterbewahrscheinlichkeit $\mathbb{P}(\boldsymbol{Z}_n \rightarrow \mathbf{0}) \stackrel{(3.99)}{=} \boldsymbol{\eta}^{\boldsymbol{Z}_0}$.

Details

Die Berechnung von `eta` gemäß der Fixpunktgleichung (3.98) erfolgt mit dem R-Paket 'FixedPoint' (Version 0.6.3 – vgl. [50]). Als Methode für den Algorithmus wird die sogenannte *Aitken acceleration* (siehe [50, Vignette, 2.2, S. 2]) und als Startpunkt $\boldsymbol{s} = \mathbf{0}$ gesetzt (damit wird der kleinste Fixpunkt in $[0, 1]^k$ gefunden). $\triangle$

Zusätzlich steht die Funktion `Mn.calc` für eine rekursive Berechnung von M^n mit $n \geq 0$ zur Verfügung. Da diese jedoch in den hier vorgestellten Routinen implizit mit aufgerufen wird, wird auf eine genauere Dokumentation verzichtet und stattdessen auf die R-Helferfunktionen unter B.2.2 verwiesen (ab der Zeile 23).

R-Funktionen für die Bestimmung der Momente der Populationsgrößen eines MGWP

<code>EZn.calc,</code> <code>VZn.calc</code>	Berechnung der Erwartungswertvektoren $\mathbb{E}\mathbf{Z}_n$ sowie der Kovarianzmatrizen $\mathbb{V}\mathbf{Z}_n$ (gemäß dem Satz 3.31).
---	--

Verwendung

```
EZn.calc(f.str = NULL, n, ZO, M = M.calc(f.str)$M, all = FALSE,
         include.EZO = FALSE | (n == 0), simplify = all)
VZn.calc(f.str = NULL, n, ZO, M = M.calc(f.str)$M)
```

Argumente

<code>f.str</code>	WeF f als character-Vektor (siehe <code>wef.create</code>).
<code>n</code>	Nichtnegativer integer – Generation n (positiv für <code>VZn.calc</code>).
<code>ZO</code>	Nichtnegativer integer-Vektor, Länge k – init. Populationsgröße $\mathbf{Z}_0$.
<code>M</code>	Alternative zu <code>f.str</code> : Eingabe von $\mathbf{M}$ als $k \times k$ -Matrix (vorzugsweise $\$M$ aus <code>M.calc</code>).
<code>all</code>	logical – wenn <code>TRUE</code> , werden $\mathbb{E}\mathbf{Z}_1$ bis $\mathbb{E}\mathbf{Z}_n$ ausgegeben, ansonsten nur $\mathbb{E}\mathbf{Z}_n$ – siehe Ausgabewert.
<code>include.EZO</code>	logical – wenn <code>TRUE</code> , inkl. $\mathbb{E}\mathbf{Z}_0$ (nur mit <code>all</code> oder <code>n = 0</code> relevant).
<code>simplify</code>	logical – siehe Ausgabewert (nur mit <code>all</code> relevant).

Ausgabewert

Abhängig von der gewählten Funktion bzw. Konfiguration.

`EZn.calc` mit `all = FALSE`: Erwartungswertvektor $\mathbb{E}\mathbf{Z}_n$ als Vektor der Länge k .

`EZn.calc` mit `all = TRUE`:

Falls `simplify = TRUE`: Matrix mit $\mathbb{E}\mathbf{Z}_1$ (oder $\mathbb{E}\mathbf{Z}_0$ mit `include.EZO`) bis $\mathbb{E}\mathbf{Z}_n$ als Spalten (also n oder $n + 1$ Spalten und k Reihen).

Falls `simplify = FALSE`: list mit n (oder $n + 1$) Einträgen – jeweils $\mathbb{E}\mathbf{Z}_n$ als Vektor.

`VZn.calc`: Kovarianzmatrix $\mathbb{V}\mathbf{Z}_n$ als $k \times k$ -Matrix.

Details

Bei `EZn.calc` handelt es sich um einen *Wrapper* um die bereits erwähnte rekursive Funktion `Mn.calc`. Die Option `simplify = FALSE` wird hauptsächlich für den internen Gebrauch benötigt (sprich für `VZn.calc`). Die Berechnungen erfolgen gemäß (3.75). $\triangle$

Zusätzlich stehen die Funktionen `covXj_il.calc` bzw. `Vj.list.calc` zur Verfügung – für die Berechnung von $Cov(X_i^{(j)} X_l^{(j)})$ bzw. $\mathbf{V}_j$ mit $1 \leq i, j, l \leq k$ – vgl. die Def. 3.29. Ein Aufruf dieser ist aber nur bei spezifischem Interesse für diese Größen nötig. Es sei auf die R-Helferfunktionen unter B.2.2 (ab den Zeilen 37 bzw. 43) und die Bem. 4.35 verwiesen.

R-Funktionen für die Simulation eines MGWP

Für Simulationen unterschiedlicher MGWP kommt grundsätzlich die R-Funktion `rBGWM` (vgl. [57, S. 14, ff.]) aus dem eingangs zitierten R-Paket 'Branching' zum Einsatz.

Die nun folgenden R-Funktionen ermöglichen es in erster Linie, direkt entsprechende Simulationen starten zu können. Weiters dienen sie dazu, beliebig viele Prozesse in einem Zug zu simulieren, wesentliche Ergebnisse zu speichern bzw. wieder zu laden sowie grundlegende Kenngrößen zu schätzen.

<code>dists.sim.create</code>	Aufbereitung der Reproduktionsverteilungen für die Simulation mit <code>rBGWM</code> – gegeben ihrer W-Funktionen oder aller $\text{weF } f^{(j)}$.
-------------------------------	--

Verwendung

```
dists.sim.create(f.str = NULL, wfkt = wfkt.calc(f.str))
```

Argumente

<code>f.str</code>	WeF f als <code>character</code> -Vektor (siehe <code>wef.create</code>).
<code>wfkt</code>	Alternative zu <code>f.str</code> : <code>list</code> mit den Matrizen <code>\$pr</code> und <code>\$r.matrix</code> als Komponenten – vorzugsweise die Ausgabe von <code>wfkt.calc</code> (siehe Details).

Ausgabewert

Eine Liste von `list`-Objekten, wie von der R-Funktion `rBGWM` des R-Pakets 'Branching' als Argument `dists` erfordert (mit `type = "general"` – siehe [57, S. 17, f.]).

Details

Es wird nur entweder `f.str` oder `wfkt` als Eingabe benötigt. (Letztere wird priorisiert, wenn beide eingegeben werden. Dabei ist es nicht erforderlich, sich mit der Struktur dieses Arguments – siehe dafür `wfkt.calc` – oder der der Ausgabe auseinanderzusetzen.) Die Ausgabe kann direkt für die gleich folgende Funktion `MGWP.sim` verwendet werden – siehe z.B. die R-Ausgabe 4.5. △

<code>MGWP.sim,</code>	Simulation einer gewissen Anzahl an Realisierungen eines MGWP
<code>MGWP.sim.load</code>	inkl. optionalem Speichern bzw. Laden der Ergebnisse.

Verwendung

```
MGWP.sim(iter, dists, max.gen = 10, Z0 = NULL, only.eta = FALSE,
         savepath = NULL, type = "general", time = TRUE)
MGWP.sim.load(iter, max.gen = 10, Z0, loadpath)
```

Argumente

<code>iter</code>	Anzahl der simulierten Prozesse (bzw. Wiederholungen oder <i>Iterationen</i>).
<code>dists</code>	Reproduktionsverteilungen $p^{(1)}, \dots, p^{(k)}$ als passende <code>list</code> – vorzugsweise die Ausgabe von <code>dists.sim.create</code> (siehe Details).
<code>max.gen</code>	Positiver <code>integer</code> – Generation n , bis zu welcher jeweils simuliert wird.

<code>Z0</code>	Nichtnegativer <code>integer</code> -Vektor der Länge k – initiale Populationsgröße Z_0 (nur relevant, wenn <code>only.eta</code> = <code>FALSE</code> ist – siehe Details).
<code>only.eta</code>	<code>logical</code> – wenn <code>TRUE</code> , wird lediglich für jedes $1 \leq j \leq k$ ein Schätzer für $\eta^{(j)}$ ermittelt (d.h. <code>iter</code> · k Simulationen); wenn <code>FALSE</code> , werden stattdessen die Trajektorien für die Populationsgrößen von <code>iter</code> Wiederholungen mit $Z_0 = Z0$ berichtet (siehe Details).
<code>savepath</code> , <code>loadpath</code>	<code>character</code> -Variable – Datenverzeichnis. Bei <code>MGWP.sim</code> nur für <code>only.eta</code> = <code>FALSE</code> relevant. Dann wird die Ausgabe als <code>.RData</code> -Datei im <code>savepath</code> gespeichert, sofern dieser nicht <code>NULL</code> ist. Mit <code>savepath</code> = <code>NULL</code> wird die benötigte Zeit in der Konsole berichtet und nur das erste Element der unten stehenden Liste direkt ausgegeben.
<code>type</code>	Verteilungsklasse – in der Regel <code>"general"</code> (siehe Details).
<code>time</code>	<code>logical</code> – wenn <code>TRUE</code> , wird die benötigte Zeit (in Sekunden) berichtet und der Fortschritt (in Prozent) angezeigt.

Ausgabewert

Abhängig von der gewählten Funktion bzw. Konfiguration.

`MGWP.sim` mit `only.eta` = `TRUE`: Vektor der Länge k – Schätzer für $\boldsymbol{\eta}^{(3.96)} = (\eta^{(1)}, \dots, \eta^{(k)})^T$.

`MGWP.sim(.load)` mit `only.eta` = `FALSE`: list mit den folgenden Komponenten:

<code>traj</code>	list der Länge <code>iter</code> , je Eintrag eine Matrix mit k Spalten (je Typ) und $(\text{maxgen}+1)$ Zeilen (von Z_0 bis Z_n mit $n = \text{max.gen}$).
<code>timer</code>	Benötigte Zeit in Sekunden (<code>NULL</code> , wenn <code>time</code> = <code>FALSE</code>).

Details

Bei `MGWP.sim` handelt es sich um einen *Wrapper* um `rBGWM`. Für Verteilungen `dists`, die mit `dists.sim.create` erstellt wurden, sollte `type` auf dem Standardwert belassen werden – bzgl. der Weitergabe anderer Verteilungsklassen an `rBGWM` siehe [57, *Details*, S. 15] (oder insbesondere für `type` = `"independents"` etwa das Beispiel 4.38).

Mit `only.eta` = `TRUE` werden für jedes $1 \leq j \leq k$ `iter` Durchläufe mit $Z_0 = e_j$ gestartet. Der jeweilige Anteil an Realisierungen, welche nach `max.gen` Generationen ausgestorben sind, bildet dann den Schätzwert $\hat{\eta}^{(j)}$ für $\eta^{(j)}$ (sehr relevant für die Aussagekraft dieses Schätzers ist die 'Rest-Aussterbewahrscheinlichkeit' der in der Generation `max.gen` noch lebenden Pfade – vgl. die Bemerkung 4.33).

Der Dateiname für die `.RData`-Datei (für die Speicherung bzw. das Laden) wird aus den Eingaben generiert. △

<code>MGWP.sim.est</code>	Schätzen von Kenngrößen eines MGWP (etwa auf Grundlage simulierter Daten, welche mit <code>MGWP.sim</code> erzeugt wurden).
---------------------------	---

Verwendung

```
MGWP.sim.est(traj, n = NULL)
```

Argumente

<code>traj</code>	<code>list</code> – Struktur wie die entsprechende Ausgabe von <code>MGWP.sim</code> .
<code>n</code>	Nichtnegativer <code>integer</code> (optional) – Generation n , für welche die ersten drei Komponenten der unten stehenden Liste berichtet werden.

Ausgabewert

Eine `list` bestehend aus folgenden Komponenten (anstelle von `n` steht im Variablennamen jeweils der entsprechende Wert – mit `n = NULL` wird lediglich der letzte Listeneintrag als `numeric`-Variable zurückgegeben):

<code>Zn.mean</code>	Stichprobenmittel als Schätzer für $\mathbb{E}\mathbf{Z}_n$.
<code>Zn.cov</code>	(Korrigierte) Stichprobenkovarianzmatrix als Schätzer für $\mathbb{V}\mathbf{Z}_n$.
<code>Zn.sample</code>	Realisierte Populationsgrößen $\mathbf{Z}_n$ der Stichprobe <code>traj</code> als Matrix (eine Zeile je Typ, eine Spalte je Realisierung).
<code>AWK</code>	Schätzwert für die AWK $\mathbb{P}(\mathbf{Z}_n \rightarrow \mathbf{0})$, unabhängig von <code>n</code> (relativer Anteil der Realisierungen in <code>traj</code> , welche bereits in der letzten Generation ausgestorben sind – also in der Generation $n = \text{max.gen}$; siehe die Details unter <code>MGWP.sim</code> bzgl. der Aussagekraft dieses Schätzers). △

Die letzte der dokumentierten Routinen – `MGWP.sim.est` – kann prinzipiell für beliebige Daten herangezogen werden. Sie ist einstweilen aber hauptsächlich auf wie vorher beschrieben erzeugte Simulationen ausgelegt. Für diverse Schätzverfahren auf der Grundlage von Beobachtungen werden etwa im Paket '`Branching`' bereits spezialisiere Funktionen angeboten.

Alle hier angeführten Funktionen wurden mehrfach in unterschiedlichen Szenarien getestet. Weiters wurden Laufzeitvergleiche und -optimierungen durchgeführt. Natürlich ist dennoch Verbesserungspotential vorhanden.

V.a. bei langen Simulationsdauern angesichts explodierender Populationszahlen in superkritischen Fällen (vgl. etwa das Bsp. 4.9) ist dieses Potential – zumindest so wie `MGWP.sim` gebaut ist – mehr oder weniger ausgeschöpft. So gut wie die gesamte benötigte Zeit wird von den mehrfachen Aufrufen der Kernfunktion `rBGWM` verbraucht.

4.2.2 Anwendung: Numerisches Beispiel

Zur exemplarischen Veranschaulichung der Funktionalität der im vorherigen Abschnitt vorgestellten $\mathbf{R}$ -Routinen wird ein MGWP anhand eines *numerischen Beispiels* durchexerziert. Dieses ist [19, II.8, S. 43] entnommen.

Bemerkung. Den durchgängigen und ausführbaren Quellcode für die rechnergestützten Analysen in diesem Unterabschnitt (und auch jener für die in 4.2.3) findet sich im Appendix unter B.2.1. $\triangle$

Es gebe 3 unterschiedliche Typen, es sei also $k = 3$. Die Reproduktionsverteilungen der einzelnen Objekte bzw. von $X^{(j)}$ mit $1 \leq j \leq 3$ seien gegeben durch ihre jeweilige weF $f^{(j)}$. Diese seien definiert als

$$\begin{aligned} f^{(1)}(\mathbf{s}) &\stackrel{(3.32)}{=} f^{(1)}(s_1, s_2, s_3) := \frac{4}{7}s_1^2 + \frac{3}{7}s_2, \\ f^{(2)}(\mathbf{s}) &\stackrel{(3.32)}{=} f^{(2)}(s_1, s_2, s_3) := \frac{1}{35} + \frac{16}{35}s_1s_3 + \frac{18}{35}s_2^2 \quad \text{und} \\ f^{(3)}(\mathbf{s}) &\stackrel{(3.32)}{=} f^{(3)}(s_1, s_2, s_3) := \frac{3}{7}s_2 + \frac{4}{7}s_3^2. \end{aligned} \quad (4.67)$$

Analyse von $\mathbf{f} \stackrel{(3.38)}{=} (f^{(1)}, f^{(2)}, f^{(3)})^T$ bzw. der Erwartungswertmatrix $\mathbf{M}$.

Um aus (4.67) die Verteilung der $X^{(j)}$ zu bestimmen, müssen gemäß dem Satz 3.10 (b) diverse partielle Ableitungen bestimmt werden – im vorliegenden Fall maximal bis zur zweiten Ordnung. Für alle $\mathbf{s} \in [0, 1]$ gilt $\frac{\partial f^{(1)}}{\partial s_1}(\mathbf{s}) = \frac{8}{7}s_1$, $\frac{\partial f^{(1)}}{\partial s_2}(\mathbf{s}) = \frac{3}{7}$, $\frac{\partial^2 f^{(1)}}{\partial s_1^2}(\mathbf{s}) = \frac{8}{7}$, weiters $\frac{\partial f^{(2)}}{\partial s_1}(\mathbf{s}) = \frac{16}{35}s_3$, $\frac{\partial f^{(2)}}{\partial s_2}(\mathbf{s}) = \frac{36}{35}s_2$, $\frac{\partial f^{(2)}}{\partial s_3}(\mathbf{s}) = \frac{16}{35}s_1$, $\frac{\partial^2 f^{(2)}}{\partial s_1 \partial s_3}(\mathbf{s}) = \frac{16}{35}$, $\frac{\partial^2 f^{(2)}}{\partial s_2^2}(\mathbf{s}) = \frac{36}{35}$ und $\frac{\partial f^{(3)}}{\partial s_2}(\mathbf{s}) = \frac{3}{7}$, $\frac{\partial f^{(3)}}{\partial s_3}(\mathbf{s}) = \frac{8}{7}s_3$, $\frac{\partial^2 f^{(3)}}{\partial s_3^2}(\mathbf{s}) = \frac{8}{7}$. Alle anderen partiellen Ableitungen der Form $\frac{\partial^{r_1+r_2+r_3} f^{(j)}}{\partial s_1^{r_1} \partial s_2^{r_2} \partial s_3^{r_3}}(\mathbf{s})$ für $1 \leq j \leq 3$ und $r_1, r_2, r_3 \geq 0$ ($\mathbf{r} \neq \mathbf{0}$) sind 0.

Mit dem genannten Satz ergeben sich daraus die W-Funktionen, welche laut (3.38) als

$$\mathbf{p}(\mathbf{r}) = \begin{pmatrix} p^{(1)}(\mathbf{r}) \\ p^{(2)}(\mathbf{r}) \\ p^{(3)}(\mathbf{r}) \end{pmatrix} = \begin{pmatrix} \mathbb{P}(X^{(1)} = \mathbf{r}) \\ \mathbb{P}(X^{(2)} = \mathbf{r}) \\ \mathbb{P}(X^{(3)} = \mathbf{r}) \end{pmatrix} = \begin{cases} (0, \frac{1}{35}, 0)^T & \text{für } \mathbf{r} = \mathbf{0}, \\ (\frac{4}{7}, 0, 0)^T & \text{für } \mathbf{r} = (2, 0, 0)^T, \\ (\frac{3}{7}, 0, \frac{3}{7})^T & \text{für } \mathbf{r} = (0, 1, 0)^T, \\ (0, \frac{18}{35}, 0)^T & \text{für } \mathbf{r} = (0, 2, 0)^T, \\ (0, \frac{16}{35}, 0)^T & \text{für } \mathbf{r} = (1, 0, 1)^T, \\ (0, 0, \frac{4}{7})^T & \text{für } \mathbf{r} = (0, 0, 2)^T, \\ \mathbf{0} & \text{sonst} \end{cases} \quad (4.68)$$

geschrieben werden können.

Dabei ist anhand von (4.68) auf einen Blick die Nichtsingularität des Prozesses feststellbar (siehe die Definition 3.32).

In der $\mathbf{R}$ -Ausgabe 4.1 ist ersichtlich, wie die Eingabe von (4.67) als Zeichenkette erfolgen sollte (Zeile 1). Anschließend wird die W-Funktion mit `wfkt.calc` bestimmt (wobei mit `loop = TRUE` weniger Zeit in Anspruch genommen wird). Das Ergebnis für `pr` stimmt mit dem von (4.68) überein.

4 Anwendungsbeispiele

```

1 > f.str<-c("4/7*s1^2+3/7*s2","1/35+16/35*s1*s3+18/35*s2^2","3/7*s2+4/7*s3^2")
2 > invisible(wfkt.calc(f.str,time=T,loop=F))
3 pr-Berechnung: 1.32 sec elapsed
4 > wfkt <- wfkt.calc(f.str,time=T,loop=T)
5 0.27 sec elapsed
6 > (pr <- wfkt$pr)
7
8           p1           p2           p3
9 r=(0,0,0) 0.0000000 0.02857143 0.0000000
10 r=(2,0,0) 0.5714286 0.00000000 0.0000000
11 r=(0,1,0) 0.4285714 0.00000000 0.4285714
12 r=(0,2,0) 0.0000000 0.51428571 0.0000000
13 r=(1,0,1) 0.0000000 0.45714286 0.0000000
14 r=(0,0,2) 0.0000000 0.00000000 0.5714286

```

R-Ausgabe 4.1: Berechnung der Reproduktionsverteilungen $p^{(1)}$, $p^{(2)}$ und $p^{(3)}$ aus (4.67).

Werden die (oben bereits mitbestimmten) partiellen Ableitungen erster Ordnung an der Stelle $\mathbf{s} = \mathbf{1}$ ausgewertet, ergibt sich mit (3.53) direkt

$$\mathbf{M} = \begin{pmatrix} \frac{8}{7} & \frac{3}{7} & 0 \\ \frac{16}{35} & \frac{36}{35} & \frac{16}{35} \\ 0 & \frac{3}{7} & \frac{8}{7} \end{pmatrix} \Rightarrow \mathbf{M}^2 = \begin{pmatrix} \frac{368}{245} & \frac{228}{245} & \frac{48}{245} \\ \frac{1216}{1225} & \frac{1776}{1225} & \frac{1216}{1225} \\ \frac{48}{245} & \frac{228}{245} & \frac{368}{245} \end{pmatrix} > \mathbf{0}. \quad (4.69)$$

Auf der rechten Seite von (4.69) erkennt man, dass die Bedingung (3.81) (siehe das Lemma 3.35) für die positive Regularität des MGWP bereits in der zweiten Generation erfüllt ist. In der Nebenrechnung NR B.15 werden die *Perron-Wurzel* und der dazugehörige linke bzw. rechte Eigenvektor $\mathbf{u}$ bzw. $\mathbf{v}$ (siehe den Satz 3.39 von Perron-Frobenius) bestimmt:

$$\rho = \frac{12}{7} = 1,7143 > 1 \quad \xRightarrow{\text{Satz 3.50 (ii)}} \eta < 1 \text{ und der Prozess ist superkritisch.}$$

```

1 > M.calc(f.str,check.PR=T)[c("M","rho","Klasse")]
2 [1] "M ist garantiert positiv regulär."
3 $M
4           [,1]      [,2]      [,3]
5 [1,] 1.1428571 0.4285714 0.0000000
6 [2,] 0.4571429 1.0285714 0.4571429
7 [3,] 0.0000000 0.4285714 1.1428571
8 $rho
9 [1] 1.714286
10 $Klasse
11 [1] "superkritisch"

```

R-Ausgabe 4.2: Bestimmung von $\mathbf{M}$, ρ und der Kritikalität des MGWP aus (4.67).

Das gleichlautende Ergebnis von `M.calc` für das vorliegende Beispiel wird in der R-Ausgabe 4.2 angeführt (als Listenelemente 3 und 4 wären zusätzlich $\mathbf{u}$ und $\mathbf{v}$ abrufbar).

Da `check.PR = TRUE` gesetzt wurde, wird zusätzlich die Information ausgegeben, dass es sich um eine positiv reguläre Matrix handelt (Zeile 2).

Wird die vektorwertige Funktion $\mathbf{f}$ iteriert, können gemäß dem Korollar zum Satz 3.20 (ii) prinzipiell auf direkte Weise die $f_n^{(j)}$ für alle $n \geq 0$ und $1 \leq j \leq k$ erhalten werden. Dies stellt sich aber im Allgemeinen (abgesehen von gebrochen linearen Fällen² – siehe die Bem. 4.3) als sehr aufwendig dar – und zwar sowohl analytisch als auch rechnerunterstützt. Auf Ersteres wird deshalb gänzlich verzichtet. Für eine Skizzierung der (beschränkten) Möglichkeiten des softwaregestützten Umgangs mit expliziten Ausdrücken für $\mathbf{f}_n$ sei erneut auf die NR B.16 verwiesen. Dort werden über $\mathbf{f}_2$ die W-Funktion und der EW von $\mathbf{Z}_2$ ermittelt. Eine weitere Anwendung wird im nächsten Teilabschnitt angerissen.

Bestimmung des Aussterbewahrscheinlichkeitsvektors $\boldsymbol{\eta}$

Um die Frage nach der Wahrscheinlichkeit des Aussterbens dieses konkreten MGWP zu beantworten, soll nun der Vektor $\boldsymbol{\eta}$ ermittelt werden (siehe die Notation 3.48). Darin sind die AWK für jene Populationen ablesbar, welche initial nur aus je einem Objekt bestehen. Gemäß (3.99) kann daraus mühelos die AWK $\boldsymbol{\eta}^{\mathbf{Z}_0}$ für beliebige $\mathbf{Z}_0 \in \mathbb{N}_0^3$ berechnet werden.

Die theoretischen Resultate des 3. Kapitels bieten mehrere Möglichkeiten für die Bestimmung von $\boldsymbol{\eta}$. Zunächst kann direkt die Definition 3.47 herangezogen werden. In Kombination mit (3.39) und der Notation 3.48 besagt diese, dass $\{\mathbf{f}_n(\mathbf{0})\}_{n \geq 0}$ gegen $\boldsymbol{\eta}$ konvergiert. Eine Analyse dieser Konvergenz ist mit `fn.sym.calc` oder `fns.calc` möglich. Da nur die punktweise Konvergenz bei $\mathbf{s} = \mathbf{0}$ benötigt wird, ist Ersteres dabei etwas überschießend und die direkte Auswertung zu bevorzugen. In der R -Ausgabe 4.3 wurde für einen Laufzeitvergleich zunächst beides bis zur Generation 10 getestet (Zeilen 1-8 bzw. 9-10). Dabei ist die Auswertung mit `fn.sym.calc` bedeutend langsamer (vgl. die Zeile 4 mit der Zeile 10). Zudem stößt diese Variante über explizite Funktionen bereits mit $n = 11$ sowohl bzgl. des Zeit- als auch des Speicherverbrauchs an ihre Grenzen.

Die Rekursion durch `fns.calc` an der Stelle $\mathbf{s} = \mathbf{0}$ hingegen kann problemlos auch auf höhere Generationen angewendet werden. In den Zeilen 11-18 wird dies bis zur Generation 500 vorgeführt und ein Auszug des Ergebnisses berichtet. Dabei zeigt sich, dass zumindest in der Standarddarstellung auch bereits ab der zehnten Generation keine Veränderung mehr wahrnehmbar ist. Am Ende der R -Ausgabe 4.3 (Zeilen 19-23) erkennt man, dass ab $\mathbf{f}_{18}$ auch in der R -Arithmetik³ kein Unterschied mehr beobachtet werden kann.

Bemerkung 4.32 (Konvergenz der Folge $\{\mathbf{f}_n\}_{n \geq 0}$ – vgl. [19, Theorem II.7.2, S. 42])
Für die einfachen Prozesse wurde durch den Satz 2.46 gezeigt, dass die Konvergenz der iterierten weF gegen $\boldsymbol{\eta}$ nicht nur für 0, sondern für alle $s \in [0, 1)$ gilt. Analoges gilt allgemein

²Für ein Bsp. eines MGWP mit gebrochen linearen weF siehe etwa [19, II.12.3, S. 49].

³Siehe bzgl. der numerischen Genauigkeit von R etwa [26, 9, S. 163 ff.]. Genauer ist ab der Version 4.0.0 *long-double-Genauigkeit* für die Unterscheidung reeller Zahlen verfügbar, d.h. bis 1.084202e-19, abrufbar mit `.Machine$longdouble.eps` aus dem 'base'-Paket – vgl. [55] (es sind aber wesentlich kleinere Zahlen darstellbar, so ist `.Machine$double.xmin = 2.225074e-308`).

4 Anwendungsbeispiele

bzgl. nichtsingulärer, positiv regulärer Multitype GWP. D.h., es gilt $\lim_{n \rightarrow \infty} \mathbf{f}_n(\mathbf{s}) = \boldsymbol{\eta}$ für alle $\mathbf{s} \in [0, 1]^k$ mit $\mathbf{s} \neq \mathbf{1}$ – siehe dazu das oben angegebene Theorem.

Eine Überprüfung für das hier untersuchte numerische Beispiel am willkürlich gewählten Punkt $\mathbf{s} = (0,0001; 0,8; 0,5)^T$ zeigt, dass die Folge tatsächlich gegen die gleichen Werte wie in der R-Ausgabe 4.3 (Zeile 18) strebt:

```
> fns.calc(f.str,n=500,s=c(0.0001,0.8,0.5))
f500(1)(1e-04,0.8,0.5) f500(2)(1e-04,0.8,0.5) f500(3)(1e-04,0.8,0.5)
0.01255216 0.02907831 0.01255216
```

△

```
1 > fn.sym.calc(f.str,n=10,all=T,eval=T,time=T)[c(1,5,10),]
2 Verkettung (ohne Evaluierung): 129.34 sec elapsed
3 Evaluierung: 27.02 sec elapsed
4 Gesamt: 156.36 sec elapsed
5      [,1]      [,2]      [,3]
6 f1(0,0,0) 0.00000000 0.02857143 0.00000000
7 f5(0,0,0) 0.01255184 0.02907825 0.01255184
8 f10(0,0,0) 0.01255216 0.02907831 0.01255216
9 > invisible(fns.calc(f.str,n=10,time=T))
10 fn Rekursion: 0.02 sec elapsed
11 > (fns.500 <- fns.calc(f.str,n=500,all=T,time=T))[c(1,5,10,100,500),]
12 fn Rekursion: 0.07 sec elapsed
13      [,1]      [,2]      [,3]
14 f1(0,0,0) 0.00000000 0.02857143 0.00000000
15 f5(0,0,0) 0.01255184 0.02907825 0.01255184
16 f10(0,0,0) 0.01255216 0.02907831 0.01255216
17 f100(0,0,0) 0.01255216 0.02907831 0.01255216
18 f500(0,0,0) 0.01255216 0.02907831 0.01255216
19 > diff(fns.500)[c(15:17),]
20      [,1]      [,2]      [,3]
21 f16(0,0,0) 1.214306e-17 3.469447e-18 1.214306e-17
22 f17(0,0,0) 1.734723e-18 0.000000e+00 1.734723e-18
23 f18(0,0,0) 0.000000e+00 0.000000e+00 0.000000e+00
```

R-Ausgabe 4.3: Bestimmung von $\boldsymbol{\eta}$ durch Iteration der weF mit `fn.sym.calc` (bis `n=10`) bzw. `fns.calc` (bis `n=500`).

Als nächste Möglichkeit zur $\boldsymbol{\eta}$ -Bestimmung soll versucht werden, die Fixpunktgleichung (3.98) des Lemmas 3.49 zu lösen. Dies kann die R-Funktion `eta.calc` bewerkstelligen, welche wie beschrieben einen entsprechenden Fixpunktalgorithmus anwendet. Es kann direkt $\mathbf{f}$ bzw. (4.67) eingegeben werden (und wenn man nur an $\boldsymbol{\eta}$ interessiert ist, ist keine Eingabe für Z_0 nötig). In der R-Ausgabe 4.4 ist das Resultat dieser Kalkulation ersichtlich (Zeile 2-3). Dieses entspricht offensichtlich dem vorherigen Ergebnis (vgl. etwa die Zeile 18 der R-Ausgabe zuvor). In den letzten beiden Zeilen wird zusätzlich die Gleichheit der beiden Vektoren $\mathbf{f}_{500}$ und $\boldsymbol{\eta}$ überprüft (die Toleranz für diese Kontrolle mit `all.equal` liegt nahe bei $1,5 \cdot 10^{-8}$ – vgl. ggf. die zweite Spalte der Tabelle 4.7 am Ende dieses Teilschnitts für die exakte, verschwindend kleine Abweichung bereits gegenüber $\mathbf{f}_{20}(\mathbf{0})$).


```

1 > (eta <- eta.calc(f.str))
2   eta(1)    eta(2)    eta(3)
3 0.01255216 0.02907831 0.01255216
4 > all.equal(fns.500[500,],eta,check.attributes=F)
5 [1] TRUE

```

R-Ausgabe 4.4: Bestimmung von η mit `eta.calc`.

Im Folgenden wird das Resultat der R-Ausgabe 4.4 als eine 'genaue' Angabe für η angesehen (und deshalb auch als `eta` gespeichert), d.h.

$$\eta = \begin{pmatrix} \mathbb{P}(Z_n^{(1)} \rightarrow 0) \\ \mathbb{P}(Z_n^{(2)} \rightarrow 0) \\ \mathbb{P}(Z_n^{(3)} \rightarrow 0) \end{pmatrix} = \begin{pmatrix} 0,01255216 \\ 0,02907831 \\ 0,01255216 \end{pmatrix} = \begin{pmatrix} 1,255216 \cdot 10^{-2} \\ 2,907831 \cdot 10^{-2} \\ 1,255216 \cdot 10^{-2} \end{pmatrix} \approx \begin{pmatrix} 1,255 \% \\ 2,908 \% \\ 1,255 \% \end{pmatrix}. \quad (4.70)$$

Als letzte und ungenaueste Variante wird nun noch die Möglichkeit der Schätzung von η durch simulierte Daten vorgestellt. Dafür steht die R-Funktion `MGWP.sim` mit der Option `only.eta = TRUE` zur Verfügung. Wie in der entsprechenden Dokumentation beschrieben, bildet der relative Anteil der ausgestorbenen Realisierungen den Schätzwert, von nun an als $\hat{\eta}$ bezeichnet. In der R-Ausgabe 4.5 findet sich das Ergebnis dreier Simulationsversuche mit jeweils drei mal 50 000, 500 000 bzw. 5 000 000 Wiederholungen bis zur zehnten Generation (Festlegung dieser Werte in der Zeile 1). In den Zeilen 2-4 steht der entsprechende Befehl, wobei den Kern der Aufruf von `MGWP.sim` mit den Argumenten `iter = iter.vec[i]` ($1 \leq i \leq 3$), `dists = dists.sim.create(f.str)`, `max.gen = 10` und eben `only.eta = TRUE` bildet (ansonsten werden noch passende Zeilenbeschriftungen hinzugefügt und das Ergebnis – Zeilen 8-11 – gespeichert). Die benötigte Zeit pro Versuch ist in den Zeilen 5-7 ablesbar (der erste Durchgang war nach gut 5 Minuten beendet, der zweite dauerte bereits fast eine Stunde und der dritte über 8,5 Stunden).

```

1 > max.gen <- 10; iter.vec <- c(5e4,5e5,5e6)
2 > (eta.hat <- set_rownames(
3 + t(sapply(iter.vec, MGWP.sim, dists.sim.create(f.str), max.gen, only.eta=T)),
4 + paste0("iter=", iter.vec)))
5 Simulation eta.hat (3x50000): 345 sec elapsed
6 Simulation eta.hat (3x5e+05): 3167.68 sec elapsed
7 Simulation eta.hat (3x5e+06): 30732.2 sec elapsed
8   eta.hat(1) eta.hat(2) eta.hat(3)
9 iter=50000  0.0130600 0.0294800 0.012080
10 iter=5e+05  0.0127660 0.0292860 0.012722
11 iter=5e+06  0.0125202 0.0291762 0.012574

```

R-Ausgabe 4.5: Schätzungen $\hat{\eta}$ für η mit `MGWP.sim`. Drei Versuche mit 50 000, 500 000 und 5 000 000 Durchläufen pro $\eta^{(j)}$, $1 \leq j \leq 3$ (Fortschrittsanzeige entfernt).

Natürlich ist es prinzipiell denkbar, dass eine in der zehnten Generation noch lebende Population dennoch ausstirbt. Potentiell können die Schätzwerte für η dadurch verzerrt

werden, da solche Pfade als 'nicht aussterbend' gewertet werden. Dem könnte man mit größeren Werten für `max.gen` entgegenwirken. Stattdessen soll aber kurz versucht werden, die Frage nach der 'Restaussterbewahrscheinlichkeit' bedingt auf das Überleben nach einer gewissen Anzahl an Generationen zu beantworten. Diese Frage wird in der nächsten Bemerkung erörtert. Dabei zeigt sich auch, dass für das vorliegende Beispiel die Wahl von `max.gen = 10` eindeutig gerechtfertigt ist.

Bemerkung 4.33 ('Restaussterbe-WK' einer noch lebenden Realisierung)

Zuerst soll allgemein die bedingte ('Rest'-)AWK – gegeben, dass ein MGWP in einer gewissen Generation noch nicht ausgestorben ist – ermittelt werden (i). Anschließend wird diese für das konkrete in der R-Ausgabe 4.5 behandelte Szenario berechnet (ii).

- (i) Sei $m \geq 1$. Es gilt $\{\mathbf{Z}_m = \mathbf{0}\} \subseteq \{\mathbf{Z}_n \rightarrow \mathbf{0}\}$ (*)

$$\Rightarrow \mathbb{P}(\mathbf{Z}_n \rightarrow \mathbf{0}) \stackrel{(A.70)}{=} \mathbb{P}(\mathbf{Z}_n \rightarrow \mathbf{0}, \mathbf{Z}_m > \mathbf{0}) + \overbrace{\mathbb{P}(\mathbf{Z}_n \rightarrow \mathbf{0}, \mathbf{Z}_m = \mathbf{0})}^{=\mathbb{P}(\mathbf{Z}_m = \mathbf{0}) \text{ mit } (*)} \quad (**)$$

$$\Rightarrow \mathbb{P}(\mathbf{Z}_n \rightarrow \mathbf{0} | \mathbf{Z}_m > \mathbf{0}) \stackrel{(A.69)}{=} \frac{\mathbb{P}(\mathbf{Z}_n \rightarrow \mathbf{0}, \mathbf{Z}_m > \mathbf{0})}{\mathbb{P}(\mathbf{Z}_m > \mathbf{0})} \stackrel{(**)}{=} \frac{\mathbb{P}(\mathbf{Z}_n \rightarrow \mathbf{0}) - \mathbb{P}(\mathbf{Z}_m = \mathbf{0})}{1 - \mathbb{P}(\mathbf{Z}_m = \mathbf{0})}.$$
- (ii) Wegen (i) gilt insbesondere $\mathbb{P}(\mathbf{Z}_n^{(j)} \rightarrow \mathbf{0} | \mathbf{Z}_m^{(j)} > \mathbf{0}) \stackrel{(3.92)}{=} \frac{\eta^{(j)} - \eta_m^{(j)}}{1 - \eta_m^{(j)}} \stackrel{(3.92)}{=} \frac{\eta^{(j)} - f_m^{(j)}(\mathbf{0})}{1 - f_m^{(j)}(\mathbf{0})}$ für alle $1 \leq j \leq k$ und $m \geq 1$.

Somit kann die 'Rest-AWK' für in der Generation `max.gen = 10` noch lebende (von einem Objekt gestartete) Prozesse wie folgt bestimmt werden:

```
> unname((eta-fns.500[max.gen,]) / (1-fns.500[max.gen,]))
[1] 1.481346e-12 3.517431e-13 1.481346e-12
```

Es ist also bspw. $\mathbb{P}(\mathbf{Z}_n^{(2)} \rightarrow \mathbf{0} | \mathbf{Z}_{10}^{(2)} > \mathbf{0}) = 3.517431 \cdot 10^{-13}$. Somit ist es äußerst unwahrscheinlich (praktisch so gut wie unmöglich), dass die Schätzwerte durch erst nach `max.gen = 10` Generationen aussterbende Realisierungen verfälscht werden (dies liegt schlicht daran, dass auch schon $f_{10}(\mathbf{0})$ fast $\boldsymbol{\eta}$ entspricht – siehe die R-Ausgabe 4.3). In der Bemerkung 4.36 (ii) wird diese WK explizit, für einen bis zur 10. Generation realisierten Pfad, mit `eta.calc` berechnet. $\triangle$

Die Simulationsschätzungen (Zeilen 9-11 der R-Ausgabe 4.5) liegen jedenfalls in der Größenordnung des in (4.70) berichteten Wertes für $\boldsymbol{\eta}$ und erwartungsgemäß nähern sie sich mit steigender Iterationsanzahl immer mehr an.

In der Tabelle 4.7 werden alle drei besprochenen Varianten einander zusammenfassend gegenübergestellt. Dargestellt sind zuerst noch einmal $\boldsymbol{\eta}$ (ermittelt mit `eta.calc` – R-Ausgabe 4.4) und in den weiteren Spalten die Abweichungen davon (positive Werte bedeuten, dass die jeweilige Komponente um diesen Betrag größer ist). In der zweiten Spalte erkennt man, dass der Unterschied von $f_{20}(\mathbf{0})$ zu $\boldsymbol{\eta}$ (aufgrund von $f_n(\mathbf{0}) \rightarrow \boldsymbol{\eta}$) vernachlässigbar klein ist und die Vektoren als identisch angesehen werden können (vgl. die R-Ausgabe 4.3 – es wird der Wert für $n = 20$ berichtet, da ab der 18. Generation keine Veränderung dieser Folge mehr feststellbar ist, siehe die Zeilen 21-23 dieser Ausgabe). In Bezug auf $\hat{\boldsymbol{\eta}}$ ist in der Tabelle abermals gut ersichtlich, dass der Schätzfehler umso kleiner ist, je mehr Simulationsdurchläufe durchgeführt wurden.

$\boldsymbol{\eta}$ (4.70)	$\boldsymbol{f}_{20}(\mathbf{0})$	Abweichung von $\boldsymbol{\eta}$		
		$\hat{\boldsymbol{\eta}}$ (Iterationen)		
		(3 · 50 000)	(3 · 500 000)	(3 · 5 000 000)
$1,255 \cdot 10^{-2}$	$7,075 \cdot 10^{-13}$	$5,078 \cdot 10^{-4}$	$2,138 \cdot 10^{-4}$	$-3,196 \cdot 10^{-5}$
$2,908 \cdot 10^{-2}$	$5,616 \cdot 10^{-14}$	$4,017 \cdot 10^{-4}$	$2,077 \cdot 10^{-4}$	$9,789 \cdot 10^{-5}$
$1,255 \cdot 10^{-2}$	$7,075 \cdot 10^{-13}$	$-4,722 \cdot 10^{-4}$	$1,698 \cdot 10^{-4}$	$2,184 \cdot 10^{-5}$

Tabelle 4.7: Vergleich der ermittelten Werte für $\boldsymbol{\eta}$ bzw. $\hat{\boldsymbol{\eta}}$.**Bemerkung 4.34 (Vergleich mit den Ergebnissen aus [19, II.8, S. 43])**

Unter [19, II.(8.2), S. 43] werden die durch die Iteration der weF aus (4.67) an der Stelle $\boldsymbol{s} = (0,03; 0,015; 0,03)^T$ gewonnen Ungleichungen $0,0125 < \eta^{(1)} = \eta^{(3)} < 0,0126$ und $0,0290 < \eta^{(2)} < 0,0291$ angegeben. Die genauen Werte für $\boldsymbol{\eta}$ aus (4.70) liegen wie zu erwarten innerhalb dieser Schranken. Bei den Schätzern $\hat{\eta}^{(j)}$ ist dies erst beim Versuch mit 5 000 000 Wiederholungen für alle $1 \leq j \leq 3$ der Fall (siehe die R-Ausgabe 4.5). $\triangle$

Momentbestimmungen und Simulationsbsp. bzw. Analyse der Trajektorien von $\{\boldsymbol{Z}_n\}_{n \geq 0}$

Mit `EZn.calc` und `VZn.calc` können wie beschrieben $\mathbb{E}\boldsymbol{Z}_n$ und $\mathbb{V}\boldsymbol{Z}_n$ für beliebige initiale Populationsgrößen $\boldsymbol{Z}_0$ und Generationen $n \geq 1$ berechnet werden.

Um dies und zugleich die Funktionsweisen von `MGWP.sim` für und `MGWP.sim.est` zu präsentieren, wurden 500 000 beispielhafte Trajektorien der Populationsgrößen $\{\boldsymbol{Z}_n\}_{n \geq 0}$ mit $\boldsymbol{Z}_0 = (0,1,0)^T$ simuliert (wieder über 10 Generationen). Bis auf eine Ausnahme (am Ende des Beispiels 4.37) werden in diesem Teilabschnitt stets die Ergebnisse dieser Simulation behandelt.

Zuerst werden `max.gen <- 10` und `Z0 <- c(0,1,0)` gesetzt. Anschließend startet der Befehl `traj <- MGWP.sim(5e5, dists.sim.create(f.str), max.gen, Z0=Z0)` die Simulationen und speichert die Trajektorien als Liste ab (siehe den Quellcode unter B.2.1 ab der Zeile 41 – dabei ist `iter = 5e5` und `only.eta` standardmäßig `TRUE`). Alternativ kann über `savepath` ein Speicherpfad angegeben und das Ergebnis anschließend wieder mit `MGWP.sim.load` geladen werden.

Auf der linken Seite der R-Ausgabe 4.6 findet sich eine kurze Analyse eines solchen exemplarischen Simulationsversuches (welcher knapp 20 Minuten dauerte). Der Aufruf der Funktion `MGWP.sim.est` (Zeile 1 – mit `n = max.gen`, also 10) liefert wie erläutert Schätzgrößen für $\mathbb{E}\boldsymbol{Z}_{10}$ (Zeilen 4-6), $\mathbb{V}\boldsymbol{Z}_{10}$ (Zeilen 7-11) und die Aussterbewahrscheinlichkeit (Zeilen 12-13). Letztere ist im vorliegenden Szenario aufgrund von $\boldsymbol{Z}_0 = \boldsymbol{e}_2$ ein Schätzer für $\eta^{(2)}$ und stimmt hier sogar – zufällig – auf die sechste Nachkommastelle genau (vgl. etwa (4.70) bzw. die Zeile 13 der rechten Seite).

Auf der rechten Seite der R-Ausgabe werden als Vergleichsgrößen die durch den Satz 3.31 theoretisch untermauerten Momente bestimmt (Zeilen 4-11). Es zeigt sich, dass die anhand der Trajektorien geschätzten Werte für die ersten beiden Momente relativ nah an den tatsächlichen Werten liegen. Ergänzend wird die AWK mitberechnet (Zeilen 12-13 – diese stimmt hier wie erläutert mit $\eta^{(2)}$ überein).

4 Anwendungsbeispiele

<pre> 1 > est.list<-MGWP.sim.est(traj,max.gen) 2 > ### Schätzer für EW/Cov/AWK: 3 > est.list[c(1,2,4)] 4 \$Z10.mean 5 type1 type2 type3 6 79.69223 99.62385 79.72201 7 \$Z10.cov 8 type1 type2 type3 9 type1 793.1151 772.7440 621.9683 10 type2 772.7440 1127.6825 773.9985 11 type3 621.9683 773.9985 794.7597 12 \$AWK 13 [1] 0.029078 </pre>	<pre> 1 > ### 2 > ### Vergleich: Berechnete Größen 3 > ### 4 > EZn.calc(f.str,max.gen,Z0) 5 EZ10_1 EZ10_2 EZ10_3 6 79.70739 99.63463 79.70739 7 > VZn.calc(f.str,max.gen,Z0) 8 [,1] [,2] [,3] 9 [1,] 792.8972 772.8995 620.2715 10 [2,] 772.8995 1129.1138 772.8995 11 [3,] 620.2715 772.8995 792.8972 12 > eta.calc(f.str,Z0=Z0)\$AWK 13 [1] 0.02907831 </pre>
--	---

R-Ausgabe 4.6: $\mathbb{E}\mathbf{Z}_{10}$, $\mathbb{V}\mathbf{Z}_{10}$ und $\mathbb{P}(\mathbf{Z}_n \rightarrow \mathbf{0})$. Linke Seite: Ergebnisse der Schätzungen mit `MGWP.sim.est` (Grundlage: 500 000 Durchläufe mit `MGWP.sim` bis zur 10. Generation). Rechte Seite: Berechnungen der selben Größen zum Vergleich – durchgeführt mit `eta.calc`, `EZn.calc` und `VZn.calc`.

Bemerkung 4.35 (Bestimmung der Matrizen $\mathbf{V}_j$)

In der im rechten Teil der R-Ausgabe 4.6 durchgeführten Berechnung der Kovarianzmatrix $\mathbb{V}\mathbf{Z}_{10}$ werden die Kovarianzmatrizen $\mathbf{V}_j = \mathbb{V}\mathbf{X}^{(j)}$ (siehe die Def. 3.29) für $1 \leq j \leq 3$ implizit mitberechnet. Wollen diese direkt erhalten werden, kann jedoch die im Unterabschnitt 4.2.1 erwähnte (aber nicht genauer beschriebene) R-Funktion `Vj.list.calc` verwendet werden. Alternativ ist das wegen (3.27) aber etwa für $\mathbf{V}_1$ auch über den Befehl `VZn.calc(f.str,n=1,c(1,0,0))` möglich (und analog für $\mathbf{V}_2$ bzw. $\mathbf{V}_3$). Auf zwei Nachkommastellen gerundet erhält man dadurch

$$\mathbf{V}_1 = \begin{pmatrix} 0,98 & -0,49 & 0 \\ -0,49 & 0,24 & 0 \\ 0,00 & 0,00 & 0 \end{pmatrix}, \mathbf{V}_2 = \begin{pmatrix} 0,25 & -0,47 & 0,25 \\ -0,47 & 1,00 & -0,47 \\ 0,25 & -0,47 & 0,25 \end{pmatrix}, \mathbf{V}_3 = \begin{pmatrix} 0 & 0,00 & 0,00 \\ 0 & 0,24 & -0,49 \\ 0 & -0,49 & 0,98 \end{pmatrix}. \triangle$$

Bemerkung 4.36 (Größere initiale Populationsgrößen)

- (i) Die Funktionen `eta.calc` (siehe die gleich folgende Ausgabe) und `MGWP.sim(.load)` bzw. `MGWP.sim.est` (vgl. das Bsp 4.37) funktionieren wie dokumentiert auch für beliebige initiale Populationen. Selbiges trifft genauso auf `EZn.calc` und `VZn.calc` zu. Für größere $\mathbf{Z}_0$ ist die AWK in diesem Bsp. generell wesentlich niedriger (siehe (3.99)). Beispielsweise bereits für zwei ($\mathbf{Z}_0 = (1, 0, 1)^T$) oder drei ($\mathbf{Z}_0 = (1, 1, 1)^T$) Startobjekte ergeben sich die folgenden Wahrscheinlichkeiten:

```

> c(eta.calc(f.str,Z0=c(1,0,1))$AWK, eta.calc(f.str,Z0=c(1,1,1))$AWK)
[1] 1.575568e-04 4.581486e-06

```

Deshalb wurde für die Simulation in diesem Teilabschnitt als Anfangspopulation nur ein Objekt (vom Typ 2) gewählt.

- (ii) Die soeben beschriebene Funktionalität kann auch verwendet werden, um die 'Rest-AWK' (siehe die Bem. 4.33) konkreter noch lebender Pfade eines Simulationsergeb-

nisses zu bestimmen. Allgemein gilt aufgrund der Markov-Eigenschaft

$$\mathbb{P}(\mathbf{Z}_n \rightarrow \mathbf{0} | \mathbf{Z}_m = \mathbf{r}) = \mathbb{P}(\mathbf{Z}_{n+m} \rightarrow \mathbf{0} | \mathbf{Z}_m = \mathbf{r}) \stackrel{\text{ME}}{\underset{(3.89)}}{=} \mathbb{P}(\mathbf{Z}_n \rightarrow \mathbf{0} | \mathbf{Z}_0 = \mathbf{r}) \underset{(3.99)}{=} \boldsymbol{\eta}^{\mathbf{r}}$$

für alle $m \geq 0$ und $\mathbf{r} \in \mathbb{N}_0^k$. Gemäß dieser Formel wurden für das gegenständliche Simulationsergebnis alle 'Rest-AWK' ab der 10. Generation mit `eta.calc` berechnet. Dabei stellte sich heraus, dass diese WK für $\mathbf{z}_{10}^* = (3, 5, 5)^T$ unter den 485 461 zu diesem Zeitpunkt noch nicht Ausgestorbenen am größten ist. Zugleich weist dieser Pfad (mit 13) die kleinste von $\mathbf{0}$ verschiedene Gesamtpopulation in der 10. Generation auf (prinzipiell müssten dieses Maximum bzw. Minimum nicht zwingend eindeutig sein und nicht unbedingt die selbe Realisierung betreffen).

Mit dem Kommando `eta.calc(f.str, Z0=c(3,5,5))` erhält man konkret die WK $\mathbb{P}(\mathbf{Z}_n \rightarrow \mathbf{0} | \mathbf{Z}_{10} = (3, 5, 5)^T) = 1,281134 \cdot 10^{-23}$, was klarerweise die allgemeine 'Rest-AWK' (siehe den Punkt (ii) der erwähnten Bem.) noch einmal beträchtlich unterschreitet. Es kann also behauptet werden, dass alle noch lebenden Pfade auch mit größerem `max.gen` (so gut wie) fast sicher nicht mehr ausgestorben wären. $\triangle$

Abschließend zu diesem numerischen Beispiel soll versucht werden, die proportionale Aufteilung der drei unterschiedlichen Objekttypen auf längere Sicht zu untersuchen.

Beispiel 4.37 (Relative Häufigkeiten der einzelnen Objekttypen)

Das Lemma 3.41 legt nahe, dass sich die relative Häufigkeit der einzelnen Objekttypen asymptotisch durch den linken Eigenvektor $\mathbf{u}$ von $\mathbf{M}$ beschreiben lassen sollte.

Bemerkung. Gemeint ist damit jeweils die (über alle simulierten Realisierungen) gemittelte Anzahl eines Objekttyps, relativ zur mittleren gesamten Population einer Generation (und *nicht* das Mittel über die relative Häufigkeit pro Realisierung – diese Häufigkeit wäre etwa im Fall des Aussterbens nicht definiert).

Dieses Verhältnis kann beispielsweise wie folgt für die oben simulierten Trajektorien (500 000 Wiederholungen mit $\mathbf{Z}_0 = \mathbf{e}_2$) ermittelt werden:

```
> n <- c(1,2,3,5,10); prop <- NULL
> for(i in n){ mean.i <- MGWP.sim.est(traj,i)[[1]]
+           prop <- rbind(prop,mean.i / sum(mean.i)) }
> set_rownames(prop,paste0("n=",n))
      type1      type2      type3
n=1  0.2356013 0.5287973 0.2356013
n=2  0.2889611 0.4222141 0.2888248
n=3  0.3026956 0.3945819 0.3027225
n=5  0.3072095 0.3851378 0.3076527
n=10 0.3076468 0.3845915 0.3077617
```

Wie im Lemma 3.41 prognostiziert gleicht sich der relative Anteil jedes Objekttyps immer mehr dem jeweiligen Eintrag des linken Eigenvektors $\mathbf{u}$ an. Es zeigt sich, dass dieser die Proportionen bereits ab der 5. Generation gut beschreibt:

```
> M.calc(f.str)$u
      u1      u2      u3
0.3076923 0.3846154 0.3076923
```

4 Anwendungsbeispiele

Ein wesentlicher Grund ist die hohe Anzahl an Wiederholungen. Um ähnliche Verhältnisse zu erreichen, wären zwar nicht gar so viele notwendig, aber bei nur sehr wenigen Iterationen ist die Konvergenz gegen $\mathbf{u}$ in der Regel naturgemäß langsamer.

Das Resultat des Lemmas 3.41 gilt für beliebige $\mathbf{Z}_0$. Es kann also auch mit sehr ungleichen Typenverteilungen gestartet werden und dennoch stellen sich auf längere Sicht die selben Proportionen ein. So ergibt etwa die folgende Simulation:

```
> est.list.new <- MGWP.sim.est(
+ MGWP.sim(iter=1,dists.sim.create(f.str),max.gen=20,Z0=c(3,1,1000)),n=20)
Simulation Trajektorien (1): 1.21 sec elapsed
> est.list.new[[1]] / sum(est.list.new[[1]])
      type1      type2      type3
0.3073127 0.3848407 0.3078466
```

Hier wurde nur eine einzelne Realisierung mit $\mathbf{Z}_0 = (3, 1, 1000)^T$ erzeugt (dafür über 20 Generationen). Dass sich dennoch erneut die nahezu gleichen Verhältnisse ergeben, ist der *Additive Property* des Prozesses geschuldet (vgl. etwa (3.28)). Mit etwas Verzögerung tritt im Prinzip der gleiche Effekt wie durch viele Wiederholungen ein, da die einzelnen Stränge als unabhängige, eigene Prozesse betrachtet werden können. Beachte, dass dafür diese eine Realisierung nicht schon vor der 20. Generation aussterben darf – doch das ist in diesem konkreten Fall bei einer solchen initialen Population ohnehin praktisch unmöglich (siehe die Bem. 4.36 (i) – in der R-Arithmetik ergibt die Berechnung der AWK 0). $\triangle$

4.2.3 Anwendung: Weitere Multitype GWP

Die bereitgestellten R-Funktionen können – bei bekanntem weF-Vektor der Reproduktionsverteilungen – auf beliebige Anzahlen unterschiedlicher Objekttypen $k \geq 1$ angewendet werden.

Seien etwa $k = 2$ und $f^{(1)} = \frac{7}{8} + \frac{1}{16}s_1 + \frac{1}{16}s_2^2$ bzw. $f^{(2)} = \frac{15}{17} + \frac{2}{17}s_1s_2$ (also jeweils einer ähnlichen Bauweise wie im soeben unter 4.2.2 besprochenen Beispiel). Dann lässt sich schnell feststellen, dass es sich um einen *subkritischen*, positiv regulären MGWP handelt – siehe die Zeilen 1-4 der R-Ausgabe 4.7.

```
1 > f.str.A <- c("7/8+1/16*s1+1/16*s2^2", "15/17+2/17*s1*s2")
2 > c(format(M.calc(f.str.A,check.PR=T)$rho,digits=6),M.calc(f.str.A)$Klasse)
3 [1] "M ist garantiert positiv regulär."
4 [1] "0.214437"      "subkritisch"
5 > f.str.B <- function(p) c(paste0(p,"*s1+",(1-p),"*s2"),
6 + paste0("1/2*",p,"*s1+1/2*s2+1/2*",(1-p),"*s3"),
7 + paste0(p,"*s2+",(1-p),"*s3"))
8 > c(M.calc(f.str.B(0.4),check.PR=T)$rho,M.calc(f.str.B(0.4))$Klasse)
9 [1] "M ist garantiert positiv regulär."
10 [1] "1"      "kritisch"
```

R-Ausgabe 4.7: Bestimmung von ρ für zwei weitere Beispielprozesse mit `M.calc`.

In den Zeilen 5-10 wird das Gleiche für einen *kritischen*, positiv regulären MGWP getestet. Dieses Beispiel entstammt der Populationsgenetik und ist [39, Example 1.3 bzw.

3.1, S. 2 bzw. 7 f.] entlehnt. Für $k = 3$ und $0 < p < 1$ sind dabei $f^{(1)} = ps_1 + (1-p)s_2$, $f^{(2)} = \frac{1}{2}ps_1 + \frac{1}{2}s_2 + \frac{1}{2}(1-p)s_3$ und $f^{(3)} = ps_2 + (1-p)s_3$ (für die Ausgabe wurde $p = 0,4$ gewählt, der Prozess ist aber für alle in Frage kommenden p kritisch).

Wie erwähnt sind natürlich auch die einfachen GWP mit $k = 1$ berücksichtigt. In der R-Ausgabe 4.8 wird etwa vorgezeigt, wie mit `VZn.calc` die Varianzen der Populationsgrößen der ersten fünf Generationen im geometrischen Fall mit $\mu = 1,1$ berechnet werden können (beachte, dass "s1" für s geschrieben werden muss). Diese Ergebnisse sind identisch zu jenen der Tabelle 4.2 (siehe den Beginn der dritten Zeile der Tabelle – dies verdeutlicht auch, dass die Berechnungsformeln (2.57) und (3.75) für $k = 1$ klarerweise gleichwertig bzw. äquivalent sind).

```

1 > mu <- 1.1; (f.str.C <- paste0("1 / (1+",mu,"-s1*",mu,""))
2 [1] "1 / (1+1.1-s1*1.1)"
3 > sapply(1:5,function(n) VZn.calc(f.str.C,Z0=1,n))
4 [1] 2.310000 5.336100 9.251781 14.269265 20.647882

```

R-Ausgabe 4.8: Berechnung von $\mathbb{V}Z_1$ bis $\mathbb{V}Z_5$ mit `VZn.calc` – geometrischer Fall, $\mu = 1,1$.

Generell kamen für wesentliche Teile der softwaregestützten Ergebnisse bzgl. geometrischer (einfacher) Fälle wie erwähnt die unter 4.2.1 beschriebenen Funktionen zum Einsatz. Einige Berechnungen betrafen aber auch nur die expliziten Resultate dieser einen spezifischen Prozessklasse. Deshalb – und auch um Wiederholungen zu vermeiden – wird auf einen detaillierten Abdruck des entsprechenden Quellcodes verzichtet.

Etwas Vorsicht ist allgemein bei der Nutzung von `MGWP.sim` geboten, wenn die weF – als Potenzreihen betrachtet – *nicht* abbrechen. Diese Funktion benötigt als Eingabe die Verteilungen, welche dann durch `wfkt.calc` und infolgedessen auch durch `dists.sim.create` nicht exakt geliefert werden (vgl. die entsprechenden Details/Limitationen). Dies kann aber grundsätzlich in Kauf genommen werden, wenn die 'ausgelassenen' $\mathbf{r} \in \mathbb{N}_0^k$ nur sehr unwahrscheinlich sind (bzw. die betroffenen Werte von $\mathbf{p}(\mathbf{r})$ sehr klein sind).

Im Fall bekannter spezifischer Reproduktionsverteilungen ist es dennoch ratsamer, direkt mit diesen zu arbeiten. Wie in der Dokumentation zu `MGWP.sim` beschrieben ist, kann der `type` entsprechend angepasst werden – siehe das folgende Beispiel.

Beispiel 4.38. Die Simulationsergebnisse im Abschnitt 4.1 wurden etwa mit Aufrufen wie `MGWP.sim(100,dists.geom.create(mu),20,Z0=1,type="independents")` generiert. Mit diesem Kommando wird bspw. ein Datensatz von 100 Wiederholungen eines geometrischen GWP mit $\mu = \text{mu}$ bis zur 20. Generation erzeugt.

Die Hilfsfunktion `dists.geom.create` (vgl. die R-Helferfunktionen unter B.2.2 ab der Zeile 15) bereitet dabei die geometrische Nachkommensverteilung so auf wie von `rBGWM` verlangt auf. Diesfalls muss `type = "independents"` gesetzt werden (je nach μ wären aber auch mit `dists.sim.create` und `type = "general"` sehr ähnliche Ergebnisse erzielbar, wenn der Großteil der Wahrscheinlichkeitsmasse von `wfkt.calc` abgedeckt wird). $\triangle$

Die Funktionen abgesehen von `MGWP.sim(.load)` sind aber nicht auf die W-Funktion angewiesen. Deshalb sind diese auch nicht von der beschriebenen Problematik betroffen.

5 Zusammenfassung und Ausblick

Ungeachtet der Anzahl der in einem Galton-Watson-Prozess vorkommenden unterschiedlichen Typen ist es mit dem Erarbeiteten möglich, deren grundlegende stochastische Struktur nachzuvollziehen. Dabei zeigte sich das bekannte Resultat¹, dass ein Aussterben der Prozesse i.A. keineswegs sicher ist. Bei entsprechenden Nachkommens- bzw. Reproduktionsverteilungen überleben die Prozesse mit einer positiven Wahrscheinlichkeit. Diesfalls wächst deren Population über alle Schranken.

Dem Vorhaben entsprechend war es machbar, im Kern vollständige Beweisketten herzustellen. Den vorgelegten Argumenten und Herleitungen sollte durchgängig gefolgt werden können – möglichst ohne den Bedarf zusätzlicher Ressourcen. Sollte dies an manchen Stellen dennoch nicht zutreffen, wurde versucht, entsprechende Literaturhinweise mit anzugeben. Es sei auch erneut auf den mathematischen Appendix A verwiesen.

Im konkreten Fall einfacher GWP mit geometrischen Nachkommensverteilungen konnte die (im Überlebensfall stetige) Verteilung des Grenzwertes der normalisierten Populationsgrößen ermittelt werden. Selbiges wurde für den Aussterbezeitpunkt und die Gesamtpopulation erzielt (wobei es sich hier jeweils um diskrete Verteilungen handelt). Erwartungsgemäß waren die softwaregestützten Berechnungen im Einklang mit den theoretischen Erkenntnissen. Dies betraf auch die Analysen des beispielhaften Multitype GWP. Aber auch bzgl. simulierter Datensätze zeigten sich durchwegs plausible Resultate. Beides zusammengenommen kann eventuell weiter der besseren Nachvollziehbarkeit dienen.

Insgesamt konnten die reichhaltig vorhandenen Theorien über Galton-Watson-Prozesse in dieser Magisterarbeit gewiss nur angerissen werden – und jene über Verzweigungsprozesse im Allgemeinen erst recht. Es bestand aber auch nie ein anderer Anspruch. Bezüglich der technischen Voraussetzungen musste festgestellt werden, dass der Umgang mit expliziten mathematischen Funktionen mittels **R** doch eher eingeschränkt ist. Das Paket '**Ryacas**' bietet zwar interessante Anwendungen (und es existieren auch andere diesbezügliche Implementationen). Aber angesichts der Ausrichtung von **R** auf die numerische Datenanalyse wären diese Versuche wohl besser in einem 'echten' CAS aufgehoben.

Davon abgesehen könnten die bereitgestellten **R**-Funktionen auch künftig bei tiefergehenden Analysen Verwendung finden und fortlaufend dahingehend ausgebaut werden. Dies kann auch weiterhin und verstärkt in Kombination mit bereits bestehenden Paketen geschehen, deren Potential in diesem Rahmen natürlich nicht ausgeschöpft wurde.

¹Es soll noch ergänzend hinzugefügt werden, dass die Aussterbewahrscheinlichkeit in den 1920er Jahren erstmals korrekt publiziert wurde – etwa ein halbes Jahrhundert nach der Veröffentlichung von GALTON und WATSON (vgl. erneut [25, 1.1, S. 2 f.]).

A Appendix – Mathematisch

A.1 Maß- und wahrscheinlichkeitstheoretische Grundbegriffe

Wie einleitend erwähnt, werden Grundkenntnisse in Maß- und Wahrscheinlichkeitstheorie vorausgesetzt. Es sollen im Folgenden aber dennoch systematisch wesentliche Grundlagen – insbesondere auch für diskrete Wahrscheinlichkeitsräume – angeführt werden.

Vieles orientiert sich an [30] und [38], aber auch teilweise etwa an [5], [15], [21], [36] oder [42]. An gewissen Stellen wird zudem expliziter auf Definitionen oder Resultate in den genannten Quellen referenziert.

A.1.1 Mengentheorie

Notation A.1 (Mengentheorie)

- Eine allgemeine, nichtleere *Grundmenge* wird meist mit Ω bezeichnet. Für die leere Menge steht das Symbol $\emptyset$.
- Allgemein ist ein *Mengensystem* über Ω eine Menge von Mengen, bestehend aus Teilmengen von Ω . Die *Potenzmenge* $\mathcal{P}$ einer Grundmenge Ω ist ein Mengensystem, in welchem *alle* Teilmengen von Ω enthalten sind, d.h. $\mathcal{P}(\Omega) := \{A : A \subseteq \Omega\}$.
- Eine σ -*Algebra* ist ein Mengensystem über einer Grundmenge Ω , welches spezielle Eigenschaften erfüllt (siehe die Def. A.2) und wird meist mit Σ (alternativ mit Skriptbuchstaben wie etwa $\mathcal{F}$ oder $\mathcal{G}$) bezeichnet. Sie ist eine Teilmenge der entsprechenden Potenzmenge, d.h. $\Sigma \subseteq \mathcal{P}(\Omega)$.
- Die *Borelsche- σ -Algebra* $\mathcal{B}(\Omega)$ ist die kleinste σ -Algebra, die alle offenen Mengen einer Menge Ω enthält und sie enthält alle sogenannten Borel-Mengen (insbesondere enthält beispielsweise $\mathcal{B}(\mathbb{R})$ u.a. alle offenen, abgeschlossenen und kompakten Mengen in $\mathbb{R}$). Für Details hierzu siehe z.B. [38, S. 11].
- *Paarweise disjunkte Mengen* $\{A_i\}_{i \in I}$, wobei I eine beliebige Indexmenge ist, bezeichnet Mengen, bei denen $A_i \cap A_j = \emptyset$ für alle $i, j \in I$ mit $i \neq j$ gilt.
- Eine *Partition* einer nichtleeren Menge Ω besteht aus Mengen $\{A_i\}_{i \in I}$, die alle Teilmengen von Ω sind sowie die Bedingungen $A_i \cap A_j = \emptyset$ für alle $i, j \in I$ mit $i \neq j$ und $\bigcup_{i \in I} A_i = \Omega$ erfüllen. Also solche, die paarweise disjunkt sind und eine Zerlegung von Ω bilden. Für eine *endliche* Partition muss auch die Indexmenge I endlich sein (bspw. $I = \{1, \dots, n\}$). Analoges gilt für eine (*höchstens*) *abzählbare* Partition (bspw. $I = \mathbb{N}$ – siehe den übernächsten Punkt sowie die gleich folgende Bemerkung).

- Die *Mächtigkeit* (vgl. z.B. [8]) einer Menge A wird als $|A|$ geschrieben. Ist A endlich, so ist $|A|$ die Anzahl der Elemente in A .

Aber auch für unendliche Mengen können mit dem Mächtigkeitsbegriff gewissermaßen 'Größenvergleiche' angestellt werden. So gilt etwa für zwei *gleichmächtige* Mengen A, B (vgl. z.B. [22, S. 137]):

$$|A| = |B| \quad \Longleftrightarrow \quad \text{es gibt eine bijektive Funktion } \pi : A \rightarrow B. \quad (\text{A.1})$$

- Eine Menge heißt *abzählbar*, wenn sie als eine Folge unterschiedlicher Glieder geschrieben werden kann. *Höchstens abzählbar* bedeutet, dass eine Menge entweder endlich oder abzählbar ist. Letzteres ist genau dann der Fall, wenn sie gleichmächtig zu den natürlichen Zahlen $\mathbb{N}$ ist. Das trifft beispielsweise (vgl. [22, S. 139 f.]) auf $\mathbb{Z}$ oder $\mathbb{Q}$, aber auch auf alle höherdimensionalen $\mathbb{N}^k$ mit $k \geq 1$ zu (siehe die nächste Bemerkung). Allgemeiner ist eine Vereinigung höchstens abzählbar vieler höchstens abzählbarer Mengen immer noch höchstens abzählbar. Oft wird einfach von abzählbaren Mengen gesprochen, wenn eigentlich höchstens abzählbare Mengen gemeint sind (es sind dann also endliche Mengen mitgemeint). Mengen mit einer größeren Mächtigkeit als $\mathbb{N}$ sind *überabzählbar*, was z.B. für $\mathbb{R}$ der Fall ist.
- Das *Komplement* einer Menge $A \subseteq \Omega$, also die Menge aller $\{x \in \Omega : x \notin A\}$, wird als A^c geschrieben. Es gilt $A^c = \Omega \setminus A$.
- Die *Indikatorfunktion* einer Menge $A \subseteq \Omega$ werde als $\mathbb{1}_A : \Omega \rightarrow \{0, 1\}$ bezeichnet und als

$$\mathbb{1}_A(\omega) := \begin{cases} 1, & \text{wenn } \omega \in A \text{ ist,} \\ 0, & \text{wenn } \omega \notin A \text{ ist,} \end{cases} \quad \text{für alle } \omega \in \Omega \text{ definiert.} \quad (\text{A.2})$$

Für die Indikatorfunktionen von Mengen $A, B \subseteq \Omega$ lässt sich leicht zeigen, dass

$$\begin{aligned} \mathbb{1}_{A \cap B} &= \mathbb{1}_A \mathbb{1}_B, & \mathbb{1}_{A \cup B} &= \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A \mathbb{1}_B, \\ \mathbb{1}_{A \setminus B} &= \mathbb{1}_A - \mathbb{1}_A \mathbb{1}_B & \text{und } \mathbb{1}_{A^c} &= 1 - \mathbb{1}_A \quad \text{gilt.} \end{aligned} \quad (\text{A.3})$$

△

Bemerkung zur Notation A.1 (Abzählbare Mengen)

Ist eine beliebige Menge A höchstens abzählbar, dann gilt $|A| \leq |\mathbb{N}|$. Deshalb und aufgrund von (A.1) gibt es stets eine Abzählung ihrer Elemente der Gestalt $\{x_i : i \in I\}$, wobei I eine höchstens abzählbare Indexmenge mit $I \subseteq \mathbb{N}$ ist.

Als Bijektion π , welche jeder Zahl aus $\mathbb{N}$ ein Paar aus $\mathbb{N}^2$ oder ein k -Tupel aus $\mathbb{N}^k$ zuordnet, bietet sich etwa die sogenannte *Cantorsche Paarungs-* bzw. *Tupelfunktion* an – vgl. etwa [48, 2.8, S. 33]. △

Definition A.2 (σ -Algebra – vgl. z.B. [38, Def. 1.7, S. 9])

Ein Mengensystem $\Sigma \subseteq \mathcal{P}(\Omega)$ heißt σ -Algebra (auf Ω), wenn Folgendes erfüllt ist:

- (i) $\Omega \in \Sigma$,
- (ii) wenn $A \in \Sigma$ ist, so ist auch das Komplement $A^c \in \Sigma$ und
- (iii) wenn aus einer gegebenen Folge von Mengen $\{A_i\}_{i \geq 1}$ jedes $A_i \in \Sigma$ ist, so ist auch deren Vereinigung $\bigcup_{i \geq 1} A_i \in \Sigma$.

Bemerkungen zur Definition A.2.

- (i) Eine σ -Algebra ist somit abgeschlossen bezüglich Vereinigungen, aber auch bezüglich Durchschnittsbildungen, da (laut den de-morganschen Regeln, kurz *deMorgan*) $\bigcap_{i \geq 0} A_i = \left(\bigcup_{i \geq 0} A_i^c \right)^c$ gilt.
- (ii) Für eine beliebige Teilmenge $\mathcal{E}$ der Potenzmenge von Ω , also $\mathcal{E} \subseteq \mathcal{P}(\Omega)$, gibt es (vgl. z.B. [15, Bem. u. Def. (1.6), S. 11]) genau eine kleinste σ -Algebra auf Ω , welche $\mathcal{E}$ enthält. Für diese wird der sogenannte σ -Operator verwendet und $\sigma(\mathcal{E})$ geschrieben (die von $\mathcal{E}$ erzeugte σ -Algebra).
- (iii) Die kleinste denkbare σ -Algebra ist $\{\emptyset, \Omega\}$ und die größte ist die Potenzmenge $\mathcal{P}(\Omega)$. Letztere wird beispielsweise meist bei abzählbarem Ω (z.B. $\Omega = \mathbb{N}_0$, $\Sigma = \mathcal{P}(\mathbb{N}_0)$) angewendet (vgl. z.B. [15, Bsp. (1.7), S. 11 und *Diskreter Fall*, S. 12]).
- (iv) Eine Teilmengensystem $\mathcal{G}$ einer σ -Algebra Σ (auf Ω), also $\mathcal{G} \subseteq \Sigma$, wird, sofern es selbst auch eine σ -Algebra (auf Ω) ist, als Teil- oder *Sub- σ -Algebra von Σ* bezeichnet (vgl. z.B. [36, Def. 7.34, S. 92]). $\triangle$

Definition A.3 (lim inf bzw. lim sup von Mengenfolgen – vgl. z.B. [30, Def. 1.13 und Bem. 1.14, S. 5])

Sei $\{A_n\}_{n \geq 1}$ eine Folge von Teilmengen von Ω , also $A_n \subseteq \Omega$ für alle $n \geq 1$. Dann werden

$$\liminf_{n \rightarrow \infty} A_n := \bigcup_{n \geq 1} \bigcap_{m \geq n} A_m \quad \text{bzw.} \quad \limsup_{n \rightarrow \infty} A_n := \bigcap_{n \geq 1} \bigcup_{m \geq n} A_m \quad (\text{A.4})$$

als *Limes inferior* bzw. *Limes superior* der Mengenfolge bezeichnet. Es gilt

$$\liminf_{n \rightarrow \infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ für fast alle } n\} = \{\omega \in \Omega : |\{n \in \mathbb{N} : \omega \notin A_n\}| < \infty\}, \quad (\text{A.5})$$

$$\limsup_{n \rightarrow \infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ für } \infty \text{ viele } n\} = \{\omega \in \Omega : |\{n \in \mathbb{N} : \omega \in A_n\}| = \infty\} \quad (\text{A.6})$$

$$\text{und } \liminf_{n \rightarrow \infty} A_n \subseteq \limsup_{n \rightarrow \infty} A_n. \quad (\text{A.7})$$

Definition A.4 (Bild bzw. Urbild von Teilmengen unter einer Abbildung)

Für eine Abbildung $f : \Omega \rightarrow \Omega'$ und $A' \subseteq \Omega'$ bzw. $A \subseteq \Omega$ bezeichnet man die Menge

$$f^{-1}(A') = \{\omega \in \Omega : \omega \in f^{-1}(A')\} := \{\omega \in \Omega : f(\omega) \in A'\} \quad (\text{A.8})$$

als *Urbild von A' unter f* . Umgekehrt wird die Menge

$$f(A) := \{f(\omega) \in \Omega' : \omega \in A\} = \{\omega' \in \Omega' : \exists \omega \in A : f(\omega) = \omega'\} \quad (\text{A.9})$$

Bild von A unter f genannt.

Sind $A' = \{a'\}$ bzw. $A = \{a\}$ *einelementige* Mengen, so schreibt man auch

$$\begin{aligned} f^{-1}(a') &:= f^{-1}(\{a'\}) \stackrel{(\text{A.8})}{=} \{\omega \in \Omega : f(\omega) = a'\} && \text{bzw.} \\ f(a) &:= f(\{a\}) && \left(\stackrel{(\text{A.9})}{=} \{\omega' \in \Omega' : \omega' = f(a)\} \right). \end{aligned}$$

Bemerkung zur Definition A.4. Das Bild $f(a)$ einer einelementigen Mengen ist immer ebenfalls einelementig. Das Urbild $f^{-1}(a')$ muss hingegen nicht einelementig, sondern kann auch leer sein oder mehrere Elemente beinhalten. $\triangle$

Lemma A.5 (Eigenschaften von Bildern bzw. Urbildern)

Sei $f : \Omega \rightarrow \Omega'$ eine Abbildung und seien $A, B \subseteq \Omega$, $A', B' \subseteq \Omega'$. Weiters seien I eine Indexmenge und jede der Mengenfolgen $\{A_i\}_{i \in I}$ bzw. $\{A'_i\}_{i \in I}$ so, dass $A_i \subseteq \Omega$ bzw. $A'_i \subseteq \Omega'$ für alle $i \in I$ gilt. Dann gilt u.a.

- bezüglich Teilmengen der Zielmenge Ω' :
 - (a) $A' \subseteq B' \Rightarrow f^{-1}(A') \subseteq f^{-1}(B')$ und $\omega' \in A' \Rightarrow f^{-1}(\omega') \subseteq f^{-1}(A')$.
 - (b) Das Bild eines Urbildes von A' ist eine Teilmenge von A' , also $f(f^{-1}(A')) \subseteq A'$.
 - (c) $f^{-1}(\bigcup_{i \in I} A'_i) = \bigcup_{i \in I} f^{-1}(A'_i)$.
- bezüglich Teilmengen der Definitionsmenge Ω :
 - (d) $A \subseteq B \Rightarrow f(A) \subseteq f(B)$ und $\omega \in A \Rightarrow f(\omega) \in f(A)$.
 - (e) Das Urbild eines Bildes von A ist eine Obermenge von A , also $A \subseteq f^{-1}(f(A))$.
 - (f) $f(\bigcup_{i \in I} A_i) = \bigcup_{i \in I} f(A_i)$.
- und beide betreffend:
 - (g) $f(\omega) \in A' \iff \omega \in f^{-1}(A')$ für $\omega \in \Omega$, aber i.A. nur $f^{-1}(\omega') \subseteq A \Rightarrow \omega' \in f(A)$ für $\omega' \in \Omega'$.
 - (h) $A \subseteq f^{-1}(A') \iff f(A) \subseteq A'$.
 - (i) $f^{-1}(\Omega') = \Omega$ und $f(\Omega) \subseteq \Omega'$.

Beweis. Da teilweise Resultate aus anderen Unterpunkten benötigt werden, richtet sich die Beweisreihenfolge daran.

- (a),(d) Der erste Teil von (a) ergibt sich aus der Urbilddefinition (A.8), denn aus $A' \subseteq B'$ folgt $\{\omega : f(\omega) \in A'\} \subseteq \{\omega : f(\omega) \in B'\}$. Der zweite ist das Analogon für eine einelementige Menge $\{\omega'\}$. Sehr ähnlich wie (a) ergibt sich (d) aus (A.9), denn aus $A \subseteq B$ folgt $\{f(\omega) : \omega \in A\} \subseteq \{f(\omega) : \omega \in B\}$.
- (g) Die erste Äquivalenz ist wieder eine direkte Konsequenz der Def. des Urbildes. Für die zweite Aussage beachte, dass wegen der ersten $f(\omega) = \omega'$ für $\omega \in f^{-1}(\omega')$ gilt. Für ein solches ω gilt wegen $f^{-1}(\omega') \subseteq A$ auch $\omega \in A$, woraus mit (d) $f(\omega) \in f(A)$ folgt. Zusammen gilt also $\omega' = f(\omega) \in f(A)$.
- (b) $f(f^{-1}(A')) \stackrel{(A.8)}{=} f(\{\omega : f(\omega) \in A'\}) \stackrel{(A.9)}{=} \{\omega' : \exists \omega \in \{\omega : f(\omega) \in A'\} : f(\omega) = \omega'\}$, d.h., für ein beliebiges $\omega' \in f(f^{-1}(A'))$ gibt es ein $\omega \in \{\omega : f(\omega) \in A'\}$ mit $f(\omega) = \omega'$. Für dieses ω gilt eben auch, dass $f(\omega) \in A'$ ist, woraus $\omega' = f(\omega) \in A'$ folgt. Kurz, es gilt für alle $\omega' \in f(f^{-1}(A'))$, dass $\omega' \in A'$ ist, also ist $f(f^{-1}(A')) \subseteq A'$.
- (c) $\omega \in f^{-1}(\bigcup_{i \in I} A'_i) \stackrel{(g)}{\iff} f(\omega) \in \bigcup_{i \in I} A'_i \iff \exists i \in I : f(\omega) \in A'_i$
 $\stackrel{(g)}{\iff} \exists i \in I : \omega \in f^{-1}(A'_i) \iff \omega \in \bigcup_{i \in I} f^{-1}(A'_i).$
- (e) $\omega \in A \stackrel{(d)}{\implies} f(\omega) \in f(A) \implies A \subseteq \{\omega : f(\omega) \in f(A)\} \stackrel{(A.8)}{=} f^{-1}(f(A)).$
- (f) $\omega' \in f(\bigcup_{i \in I} A_i) \stackrel{(A.9)}{\iff} \exists \omega \in \bigcup_{i \in I} A_i : f(\omega) = \omega' \iff \exists i \in I, \omega \in A_i : f(\omega) = \omega'$
 $\iff \exists i \in I : \omega' \in f(A_i) \iff \omega' \in \bigcup_{i \in I} f(A_i).$
- (h) ' $\Rightarrow$ ': Wenn $\omega \in A$ ist, gilt laut Voraussetzung auch, dass $\omega \in f^{-1}(A')$ ist
 $\stackrel{(d)}{\implies} f(\omega) \in f(f^{-1}(A')) \stackrel{(b)}{\subseteq} A'$. D.h., für alle $\omega \in A$ gilt $f(\omega) \in A' \implies f(A) \subseteq A'$.
- ' $\Leftarrow$ ': Sei $\omega \in A \stackrel{(d)}{\implies} f(\omega) \in f(A) \stackrel{\text{Voraus.}}{\subseteq} A'$. Also gilt für alle $\omega \in A$ dass $f(\omega) \in A'$ ist.
 Wegen (g) ist das gleichbedeutend mit $\omega \in f^{-1}(A')$, weshalb $A \subseteq f^{-1}(A')$ gilt.
- (i) Definitionsgemäß (wegen $f : \Omega \rightarrow \Omega'$) gilt $f(\omega) \in \Omega'$ für alle $\omega \in \Omega$. Deshalb ist $f^{-1}(\Omega') = \{\omega : f(\omega) \in \Omega'\} = \Omega$. Umgekehrt wird (außer bei Surjektivität) im Allgemeinen nicht jedes Element der Zielmenge angenommen, weshalb nur $f(\Omega) \stackrel{(A.9)}{=} \{f(\omega) \in \Omega' : \omega \in \Omega\} \subseteq \Omega'$ gilt. $\square$

A.1.2 Maßtheorie

Notation A.6 (Maßtheorie)

- Eine nichtleere Grundmenge Ω bildet zusammen mit einer σ -Algebra Σ (siehe die Definition A.2) einen *Messraum* (Ω, Σ) . Teilmengen von Ω , welche auch in Σ enthalten sind, heißen *messbare Mengen*.
- Ist ein Messraum (Ω, Σ) gegeben, so kann man ein *Maß* $\mu : \Sigma \rightarrow \bar{\mathbb{R}}$ definieren. Es handelt sich also um eine Mengenfunktion, die jeder Menge aus Σ einen Wert zuordnet (für die an diese gestellten Bedingungen siehe beispielsweise [38, Def. 1.32, S. 20]). Der Messraum und das Maß bilden zusammen einen sogenannten *Maßraum* (Ω, Σ, μ) . In den Beispielen A.7 werden ein paar spezielle Maße auf einem Messraum angeführt.
- Eine Menge $N \subseteq \Omega$ auf einem Maßraum wird als μ -*Nullmenge* bezeichnet, wenn $\mu(N) = 0$ ist. Wenn eine den Maßraum betreffende Aussage bis auf eine μ -Nullmenge N überall (also auf N^c) gilt, sagt man, sie gilt (μ) -*fast überall* (kurz μ -f.ü.).
- Einen Integralbegriff für eine auf einem Maßraum (Ω, Σ, μ) definierte, (erweitert) reellwertige Funktion f , d.h. $f : \Omega \rightarrow \bar{\mathbb{R}}$, bietet das sogenannte *Lebesgue-Integral* (bzgl. des Maßes μ). Wenn von 'Integrierbarkeit' gesprochen wird, ist üblicherweise die Lebesgue-Integrierbarkeit (siehe die Def. A.13) gemeint. Die Schreibweisen

$$\int_{\omega \in \Omega} f(\omega) d\mu(\omega) = \int_{\Omega} f d\mu = \int f(\omega) \mu(d\omega) = \int f d\mu \quad (\text{A.10})$$

werden äquivalent verwendet. Die erste ist die ausführlichste, die zweite eine kompaktere Variante. Die letzten beiden können geschrieben werden, wenn Ω klar ist – ohne die Angabe eines Integrationsbereiches ist immer die Integration über ganz Ω gemeint. Wird nur über eine Teilmenge $A \subseteq \Omega$ integriert, soll

$$\int_{\omega \in A} f(\omega) d\mu(\omega) = \int_A f d\mu := \int_{\omega \in \Omega} \mathbb{1}_A(\omega) f(\omega) d\mu(\omega) = \int_{\Omega} \mathbb{1}_A f d\mu \quad (\text{A.11})$$

gesetzt werden. Zur Konstruktion des Integrals siehe beispielsweise [38, 2.1, S. 27 ff.] oder [30, 4.1, S. 87 ff.] sowie die Skizzierung im Rahmen der Definition A.13. $\triangle$

Beispiele A.7 (Wichtige Maße)

- (a) **Zählmaß** (vgl. z.B. [21, Bsp. 2.2(b), S. 31]): Wird auf einem beliebigen Maßraum (Ω, Σ) die Funktion $\mu_Z : \Sigma \rightarrow \bar{\mathbb{R}}$ als

$$\mu_Z(A) := \begin{cases} |A|, & \text{falls } A \text{ endlich viele Elemente hat,} \\ \infty, & \text{sonst,} \end{cases} \quad (\text{A.12})$$

definiert, so ist (Ω, Σ, μ_Z) ein Maßraum und das Maß μ_Z wird als *Zählmaß* bezeichnet. Es findet beispielsweise in der *diskreten* Wahrscheinlichkeitstheorie (A.1.4) Anwendung. Insbesondere hilft es dort, Integrale in Summen überzuführen.

- (b) **Lebesgue-Stieltjes-** bzw. **Lebesgue-Maß**: Siehe etwa [38, Theorem 1.40, S. 24].
(c) **Wahrscheinlichkeitsmaße**: Siehe die Definition A.22. $\triangle$

Messbare Abbildungen

Für den *Messbarkeitsbegriff* bzw. *messbare Abbildungen* und ihre Eigenschaften sei etwa auf [38, Def. 1.38, S. 17] und [30, 1.4, S. 34 ff.] verwiesen. Eine grundlegende Definition und einen Auszug diesbezüglicher Resultate wird zusätzlich hier angeführt.

Definition A.8 (Messbare Abbildungen – vgl. z.B. [30, Def. 1.76, S. 34])

Seien zwei Messräume (Ω, Σ) und (Ω', Σ') gegeben. Eine Abbildung bzw. Funktion $f : \Omega \rightarrow \Omega'$ heißt $\Sigma - \Sigma'$ *messbar*, wenn

$$f^{-1}(A') \in \Sigma \quad \text{für alle } A' \in \Sigma' \text{ gilt.} \quad (\text{A.13})$$

Das bedeutet in anderen Worten, dass die Urbilder (Def. A.4) aller messbaren Mengen aus (Ω', Σ') messbare Mengen in (Ω, Σ) sind.

Soll nur die entsprechende σ -Algebra *einer* der Räume betont werden, wird von der *Messbarkeit von f bezüglich Σ oder Σ'* gesprochen. Sind beide Messräume klar, kann f auch kurz nur als *messbar* bezeichnet werden.

Bemerkungen zur Definition A.8. Für allgemein definierte (vgl. z.B. [30, Satz 1.88, S. 38]) und reellwertige (vgl. z.B. [15, Bsp. 1.27, S. 21]) Abbildungen $f : \Omega \rightarrow \Omega'$ ist Stetigkeit eine hinreichende Bedingung für die $\mathcal{B}(\Omega) - \mathcal{B}(\Omega')$ Messbarkeit.

Insbesondere sind konstante Abbildungen $f : \Omega \rightarrow \Omega'$ mit $f \equiv c \in \Omega'$ immer messbar, denn für alle $B' \in \Sigma'$ ist $f^{-1}(B')$ entweder Ω (wenn $c \in B'$ ist) oder die leere Menge $\emptyset$ (wenn $c \notin B'$ ist) und somit jeweils in Σ .

Ist Ω abzählbar und $\Sigma = \mathcal{P}(\Omega)$ (dann handelt es sich um einen diskreten Messraum) sind sogar alle Abbildungen $f : \Omega \rightarrow \Omega'$ messbar (vgl. z.B. [15, Bsp. 1.24, S. 21]). $\triangle$

Satz A.9 (Messbarkeit arithmetischer Operationen – vgl. [38, Satz 1.29, S. 18]; ohne Beweis). Seien ein Messraum (Ω, Σ) , messbare Abbildungen $f, g : \Omega \rightarrow \mathbb{R}$ und $\alpha, \beta \in \mathbb{R}$ gegeben. Dann sind sowohl $\alpha f + \beta g$ als auch $f \cdot g$ messbar. Außerdem ist, falls $g(\omega) \neq 0$ für alle $\omega \in \Omega$ gilt, $\frac{f}{g}$ ebenfalls messbar.

Satz A.10 (Messbarkeit von Koordinatenabbildungen – vgl. [30, Satz 1.90, S. 38]; ohne Beweis). Seien auf einem Messraum (Ω, Σ) die Abbildung $\mathbf{f} : \Omega \rightarrow \mathbb{R}^k$ sowie für $1 \leq i \leq k$ und $k \geq 1$ die Abbildungen $f_i : \Omega \rightarrow \mathbb{R}$ definiert. Dann gilt, dass

$$\mathbf{f} : \Sigma - \mathcal{B}(\mathbb{R}^k) \text{ messbar ist} \iff f_i : \Sigma - \mathcal{B}(\mathbb{R}) \text{ messbar für jedes } 1 \leq i \leq k \text{ ist.}$$

Satz A.11 (Von Abbildungen erzeugte σ -Algebren – vgl. [30, Satz 1.78 und Def. 1.79, S. 34 f.]; ohne Beweis). Seien Ω nichtleer, I eine beliebige Indexmenge und (Ω', Σ') sowie (Ω'_i, Σ'_i) für jedes $i \in I$ je ein Messraum. Weiters seien auf diesen Messräumen die Abbildungen $f : \Omega \rightarrow \Omega'$ bzw. $f_i : \Omega \rightarrow \Omega'_i$ definiert.

(i) Die von einer Abbildung f erzeugte σ -Algebra auf Ω sei – gemeinsam mit der Menge aller Urbilder von Mengen in Σ' , dem Urbild $f^{-1}(\Sigma')$ – als

$$\sigma(f) := f^{-1}(\Sigma') := \{f^{-1}(A') : A' \in \Sigma'\} \quad \text{definiert.} \quad (\text{A.14})$$

Das Urbild $f^{-1}(\Sigma')$ bzw. $\sigma(f)$ ist die kleinste σ -Algebra, bezüglich der f messbar ist.

(ii) Die von mehreren Abbildungen $\{f_i\}_{i \in I}$ erzeugte σ -Algebra auf Ω sei definiert als

$$\sigma(\{f_i\}_{i \in I}) := \sigma\left(\bigcup_{i \in I} \sigma(f_i)\right) \stackrel{(A.14)}{=} \sigma\left(\bigcup_{i \in I} f_i^{-1}(\Sigma'_i)\right). \quad (A.15)$$

Sie ist die kleinste σ -Algebra, bezüglich der alle f_i messbar sind.

Lemma A.12. Seien (Ω, Σ) und (Ω', Σ') Messräume und sei $\mathcal{G} \subseteq \Sigma$, also eine Sub- σ -Algebra von Σ (siehe die Bem. zur Def. A.2 (iv)). Ist $f : \Omega \rightarrow \Omega'$ messbar bzgl. $\mathcal{G}$, dann ist f auch messbar bzgl. Σ , denn es gilt für alle $B' \in \Sigma'$ $f^{-1}(B') \in \mathcal{G} \subseteq \Sigma \Rightarrow f^{-1}(B') \in \Sigma$.

Lebesgue-Integral

Definition A.13 (Lebesgue-Integrierbarkeit)

Seien (Ω, Σ, μ) ein Maßraum und $f : \Omega \rightarrow \overline{\mathbb{R}}$ messbar. Das Lebesgue-Integral wird sukzessive, für immer größere Klassen an messbaren Abbildungen f auf dem Messraum (Ω, Σ) definiert (Literaturhinweise zur Konstruktion werden in der Notation A.6 angegeben – diese können auch als Referenz für die vorliegende Definition herangezogen werden).

1. Schritt: Für nichtnegative, sogenannte *einfache Funktionen* der Form

$f(\omega) = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}(\omega)$ mit $\alpha_i \in \overline{\mathbb{R}}_+$, $A_i \in \Sigma$ und $n \geq 1$ (die A_i können immer auch so gewählt werden, dass sie eine Partition von Ω bilden) wird das Lebesgue-Integral von f bezüglich μ als $\int f d\mu := \sum_{i=1}^n \alpha_i \mu(A_i)$ definiert. Insbesondere gilt demnach (setze $n = 2$, $A_1 = A$, $A_2 = A^c$, $\alpha_1 = 1$ und $\alpha_2 = 0$)

$$\int \mathbb{1}_A d\mu = \mu(A) \quad \text{für alle } A \in \Sigma. \quad (A.16)$$

2. Schritt: Ist $f \geq 0$, also nichtnegativ, dann gibt es eine Folge nichtnegativer, einfacher Funktionen $\{f_n\}_{n \geq 1}$ wie im 1. Schritt, sodass f_n von unten gegen f strebt. D.h., es gilt $f_1(\omega) \leq f_2(\omega) \leq \dots \leq \lim_{n \rightarrow \infty} f_n = f$ für alle $\omega \in \Omega$, kurz $f_n \nearrow f$. Dann lautet die Integraldefinition $\int f d\mu := \lim_{n \rightarrow \infty} \int f_n d\mu$.

3. Schritt: Für allgemeine Funktionen $f : \Omega \rightarrow \overline{\mathbb{R}}$ wird f in einen Positivteil

$f^+(\omega) := \max\{0, f(\omega)\} = f(\omega) \mathbb{1}_{\{f(\omega) \geq 0\}}$ für alle $\omega \in \Omega$ und einen Negativteil

$f^-(\omega) := \max\{0, -f(\omega)\} = -f(\omega) \mathbb{1}_{\{f(\omega) < 0\}}$ für alle $\omega \in \Omega$ aufgeteilt.

Es gilt dann $f = f^+ - f^-$ und f^+ bzw. f^- sind nichtnegative Funktionen wie im 2. Schritt. Deshalb lautet die Def. für diesen Fall $\int f d\mu := \int f^+ d\mu - \int f^- d\mu$.

Nun heißt f bezüglich μ (*Lebesgue-*)*integrierbar*, falls die Integrale des Positiv- und des Negativteils beide endlich sind, d.h. $\int f^+ d\mu < \infty$ und $\int f^- d\mu < \infty$. Falls nicht beide Teile ∞ sind, heißt f bezüglich μ (*Lebesgue-*)*quasiintegrierbar*. Sind $\int f^+ d\mu = \infty$ und $\int f^- d\mu = \infty$, so ist f *nicht (Lebesgue-)integrierbar*.

Bemerkungen zur Definition A.13. Klarerweise sind integrierbare Funktionen auch quasiintegrierbar und Funktionen wie im ersten und zweiten Schritt sind immer zumindest quasiintegrierbar. Selbiges gilt für den Betrag $|f|$ von allgemeinen Funktionen f .

Das Lebesgue-Integral hat die gewöhnlichen, vom Riemann-Integral bekannten Eigenschaften wie Linearität und Monotonie (vgl. etwa [38, Satz 2.12, S. 33]). Es gilt daher, dass die Integrierbarkeit von $|f|$ äquivalent mit der Integrierbarkeit von f ist (da $-|f| \leq f \leq |f|$ und $|f| = f^+ + f^-$ ist). D.h., es gilt

$$-\infty < \int f d\mu < \infty \iff \int |f| d\mu < \infty \quad \text{und außerdem (A.17)}$$

$$\int f d\mu = \int_A f d\mu + \int_{A^c} f d\mu \quad \text{für } A \in \Sigma \quad \text{sowie} \quad \left| \int f d\mu \right| \leq \int |f| d\mu. \quad (\text{A.18})$$

Die Monotonieeigenschaft des Integrals bleibt auch erhalten, wenn die Beziehungen nur (μ) -f.ü. erfüllt sind (vgl. z.B. [30, Satz 4.9, S. 91]), weshalb u.a. das Folgende gilt:

$$f \underset{(\equiv)}{\leq} g \text{ f.ü.} \Rightarrow \int f d\mu \underset{(\equiv)}{\leq} \int g d\mu, \quad \text{insbes. } f = 0 \text{ f.ü.} \Rightarrow \int f d\mu = 0. \quad (\text{A.19})$$

Ist f nichtnegativ, dann gilt die letzte Implikation in beide Richtungen (siehe etwa [30, Satz 4.8(i), S. 90] – Analoges gilt klarerweise für $f \leq 0$), d.h.,

$$\text{wenn } f \geq 0 \text{ vorausgesetzt wird, gilt:} \quad f = 0 \text{ f.ü.} \iff \int f d\mu = 0. \quad (\text{A.20})$$

△

Die folgenden Resultate beschreiben weitere Eigenschaften des Lebesgue-Integrals.

Satz A.14 (Monotone Konvergenz (MON)) – vgl. z.B. [36, Sätze 9.20 und 9.31, S. 130 ff.]; ohne Beweis). Sei (Ω, Σ, μ) ein Maßraum. Seien weiters $f, \{f_n\}_{n \geq 1} : \Omega \rightarrow \overline{\mathbb{R}}$ so, dass alle $(f, \text{ü.})$ messbar sind und $g \leq f_1 \leq f_2 \leq \dots \leq \lim_{n \rightarrow \infty} f_n = f$ μ -fast überall (kurz: $g \leq f_n \nearrow f$ f.ü.) für ein $(f, \text{ü.})$ messbares $g : \Omega \rightarrow \overline{\mathbb{R}}$ mit $\int g^- d\mu < \infty$ gilt (Letzteres ist insbesondere für $g \equiv 0$ erfüllt). Dann gilt

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int \lim_{n \rightarrow \infty} f_n d\mu = \int f d\mu. \quad (\text{A.21})$$

Satz A.15 (Dominierte Konvergenz) – vgl. z.B. [36, Satz 9.33, S. 134]; ohne Beweis) Auf einem Maßraum (Ω, Σ, μ) seien $(f, \text{ü.})$ messbare $f, \{f_n\}_{n \geq 1} : \Omega \rightarrow \overline{\mathbb{R}}$ so definiert, dass $f_n \rightarrow f$ μ -f.ü. gilt. Gibt es eine integrierbare Funktion $g : \Omega \rightarrow \overline{\mathbb{R}}$, sodass $|f_n| \leq g$ für alle $n \geq 1$ und μ -f.ü. gilt, dann ist f (und jedes f_n) ebenso integrierbar und es gilt

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int \lim_{n \rightarrow \infty} f_n d\mu = \int f d\mu \quad \text{und} \quad \lim_{n \rightarrow \infty} \int |f_n - f| d\mu = 0. \quad (\text{A.22})$$

Satz A.16 (Lemma von Fatou (nichtnegativ)) – vgl. z.B. [30, Satz 4.21, S. 95]; ohne Beweis). Seien (Ω, Σ, μ) ein Maßraum und $\{f_n\}_{n \geq 1} : \Omega \rightarrow \overline{\mathbb{R}}_+$ (also nichtnegativ) so, dass alle f_n messbar sind. Dann gilt

$$\int \liminf_{n \rightarrow \infty} f_n d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu. \quad (\text{A.23})$$

Lemma A.17 (Integral über eine Nullmenge)

Seien (Ω, Σ, μ) ein Maßraum und $f : \Omega \rightarrow \mathbb{R}$ integrierbar oder nichtnegativ. Ist N eine μ -Nullmenge (also $\mu(N) = 0$), dann gilt

$$\int_N f d\mu = 0 \quad \text{und somit auch} \quad \int f d\mu = \int_{N^c} f d\mu. \quad (\text{A.24})$$

Beweis. Für alle $\omega \in N$ ist $\mathbb{1}_N(\omega)f(\omega) = f(\omega)$ und für alle $\omega \in N^c$ gilt $\mathbb{1}_N(\omega)f(\omega) = 0$. Daraus lässt sich $\mathbb{1}_N f = 0$ f.ü. schließen. Lt. (A.19) ist deshalb $\int_N f d\mu \stackrel{(\text{A.11})}{=} \int \mathbb{1}_N f d\mu = 0$, womit auch $\int f d\mu \stackrel{(\text{A.18})}{=} \int_N f d\mu + \int_{N^c} f d\mu = \int_{N^c} f d\mu$ folgt. $\square$

Lemma A.18 (Integral nichtnegativer Funktionen mit abzählbarem Träger – vgl. [38, Satz 2.13, S. 34])

Sei (Ω, Σ, μ_Z) ein Maßraum (μ_Z das Zählmaß wie im Bsp. A.7 (a)). Sei weiters eine messbare Funktion $f : \Omega \rightarrow \mathbb{R}$ mit $f \geq 0$ so, dass die Menge $\{f > 0\} (= \{\omega \in \Omega : f(\omega) > 0\})$ abzählbar ist (d.h., $f(\omega) > 0$ ist nur für abzählbar viele $\omega \in \Omega$ größer als 0). Dann gilt

$$\int_A f d\mu_Z = \sum_{\omega \in A} f(\omega) \quad \text{für alle } A \in \Sigma. \quad (\text{A.25})$$

Beweisskizze. Zunächst macht der Ausdruck auf der rechten Seite der Gleichung in (A.25) Sinn, da nur abzählbar viele Summanden $f(\omega)$ größer als 0 sind. Das Lemma wird hier nur für Indikatorfunktionen, also $f = \mathbb{1}_B$ mit $B \in \Sigma$, gezeigt. Darauf aufbauend kann die Aussage dann für einfache Funktionen und anschließend für nichtnegative Funktionen (wie im 1. und 2. Schritt der Def. A.13 und unter Verwendung der Integraldefinitionen ebendieser) gezeigt werden. Auf diese Weise lassen sich oft Resultate bezüglich des Lebesgue-Integrals beweisen. In [38, S. 34 f.] wird diese Vorgangsweise deshalb auch als 'Standardprozedur' bezeichnet und an der genannten Stelle findet sich auch ein vollständiger Beweis dieses Lemmas.

Seien nun also $B \in \Sigma$ und $f := \mathbb{1}_B$. Laut Voraussetzung muss die Menge $\{f > 0\}$ abzählbar sein. Das ist in diesem Fall gleichbedeutend damit, dass B abzählbar ist (denn $\{f > 0\} = \{\mathbb{1}_B > 0\} \stackrel{(\text{A.2})}{=} \{\omega \in \Omega : \mathbb{1}_B(\omega) = 1\} = \{\omega \in \Omega : \omega \in B\} = B$). Deshalb gilt

$$\begin{aligned} \int_A f d\mu_Z &= \int_{\omega \in A} \mathbb{1}_B(\omega) d\mu_Z(\omega) \stackrel{(\text{A.11})}{=} \int_{\omega \in \Omega} \overbrace{\mathbb{1}_A(\omega) \mathbb{1}_B(\omega)}^{= \mathbb{1}_{A \cap B}(\omega) \text{ (A.3)}} d\mu_Z(\omega) \stackrel{(\text{A.16})}{=} \mu_Z(A \cap B) \\ &\stackrel{(\text{A.12})}{=} |A \cap B| \stackrel{\text{zähle}}{\underset{\text{Elem.}}{=}} \sum_{\omega \in \Omega} \mathbb{1}_{A \cap B}(\omega) = \sum_{\omega \in \Omega} \underbrace{\mathbb{1}_A(\omega) \mathbb{1}_B(\omega)}_{\substack{\text{sinnvoll, da } B \text{ abzählbar ist} \\ = \begin{cases} \mathbb{1}_B(\omega) & \forall \omega \in A, \\ 0 & \forall \omega \notin A. \end{cases}}} = \sum_{\omega \in A} \mathbb{1}_B(\omega) = \sum_{\omega \in A} f(\omega). \end{aligned} \quad \square$$

Lemma A.19 (Maß mit Dichte – vgl. [38, Def. 2.35, Lemma 2.34, S. 50]; ohne Bew.)

Seien (Ω, Σ, μ) ein Maßraum und $f : \Omega \rightarrow \mathbb{R}$, $f \geq 0$ und messbar. Dann heißt $\mu_f : \Sigma \rightarrow \mathbb{R}$, welches als $\mu_f(A) := \int_A f d\mu$ für jedes $A \in \Sigma$ definiert ist, Maß mit Dichte f (bzgl. μ). Die Funktion f selbst ist die Dichte des Maßes μ_f (bzgl. μ).

Das so definierte μ_f ist tatsächlich für jedes solche f ein Maß. Weiters gilt für jede nichtnegative, messbare Funktion $g : \Omega \rightarrow \bar{\mathbb{R}}$

$$\int_A g d\mu_f = \int_A gf d\mu \quad \text{für alle } A \in \Sigma. \quad (\text{A.26})$$

Satz A.20 (Transformationssatz / Substitutionsregel – vgl. z.B. [42, Satz 9.4.1 bzw. Aufgabe 9.4.A, S. 163 f.]; ohne Beweis)

Seien (Ω, Σ, μ) ein Maßraum und (Ω', Σ') ein Messraum. Weiters sei

- $f : \Omega \rightarrow \Omega'$ messbar (genauer $\Sigma - \Sigma'$ messbar).
- $h : \Omega' \rightarrow \bar{\mathbb{R}}$ messbar (genauer $\Sigma' - \mathcal{B}(\bar{\mathbb{R}})$ messbar).
- $\mu_f : \Sigma' \rightarrow [0, \infty]$ das sog. Bildmaß von μ unter f (vgl. [42, 6.1, S. 79]), def. als

$$\mu_f(A') := \mu(f^{-1}(A')) \quad \text{für alle } A' \in \Sigma'.$$

Wenn h entweder nichtnegativ oder so ist, dass $h \circ f$ μ -(quasi-)integrierbar ist (das ist genau dann der Fall, wenn μ_f -(quasi-)integrierbar ist), gilt für alle $A' \in \Sigma'$

$$\int_{f^{-1}(A')} (h \circ f)(\omega) d\mu(\omega) = \int_{A'} h(\omega') d\mu_f(\omega'). \quad \text{Falls } A' = \Omega' \text{ ist, gilt } f^{-1}(A') \stackrel{\text{Lemma}}{\underset{\text{A.5(i)}}{=}} \Omega.$$

A.1.3 Wahrscheinlichkeitstheorie – Allgemein

Wahrscheinlichkeitsräume, Verteilungen und Verteilungsfunktionen

Definition A.21 (Wahrscheinlichkeitsraum (W-Raum)) – angelehnt an [38, Definition 3.2, S. 59])

Für ein Zufallsexperiment seien definiert:

- (i) **Ergebnisse und Ergebnismenge:** Alle möglichen (nicht notwendigerweise näher spezifizierten) *Ergebnisse* des Experiments werden in der nichtleeren *Ergebnismenge* Ω zusammengefasst. Alle Ergebnisse $\omega \in \Omega$ schließen sich bei einmaliger Durchführung gegenseitig aus und sie werden auch als *Realisierungen* des Experiments bezeichnet.
- (ii) **Ereignisse und Ereignisraum:** Jede Teilmenge von Ω wird als *Ereignis* bezeichnet. Alle Ereignisse, denen Wahrscheinlichkeiten zugeordnet werden, bilden den sogenannten *Ereignisraum* Σ (auf Ω), welcher eine σ -Algebra (siehe die Notation A.1 bzw. die Definition A.2) sein soll.
- (iii) **Wahrscheinlichkeitsmaß:** Jedem Ereignis $A \in \Sigma$ wird durch das *Wahrscheinlichkeitsmaß* $\mathbb{P}$ (siehe die Definition A.22 – kurz *W-Maß*) seine Wahrscheinlichkeit $\mathbb{P}(A)$ zugeordnet. Deshalb wird es synonym auch als *Wahrscheinlichkeitsverteilung* oder kurz *Verteilung* bezeichnet (vgl. etwa [15, 1.1.3, S. 13] - nicht zu verwechseln mit Verteilungsfunktionen etwa der Definition A.26).

Dann bilden die Ergebnismenge Ω (i), der Ereignisraum Σ (ii) und das Wahrscheinlichkeitsmaß $\mathbb{P}$ (iii) zusammen den *Wahrscheinlichkeitsraum* $(\Omega, \Sigma, \mathbb{P})$, durch welchen das Zufallsexperiment modelliert wird. Er wird zukünftig auch als W-Raum abgekürzt. Der Messraum (Ω, Σ) wird also, ausgestattet mit einem W-Maß $\mathbb{P}$, zu einem speziellen Maßraum, einen W-Raum $(\Omega, \Sigma, \mathbb{P})$.

Bemerkungen zur Definition A.21.

- (i) Ein Ereignis $A \subseteq \Omega$ tritt ein, wenn ein $\omega \in A$ das beobachtete Ergebnis des Experiments ist. Ereignisse werden meist in Großbuchstaben aus dem vorderen Teil des Alphabets wie $A, B, \dots$, ggf. versehen mit Indizes, geschrieben. Auch die leere Menge $\emptyset$ ist ein (unmögliches) Ereignis und Ω selbst ist das sichere Ereignis. Da die Ereignisse Mengen sind, können die üblichen mengentheoretischen Operationen mit ihnen durchgeführt werden – insbesondere ist A^c das Ereignis, dass A nicht eintritt.
- (ii) Ausgewählte Eigenschaften von Wahrscheinlichkeitsmaßen finden sich im Satz A.23. Im diskreten Fall wird die Verteilung meist durch eine sogenannte *Wahrscheinlichkeitsfunktion* (siehe die Definition A.40) definiert. $\triangle$

Definition A.22 (Wahrscheinlichkeitsmaß – vgl. z.B. [17, 1.3 (1), S. 5])

Seien Ω eine Ergebnismenge und Σ ein Ereignisraum (wie in der Def. A.21 (i) und (ii)). Dann ist das *Wahrscheinlichkeitsmaß* $\mathbb{P}$ auf (Ω, Σ) , kurz *W-Maß*, eine Mengenfunktion $\mathbb{P} : \Sigma \rightarrow [0, 1]$, welche die folgenden Eigenschaften erfüllt:

- (i) *Normiertheit:* $\mathbb{P}(\Omega) = 1$ und $\mathbb{P}(\emptyset) = 0$.
- (ii) *σ -Additivität:* Für jede Folge paarweise disjunkter Ereignisse $\{A_i\}_{i \geq 1}$ mit $A_i \in \Sigma$ für alle $i \geq 1$ gilt

$$\mathbb{P}\left(\bigcup_{i \geq 1} A_i\right) = \sum_{i \geq 1} \mathbb{P}(A_i). \quad (\text{A.27})$$

Bemerkung zur Definition A.22. Maßtheoretisch gesprochen ist ein Wahrscheinlichkeitsmaß ein durch die Normiertheit (i) auf $\mathbb{P}(\Omega) = 1$ spezifiziertes (σ -endliches) Maß. Die Forderung $\mathbb{P}(\emptyset) = 0$ ergäbe sich eigentlich aus (ii) mit $A_1 = \Omega$ und $A_i = \emptyset \ \forall i > 1$. $\triangle$

Satz A.23 (Eigenschaften von Wahrscheinlichkeitsmaßen – vgl. z.B. [15, Satz 1.11, S. 14 f.]; ohne Beweis). *Seien ein W-Raum $(\Omega, \Sigma, \mathbb{P})$ und Ereignisse $A, B \in \Sigma$ sowie $\{A_i\}_{i \geq 1}$, wobei ebenso jedes $A_i \in \Sigma$ ist, gegeben. $\mathbb{P}$ erfüllt u.a. die folgenden Eigenschaften:*

- (i) **Endliche Additivität:** Aus (A.27) folgt für paarweise disjunkte $\{A_i\}_{i=1}^n$ sofort

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i). \quad (\text{A.28})$$

Deshalb (und da A und A^c disjunkt sind sowie $A \cup A^c = \Omega$ ist) folgt

$$1 \stackrel{\text{Def.}}{\underset{\text{A.22(i)}}{=}} \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) \underset{(\text{A.28})}{=} \mathbb{P}(A) + \mathbb{P}(A^c) \Rightarrow \mathbb{P}(A) = 1 - \mathbb{P}(A^c). \quad (\text{A.29})$$

- (ii) **Subtraktivität und Monotonie:**

Ist $A \subseteq B$, so gilt $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$ und deshalb auch $\mathbb{P}(A) \leq \mathbb{P}(B)$.

- (iii) **σ -Subadditivität:** Für die (nicht notwendigerweise paarw. disj.) $\{A_i\}_{i \geq 1}$ gilt

$$\mathbb{P}\left(\bigcup_{i \geq 1} A_i\right) \leq \sum_{i \geq 1} \mathbb{P}(A_i). \quad (\text{A.30})$$

- (iv) **Stetigkeit von unten:** Aus $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$ folgt

$$\mathbb{P}\left(\bigcup_{i \geq 1} A_i\right) = \lim_{i \rightarrow \infty} \mathbb{P}(A_i). \quad (\text{A.31})$$

- (v) **Stetigkeit von oben:** Aus $A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots$ folgt hingegen

$$\mathbb{P}\left(\bigcap_{i \geq 1} A_i\right) = \lim_{i \rightarrow \infty} \mathbb{P}(A_i). \quad (\text{A.32})$$

Bemerkung zum Satz A.23. Anstatt $(\mathbb{P})$ -f.ü. sagt man bei einem W-Maß $(\mathbb{P})$ -*fast sicher* (kurz $\mathbb{P}$ -f.s. oder nur f.s.). Dementsprechend erfüllt ein *fast sicheres* Ereignis E , dass $\mathbb{P}(E) = 1$ ist (vgl. (A.29) – deshalb kann E auch als eine $(\mathbb{P})$ -Einsmenge bezeichnet werden). Umgekehrt ist eine $\mathbb{P}$ -Nullmenge ein *unmögliches Ereignis*. $\triangle$

Lemma A.24 (Erstes Borel-Cantelli-Lemma – vgl. z.B. [30, Satz 2.7 (i), S. 53]; ohne Beweis). *Seien $(\Omega, \Sigma, \mathbb{P})$ ein W-Raum und $\{A_n\}_{n \geq 1}$ eine Folge von Ereignissen mit $A_n \in \Sigma$ für alle $n \geq 1$. Dann folgt aus $\sum_{n \geq 1} \mathbb{P}(A_n) < \infty$ auch, dass $\mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 0$ ist. I.a.W.: Die WK dafür, dass unendlich viele der Ereignisse A_n eintreten, ist gleich 0.*

Die nächste Aussage ist eigentlich intuitiv klar, soll aber kurz formal gezeigt werden.

Lemma A.25 (Schnitt mit einer Eins- oder einer Nullmenge)

Sei $(\Omega, \Sigma, \mathbb{P})$ ein W-Raum und seien $A, E, N \in \Sigma$. Ist E ein fast sicheres Ereignis, also $\mathbb{P}(E) = 1$ (E ist also eine $\mathbb{P}$ -Einsmenge), dann gilt

$$\mathbb{P}(A \cap E) = \mathbb{P}(A) \quad \text{und insbesondere gilt} \quad \mathbb{P}(A \cap \Omega) = \mathbb{P}(A). \quad (\text{A.33})$$

Ist N ein fast unmögliches Ereignis, also $\mathbb{P}(N) = 0$ (N ist $\mathbb{P}$ -Nullmenge), so gilt

$$\mathbb{P}(A \cap N) = 0 \quad \text{und insbesondere gilt} \quad \mathbb{P}(A \cap \emptyset) = 0. \quad (\text{A.34})$$

Beweis. Für $B, C \in \Sigma$ gilt allgemein $B \cap C \subseteq B$, woraus (wegen des Satzes A.23 (ii)) $\mathbb{P}(B \cap C) \leq \mathbb{P}(B)$ folgt. Daher gilt $\mathbb{P}(A \cap N) \leq \mathbb{P}(N) = 0$, womit der erste Teil von (A.34) gezeigt ist, aber auch $\mathbb{P}(A \cap E) \leq \mathbb{P}(A)$. Andererseits ist $\mathbb{P}(A \cap E) \stackrel{=0, \text{ da } \mathbb{P}(E)=1 \text{ ist}}{=} \mathbb{P}(A)$.
 $1 - \mathbb{P}(A \cap E) \stackrel{(\text{A.29})}{=} \mathbb{P}((A \cap E)^c) \stackrel{\text{deMorgan-Regeln}}{=} \mathbb{P}(A^c \cup E^c) \stackrel{(\text{A.30})}{\leq} \mathbb{P}(A^c) + \overbrace{\mathbb{P}(E^c)}^{=0, \text{ da } \mathbb{P}(E)=1 \text{ ist}} \stackrel{(\text{A.29})}{=} 1 - \mathbb{P}(A)$
 $\Rightarrow \mathbb{P}(A) \leq \mathbb{P}(A \cap E)$. Zusammengenommen ergibt sich der erste Teil von (A.33). Für die jeweiligen Sonderfälle verwende $\mathbb{P}(\Omega) = 1$ bzw. $\mathbb{P}(\emptyset) = 0$ (siehe die Def. A.22 (i)). $\square$

Für *reelle* Wahrscheinlichkeitsräume $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ lassen sich sogenannte (Wahrscheinlichkeits-) *Verteilungsfunktionen* $F_{\mathbb{P}}$ definieren. Zu jedem $\mathbb{P}$ gibt es eine eindeutig zugehörige $F_{\mathbb{P}}$ und umgekehrt (vgl. etwa [42, Korrespondenzsatz 12.1.1, S. 246]). Somit besteht eine Eins-zu-eins-Entsprechung zwischen W-Maßen und Verteilungsfunktionen (beide modellieren ein und dasselbe Zufallsexperiment).

Definition A.26 (Verteilungsfunktionen von $\mathbb{P}$ (auf $\mathbb{R}$) - angelehnt an [38, Def. 3.12, S. 69]). Für W-Maße bzw. Verteilungen $\mathbb{P}$ auf $\mathbb{R}$ (also ist $\mathbb{P} : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$), sprich gegeben einen W-Raumes $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$, definiere $F_{\mathbb{P}} : \mathbb{R} \rightarrow [0, 1]$ als

$$F_{\mathbb{P}}(x) := \mathbb{P}((-\infty, x]) \quad \text{für } x \in \mathbb{R}. \quad (\text{A.35})$$

$F_{\mathbb{P}}$ wird als *Verteilungsfunktion von $\mathbb{P}$* bezeichnet.

Zufallsvariablen allgemein

Definition A.27 (Zufallsvariable bzw. -vektor – angelehnt an [38, Def. 4.1, S. 90]) Seien $(\Omega, \Sigma, \mathbb{P})$ ein W-Raum und (Ω', Σ') ein Messraum, auch *Bildraum* genannt. Jede Abbildung $X : \Omega \rightarrow \Omega'$ mit der Eigenschaft

$$X^{-1}(A') \in \Sigma \quad \text{für alle } A' \in \Sigma' \quad (\text{A.36})$$

wird als *Zufallsvariable* (kurz *ZV*) bezeichnet. Ist $(\Omega', \Sigma') = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, so nennt man X eine *reelle* oder *reellwertige* ZV (erweitert reellwertig im Fall $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$).

Wenn der Bildraum $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ oder $(\overline{\mathbb{R}}^k, \mathcal{B}(\overline{\mathbb{R}}^k))$ mit $k > 1$ ist, wird auch der Begriff (*k-dimensionaler*) *Zufallsvektor* (kurz *Z-Vektor*) verwendet.

Bemerkungen zur Definition A.27.

- (i) Die Beziehung (A.36) garantiert, dass eine Zufallsvariable X eine *messbare Abbildung* (gemäß der Def. A.8 – genauer $\Sigma - \Sigma'$ messbar) ist. Damit sind die üblichen arithmetischen Operationen durchführbar (auf diese wird nicht genauer eingegangen, sondern etwa auf den Satz A.9, [20, 3.3, S. 14] oder [5, Satz 3.1.4, S. 74] verwiesen). Ist eine Abbildung also nicht messbar, ist sie keine Zufallsvariable.
- (ii) Das Urbild $X^{-1}(A')$ (vgl. die Def. A.4) von A' in (A.36) soll auch als

$$X^{-1}(A') \stackrel{(\text{A.8})}{=} \{\omega \in \Omega : X(\omega) \in A'\} =: \{X \in A'\} \quad (\text{A.37})$$

geschrieben werden.

- (iii) Für einen Z-Vektor $\mathbf{X} := (X_1, \dots, X_k)^T : \Omega \rightarrow \overline{\mathbb{R}}^k$ ist (A.36) laut dem Satz A.10 gleichbedeutend damit, dass alle X_i mit $1 \leq i \leq k$ $\mathcal{B}(\overline{\mathbb{R}})$ -messbar sind. Dementsprechend ist $\mathbf{X}$ genau dann ein Zufallsvektor, wenn seine Komponenten (erweitert) reellwertige Zufallsvariablen sind (vgl. den Unterabschnitt 3.1.3). $\triangle$

Entsprechend dem Satz A.11 ist die kleinste σ -Algebra, bezüglich der eine ZV X messbar (und somit noch eine ZV) ist, als

$$\sigma(X) \stackrel{(\text{A.14})}{=} X^{-1}(\Sigma') \stackrel{(\text{A.14})}{=} \{X^{-1}(A') : A' \in \Sigma'\} \stackrel{(\text{A.37})}{=} \{\{X \in A'\} : A' \in \Sigma'\} \quad (\text{A.38})$$

gegeben. Sie wird deshalb als die von X *erzeugte* σ -Algebra bezeichnet. Analoges (siehe (A.15)) gilt für die von mehreren Zufallsvariablen $\{X_i\}_{i \in I}$ erzeugte σ -Algebra $\sigma(\{X_i\}_{i \in I})$, welche somit die kleinste ist, bezüglich der alle X_i messbar sind. Schreibe, falls etwa $I \subseteq \mathbb{N}_0$ und $n \geq 0$ ist, beispielsweise $\sigma(X_0, \dots, X_n) := \sigma(\{X_i\}_{i=0}^n)$.

Nun soll ein Wahrscheinlichkeitsmaß auf dem Bildraum definiert werden.

Satz A.28 (Bildmaß bzw. Verteilung einer ZV – vgl. z.B. [15, Satz 1.28, S. 21])

Sei eine Zufallsvariable X , welche von einem W-Raum $(\Omega, \Sigma, \mathbb{P})$ in einen Bildraum (Ω', Σ') abbildet, gegeben. Für alle $A' \in \Sigma'$ wird das sogenannte *Bildmaß* $\mathbb{P}_X : \Sigma' \rightarrow [0, 1]$ (oder das *Bildmaß* von $\mathbb{P}$ unter X) als

$$\mathbb{P}_X(A') := \mathbb{P}(X^{-1}(A')) \stackrel{(\text{A.37})}{=} \mathbb{P}(\{X \in A'\}) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A'\}) \quad (\text{A.39})$$

definiert. Dann ist $\mathbb{P}_X$ ein W-Maß auf dem Bildraum und $(\Omega', \Sigma', \mathbb{P}_X)$ somit ein W-Raum. Insbesondere gilt $\mathbb{P}_X(\Omega') \stackrel{(A.39)}{=} \mathbb{P}(X^{-1}(\Omega')) \stackrel{\text{Lemma A.5(i)}}{=} \mathbb{P}(\Omega) \stackrel{\text{Def. A.22(i)}}{=} 1$. $\mathbb{P}_X$ heißt auch Verteilung von X (analog zur Def. A.21 (iii)).

Beweis. Wegen (A.36) kann $\mathbb{P}_X$ für $A' \in \Sigma$ wie in (A.39) definiert werden. Um zu zeigen, dass $\mathbb{P}_X$ ein W-Maß ist, müssen (gemäß der Def. A.22) die Normiertheit (i) und die σ -Additivität (ii) gezeigt werden. In der Formulierung des Satzes wurde (i) bereits gezeigt. Für (ii) betrachte paarweise disjunkte Ereignisse $A'_i \in \Sigma'$ für $i \geq 1$. Wenn dann für ein ω und $i \neq j$ gilt, dass $X(\omega) \in A'_i$ ist, kann nicht gleichzeitig $X(\omega) \in A'_j$ sein ($A'_i \cap A'_j = \emptyset$). Demnach sind auch die Ereignisse $\{X \in A'_i\}$ und $\{X \in A'_j\}$ disjunkt und somit sind dann auch alle Urbilder $\{X^{-1}(A'_i)\}_{i \geq 1}$ (vgl. (A.37)) paarweise disjunkt. Also folgt

$$\begin{aligned} \mathbb{P}_X\left(\bigcup_{i \geq 1} A'_i\right) &\stackrel{(A.39)}{=} \mathbb{P}\left(X^{-1}\left(\bigcup_{i \geq 1} A'_i\right)\right) \stackrel{\text{Lemma A.5 (c)}}{=} \mathbb{P}\left(\bigcup_{i \geq 1} X^{-1}(A'_i)\right) \stackrel{\sigma\text{-Add. (A.27)}}{=} \sum_{i \geq 1} \mathbb{P}(X^{-1}(A'_i)) \\ &\stackrel{(A.39)}{=} \sum_{i \geq 1} \mathbb{P}_X(A'_i), \end{aligned} \quad \text{womit die } \sigma\text{-Add. auch für } \mathbb{P}_X \text{ erfüllt ist.} \quad \square$$

Im Beispiel A.32 (i) wird auf die Bildmaßdefinition in (A.39) zur zusätzlichen Überprüfung der sogenannte Transformationssatz (Lemma A.31) angewendet.

Die nächste Definition erläutert erneut den Begriff der Verteilung einer Zufallsvariablen und weitere damit in Zusammenhang stehende Begrifflichkeiten.

Definition A.29 (Begrifflichkeiten bzgl. ZV – angelehnt an [30, S. 43])

Sei X eine Zufallsvariable (wie in der Definition A.27).

- (i) **Verteilung von X :** Die Bezeichnung *Verteilung* von X (auf dem Bildraum (Ω', Σ')) für das W-Maß $\mathbb{P}_X$ macht Sinn, da es jedem Ereignis $A' \in \Sigma'$ seine Wahrscheinlichkeit zuordnet. Der Satz A.28 ($\mathbb{P}_X = \mathbb{P} \circ X^{-1}$) liefert den Zusammenhang zur Verteilung $\mathbb{P}$ (auf dem Urbildraum $(\Omega, \Sigma, \mathbb{P})$). Für $\mu = \mathbb{P}_X$ schreibt man $X \sim \mu$ und sagt, dass X nach μ verteilt ist (handelt es sich um eine bekannte Verteilung, z.B. die geometrische, schreibt man $X \sim \text{Geom}(p)$ und sagt X ist geometrisch verteilt).
- (ii) **Integrierbarkeit von X :** Für (erweitert) reellwertige X seien der Positivteil X^+ und der Negativteil X^- analog zum 3. Schritt der Def. A.13 definiert. Der Begriff der (Quasi-)Integrierbarkeit von X (bzgl. $\mathbb{P}$) richtet sich ebenfalls an der genannten Definition und wird in der Definition A.36 nochmals präzisiert.
- (iii) **Identisch verteilt/Kopien von X :** Sei I eine beliebige (höchstens) abzählbare Indexmenge. Falls für eine Familie von Zufallsvariablen $\{X_i\}_{i \in I}$ für alle $i, j \in I$ gilt, dass $\mathbb{P}_{X_i} = \mathbb{P}_{X_j}$ ist, so heißen sie *identisch verteilt*. Für zwei identisch verteilte Zufallsvariablen X_1 und X_2 (es gilt also $\mathbb{P}_{X_1} = \mathbb{P}_{X_2}$) schreibt man $X_1 \stackrel{d}{=} X_2$. Gilt $X_i \stackrel{d}{=} X$ für alle $i \in I$, spricht man davon, dass $\{X_i\}_{i \in I}$ *Kopien von X* sind. Sind sie zudem unabhängig (siehe die Def. A.33 (iii)), nennt man sie *i.i.d.-Kopien von X* (*i.i.d.* für *independent and identically distributed*).

(iv) **Verteilungsfunktion von X :** Ist X reellwertig (also $(\Omega', \Sigma') = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$), definiere (analog zur Def. A.26) $F_X : \mathbb{R} \rightarrow [0, 1]$ als $F_X(x) := \mathbb{P}_X((-\infty, x])$ für $x \in \mathbb{R}$. F_X wird als *Verteilungsfunktion von X* bezeichnet (wenngleich es eigentlich die Verteilungsfunktion von $\mathbb{P}_X$ ist, da definitionsgemäß $F_X \stackrel{(A.35)}{=} F_{\mathbb{P}_X} \stackrel{(A.39)}{=} F_{\mathbb{P} \circ X^{-1}}$ gilt).

Notation A.30 (Zufallsvariablen)

- Zufallsvariablen werden meist in Großbuchstaben aus dem hinteren Teil des Alphabets wie $X, Y, \dots$ (ggf. versehen mit Indizes) geschrieben. Für ein beobachtetes Ergebnis $\omega \in \Omega$ ist $X(\omega)$ eine *Realisierung von X* und es werden meist Kleinbuchstaben dafür verwendet, bspw. $X(\omega) = x$.
- Wie bereits in der Definition A.29 praktiziert, wird der zugrundeliegende W-Raum $(\Omega, \Sigma, \mathbb{P})$ bei der Einführung einer oder mehrerer Zufallsvariablen oft nicht mit angegeben. Eine genaue Kenntnis über ihn ist häufig nicht unbedingt notwendig (er kann wesentlich komplizierter als die Beobachtungen der ZV sein) und seine konkrete Form wird dann meist offen gelassen. Auch wird vielfach aus dem Kontext des Zufallsexperiments klar, welcher gemeint ist. Allenfalls wird durch die Angabe der 'Wertigkeit' (z.B. $\mathbb{N}_0$ -wertig) der Bildraum (Ω', Σ') implizit mit angegeben.
- Um die verschiedenen Schreibweisen für die Verteilungen von ZV zusammenfassend darzustellen, sind diese hier kurz angeführt (links ausgedrückt mit dem W-Maß $\mathbb{P}_X$, rechts mit dem W-Maß $\mathbb{P}$ – die jeweils erste Gleichheit beruht auf dem Satz A.28).

$$\begin{aligned}
 & \mathbb{P}_X : \Sigma' \rightarrow [0, 1] & \mathbb{P} : \Sigma \rightarrow [0, 1] \\
 & \underbrace{\mathbb{P}_X(\{\omega' : \omega' \in A'\})}_{=\mathbb{P}_X(A')} = \underbrace{\mathbb{P}(\{\omega : \omega \in X^{-1}(A')\})}_{=\mathbb{P}(X^{-1}(A'))} = \underbrace{\mathbb{P}(\{\omega : X(\omega) \in A'\})}_{=\mathbb{P}(\{X \in A'\})} \quad \forall A' \subseteq \Omega' \quad (A.37) \\
 & \underbrace{\mathbb{P}_X(\{\omega' : \omega' = c\})}_{=\mathbb{P}_X(\{c\})} = \underbrace{\mathbb{P}(\{\omega : \omega \in X^{-1}(c)\})}_{=\mathbb{P}(X^{-1}(c))} = \underbrace{\mathbb{P}(\{\omega : X(\omega) = c\})}_{=\mathbb{P}(\{X=c\})} \quad \forall c \in \Omega' \quad (A.37)
 \end{aligned}$$

Oft werden die geschwungenen Klammern bei den Ausdrücken unter der Klammer ganz rechts – zur einfacheren Schreibweise – weggelassen und es wird meist nur $\mathbb{P}(X \in A')$, $\mathbb{P}(X = c)$, $\mathbb{P}(X \leq x)$, etc. geschrieben. Ähnlich wird für Schnittmengen verfahren und bspw. $\mathbb{P}(\{X \in A'\} \cap \{X \in B'\})$ zu $\mathbb{P}(X \in A', X \in B')$ vereinfacht. $\triangle$

Direkt aus dem Satz A.20 folgt das nächste Lemma.

Lemma A.31 (Transformationssatz für ZV/ Integration bzgl. des Bildmaßes – vgl. z.B. [38, Satz 4.3, S. 91])

Seien $X : \Omega \rightarrow \Omega'$ eine Zufallsvariable und $h : \Omega' \rightarrow \bar{\mathbb{R}}$ messbar und so, dass entweder h nichtnegativ oder $h \circ X : \Omega \rightarrow \bar{\mathbb{R}}$ $\mathbb{P}$ -(quasi-)integrierbar ist ($\iff h$ $\mathbb{P}_X$ -(quasi-)integrierbar ist). Dann gilt für alle $A' \in \Sigma'$

$$\int_{X^{-1}(A')} (h \circ X) d\mathbb{P} = \int_{A'} h d\mathbb{P}_X \quad \text{bzw.} \quad \int_{\Omega} (h \circ X) d\mathbb{P} = \int_{\Omega'} h d\mathbb{P}_X. \quad (A.40)$$

Beispiel A.32 (Transformationssatz / Bildmaß). Sei alles wie im Lemma A.31.

- (i) Mit $h(\omega') = 1$ für alle $\omega' \in \Omega'$ sieht man, dass die Definition des Bildmaßes wie in (A.39) dem Transformationssatz (A.40) genügt. Es gilt nämlich für $A' \in \Sigma'$

$$\begin{aligned} \mathbb{P}(X^{-1}(A')) &\stackrel{(A.16)}{=} \int_{\Omega} \mathbb{1}_{X^{-1}(A')} d\mathbb{P} \stackrel{(A.11)}{=} \int_{X^{-1}(A')} 1 d\mathbb{P} \stackrel{1=h(X(\omega))}{\stackrel{\forall \omega}{=}} \int_{X^{-1}(A')} (h \circ X) d\mathbb{P} \\ &\stackrel{(A.40)}{=} \int_{A'} h d\mathbb{P}_X \stackrel{h(\omega')=1}{\stackrel{\forall \omega'}{=}} \int_{A'} 1 d\mathbb{P}_X \stackrel{(A.11)}{=} \int_{\Omega'} \mathbb{1}_{A'} d\mathbb{P}_X \stackrel{(A.16)}{=} \mathbb{P}_X(A'). \end{aligned}$$

- (ii) Mit $h(\omega') = \omega'$ für alle $\omega' \in \Omega'$ kann man sich für (quasi-)integrierbare X die letzte Gleichheit in (A.43) der Erwartungswertdef. A.35 plausibel machen, denn

$$\int_{\omega \in \Omega} X(\omega) d\mathbb{P}(\omega) \stackrel{h(X(\omega))}{=} \int_{\Omega} (h \circ X) d\mathbb{P} \stackrel{(A.40)}{=} \int_{\Omega'} h d\mathbb{P}_X \stackrel{h(\omega')=\omega'}{\stackrel{\forall \omega'}{=}} \int_{\omega' \in \Omega'} \omega' d\mathbb{P}_X(\omega'). \quad \triangle$$

Unabhängigkeit

Es wird noch kurz auf die stochastische Unabhängigkeit eingegangen.

Definition A.33 ((stochastische) Unabhängigkeit – vgl. [38, Def. 5.6, S. 123])

Seien I eine beliebige (nichtleere) Indexmenge und $(\Omega, \Sigma, \mathbb{P})$ ein W-Raum.

- (i) Ereignisse $\{A_i\}_{i \in I}$, wobei jeweils $A_i \in \Sigma$ für alle $i \in I$ sei, heißen *unabhängig*, wenn für jede endliche (nichtleere) Teilmenge $J \subseteq I$ gilt, dass

$$\mathbb{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbb{P}(A_j) \quad \text{ist.} \tag{A.41}$$

- (ii) Eine *Familie von Mengensystemen* $\{\Sigma_i\}_{i \in I}$ mit $\Sigma_i \subseteq \Sigma$ für alle $i \in I$ heißt *unabhängig*, wenn jede Wahl einer Familie von Ereignissen $\{A_i\}_{i \in I}$, wobei jeweils $A_i \in \Sigma_i$ ist, unabhängig (wie in (i) definiert) ist.
- (iii) Eine *Familie von Zufallsvariablen* $\{X_i\}_{i \in I}$ auf dem W-Raum heißt *unabhängig*, wenn die Familie der erzeugten σ -Algebren $\{\sigma(X_i)\}_{i \in I}$ (siehe (A.38)) unabhängig (wie in (ii) definiert) ist.

Wichtig ist, dass die *paarweise Unabhängigkeit* alleine (z.B. im Falle der Def. A.33 (i) müsste dafür nur $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$ für alle $i, j \in I$ mit $i \neq j$ gelten) nicht ausreicht, sondern eben (A.41) für alle endlichen Teilmengen von I gelten muss.

Bezüglich Zufallsvariablen sind die in der Def. A.33 (iii) zu betrachtenden Mengensysteme $\sigma(X_i) \stackrel{(A.38)}{=} \{\{X_i \in A'\} : A' \in \Sigma'_i\}$, weshalb sich die Definition auch folgendermaßen schreiben lässt:

Für die Unabhängigkeit der ZV Zufallsvariablen $\{X_i\}_{i \in I}$ muss für jede Wahl von Ereignissen $\{A'_i\}_{i \in I}$, wobei jeweils A'_i aus Σ'_i ist, und alle nichtleeren endlichen Teilmengen $J \subseteq I$

$$\mathbb{P}\left(\bigcap_{j \in J} \{X_j \in A'_j\}\right) = \prod_{j \in J} \mathbb{P}(X_j \in A'_j) \quad \text{gelten.} \quad (\text{A.42})$$

Ist eine Familie von Zufallsvariablen unabhängig, so sind es auch (messbare) Funktionen davon, wie das nächste Lemma zeigt.

Lemma A.34 (Funktionen unabhängiger Zufallsvariablen)

Seien I eine beliebige (nichtleere) Indexmenge und $\{X_i\}_{i \in I}$ eine Familie unabhängiger ZV auf einem W -Raum $(\Omega, \Sigma, \mathbb{P})$ mit jeweiligem Messraum $(\Omega'_{X_i}, \Sigma'_{X_i})$. Für jedes $i \in I$ seien ein weiterer Messraum $(\Omega'_{Y_i}, \Sigma'_{Y_i})$ und eine messbare Abbildung $g_i : \Omega'_{X_i} \rightarrow \Omega'_{Y_i}$ definiert. Dann sind die ZV $\{Y_i\}_{i \in I}$ mit $Y_i := g_i \circ X_i$ auch unabhängig.

Beweis. Sei die Ereignisfamilie $\{B'_i\}_{i \in I}$ eine beliebige Wahl, jeweils mit $B'_i \in \Sigma'_{Y_i}$. Für jedes $i \in I$ sei dazu die Menge $A'_i := \{\omega' \in \Omega'_{X_i} : g_i(\omega') \in B'_i\}$ definiert. Diese lässt sich mit dem Lemma A.5 (g) umschreiben zu $A'_i = \{\omega' \in \Omega'_{X_i} : \omega' \in g_i^{-1}(B'_i)\} = g_i^{-1}(B'_i)$ und ist wegen der Messbarkeit der g_i jeweils ein Element aus Σ'_{X_i} (die Messbarkeit heißt hier genau, dass $g_i^{-1}(B'_i) \in \Sigma'_{X_i}$ für alle $B'_i \in \Sigma'_{Y_i}$ gilt – vgl. (A.13) bzw. (A.36)).

Für alle $i \in I$ gilt für die Zufallsvariable $Y_i = g_i \circ X_i = g_i(X_i)$, dass

$$\{Y_i \in B'_i\} = \{g_i(X_i) \in B'_i\} \stackrel{\text{Lemma A.5 (g)}}{=} \{X_i \in g_i^{-1}(B'_i)\} \stackrel{g_i^{-1}(B'_i)=A'_i}{=} \{X_i \in A'_i\} \quad \text{ist.} \quad (\star)$$

Deshalb gilt auch für jede endliche Teilmenge $J \subseteq I$

$$\mathbb{P}\left(\bigcap_{j \in J} \{Y_j \in B'_j\}\right) \stackrel{(\star)}{=} \mathbb{P}\left(\bigcap_{j \in J} \{X_j \in A'_j\}\right) \stackrel{\text{uabh. (A.42)}}{=} \prod_{j \in J} \mathbb{P}(X_j \in A'_j) \stackrel{(\star)}{=} \prod_{j \in J} \mathbb{P}(Y_j \in B'_j).$$

Genau das ist für die Unabhängigkeit der Zufallsvariablen $\{Y_i\}_{i \in I}$ ($Y_i : \Omega \rightarrow \Omega'_{Y_i}$) laut der Def. A.33 (iii) gefordert. $\square$

Die Beziehung (A.42) kann schwierig zu überprüfen sein, weshalb vereinfachende Kriterien für die Unabhängigkeit zur Verfügung stehen. Für *endlich viele reelle* Zufallsvariablen ($I = \{1, \dots, n\}$) ist es beispielsweise ausreichend, für jede Ereignisfamilienwahl alle simultan zu betrachten und auf die Teilmengen $J \subset I$ zu verzichten (vgl. z.B. [38, (5.2), S. 125]). Aber auch diese Überprüfung lässt sich u.U. weiter vereinfachen (siehe etwa den Satz A.45 am Ende des Unterabschnitts A.1.4 für den Fall diskreter Zufallsvariablen).

Erwartungswerte und Momente – allgemein

Es soll zuerst die allgemeine Definition des Erwartungswertes (kurz EW) für Zufallsvariablen mittels des Lebesgue-Integrals angeführt werden.

Definition A.35 (Erwartungswert allgemein – vgl. z.B. [38, Def. 4.4, S. 92])

Für eine quasiintegrierbare (erweitert) reellwertige Zufallsvariable X definiere den *Erwartungswert von X* als

$$\mathbb{E}X = \mathbb{E}[X] := \mathbb{E}_{\mathbb{P}}[X] := \int_{\Omega} X d\mathbb{P} \stackrel{(*)}{=} \int_{x \in \mathbb{R}} x d\mathbb{P}_X(x). \quad (\text{A.43})$$

($\star$) Hier kommt der Transformationssatz (Lemma A.31) mit $\Omega' = \overline{\mathbb{R}}$ zur Anwendung (siehe auch das Bsp. A.32 (ii)).

Die Schreibweisen $\mathbb{E}X$ und $\int X d\mathbb{P}$ können also äquivalent verwendet werden. Analog zum 1. Schritt der Definition A.13 gilt somit für ein Ereignis $A \in \Sigma$

$$\mathbb{E}\mathbb{1}_A \stackrel{(\text{A.43})}{=} \int \mathbb{1}_A d\mathbb{P} \stackrel{(\text{A.16})}{=} \mathbb{P}(A). \quad (\text{A.44})$$

Als Nächstes soll kurz auf die Begrifflichkeiten von *Existenz* bzw. *Endlichkeit* von Erwartungswerten eingegangen werden, da v.a. der Existenzbegriff in der Literatur nicht immer einheitlich verwendet wird.

Definition A.36 (Existenz bzw. Endlichkeit von Erwartungswerten – angelehnt an [42, Folgerung 12.3.4, S. 276])

Sei X eine (erweitert) reellwertige ZV. Es soll davon gesprochen werden, dass

- (i) X *einen Erwartungswert besitzt*, wenn X zumindest quasiintegrierbar (bzgl. $\mathbb{P}$) ist (siehe die Def. A.13 bzw. A.29 (ii)). Das ist gleichbedeutend damit, dass entweder $\mathbb{E}X^+ = \int_{x \in \mathbb{R}_+} x d\mathbb{P}_X(x) < \infty$ oder $\mathbb{E}X^- = \int_{x \in \mathbb{R}_-} x d\mathbb{P}_X(x) < \infty$ ist (oder beide) und der Erwartungswert von X kann dann wie in (A.43) definiert werden.
- (ii) der *Erwartungswert von X existiert*, wenn X einen endlichen Erwartungswert besitzt, d.h., wenn sowohl $\mathbb{E}X^+ < \infty$ als auch $\mathbb{E}X^- < \infty$ gilt (X ist $\mathbb{P}$ -integrierbar).

Wie in der Bem. zur Def. A.13 für (A.17) begründet ist (ii) genau dann der Fall, wenn $\mathbb{E}|X| = \int_{x \in \mathbb{R}} |x| d\mathbb{P}_X(x) < \infty$ ist. Zusammengefasst gilt also:

$$X \text{ integrierbar} \iff |X| \text{ integrierbar} \iff -\infty < \mathbb{E}X < \infty \iff \mathbb{E}|X| < \infty. \quad (\text{A.45})$$

Bemerkung zur Definition A.36. Der Punkt (ii), die Existenz, impliziert also (i), dass X einen EW besitzt. Sind $\mathbb{E}X^+$ und $\mathbb{E}X^-$ beide ∞ , *besitzt X keinen Erwartungswert*.

In diesem Kontext besitzt $|X|$ (oder allgemeiner eine nichtnegative ZV) also *immer* einen Erwartungswert ($\mathbb{E}|X|$, das erste absolute Moment von X) in den erweiterten reellen Zahlen $\overline{\mathbb{R}}$ und ist somit immer zumindest quasiintegrierbar. Dieser EW ist entweder $+\infty$ oder in $\mathbb{R}_+$. Im letzteren Fall ist er endlich und man sagt er *existiert*. $\triangle$

Am letzten Ausdruck in (A.43) in der allgemeinen EW-Definition erkennt man gut, dass der Erwartungswert einer reellwertigen integrierbaren ZV X eigentlich nur von deren Verteilung $\mathbb{P}_X$ abhängt. Daraus folgt direkt, dass *identisch verteilte* Zufallsvariablen (siehe die Def. A.29 (iii)) den selben EW besitzen. Wenn also zwei Zufallsvariablen $X, Y : \Omega \rightarrow \bar{\mathbb{R}}$ integrierbar und identisch verteilt sind, d.h. $X \stackrel{d}{=} Y$ ($\iff \mathbb{P}_X = \mathbb{P}_Y$), dann gilt

$$\mathbb{E}X \stackrel{(A.43)}{=} \int_{x \in \mathbb{R}} x d\mathbb{P}_X(x) = \int_{x \in \mathbb{R}} x d\mathbb{P}_Y(x) = \int_{y \in \mathbb{R}} y d\mathbb{P}_Y(y) \stackrel{(A.43)}{=} \mathbb{E}Y.$$

Mit dem Transformationssatz für ZV – Lemma A.31 – gilt das auch für die höheren Momente (sofern sie existieren) oder allgemein für passende Funktionen der ZV. D.h., wenn $X, Y : \Omega \rightarrow \Omega'$, $h : \Omega' \rightarrow \mathbb{R}$ messbar und entweder $h \geq 0$ oder $h(X)$ und $h(Y)$ (quasi-)integrierbar sind, gilt:

$$\text{Aus } X \stackrel{d}{=} Y \text{ (} \iff \mathbb{P}_X = \mathbb{P}_Y \text{)} \quad \text{folgt } \mathbb{E}h(X) = \mathbb{E}h(Y). \quad (A.46)$$

Eine manchmal nützliche Abschätzung nach oben bietet der folgende Satz.

Satz A.37 (Markov-Ungleichung – vgl. z.B. [38, S. 582]; ohne Beweis)

Seien X eine reelle Zufallsvariable und $\varepsilon \in \mathbb{R}$ mit $\varepsilon > 0$. Dann gilt

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}|X|}{\varepsilon}. \quad (A.47)$$

Für beliebige Momente, d.h. $\mathbb{E}X^r$ mit $r \in \mathbb{R}_+$ (bzw. insbesondere die k -ten Momente mit $k \in \mathbb{N}$) gilt das nächste Resultat.

Lemma A.38 (Höhere Momente – vgl. z.B. [37, Sec. 9.3 a., S. 157])

Sei X eine reellwertige ZV und seien $r, s \in \mathbb{R}_+$ so, dass $0 < r \leq s$ ist. Dann folgt aus $\mathbb{E}|X|^s < \infty$ auch $\mathbb{E}|X|^r < \infty$.

Insbesondere folgt für $j, k \in \mathbb{N}$ mit $j \leq k$ aus der Existenz des k -ten Moments die Existenz des j -ten Moments. I.a.W.: Aus der Integrierbarkeit von X^k folgt auch jene von X^j .

Beweis. Es ist $0 < r \leq s$, deshalb gilt

$$\begin{aligned} |X|^r &\leq \max(1, \underbrace{|X|^r}_{\leq |X|^s, \text{ falls } |X| \geq 1 \text{ ist}}) \leq \max(1, |X|^s) \leq 1 + |X|^s && \text{<}\infty, \text{ lt. Vorauss.} \\ &\Rightarrow \mathbb{E}|X|^r \leq \mathbb{E}[1 + |X|^s] = 1 + \underbrace{\mathbb{E}|X|^s}_{< \infty} < \infty. \end{aligned}$$

Beachte weiters, dass $|X(\omega)|^k = |X(\omega)^k|$ für alle $k \in \mathbb{N}$ und $\omega \in \Omega$ gilt. Deshalb ist auch $|X|^k$ genau dann integrierbar, wenn $|X^k|$ integrierbar bzw. (wegen (A.45) äquivalent dazu) X^k integrierbar ist. Dies zeigt, kombiniert mit der ersten, die zweite Aussage. $\square$

A.1.4 Wahrscheinlichkeitstheorie - Diskret

Eine spezifische Klasse von allgemeinen W-Räumen der Def. A.21 bzw. W-Maßen der Def. A.22 bilden die *diskreten W-Räume* bzw. *diskrete Verteilungen*. Solche liegen vor, wenn sich die verteilte Wahrscheinlichkeit auf abzählbare Teilmengen der Ergebnismenge Ω 'konzentriert' oder Ω selbst bereits abzählbar ist. Auch wenn in der vorliegenden Arbeit überwiegend Letzteres benötigt wird (wichtige Spezialfälle davon bilden etwa W-Räume mit $\Omega = \mathbb{N}_0^k$ oder $\Omega = \mathbb{N}_0$), werden in diesem Appendix auch überabzählbare Ergebnismengen zugelassen, da somit allgemeinere Aussagen getroffen werden können.

Dazu werden im folgenden Teilabschnitt zunächst allgemein diskrete W-Räume bzw. Verteilungen und der Begriff der Wahrscheinlichkeitsfunktion eingeführt, sowie im nächsten diskrete Zufallsvariablen definiert (wobei jeweils der Fall mit abzählbaren Ergebnismengen als Spezialfall besprochen wird). Anschließend werden ihre Erwartungswerte behandelt.

Diskrete Wahrscheinlichkeitsräume, Verteilungen/Wahrscheinlichkeitsfunktionen/Zähldichten

Definition A.39 (W-Raum / W-Maß – diskret – angelehnt an [38, Def. 3.6, S. 63])
Sei $(\Omega, \Sigma, \mathbb{P})$ ein W-Raum. Existiert eine abzählbare Menge

$$T \in \Sigma \quad (\Rightarrow T \subseteq \Omega) \quad \text{mit } \mathbb{P}(T) = 1, \quad (\text{A.48})$$

so wird T als ein *abzählbarer Träger von $\mathbb{P}$* bezeichnet (T ist i.A. nicht eindeutig).

Ein W-Raum mit abzählbarem Träger T wird als *diskreter Wahrscheinlichkeitsraum* bezeichnet. $\mathbb{P}$ heißt dann *diskretes W-Maß* oder, synonym, *diskrete Verteilung*.

Definition A.40 (Wahrscheinlichkeitsfunktion bzw. Zähldichte diskreter $\mathbb{P}$ – angelehnt an [38, Def. 3.7, S. 63])

Gegeben eines W-Raumes $(\Omega, \Sigma, \mathbb{P})$ mit abzählbarem Träger T (also eines diskreten W-Raumes), bezeichnet man die Funktion $p : \Omega \rightarrow [0, 1]$ mit

$$p(\omega) := \begin{cases} \mathbb{P}(\{\omega\}) & \text{für } \omega \in T, \\ 0 & \text{sonst} \end{cases} \quad \text{als Wahrscheinlichkeitsfunktion,} \quad (\text{A.49})$$

kurz *W-Funktion*, oder *Zähldichte von $\mathbb{P}$* . Es gilt

$$\sum_{\omega \in \Omega} p(\omega) \stackrel{p(\omega)=0}{=} \sum_{\text{für } \omega \notin T} p(\omega) \stackrel{(\text{A.49})}{=} \sum_{\omega \in T} \mathbb{P}(\{\omega\}) \stackrel{\sigma\text{-Add.}}{=} \mathbb{P}\left(\overbrace{\bigcup_{\omega \in T} \{\omega\}}^{=T}\right) = \mathbb{P}(T) \stackrel{(\text{A.48})}{=} 1 \quad (\text{A.50})$$

und der Ausdruck ganz links ist auch für ein überabzählbares Ω sinnvoll, da aufgrund des abzählbaren Trägers nur (höchstens) abzählbar viele Summanden ungleich 0 sind.

Bemerkungen zu den Def. A.39 und A.40 (Diskrete abzählbare W-Räume)

Ist Ω selbst abzählbar, kann direkt $T = \Omega$ gesetzt (da ja lt. der Def. A.22 (i) $\mathbb{P}(\Omega) = 1$ gilt) und auf den Begriff des abzählbaren Trägers verzichtet werden. Dann gilt:

- (i) Die kleinste σ -Algebra, die auch alle Einzelergebnisse $\{\omega\}$ mit $\omega \in \Omega$ enthält, ist deren Potenzmenge $\mathcal{P}(\Omega)$ (und damit auch die größte – siehe die Bem. zur Def. A.2 (iii)). Weiters ist $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ ein diskreter (abzählbarer) W-Raum mit dem diskreten W-Maß $\mathbb{P}$.
- (ii) Weiters gilt dann (lt. (A.49)) $p(\omega) = \mathbb{P}(\{\omega\})$ für alle $\omega \in \Omega$. Es fallen also die Begriffe *W-Funktion* (bzw. *Zähldichte*) und (*Wahrscheinlichkeits-*)*Verteilung* (bzw. *W-Maß*) zusammen und können synonym verwendet werden. Die W-Funktion kann dann auch als Folge $\{p_\omega\}_{\omega \in \Omega}$ mit $p_\omega := p(\omega)$ für alle $\omega \in \Omega$ geschrieben werden. $\triangle$

Anknüpfend zum Punkt (ii) der vorherigen Bemerkung gibt es aber auch für diskrete W-Räume mit überabzählbarem Ω einen direkten Zusammenhang zwischen W-Maßen bzw. Verteilungen und W-Funktionen, wie der nächste Satz zeigt.

Satz A.41 (Zusammenhang von W-Maßen und W-Funktionen durch Integration auf diskreten W-Räumen – angelehnt an [38, Satz 3.8, S. 63 f.]

- (i) Seien $(\Omega, \Sigma, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum (Def. A.39), also mit abzählbarem Träger T , p die zu $\mathbb{P}$ gehörige W-Funktion (Def. A.40) und μ_Z das Zählmaß (Bsp. A.7 (a)). Dann gilt für alle $A \in \Sigma$

$$\mathbb{P}(A) = \int_A p d\mu_Z = \sum_{\omega \in A \cap T} p(\omega). \quad (\text{A.51})$$

- (ii) Seien umgekehrt (Ω, Σ) ein Messraum, $T \subseteq \Omega$ abzählbar und eine Folge $\{p_\omega\}_{\omega \in T}$ so, dass $0 \leq p_\omega \leq 1$ für alle $\omega \in T$ und $\sum_{\omega \in T} p_\omega = 1$ ist. Definiere $p : \Omega \rightarrow [0, 1]$ für alle $\omega \in T$ durch $p(\omega) := p_\omega$ und ansonsten durch $p(\omega) := 0$. Dann gibt es genau ein diskretes W-Maß $\mathbb{P}$, sodass (A.51) für alle $A \in \Sigma$ gilt.

Beweis.

- (i) Da $\mathbb{P}(T) \stackrel{(\text{A.48})}{=} 1$ und somit T also eine $\mathbb{P}$ -Einsmenge ist, gilt

$$\mathbb{P}(A) \stackrel{(\text{A.33})}{=} \mathbb{P}(\underbrace{A \cap T}_{\text{abzählbar}}) = \mathbb{P}\left(\bigcup_{\omega \in A \cap T} \{\omega\}\right) \stackrel{\sigma\text{-Add. (A.27)}}{=} \sum_{\omega \in A \cap T} \mathbb{P}(\{\omega\}) \stackrel{(\text{A.49})}{=} \sum_{\omega \in A \cap T} p(\omega).$$

Für das Integral beachte, dass $p(\omega) = 0$ für $\omega \notin T$ gilt, weswegen die Menge $\{p > 0\} \subseteq T$ ebenso (wie T) abzählbar ist. Somit besagt das Lemma A.18 ($A, T \in \Sigma \Rightarrow A \cap T \in \Sigma$), dass

$$\sum_{\omega \in A \cap T} p(\omega) \stackrel{(\text{A.25})}{=} \int_{A \cap T} p d\mu_Z \stackrel{(\text{A.11})}{=} \int_{\Omega} \underbrace{\mathbb{1}_A \mathbb{1}_T}_{\stackrel{(\text{A.3})}{=} \mathbb{1}_A \mathbb{1}_T} p d\mu_Z = \int_{\Omega} \underbrace{p \mathbb{1}_T}_{\stackrel{(\star)}{=} p} \mathbb{1}_A d\mu_Z \stackrel{(\text{A.11})}{=} \int_A p d\mu_Z \text{ ist.}$$

($\star$) : Es gilt $p(\omega) \mathbb{1}_T(\omega) = p(\omega)$ für alle $\omega \in \Omega$, da $p(\omega) \stackrel{(\text{A.49})}{=} 0$ für $\omega \notin T$ gilt.

- (ii) Wenn $\mathbb{P} : \Sigma \rightarrow [0, 1]$ für $A \in \Sigma$ durch $\mathbb{P}(A) := \int_A p d\mu_Z$ definiert wird, ist $\mathbb{P}$ laut dem Lemma A.19 ein (eindeutiges) Maß und die erste Gleichheit in (A.51) gilt per Definition. Die zweite erhält man analog zur letzten Zeile des Beweises von (i) in umgekehrter Reihenfolge. Dass $\mathbb{P}$ dann auch ein W-Maß ist, ergibt sich durch $\mathbb{P}(\Omega) \stackrel{(A.51)}{=} \sum_{\omega \in \Omega \cap T} p(\omega) \stackrel{T \subseteq \Omega}{=} \sum_{\omega \in T} p(\omega) = 1$ (laut Voraussetzung). $\square$

Bemerkungen zum Satz A.41.

- (i) Im diskreten Fall besteht also eine direkte Entsprechung bzw. Korrespondenz zwischen Verteilungen und W-Funktionen. Wahrscheinlichkeitsverteilungen bestimmen in eindeutiger Weise Wahrscheinlichkeitsfunktionen und umgekehrt. Bei abzählbaren Ergebnismengen bzw. W-Räumen sind sie, wie in der Bemerkung zur Def. A.40 (ii) angeführt, sogar das Gleiche.
- (ii) Maßtheoretisch gesprochen zeigt (A.51), dass p eine Dichte des (W-)Maßes $\mathbb{P}$ bezüglich des Zählmaßes μ_Z ist (vgl. die entsprechende Definition im Lemma A.19). Daher rührt auch der Begriff 'Zähldichte'. $\triangle$

Diskrete Zufallsvariablen

Definition A.42 (Diskrete Zufallsvariable bzw. -vektor)

Eine Zufallsvariable heißt *diskret*, wenn ihre Verteilung $\mathbb{P}_X$ diskret ist (vgl. die Def. A.39). D.h., es gibt eine abzählbare Menge $T_X \in \Sigma'$ mit $\mathbb{P}_X(T_X) = \mathbb{P}(X \in T_X) = 1$ und (gemäß der Bem. zur Not. A.1) eine Indexmenge $I \subseteq \mathbb{N}$, mit welcher sich eine Abzählung der Elemente aus T_X der Form

$$\{x_i : i \in I\} := T_X = \{\omega' \in \Omega : \omega' \in T_X\} \quad \text{formulieren lässt.} \quad (\text{A.52})$$

Weiters kann analog zur Def. A.40 eine *W-Funktion* $p_X : \Omega' \rightarrow [0, 1]$ von $\mathbb{P}_X$ (synonym als *W-Funktion von X* bezeichnet) definiert werden als

$$p_X(\omega') := \begin{cases} \mathbb{P}_X(\{\omega'\}) \left(\stackrel{(A.39)}{=} \mathbb{P}(X = \omega') \right) & \text{für } \omega' \in T_X, \\ 0 & \text{sonst.} \end{cases} \quad (\text{A.53})$$

Wenn Ω' selbst abzählbar ist (dann kann $T_X = \Omega'$ gesetzt werden), ist X diskret.

Bemerkung zur Def. A.42 (Diskrete reellwertige ZV mit abzählbarem Wertebereich). Gibt es für eine ZV (bzw. einen Z-Vektor) $X : \Omega \rightarrow \bar{\mathbb{R}}^k$ mit $k \geq 1$ eine abzählbare Ergebnismenge Ω' , sodass $X(\omega) \in \Omega'$ für alle $\omega \in \Omega$ gilt, dann hat X einen *abzählbaren Wertebereich* (welcher Ω' ist) und man kann gleich $X : \Omega \rightarrow \Omega'$ definieren.

Gemäß der Definition ist so eine ZV X diskret und es existiert (mit $I \subseteq \mathbb{N}$) eine Abzählung $\{x_i : i \in I\}$ der Elemente aus Ω' , für die $\bigcup_{i \in I} \{x_i\} = \Omega'$ gilt. $\triangle$

Bemerkung A.43 (Diskreter Zufallsvektor). Seien für $1 \leq j \leq k$ mit $k \geq 1$ diskrete reellwertige Zufallsvariablen $X_j : \Omega \rightarrow \bar{\mathbb{R}}$ gegeben – also besitzt jedes X_j einen abzählbaren Träger $T_{X_j} \subset \bar{\mathbb{R}}$ mit $\mathbb{P}(X_j \in T_{X_j}) = 1$, zu welchem jeweils eine Abzählung $\{x_{i_j} : i_j \in I_j\}$ (mit $I_j \subseteq \mathbb{N}$) der Elemente aus T_{X_j} existiert.

A.1 Maß- und wahrscheinlichkeitstheoretische Grundbegriffe

Dann ist der (reelle) Zufallsvektor $\mathbf{X} := (X_1, \dots, X_k)^T : \Omega \rightarrow \bar{\mathbb{R}}^k$ ebenfalls diskret mit abzählbarem Träger $T_{\mathbf{X}} = T_{X_1} \times \dots \times T_{X_k} \in \bar{\mathbb{R}}^k$ (und $\mathbb{P}(\mathbf{X} \in T_{\mathbf{X}}) = 1$) sowie einer Abzählung $\{x_i : i \in I\}$ der Elemente aus $T_{\mathbf{X}}$ (mit $I \subseteq \mathbb{N}$), denn:

Da $I_j \subseteq \mathbb{N}$ für jedes $1 \leq j \leq k$ gilt, kann die unter der Bem. zur Not. A.1 erwähnte Bijektion $\pi : \mathbb{N}^k \rightarrow \mathbb{N}$ verwendet werden, um jedem k -Tupel $(i_1, \dots, i_k)^T \in \mathbb{N}^k$ genau ein $i \in \mathbb{N}$ zuzuordnen. Mit deren Hilfe definiere eine neue Indexmenge

$$I := \{\pi(i_1, \dots, i_k) : i_1 \in I_1, \dots, i_k \in I_k\} \quad \text{und} \quad (\star)$$

$$\begin{aligned} \mathbf{x}_i &:= \mathbf{x}_{\pi(i_1, \dots, i_k)} := (x_{i_1}, \dots, x_{i_k})^T \quad \text{für alle } x_{i_j} \in T_{X_j}, i_j \in I_j, 1 \leq j \leq k \quad (\star\star) \\ &\Rightarrow i = \pi(i_1, \dots, i_k) \in I \subseteq \mathbb{N}. \end{aligned}$$

Damit ist garantiert, dass I und auch $T_{\mathbf{X}} := \{\mathbf{x}_i : i \in I\}$ abzählbar sind. Es ist

$$\begin{aligned} T_{\mathbf{X}} &\stackrel{(\star\star)}{=} \{(x_{i_1}, \dots, x_{i_k})^T : \pi(i_1, \dots, i_k) \in I\} \stackrel{(\star)}{=} \{(x_{i_1}, \dots, x_{i_k})^T : i_1 \in I_1, \dots, i_k \in I_k\} \\ &\stackrel{(A.52)}{=} \{(x_{i_1}, \dots, x_{i_k})^T : x_{i_1} \in T_{X_1}, \dots, x_{i_k} \in T_{X_k}\} \stackrel{(3.4)}{=} T_{X_1} \times \dots \times T_{X_k} \quad (\star\star\star) \end{aligned}$$

und es gilt für $\mathbf{X} := (X_1, \dots, X_k)^T$, dass

$$\begin{aligned} \mathbb{P}(\mathbf{X} \in T_{\mathbf{X}}) &\stackrel{(\star\star\star)}{=} \mathbb{P}((X_1, \dots, X_k)^T \in T_{X_1} \times \dots \times T_{X_k}) \stackrel{(3.5)}{=} \mathbb{P}(X_1 \in T_{X_1}, \dots, X_k \in T_{X_k}) \\ &\stackrel{(A.37)}{=} \mathbb{P}(\underbrace{X_1^{-1}(T_{X_1}) \cap \dots \cap X_k^{-1}(T_{X_k})}_{\text{Einsmenge (Def. A.42)}}) \stackrel{(A.33)}{=} 1 \quad \text{ist.} \end{aligned}$$

Damit ist also $T_{\mathbf{X}} = \{\mathbf{x}_i : i \in I\}$ ein abzählbarer Träger von $\mathbf{X}$ (und $I \subseteq \mathbb{N}$), womit $\mathbf{X}$ gemäß der Definition A.42 diskret ist. $\triangle$

Lemma A.44 (Integration von Funktionen bzgl. des Bildmaßes diskreter ZV)
Seien $X : \Omega \rightarrow \Omega'$ eine diskrete ZV und $g : \Omega' \rightarrow \mathbb{R}$ eine messbare Funktion, die entweder nichtnegativ oder $\mathbb{P}_X$ -integrierbar ist. Dann gilt

$$\int_{\Omega'} g d\mathbb{P}_X \stackrel{(A.24)}{=} \int_{T_X} g d\mathbb{P}_X = \int_{T_X} g p_X d\mu_Z = \sum_{\omega' \in T_X} g(\omega') p_X(\omega'), \quad (\text{A.54})$$

wobei p_X die W-Funktion bzw. T_X ein abzählbarer Träger von $\mathbb{P}_X$ (siehe die Def. A.42) und μ_Z das Zählmaß (siehe das Bsp. A.7 (a)) ist.

Beweis. Für die erste Gleichheit beachte, dass T_X^c eine $\mathbb{P}_X$ -Nullmenge ist. Anschließend kann der Satz A.41 auf den diskreten W-Raum $(\Omega', \Sigma', \mathbb{P}_X)$ angewendet werden, um für alle $A' \in \Sigma'$ zu sehen, dass $\mathbb{P}_X(A') = \int_{A'} p_X d\mu_Z$ gilt (siehe (A.51)). Dementsprechend ist also $\mathbb{P}_X$ ein W-Maß mit Dichte p_X bezüglich des Zählmaßes μ_Z wie im Lemma A.19 beschrieben, aus welchem (A.26) die zweite Gleichheit in (A.54) zeigt.

Für die dritte Gleichheit betrachte zuerst den Fall mit $g \geq 0$. Wegen (A.53) gilt in diesem $\{g p_X > 0\} \subseteq \{p_X > 0\} \subseteq T_X$. Da T_X aber abzählbar ist, ist es somit die Menge $\{g p_X > 0\}$ ebenso, weshalb die Aussage (A.25) aus dem Lemma A.18 (etwa mit $f := g p_X$ und $A := T_X$) das entsprechende Argument liefert.

Für allgemeinere, $\mathbb{P}_X$ -integrierbare Funktionen g (dann ist auch $|g|$ integrierbar – siehe (A.17)) wird eine Beweisführung angewendet, welche dem 3. Schritt der Def. A.13 ähnelt. Spaltet man (analog dazu) g in $g = g^+ - g^-$ mit $g^+ = \mathbb{1}_{\{g \geq 0\}}g$ und $g^- = -\mathbb{1}_{\{g < 0\}}g$ auf, gilt sowohl $g^+ \geq 0$ als auch $g^- \geq 0$ (und, da $|g| = g^+ + g^-$ ist, sind wegen der Linearität und der Monotonie des Integrals beide $\mathbb{P}_X$ -integrierbar). Deshalb – und mit dem bisher Bewiesenen – gilt

$$\begin{aligned} \int_{\Omega'} g^+ d\mathbb{P}_X &\stackrel{(A.54)}{=} \sum_{\omega' \in T_X} g^+(\omega') p_X(\omega') \text{ und } \int_{\Omega'} g^- d\mathbb{P}_X \stackrel{(A.54)}{=} \sum_{\omega' \in T_X} g^-(\omega') p_X(\omega') \quad (\Delta) \\ \Rightarrow \int_{\Omega'} g d\mathbb{P}_X &= \underbrace{\int_{\Omega'} g^+ d\mathbb{P}_X}_{<\infty} - \underbrace{\int_{\Omega'} g^- d\mathbb{P}_X}_{<\infty} \stackrel{(\Delta)}{=} \sum_{\omega' \in T_X} g^+(\omega') p_X(\omega') - \sum_{\omega' \in T_X} g^-(\omega') p_X(\omega') \\ &\stackrel{(\star)}{=} \sum_{\omega' \in T_X} (g^+(\omega') - g^-(\omega')) p_X(\omega') = \sum_{\omega' \in T_X} g(\omega') p_X(\omega'). \end{aligned}$$

($\star$) Diese Zusammenfassung ist aufgrund der Konvergenz beider Reihen gestattet. $\square$

Korollar zum Lemma A.44. Für X und Funktionen g wie im Lemma gilt:

$$g \text{ } \mathbb{P}_X\text{-integrierbar} \iff g \circ X \text{ } \mathbb{P}\text{-integrierbar} \iff \sum_{\omega' \in T_X} |g(\omega')| p_X(\omega') < \infty.$$

Beweis. Die erste Äquivalenz folgt aus dem Lemma A.31. Für die zweite beachte, dass (A.54) für $|g|$ (da nichtnegativ) jedenfalls gilt. Somit gilt: $\sum_{\omega' \in T_X} |g(\omega')| p_X(\omega') < \infty \iff \int_{\Omega'} |g| d\mathbb{P}_X < \infty$. Letzteres ist gleichbedeutend mit der $\mathbb{P}_X$ -Integrierbarkeit von g . $\square$

Wie am Ende des Unterabschnitts A.1.3 angekündigt, liefert der nächste Satz ein einfacheres Kriterium als (A.42) für die Überprüfung der Unabhängigkeit endlich vieler, diskreter reeller Zufallsvariablen.

Satz A.45 (Unabhängigkeit diskreter reeller Zufallsvariablen – vgl. z.B. [38, Satz 5.10, S. 125 f.]; ohne Beweis)

Sei $n \geq 1$. Reelle diskrete ZV $\{X_i\}_{i=1}^n$, also jeweils mit abzählbarem Träger $T_{X_i} \subset \mathbb{R}$, sind genau dann unabhängig, wenn für jedes $(x_1, \dots, x_n)^T \in T_{X_1} \times \dots \times T_{X_n}$ gilt, dass

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n \mathbb{P}(X_i = x_i) \quad \text{ist.} \quad (\text{A.55})$$

Erwartungswerte – diskreter Fall

Zunächst wird ein Satz für die Erwartungswerte von Funktionen diskreter Zufallsvariablen formuliert. Damit kann einerseits gesehen werden, dass die bekannte Definition des Erwartungswertes diskreter reeller ZV über eine Summe (Def. A.47) eigentlich aus der allgemeinen Erwartungswertdefinition (Def. A.35) hervorgeht. Andererseits lassen sich damit eine große Klasse von EW von Transformationen diskreter ZV, insbesondere auch die Varianz und höhere Momente, bestimmen.

Satz A.46 (Erwartungswert von Funktionen diskreter Zufallsvariablen)

Sei X eine diskrete Zufallsvariable, also $X : \Omega \rightarrow \Omega'$ mit W-Funktion p_X und abzählbarem Träger T_X (von $\mathbb{P}_X$) sowie einer Abzählung $\{x_i : i \in I\}$ mit $I \subseteq \mathbb{N}$ der Elemente aus T_X . Weiter sei eine messbare Funktion $g : \Omega' \rightarrow \mathbb{R}$ entweder nichtnegativ oder so, dass

$$\sum_{i \in I} |g(x_i)| p_X(x_i) < \infty \quad \text{gilt.} \quad (\text{A.56})$$

Dann ist $g \circ X = g(X) : \Omega \rightarrow \bar{\mathbb{R}}$ eine (erweitert) reellwertige quasiintegrierbare (bzw. integrierbare, sofern (A.56) erfüllt ist) Zufallsvariable. Für ihren Erwartungswert gilt

$$\mathbb{E}g(X) = \sum_{\omega' \in T_X} g(\omega') p_X(\omega') \stackrel{(\text{A.52})}{=} \sum_{i \in I} g(x_i) p_X(x_i) \stackrel{(\text{A.53})}{=} \sum_{i \in I} g(x_i) \mathbb{P}(X = x_i). \quad (\text{A.57})$$

Bemerkung zum Satz A.46. Es wird nichts über die Summationsreihenfolge ausgesagt. Im Fall, dass (A.56) gilt, konvergiert die Reihe im letzten Ausdruck in (A.57) aber absolut und Umordnungen spielen laut dem Satz A.96 (ii) keine Rolle (und der Erwartungswert existiert). Ist g nichtnegativ, besagt der Satz A.97, dass Umordnungen sogar bei Divergenz erlaubt sind. Dann ist auch der Fall $\mathbb{E}g(X) = +\infty$ durch (A.57) abgedeckt ($g(X)$ besitzt bei nichtnegativem g also immer einen Erwartungswert). $\triangle$

Beweis des Satzes A.46. Bezüglich der Integrierbarkeit beachte, dass im Fall $g \geq 0$ auch $g \circ X \geq 0$ und somit (gemäß der Bem. zur Def. A.36) zumindest $\mathbb{P}$ -quasiintegrierbar ist. Ist g so, dass (A.56) gilt, dann ist $g \circ X$ (laut dem Korollar zum Lemma A.44) sogar $\mathbb{P}$ -integrierbar. In beiden Fällen sind also alle Voraussetzungen für das Lemma A.31 (Transformationssatz für ZV) und das Lemma A.44 (Integration von Funktionen diskreter Zufallsvariablen) erfüllt, weshalb deren Anwendung – unter Zuhilfenahme der allg. Erwartungswertdef. A.35 – die einzige bis dato unbegründete Gleichheit in (A.57) ergibt:

$$\mathbb{E}g(X) = \mathbb{E}g \circ X \stackrel{(\text{A.43})}{=} \int_{\Omega} g \circ X \, d\mathbb{P} \stackrel{(\text{A.40})}{=} \int_{\Omega'} g \, d\mathbb{P}_X \stackrel{(\text{A.54})}{=} \sum_{\omega' \in T_X} g(\omega') p_X(\omega'). \quad \square$$

Wie erwähnt ist es oft üblich, den Erwartungswert für diskrete reellwertige Zufallsvariablen direkt über Summen zu definieren (vgl. z.B. [17, 3.3 (1), S. 50]). Über den Satz A.46 geht diese Definition aber auch direkt aus der allgemeinen Erwartungswertdefinition (mit $g : \mathbb{R} \rightarrow \mathbb{R}$, $g(x) := x$) hervor.

Definition A.47 (Erwartungswert diskreter reeller Zufallsvariablen)

Sei eine diskrete ZV $X : \Omega \rightarrow \mathbb{R}$ (mit W-Funktion p_X , abzählbarem Träger $T_X \subset \mathbb{R}$ und einer entsprechenden Abzählung $\{x_i : i \in I\}$, $I \subseteq \mathbb{N}$, der Elemente aus T_X) integrierbar oder nichtnegativ. Dann kann der *Erwartungswert* von X definiert werden als

$$\mathbb{E}X := \sum_{\omega' \in T_X} \omega' p_X(\omega') \stackrel{(\text{A.52})}{=} \sum_{i \in I} x_i p_X(x_i) \stackrel{(\text{A.53})}{=} \sum_{i \in I} x_i \mathbb{P}(X = x_i).$$

A.2 Weitere Resultate und Begrifflichkeiten – Wahrscheinlichkeitstheorie

Im vorliegenden Abschnitt dieses Appendix werden unterschiedliche Themengebiete aus der Wahrscheinlichkeitsrechnung und Wahrscheinlichkeitstheorie behandelt. Vieles davon kommt im Hauptteil dieser Arbeit zur Anwendung, wo ggf. auch an die jeweils entsprechende Stelle verwiesen wird.

Zunächst wird auf den Faltungsbegriff bzgl. der Verteilung $\mathbb{N}_0$ -wertiger ZV (A.2.1) eingegangen. Anschließend werden allgemein EW und (Ko-)Varianzen (A.2.2), die bedingte WK und Erwartung (A.2.3), Markov-Ketten (A.2.4), Konvergenzarten von ZV (A.2.5), Laplace-Transformationen (A.2.6) und Martingale (A.2.7) besprochen.

Sei dafür stets ein Wahrscheinlichkeitsraum $(\Omega, \Sigma, \mathbb{P})$ gegeben.

A.2.1 Diskrete Faltung der Verteilungen $\mathbb{N}_0$ -wertiger Zufallsvariablen

Die Verteilungen bzw. Wahrscheinlichkeitsfunktionen von diskreten Zufallsvariablen mit Werten in $\mathbb{N}_0$ (siehe die Def. A.42 und die Bem. dazu sowie insbesondere die Def. 2.2) können jeweils als Folge betrachtet werden. Werden diese Folgen (gemäß der Definition A.98) gefaltet, erhält man die Verteilung der Summe dieser Zufallsvariablen – vorausgesetzt sie sind *unabhängig*. Dies wird im nächsten Teilabschnitt für zwei solche Zufallsvariablen vorgeführt. Im darauf folgenden wird die Verteilung der Summe mehrerer $\mathbb{N}_0$ -wertiger *i. i. d.*-Zufallsvariablen durch die mehrfache Faltung beschrieben.

Faltung der Verteilungen zweier unabhängiger, $\mathbb{N}_0$ -wertiger Zufallsvariablen

Seien für alles in diesem Teilabschnitt X und Y zwei unabhängige, $\mathbb{N}_0$ -wertige Zufallsvariablen. Die dazugehörigen Wahrscheinlichkeitsfunktionen sollen als $p = \{p_k\}_{k \geq 0}$ und $q = \{q_k\}_{k \geq 0}$ bezeichnet werden, wobei jeweils $p_k := \mathbb{P}(X = k)$ und $q_k := \mathbb{P}(Y = k)$ für jedes $k \geq 0$ ist.

Satz A.48 (Faltung der Verteilungen bzw. Verteilung der Summe zweier unabhängiger, $\mathbb{N}_0$ -wertiger ZV – vgl. [35, Satz 5.6, S. 85 f.]). *Es gilt*

$$\mathbb{P}(X + Y = k) = \sum_{i=0}^k \mathbb{P}(X = i) \mathbb{P}(Y = k - i) \quad \text{für alle } k \geq 0. \quad (\text{A.58})$$

Als Faltung von p und q bezeichnet man

$$p * q = \{p * q\}_{k \geq 0} \quad \text{mit } (p * q)_k \stackrel{\text{Def. A.98}}{=} \sum_{i=0}^k p_i q_{k-i} = \sum_{\substack{l_1, l_2 \geq 0: \\ l_1 + l_2 = k}} p_{l_1} q_{l_2}. \quad (\text{A.59})$$

Bei Unabhängigkeit von X und Y ist demnach die Faltung der Verteilung von X mit der von Y die Verteilung der Summe $X + Y$. Zusammenfassend gilt für alle $k \geq 0$

$$\mathbb{P}(X + Y = k) \stackrel{(\text{A.58})}{=} \sum_{i=0}^k \mathbb{P}(X = i) \mathbb{P}(Y = k - i) = \sum_{i=0}^k p_i q_{k-i} \stackrel{(\text{A.59})}{=} (p * q)_k. \quad (\text{A.60})$$

Beweis. Für jedes $k \geq 0$ impliziert das Ereignis $\{X + Y = k\}$, dass $X \leq k$ ist. Deshalb gilt $\{X + Y = k\} = \{X + Y = k\} \cap \{X \leq k\} = \{X + Y = k\} \cap \bigcup_{i=1}^k \{X = i\} \stackrel{\text{Distr.-gesetz}}{=} \bigcup_{i=1}^k (\{X + Y = k\} \cap \{X = i\})$. Daraus folgt

$$\begin{aligned} \mathbb{P}(X + Y = k) &= \mathbb{P}\left(\bigcup_{i=0}^k \overbrace{(\{X = i\} \cap \{X + Y = k\})}^{\substack{=\emptyset, \text{ falls } X \neq i \text{ ist} \\ =\{i+Y=k\} \text{ für } X=i}}\right) \\ &\stackrel{\text{endl. Add. (A.28)}}{=} \sum_{i=0}^k \mathbb{P}(X = i, \underbrace{i + Y = k}_{\iff Y=k-i}) = \sum_{i=0}^k \mathbb{P}(X = i, Y = k - i) \\ &\stackrel{\text{uabh. (A.55)}}{=} \sum_{i=0}^k \mathbb{P}(X = i) \mathbb{P}(Y = k - i) \quad \left(= \sum_{i=0}^k p_i q_{k-i} \right). \end{aligned}$$

Für die letzte Gleichheit in (A.59) setze $l_1 = i$ und $l_2 = k - i$ (dann ist $l_1 + l_2 = k$). $\square$

Bemerkung zum Satz A.48. Der Satz gilt, leicht adaptiert, auch allgemeiner für ganzzahlige Zufallsvariablen – vgl. etwa die angegebene Referenz. $\triangle$

Das Resultat des Satzes kann verwendet werden, um die wahrscheinlichkeitserzeugende Funktion (kurz weF – siehe die Def. 2.4) von $X + Y$ zu bestimmen. Dies stellt eine Alternative zum Beweis des Lemmas 2.10, welcher sich des Produktsatzes für Erwartungswerte bedient, dar.

Korollar zum Satz A.48 (WeF der Summe unabhängiger, $\mathbb{N}_0$ -wertiger ZV – angelehnt an [17, 5.1 (6), S. 149]). *Seien f_X bzw. f_Y die wahrscheinlichkeitserzeugenden Funktionen von X und Y . Aufgrund der Unabhängigkeit gilt für die weF von $X + Y$*

$$f_{X+Y} = f_X f_Y. \quad (\text{A.61})$$

Beweis. Die Wahrscheinlichkeiten $(p * q)_k$ für die Summe der beiden unabhängigen Zufallsvariablen wurden im vorherigen Satz A.48 ermittelt. Werden diese in die Definition der weF eingesetzt, erhält man für alle $s \in [0, 1]$ (dort sind alle vorkommenden Reihen absolut konvergent – siehe die Bemerkungen zum Satz 2.5)

$$\begin{aligned} f_{X+Y}(s) &\stackrel{(2.6)}{=} \sum_{k \geq 0} \mathbb{P}(X + Y = k) s^k \stackrel{(\text{A.60})}{=} \sum_{k \geq 0} \left(\sum_{i=0}^k p_i q_{k-i} \right) s^k \\ &= \sum_{k \geq 0} \left(\sum_{i=0}^k p_i s^i q_{k-i} s^{k-i} \right) \stackrel{(\star)}{\stackrel{(\text{A.107})}{=}} \left(\sum_{k \geq 0} p_k s^k \right) \left(\sum_{k \geq 0} q_k s^k \right) \stackrel{(2.6)}{=} f_X(s) f_Y(s). \end{aligned}$$

($\star$) An dieser Stelle kommt die Cauchy-Produktformel zum Tragen (Satz A.99 mit $\sum_{i=0}^k c_k = \sum_{i=0}^k a_i b_{k-i} = \sum_{i=0}^k p_i s^i q_{k-i} s^{k-i}$ und $a_k = p_k s^k$ sowie $b_k = q_k s^k$). $\square$

Mehrfache Faltung der Verteilung $\mathbb{N}_0$ -wertiger i.i.d.-Zufallsvariablen

Sei für diesen Teilabschnitt wieder $p = \{p_k\}_{k \geq 0}$ die W-Funktion einer $\mathbb{N}_0$ -wertigen Zufallsvariablen X .

Definition A.49 (*n-fache Faltung oder Faltungspotenzen der W-Funktion einer $\mathbb{N}_0$ -wertigen ZV* – vgl. z.B. [30, Def. 2.29 oder 2.32, S. 61 f.]). Definiere die *n-fache Faltung* oder die *n-te Faltungspotenz* von p , $p^{*n} = \{p_k^{*n}\}_{k \geq 0}$, rekursiv durch

$$p^{*0} := \{\delta_{0k}\}_{k \geq 0} \stackrel{(1.1)}{=} \delta_0 \quad \text{und für } n \geq 0 \text{ durch} \quad p^{*(n+1)} := p^{*n} * p \quad (\text{A.62})$$

$$\Rightarrow p_k^{*n} \stackrel{(\text{A.62})}{=} (p^{*(n-1)} * p)_k \stackrel{(\text{A.59})}{=} \sum_{l_1, l_2 \geq 0: l_1 + l_2 = k} p_{l_1}^{*(n-1)} p_{l_2} \quad \text{für alle } n \geq 1 \text{ und } k \geq 0. \quad (\text{A.63})$$

Lemma A.50. Es gilt $p^{*1} = p$ sowie $p^{*2} = p * p$ und

$$p_k^{*n} = \sum_{\substack{l_1, \dots, l_n \geq 0: \\ l_1 + \dots + l_n = k}} p_{l_1} \cdots p_{l_n} \quad \text{für alle } n \geq 2 \text{ und } k \geq 0. \quad (\text{A.64})$$

Beweis. Sei $k \geq 0$ beliebig. Die zweite Aussage folgt mit (A.62) aus der ersten. Für die erste betrachte

$$p_k^{*1} \stackrel{(\text{A.62})}{=} (p^{*0} * p)_k \stackrel{(\text{A.59})}{=} \sum_{l_1, l_2 \geq 0: l_1 + l_2 = k} \widehat{\delta_{0l_1}} p_{l_2} \stackrel{\substack{=0 \text{ f. } l_1 \neq 0 \\ \text{f. } l_1 = 0, l_2 = k}}{=} p_k \Rightarrow p^{*1} = \{p_k^{*1}\}_{k \geq 0} = \{p_k\}_{k \geq 0} = p.$$

Nun soll (A.64) induktiv (mit $n+1$) für $n \geq 2$ gezeigt werden. Es ist (alle vorkommenden Mehrfachreihen haben nichtnegative Summanden und sind sogar absolut konvergent – vgl. die Bemerkung zum Doppelreihensatz A.101 (iii))

$$\begin{aligned} p_k^{*(n+1)} &\stackrel{(\text{A.63})}{=} \sum_{\substack{j_1, j_2 \geq 0: \\ j_1 + j_2 = k}} p_{j_1}^{*n} p_{j_2} \stackrel{\text{Ind.-Ann.}}{=} \sum_{\substack{j_1, j_2 \geq 0: \\ j_1 + j_2 = k}} \left(\sum_{\substack{l_1, \dots, l_n \geq 0: \\ l_1 + \dots + l_n = j_1}} p_{l_1} \cdots p_{l_n} \right) p_{j_2} \stackrel{(\star)}{=} \sum_{\substack{l_1, \dots, l_n, j_2 \geq 0: \\ l_1 + \dots + l_n + j_2 = k}} p_{l_1} \cdots p_{l_n} p_{j_2} \\ &\stackrel{\text{setze } l_{n+1} = j_2}{=} \sum_{\substack{l_1, \dots, l_n, l_{n+1} \geq 0: \\ l_1 + \dots + l_n + l_{n+1} = k}} p_{l_1} \cdots p_{l_n} p_{l_{n+1}}. \end{aligned} \quad \square$$

Damit kann das folgende Resultat formuliert werden.

Satz A.51 (*n-fache Faltung der Verteilungen bzw. Verteilung der Summe von n i.i.d.-Kopien einer $\mathbb{N}_0$ -wertigen ZV*). Seien $\{X_i\}_{i \geq 1}$ i.i.d.-Kopien von X (d.h., sie sind unabhängig und haben alle die W-Funktion $p = \{p_k\}_{k \geq 0}$). Für $n, k \geq 0$ gilt

$$\mathbb{P}\left(\sum_{i=1}^n X_i = k\right) = p_k^{*n} \stackrel{(\text{A.64})}{=} \sum_{\substack{l_1, \dots, l_n \geq 0: \\ l_1 + \dots + l_n = k}} p_{l_1} \cdots p_{l_n} \quad \text{bzw.} \quad \left\{ \mathbb{P}\left(\sum_{i=1}^n X_i = k\right) \right\}_{k \geq 0} = p^{*n}. \quad (\text{A.65})$$

Beweis. Sei $k \geq 0$ beliebig. Für $n = 0$ ist $\mathbb{P}(\sum_{i=1}^0 X_i = k) \stackrel{\text{leere Summe}}{=} \begin{cases} 1, & \text{wenn } k = 0 \text{ ist,} \\ 0, & \text{wenn } k \neq 0 \text{ ist.} \end{cases}$
Somit gilt, wie behauptet, $\mathbb{P}(\sum_{i=1}^0 X_i = k) \stackrel{(1.1)}{=} \delta_{0k} \stackrel{(A.62)}{=} p_k^{*0}$.

Für $n = 1$ ist $\mathbb{P}(X_1 = k) \stackrel{X_1 \stackrel{d}{=} X}{=} \mathbb{P}(X = k) = p_k \stackrel{\text{Lemma A.50}}{=} p_k^{*1}$ und für $n = 2$ kann der Satz A.48 herangezogen werden um zu sehen, dass $\mathbb{P}(X_1 + X_2 = k) = (p * p)_k = p_k^{*2}$ gilt. Der selbe Satz zeigt auch einen Induktionsschritt für ein beliebiges $n \geq 2$, denn

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{n+1} X_i = k\right) &= \mathbb{P}\left(\sum_{i=1}^n X_i + X_{n+1} = k\right) \stackrel{(A.58)}{=} \sum_{j=0}^k \underbrace{\mathbb{P}\left(\sum_{i=1}^n X_i = j\right)}_{=p_j^{*n} \text{ (Ind.-Ann.)}} \underbrace{\mathbb{P}(X_{n+1} = k - j)}_{=\mathbb{P}(X=k-j)=p_{k-j}} \\ &\stackrel{(A.60)}{=} (p^{*n} * p)_k \stackrel{(A.62)}{=} p_k^{*(n+1)}. \quad \square \end{aligned}$$

Bemerkung zum Satz A.51. Die n -fache Faltung von p (also p^{*n}) ist ebenfalls eine W-Funktion und zwar die der $\mathbb{N}_0$ -wertigen ZV $\sum_{i=1}^n X_i$, was für $n \geq 1$ durch die zweite Aussage auf der rechten Seite in (A.65) deutlich wird. Und auch im Fall $n = 0$ ist (wegen $\sum_{k \geq 0} \delta_{0k} = 1 + 0 + 0 + \dots = 1$) $p^{*0} = \delta_0$ eine W-Funktion. Es gilt also für alle $n \geq 0$, dass $\sum_{k \geq 0} p_k^{*n} = 1$ ist und gemäß der Bemerkung zur Definition 2.1 ist $\mathbb{P}_{\sum_{i=1}^n X_i}$ durch p^{*n} eindeutig bestimmt. $\triangle$

A.2.2 Erwartungswerte, Varianzen und Kovarianzen

Die grundlegenden Rechenregeln bzgl. Erwartungswerten – wie etwa Linearität und Monotonie, welche eine direkte Konsequenz der Definition über das Integral sind – werden als bekannt vorausgesetzt (es sei allenfalls beispielsweise auf [30, Abschnitt 5.1 (insbesondere den Satz 5.3), S. 103 ff.] verwiesen).

Ergänzend werden noch der EW des Produkts unabhängiger ZV und grundlegende Eigenschaften der (Ko-)Varianz, insbesondere die Varianz einer Summe von ZV, behandelt.

Satz A.52 (Produktsatz für Erwartungswerte unabhängiger ZV – angelehnt an [38, Theorem 5.13, S. 127]). *Seien $\{X_i\}_{i=1}^n$ unabhängige und integrierbare ZV. Dann gilt*

$$\mathbb{E} \left[\prod_{i=1}^n X_i \right] = \prod_{i=1}^n \mathbb{E} X_i \quad \text{und insbes.} \quad \mathbb{E}[XY] = \mathbb{E}X \mathbb{E}Y, \quad (\text{A.66})$$

wobei für die zweite Formel $n = 2$, $X := X_1$ und $Y := X_2$ gesetzt wird.

Beweis(-skizze). Das Resultat soll nur für unabhängige, integrierbare und *diskrete* ZV (siehe die Def. A.42) gezeigt werden – für einen allgemeineren Beweis mittels *Produktmaßen* und dem sogenannten *Satz von Fubini* siehe die oben angegebene Referenz.

Es genügt die zweite Formel zu zeigen, Induktion über n liefert dann die erste. Seien p_X bzw. p_Y die W-Funktionen und $T_X = \{x_i : i \in I\} \subset \mathbb{R}$ bzw. $T_Y = \{y_j : j \in J\} \subset \mathbb{R}$ die abzählbaren Träger von X und Y (dabei sind $I, J \subseteq \mathbb{N}$). Definiere zudem einen $\mathbb{Z}$ -Vektor $\mathbf{Z} : \Omega \rightarrow \mathbb{R}^2$ mit $\mathbf{Z} := (X, Y)^T$, welcher (siehe die Bem. A.43) ebenso diskret mit abzählbarem Träger $T_{\mathbf{Z}} = \{\mathbf{z}_k : k \in K\} \subset \mathbb{R}^2$ ($K \subseteq \mathbb{N}$) ist. Definiere weiters die Funktion $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ als $g((x, y)^T) := g(x, y) := |xy|$. Es gilt dann $|XY| = g(X, Y) = g(\mathbf{Z})$ und

$$\begin{aligned} \infty &> \underbrace{\mathbb{E}|X|}_{<\infty} \underbrace{\mathbb{E}|Y|}_{<\infty} \stackrel{(\text{A.57})}{=} \left(\sum_{i \in I} |x_i| p_X(x_i) \right) \left(\sum_{j \in J} |y_j| p_Y(y_j) \right) \\ &= \sum_{i \in I} \left[|x_i| p_X(x_i) \left(\sum_{j \in J} |y_j| p_Y(y_j) \right) \right] = \sum_{i \in I} \left(\sum_{j \in J} |x_i| p_X(x_i) |y_j| p_Y(y_j) \right) \\ &\stackrel{(\text{A.108})}{=} \sum_{i \in I, j \in J} |x_i| |y_j| p_X(x_i) p_Y(y_j) \stackrel{(\text{A.53})}{=} \sum_{i \in I, j \in J} |x_i y_j| \mathbb{P}(X = x_i) \mathbb{P}(Y = y_j) \\ &\stackrel{(\text{A.55})}{=} \sum_{i \in I, j \in J} |x_i y_j| \mathbb{P}(X = x_i, Y = y_j) \stackrel{(3.6)}{=} \sum_{i \in I, j \in J} g(x_i, y_j) \mathbb{P}(\mathbf{Z} = (x_i, y_j)^T) \\ &\stackrel{(\star)}{=} \sum_{k \in K} g(\mathbf{z}_k) \mathbb{P}(\mathbf{Z} = \mathbf{z}_k) \stackrel{(\text{A.57})}{=} \mathbb{E}g(\mathbf{Z}) = \mathbb{E}|XY| \stackrel{(\text{A.45})}{\Rightarrow} XY \text{ ist integrierbar.} \end{aligned}$$

($\star$) Für diese Darstellung als einfache Reihe siehe die Bem. zum Satz A.101 (i).

Dabei sind alle vorkommenden Reihen absolut konvergent, was die entsprechenden Umformungen rechtfertigt (vgl. den Unterabschnitt A.3.2 und insbesondere den Doppelreihensatz A.101). Deshalb kann die gesamte Rechnung auch ohne Betragsstriche wiederholt werden (alles bleibt dabei absolut konvergent), was zum zweiten Teil von (A.66) führt. $\square$

Definition A.53 (Kovarianz und Varianz – vgl. z.B. [38, Def. 5.15, S. 128])

Für zwei integrierbare ZV $X, Y : \Omega \rightarrow \mathbb{R}$ heißt

$$\text{Cov}[X, Y] := \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y \quad (\text{A.67})$$

die *Kovarianz von X und Y* . Ist die $\text{Cov}[X, Y] = 0$, so heißen X und Y *unkorreliert*.
 $\mathbb{V}X := \text{Cov}[X, X] \stackrel{(\text{A.67})}{=} \mathbb{E}[(X - \mathbb{E}X)^2] \stackrel{(\text{A.67})}{=} \mathbb{E}X^2 - (\mathbb{E}X)^2$ ist die *Varianz von X* .

Satz A.54 (Eigenschaft der Kovarianz – vgl. [38, Satz 5.16, S. 129])

Seien die ZV X und Y wie in der Definition A.53. Es gilt u.a.:

- (i) Falls X und Y unabhängig sind, ist die $\text{Cov}[X, Y] = 0 \Rightarrow$ sie sind unkorreliert.
- (ii) $\text{Cov}[X, Y] = \text{Cov}[Y, X]$ und für $a, b, c, d \in \mathbb{R}$ gilt $\text{Cov}[a+bX, c+dY] = bd \text{Cov}[X, Y]$ sowie insbesondere $\mathbb{V}[a+bX] = b^2\mathbb{V}X$.

Beweis. (i) ist eine Konsequenz aus (A.66) (es gilt $\mathbb{E}[XY] \stackrel{\text{uabh.}}{=} \mathbb{E}X\mathbb{E}Y$) bzw. der Def. der Kovarianz und der erste Teil von (ii) gilt ebenfalls definitionsgemäß. Für den zweiten betrachte $\text{Cov}[a+bX, c+dY] \stackrel{(\text{A.67})}{=} \mathbb{E}[(a+bX - \mathbb{E}[a+bX])(c+dY - \mathbb{E}[c+dY])]$
 $= \mathbb{E}[b(X - \mathbb{E}X)d(Y - \mathbb{E}Y)] = bd \text{Cov}[X, Y]. \quad \square$

Satz A.55 (Bienaymé-Gleichung – vgl. [30, Satz 5.7, S. 106])

Sei $n \geq 1$. Für integrierbare $\{X_i\}_{i=1}^n$ gilt, vorausgesetzt für alle $1 \leq i \leq n$ ist $\mathbb{E}X_i^2 < \infty$,

$$\mathbb{V}\left[\sum_{i=1}^n X_i\right] = \sum_{i,j=1}^n \text{Cov}[X_i, X_j] = \sum_{i=1}^n \mathbb{V}X_i + \sum_{i,j=1, i \neq j}^n \text{Cov}[X_i, X_j].$$

Falls die $\{X_i\}_{i=1}^n$ unkorreliert sind (hinreichend dafür ist gemäß dem Satz A.54 (i) ihre Unabhängigkeit), gilt wegen $\text{Cov}[X_i, X_j] = 0$ für $i \neq j$ insbesondere

$$\mathbb{V}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{V}X_i. \quad (\text{A.68})$$

Beweis. Verwende, dass allgemein $(\sum_{i=1}^n a_i)^2 = \sum_{i,j=1}^n a_i a_j$ gilt ($\star$). Berechne dann

$$\begin{aligned} \mathbb{V}\left[\sum_{i=1}^n X_i\right] &= \mathbb{E}\left[\left(\sum_{i=1}^n X_i - \mathbb{E}\left[\sum_{i=1}^n X_i\right]\right)^2\right] \stackrel{\text{endl. Summe}}{=} \mathbb{E}\left[\left(\sum_{i=1}^n (X_i - \mathbb{E}X_i)\right)^2\right] \\ &\stackrel{(\star)}{=} \mathbb{E}\left[\sum_{i,j=1}^n (X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)\right] = \sum_{i,j=1}^n \underbrace{\mathbb{E}[(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)]}_{\text{Cov}[X_i, X_j] \text{ (A.67)}} \end{aligned}$$

und verwende $\text{Cov}[X_i, X_i] = \mathbb{V}X_i$ (aus der ersten Gleichung folgt dann (A.68)). $\square$

A.2.3 Bedingte Wahrscheinlichkeit und Erwartung

Bedingte Wahrscheinlichkeit

Es soll die gleich folgende Def. bzw. Notation für die bedingte WK verwendet werden.

Definition A.56 ((Elementare) Bedingte Wahrscheinlichkeit auf ein Ereignis)
Für ein Ereignis $B \in \Sigma$ werde $\mathbb{P}^B : \Sigma \rightarrow [0, 1]$ (für alle $A \in \Sigma$) definiert durch

$$\mathbb{P}^B(A) := \mathbb{P}(A|B) := \begin{cases} \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, & \text{wenn } \mathbb{P}(B) > 0 \text{ ist,} \\ 0, & \text{wenn } \mathbb{P}(B) = 0 \text{ ist,} \end{cases} \quad (\text{A.69})$$

als die *bedingte Wahrscheinlichkeit von A gegeben B* bezeichnet.

Bemerkungen zur Definition A.56.

- (i) Für den Fall, dass B eine $\mathbb{P}$ -Nullmenge ist, also für $\mathbb{P}(B) = 0$, ist $\mathbb{P}^B$ willkürlich auf 0 festgelegt. Diese Def. richtet sich an [30, Def. 8.2, S. 174], oft bleibt dieser Fall auch undefiniert. So ist aber garantiert, dass die Multiplikationsformel

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B), \quad (\text{A.70})$$

welche direkt aus (A.69) erhalten werden kann, für alle Ereignisse $B \in \Sigma$ gültig ist.

- (ii) Im Fall $\mathbb{P}(B) > 0$ ist $\mathbb{P}^B$ ein W-Maß auf (Ω, Σ) , die sogenannte *durch B bedingte Wahrscheinlichkeitsverteilung* (vgl. auch [30, Satz 8.4, S. 174]), denn es erfüllt dann die geforderten Eigenschaften der Definition A.22, nämlich:

- *Normiertheit (i)*: $\mathbb{P}^B(\Omega) \stackrel{(\text{A.69})}{=} \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} \stackrel{(\text{A.33})}{=} \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$.
- *σ -Add. (ii)*: Wenn $\{A_i\}_{i \geq 1}$ mit $A_i \in \Sigma$ paarweise disjunkt sind, dann sind es die $\{A_i \cap B\}_{i \geq 1}$ ebenso. Deshalb gilt, dass

$$\begin{aligned} \mathbb{P}^B(\bigcup_{i \geq 1} A_i) &\stackrel{(\text{A.69})}{=} \frac{\mathbb{P}(B \cap \bigcup_{i \geq 1} A_i)}{\mathbb{P}(B)} \stackrel{\text{Distr.-gesetz}}{=} \frac{\mathbb{P}(\bigcup_{i \geq 1} (A_i \cap B))}{\mathbb{P}(B)} \\ &\stackrel{\sigma\text{-Add. (A.27)}}{=} \frac{\sum_{i \geq 1} \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_{i \geq 1} \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} \stackrel{(\text{A.69})}{=} \sum_{i \geq 1} \mathbb{P}^B(A_i) \quad \text{ist.} \end{aligned}$$

- (iii) Sind $A, B \in \Sigma$ *unabhängig*, so gilt $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ und somit wird (A.69) für den Fall $\mathbb{P}(B) > 0$ zu $\mathbb{P}(A|B) = \mathbb{P}(A)$. $\triangle$ (A.41)

Unter anderem gilt bezüglich der bedingten WK das folgende wichtige Resultat.

Satz A.57 (Totale Wahrscheinlichkeit – vgl. z.B. [30, Satz 8.6, S. 174])

Sei I eine abzählbare Indexmenge. Wenn $\{B_i\}_{i \in I}$ mit $B_i \in \Sigma$ entweder eine Partition von Ω bilden oder zumindest paarw. disj. und so sind, dass $\mathbb{P}(\bigcup_{i \in I} B_i) = 1$ ist, gilt

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i) \quad \text{für alle } A \in \Sigma. \quad (\text{A.71})$$

Beweis. Da $\bigcup_{i \in I} B_i$ entweder gleich Ω oder zumindest eine $\mathbb{P}$ -Einsmenge ist, gilt

$$\begin{aligned} \mathbb{P}(A) &\stackrel{(A.33)}{=} \mathbb{P}\left(A \cap \bigcup_{i \in I} B_i\right) \stackrel{\text{Distr.-gesetz}}{=} \mathbb{P}\left(\bigcup_{i \in I} \underbrace{(A \cap B_i)}_{\text{paarw. disj.}}\right) \stackrel{\sigma\text{-Add. (A.27)}}{=} \sum_{i \in I} \mathbb{P}(A \cap B_i) \\ &\stackrel{(A.70)}{=} \sum_{i \in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i). \end{aligned} \quad \square$$

Bemerkung A.58 (Bedingte Verteilung einer diskreten ZV)

Die *bedingte Verteilung* $\mathbb{P}_X^B$ einer diskreten ZV X gegeben eines Ereignisses $B \in \Omega$ bzw. deren W-Funktion $p_{X|B}$ ($X|B$ ist auch eine diskrete ZV) kann für $\mathbb{P}(B) > 0$ wie folgt definiert werden:

$$p_{X|B}(\omega') \stackrel{\text{vgl. (A.53)}}{=} \begin{cases} \mathbb{P}_X^B(\{\omega'\}) := \mathbb{P}_{X|B}(\{\omega'\}) := \mathbb{P}(X = \omega'|B) & \text{für } \omega' \in T_X, \\ 0 & \text{sonst.} \end{cases} \quad (\text{A.72})$$

Ist $\mathbb{P}(B) = 0$, funktioniert die Definition genauso (siehe die Def. A.56), $p_{X|B}(\omega')$ bzw. $\mathbb{P}_X^B(\{\omega'\})$ sind dann aber immer (für alle $\omega' \in \Omega'$) gleich 0 und es handelt sich um keine Wahrscheinlichkeitsverteilung bzw. -funktion.

Gegeben einer weiteren diskreten ZV Y und für y aus dem Bildraum von Y kann analog die *bedingte Verteilung von X gegeben $Y = y$* (für $\mathbb{P}(Y = y) > 0$ und $\omega' \in T_X$) als

$$p_{X|Y}(\omega', y) := p_{X|Y=y}(\omega') := p_{X|\{Y=y\}}(\omega') \stackrel{(A.72)}{=} \mathbb{P}_{X|\{Y=y\}}(\{\omega'\}) := \mathbb{P}_{X|Y=y}(\{\omega'\}) \quad (\text{A.73})$$

definiert werden (für $\omega' \notin T_X$ und/oder $\mathbb{P}(Y = y) = 0$ ist wiederum $p_{X|Y}(\omega', y) = 0$). $\triangle$

Bedingte Erwartung allgemein

Wird der EW unter Berücksichtigung zusätzlicher Informationen über den Ausgang eines Zufallsexperiments gebildet, so wird er als bedingte Erwartung oder bedingter EW bezeichnet. Er ist i.A. kein fester Wert, sondern eine Zufallsvariable und wird generell, etwas abstrakt wie folgt definiert. Anschließend werden wichtige Eigenschaften beleuchtet. Etwas expliziter kann im nächsten Teilabschnitt die bedingte Erwartung gegeben einer diskreten Zufallsvariable (oder gleich eines konkreten Ereignisses) angegeben werden.

Definition A.59 (Bedingte Erwartung (allg.) – angelehnt an [30, Def. 8.11, S. 177])
Seien $Z : \Omega \rightarrow \bar{\mathbb{R}}$ eine integrierbare Zufallsvariable und $\mathcal{G} \subseteq \Sigma$ eine Sub- σ -Algebra (siehe die Bemerkung zur Def. A.2 (iv)). Dann existiert eine Zufallsvariable $Y : \Omega \rightarrow \bar{\mathbb{R}}$, welche Folgendes erfüllt:

- (i) Y ist messbar bezüglich $\mathcal{G}$ (genauer: $\mathcal{G} - \mathcal{B}(\bar{\mathbb{R}})$ messbar),
- (ii) $\mathbb{E}[Y \mathbb{1}_B] = \mathbb{E}[Z \mathbb{1}_B]$ für alle $B \in \mathcal{G}$ und
- (iii) Y ist integrierbar.

So ein Y heißt *bedingte Erwartung oder bedingter Erwartungswert von Z gegeben $\mathcal{G}$* und man schreibt dafür $\mathbb{E}[Z|\mathcal{G}] := Y$.

Bemerkungen zur Definition A.59.

- (a) Y ist klarerweise auch messbar bzgl. Σ (siehe das Lemma A.12).
- (b) *Integrierbarkeit:* Der Punkt (iii) muss eigentlich nicht extra gefordert werden, sondern ergibt sich aus (ii). Da $\mathcal{G}$ auch eine σ -Algebra (auf Ω) ist, gilt $\Omega \in \mathcal{G}$ und deshalb $\mathbb{E}Y = \mathbb{E}[Y\mathbb{1}_\Omega] \stackrel{(ii)}{=} \mathbb{E}[Z\mathbb{1}_\Omega] = \mathbb{E}Z$ (und Z ist, lt. Vorauss., integrierbar).
- (c) *Quasiintegrierbarkeit:* Mit leichten Adaptionen lässt sich die bedingte Erwartung auch für eine quasiintegrierbare ZV Z definieren (siehe [30, Bem. 8.16, S. 180] – dies gilt insbesondere dann, wenn Z nichtnegativ ist). Die Punkte (i) und (ii) der Def. sind dann auch erfüllt und (iii) muss um den Zusatz 'quasi-' ergänzt werden.
- (d) *Existenz:* Dass unter den Voraussetzungen (zumindest) *ein* solches Y existiert, kann mittels des *Satzes von Radon-Nikodym* (vgl. z.B. [30, Kor. 7.34, S. 161]) gezeigt werden. Dieser Existenzbeweis findet sich bspw. unter [30, Satz 8.12, S. 177].
- (e) *Eindeutigkeit:* Es kann i.A. mehrere ZV geben, welche (i)-(iii) der Def. erfüllen, diese sind aber fast sicher gleich. D.h., wenn die Def. auf zwei Varianten Y und $\tilde{Y}$ zutrifft, dann ist $Y = \tilde{Y}$ f.s. Somit ist diese bedingte Erwartung f.s. eindeutig, wie der nächste Satz zeigt, und die Festlegung von $\mathbb{E}[Z|\mathcal{G}]$ auf *eine* Variante Y ist gerechtfertigt. $\triangle$

Satz A.60 (angelehnt an [30, Satz 8.12, S. 177]). *Die bedingte Erwartung $\mathbb{E}[Z|\mathcal{G}]$ der Definition A.59 ist fast sicher eindeutig.*

Beweis. Seien Y und $\tilde{Y}$ ZV, welche die Punkte (i)-(iii) der Def. A.59 erfüllen. Dann gilt $\mathbb{E}[Y\mathbb{1}_B] \stackrel{(ii)}{=} \mathbb{E}[Z\mathbb{1}_B] \stackrel{(ii)}{=} \mathbb{E}[\tilde{Y}\mathbb{1}_B] \quad \forall B \in \mathcal{G} \Rightarrow 0 = \mathbb{E}[Y\mathbb{1}_B] - \mathbb{E}[\tilde{Y}\mathbb{1}_B] = \mathbb{E}[(Y - \tilde{Y})\mathbb{1}_B] \quad (*)$. Da Y und $\tilde{Y}$ messbar bzgl. $\mathcal{G}$ sind, ist es gemäß dem Satz A.9 auch $Y - \tilde{Y}$, weshalb $\{Y > \tilde{Y}\} = \{Y - \tilde{Y} > 0\} \in \mathcal{G}$ ist. Also gilt $(*)$ auch für $B := \{Y > \tilde{Y}\}$, sprich $\mathbb{E}[(Y - \tilde{Y})\mathbb{1}_{\{Y > \tilde{Y}\}}] = 0$. Da $(Y - \tilde{Y})\mathbb{1}_{\{Y > \tilde{Y}\}} \geq 0$ ist, folgt daraus mit (A.20), dass $(Y - \tilde{Y})\mathbb{1}_{\{Y > \tilde{Y}\}} = 0$ f.s. ist. Nun ist die letzte Gleichheit nur für $\omega \in \Omega$ mit $Y(\omega) - \tilde{Y}(\omega) \leq 0$ erfüllbar. Daraus folgt $\mathbb{P}(Y > \tilde{Y}) = 0 \Rightarrow \mathbb{P}(Y \leq \tilde{Y}) = 1$, also $Y \leq \tilde{Y}$ f.s.

Da genauso $\mathbb{E}[(\tilde{Y} - Y)\mathbb{1}_B] = 0$ für alle $B \in \mathcal{G}$ gilt, kann analog vorgegangen werden, um $\tilde{Y} \leq Y$ f.s. zu erhalten. Zusammengenommen führt dies zu $Y = \tilde{Y}$ f.s.. $\square$

Satz A.61 (Eigenschaften der bedingten Erwartung – vgl. z.B. [38, 8.2, S. 219 ff.])
Seien $Z, Z_1, Z_2: \Omega \rightarrow \bar{\mathbb{R}}$ integrierbare ZV und sei $\mathcal{G} \subseteq \Sigma$ eine Sub- σ -Algebra. Bezüglich der bedingten Erwartung, definiert wie in der Def. A.59, gelten die folgenden Eigenschaften.

- (a) *Falls Z bereits messbar bzgl. $\mathcal{G}$ ist, gilt $\mathbb{E}[Z|\mathcal{G}] = Z$ f.s. für jede Variante von $\mathbb{E}[Z|\mathcal{G}]$.*
- (b) **Totaler Erwartungswert:** *Für jede Variante von $\mathbb{E}[Z|\mathcal{G}]$ gilt*

$$\mathbb{E}[\mathbb{E}[Z|\mathcal{G}]] = \mathbb{E}Z. \quad (\text{A.74})$$

- (c) **Linearität:** *Seien $\alpha, \beta \in \mathbb{R}$. Es gilt*

$$\mathbb{E}[\alpha Z_1 + \beta Z_2|\mathcal{G}] = \alpha \mathbb{E}[Z_1|\mathcal{G}] + \beta \mathbb{E}[Z_2|\mathcal{G}] \text{ f.s.} \quad (\text{A.75})$$

(d) **Turmeigenschaft:** Ist $\mathcal{H}$ eine weitere Sub- σ -Algebra mit $\mathcal{H} \subseteq \mathcal{G} \subseteq \Sigma$, so gilt

$$\mathbb{E}[\mathbb{E}[Z|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[Z|\mathcal{H}] \text{ f.s.} \quad (\text{A.76})$$

(e) **Unabhängigkeit:** Sind $\sigma(Z)$ und $\mathcal{G}$ unabhängig, so gilt

$$\mathbb{E}[Z|\mathcal{G}] = \mathbb{E}Z \text{ f.s.}$$

Bis auf (c) lassen sich die Voraussetzungen auf quasiintegrierbar abschwächen.

Beweis. Mit (i) bis (iii) sind durchwegs die entsprechenden Punkte der Def. A.59, also die Anforderungen an eine Variante der allg. bedingten Erwartung, gemeint. Wenn zwei Varianten einer bedingten Erwartung (i) und (ii) genügen ((iii) folgt daraus – siehe die Bem. zur Def. A.59 (b)), ist deren fast sichere Gleichheit eine Konsequenz des Satzes A.60.

- (a) Die ZV Z erfüllt (ii) sowieso. Ist sie zudem bereits $\mathcal{G}$ -messbar, gilt für sie auch (i).
- (b) Wie bereits in der Bem. zur Def. A.59 (b) gezeigt, folgt dies (da $\Omega \in \mathcal{G}$ ist) aus (ii).
- (c) Die linke Seite und $\mathbb{E}[Z_1|\mathcal{G}]$ bzw. $\mathbb{E}[Z_2|\mathcal{G}]$ erfüllen definitionsgemäß die Punkte (i) und (ii). Es soll überprüft werden, ob diese auch auf die rechte Seite zutreffen, denn dann liefert, wie eingangs erwähnt, der Satz A.60 die Aussage.
Da $\mathbb{E}[Z_1|\mathcal{G}]$ und $\mathbb{E}[Z_2|\mathcal{G}]$ beide $\mathcal{G}$ -messbar sind, ist (i) für die rechte Seite eine Konsequenz des Satzes A.9. Zudem gilt für alle $B \in \mathcal{G}$

$$\begin{aligned} \mathbb{E}[(\alpha\mathbb{E}[Z_1|\mathcal{G}] + \beta\mathbb{E}[Z_2|\mathcal{G}])\mathbb{1}_B] &\stackrel[\text{d. EW}]{\text{Linearität}} \alpha\mathbb{E}[\mathbb{E}[Z_1|\mathcal{G}]\mathbb{1}_B] + \beta\mathbb{E}[\mathbb{E}[Z_2|\mathcal{G}]\mathbb{1}_B] \\ &\stackrel[\mathbb{E}[Z_i|\mathcal{G}]]{(ii) \text{ für}} \alpha\mathbb{E}[Z_1\mathbb{1}_B] + \beta\mathbb{E}[Z_2\mathbb{1}_B] \stackrel[\text{d. EW}]{\text{Linearität}} \mathbb{E}[(\alpha Z_1 + \beta Z_2)\mathbb{1}_B]. \end{aligned}$$

Wegen (ii) entspricht der letzte Ausdruck $\mathbb{E}[(\alpha Z_1 + \beta Z_2)\mathbb{1}_B]$, womit die rechte eine Variante der linken Seite ist und sie f.s. gleich sind.

- (d) Die ZV $X := \mathbb{E}[\mathbb{E}[Z|\mathcal{G}]|\mathcal{H}]$ ist wegen (i) messbar bzgl. $\mathcal{H}$ und es gilt $X = \mathbb{E}[Z|\mathcal{G}]$ (*) aufgrund von (b). Weiters gilt für alle $B \in \mathcal{H} (\Rightarrow B \in \mathcal{G} (\star\star))$, da $\mathcal{H} \subseteq \mathcal{G}$ ist)

$$\mathbb{E}[X\mathbb{1}_B] = \mathbb{E}[\mathbb{E}[Z|\mathcal{G}]\mathbb{1}_B] \stackrel[(\star\star)]{(ii)} \mathbb{E}[Z\mathbb{1}_B] \stackrel[\text{und Satz A.60}]{\text{Def. A.59, f. } \mathbb{E}[Z|\mathcal{H}]}, X = \mathbb{E}[Z|\mathcal{H}] \text{ f.s.}$$

(siehe bspw. auch [38, Satz 8.7, S. 221] – dort 'Projektionseigenschaft' genannt).

- (e) Es soll wieder überprüft werden, ob auch $\mathbb{E}Z$ den Punkten (i) und (ii) genügt. $\mathbb{E}Z$ ist eine Zahl, somit konstant und deshalb messbar bzgl. $\mathcal{G}$ (vgl. die Bem. zur Def. A.8), weshalb (i) erfüllt ist.

Für (ii) betrachte ein $B \in \mathcal{G}$. Dann ist $\sigma(\mathbb{1}_B) \stackrel{(A.14)}{=} \{\emptyset, B, B^c, \Omega\} \subseteq \mathcal{G}$. Daraus, und weil $\mathcal{G}$ und $\sigma(Z)$ unabhängig sind, folgt gemäß der Def. A.33 (ii) die Unabhängigkeit von $\sigma(\mathbb{1}_B)$ und $\sigma(Z)$, was gleichbedeutend mit der Unabhängigkeit von $\mathbb{1}_B$ und Z ist (Def. A.33 (iii)). Deshalb gilt für alle $B \in \mathcal{G}$

$$\mathbb{E}[\mathbb{E}[Z|\mathcal{G}]\mathbb{1}_B] \stackrel{(ii)}{=} \mathbb{E}[Z\mathbb{1}_B] \stackrel[\text{(A.66)}]{\text{uabh.}} \mathbb{E}Z\mathbb{E}\mathbb{1}_B \stackrel[\text{konstant}]{\mathbb{E}Z} \mathbb{E}[\mathbb{E}[Z]\mathbb{1}_B],$$

womit auch (ii) für $\mathbb{E}Z$ erfüllt ist $\stackrel{\text{Satz A.60}}{\Rightarrow} \mathbb{E}Z = \mathbb{E}[Z|\mathcal{G}] \text{ f.s.}$

Die letzte Aussage über quasiintegrierbare ZV (nichtnegative ZV sind immer quasiintegrierbar - siehe die Bem. zur Def. A.36) lässt sich dadurch begründen, dass deren bedingte Erwartungen gegeben $\mathcal{G}$ die Def. A.59 erfüllen (siehe die Bem. zur Def. A.59 (c)). $\square$

Analog zu (A.44) kann allgemein die bedingte Wahrscheinlichkeit eines Ereignisses (gegeben einer Sub- σ -Algebra) definiert werden.

Definition A.62 (Bedingte Wahrscheinlichkeit (allgemein) – angelehnt an [30, Def. 8.11, S. 177]). Sei, wie in der Def. A.59, $\mathcal{G} \subseteq \Sigma$. Definiere für ein Ereignis $A \in \Sigma$ (setze $Z := \mathbb{1}_A$) die *bedingte Wahrscheinlichkeit von A gegeben $\mathcal{G}$* als $\mathbb{P}(A|\mathcal{G}) := \mathbb{E}[\mathbb{1}_A|\mathcal{G}]$.

Nun soll nicht nur auf beliebige Sub- σ -Algebren, sondern auch auf Zufallsvariablen bedingt werden können. Die folgenden Definition sorgt dafür, dass dies gleichbedeutend mit der Bedingung auf die von einer oder mehreren ZV erzeugte σ -Algebra ist (d.h. auf $\sigma(X)$ für eine ZV X bzw. $\sigma(X_1, \dots, X_n)$ für mehrere ZV $X_1, \dots, X_n$ – wie im Satz A.11 definiert). Somit kann dann die allgemeine Definition A.59 herangezogen werden.

Definition A.63 (Auf eine ZV bedingte Erwartung (allgemein) – angelehnt an [30, Def. 8.13, S. 178]). Sei $Z : \Omega \rightarrow \bar{\mathbb{R}}$ eine integrierbare ZV. Für eine weitere ZV X ist $\sigma(X) \subseteq \Sigma$. Somit lässt sich die *bedingte Erwartung von Z gegeben X* als

$$\mathbb{E}[Z|X] := \mathbb{E}[Z|\sigma(X)] \quad \text{definieren.}$$

Für mehrere ZV $X_1, \dots, X_n$ ist wiederum $\sigma(X_1, \dots, X_n) \subseteq \Sigma$ und die *bedingte Erwartung von Z gegeben $X_1, \dots, X_n$* sei definiert als

$$\mathbb{E}[Z|X_1, \dots, X_n] := \mathbb{E}[Z|\sigma(X_1, \dots, X_n)]. \quad (\text{A.77})$$

Bedingte Erwartung gegeben einer diskreten Zufallsvariablen

Ausgehend von der elementaren bedingten Wahrscheinlichkeit der Def. A.56 soll nun zuerst die bedingte Erwartung gegeben einzelner Ereignisse definiert werden. Darauf aufbauend lässt sich die bedingte Erwartung gegeben einer diskreten ZV bestimmen. Dabei wird ausgeführt, dass diese Begriffe konsistent mit der allgemeinen Def. A.59 sind.

Zunächst wird jedoch das folgende Hilfsresultat formuliert.

Lemma A.64 (angelehnt an [36, Lemma 14.2, S. 225]). *Ist $Z : \Omega \rightarrow \bar{\mathbb{R}}$ $\mathbb{P}$ -(quasi-)integrierbar, so ist sie für $B \in \Sigma$ auch $\mathbb{P}^B$ -(quasi-)integrierbar und es gilt für $\mathbb{P}(B) > 0$*

$$\mathbb{E}_{\mathbb{P}^B} Z \stackrel{(\text{A.43})}{=} \int Z d\mathbb{P}^B \stackrel{(\text{A.10})}{=} \int Z(\omega) \mathbb{P}(d\omega|B) \stackrel{(\text{A.69})}{=} \frac{\int_B Z d\mathbb{P}}{\mathbb{P}(B)} \stackrel{(\text{A.11})}{=} \frac{\mathbb{E}[\mathbb{1}_B Z]}{\mathbb{P}(B)}. \quad (\text{A.78})$$

Ist $\mathbb{P}(B) = 0$, so gilt $\mathbb{E}_{\mathbb{P}^B} Z = 0$.

Beweisskizze. Ist $\mathbb{P}(B) = 0$, so gilt $\mathbb{P}^B(\omega) = \mathbb{P}(\{\omega\}|B) \stackrel{(\text{A.69})}{=} 0$ für alle $\omega \in \Omega$. Gemäß der Konstruktion des Integrals (Def. A.13) gilt deshalb $\int Z(\omega) \mathbb{P}(d\omega|B) = 0$.

Für $\mathbb{P}(B) > 0$ kann die einzige unbegründete Gleichheit mit der bereits in der Beweisskizze des Lemmas A.18 erwähnten 'Standardprozedur', welche sich an der Integralkonstruktion richtet, gezeigt werden. Dies wird etwa im Beweis der obigen Referenz durchgeführt (siehe auch [38, S. 215 f.]). Die (Quasi-)Integrierbarkeit ist für $\mathbb{P}(B)=0$ klar. Sonst folgt sie aus der Monotonie des EW und (A.78), denn $\mathbb{1}_B|Z| \leq |Z| \Rightarrow \mathbb{E}[\mathbb{1}_B Z] \leq \mathbb{E}[Z]$. $\square$

Das Lemma A.64 motiviert dazu, die auf ein Ereignis aus dem Ereignisraum Σ bedingte Erwartung einer Zufallsvariablen zu definieren.

Definition A.65 (Bedingte Erwartung gegeben eines Ereignisses – angelehnt an [30, Def. 8.9, S. 175])

Seien $Z : \Omega \rightarrow \mathbb{R}$ eine (quasi-)integrierbare ZV und $B \in \Sigma$ ein Ereignis. Dann ist die Größe

$$\mathbb{E}[Z|B] := \mathbb{E}_{\mathbb{P}_B} Z \stackrel{(A.78)}{=} \begin{cases} \frac{\mathbb{E}[\mathbb{1}_B Z]}{\mathbb{P}(B)} & \text{für } \mathbb{P}(B) > 0, \\ 0 & \text{sonst} \end{cases} \quad (A.79)$$

der *bedingte Erwartungswert von Z gegeben B* .

Bei einer diskreten ZV X gibt es nur höchstens abzählbar viele Ereignisse der Form $\{X = x\}$ mit positiver Wahrscheinlichkeit. Deshalb kann eine explizite Formel für den auf eine solche bedingten EW – basierend auf der soeben gegebenen Def. – angegeben werden.

Definition A.66 (Auf eine diskrete Zufallsvariable bedingte Erwartung – angelehnt an [38, (8.1), S. 215])

Seien $Z : \Omega \rightarrow \mathbb{R}$ eine (quasi-)integrierbare ZV, (Ω', Σ') ein Messraum und $X : \Omega \rightarrow \Omega'$ eine diskrete ZV. Weiters sei $g : \Omega' \rightarrow \mathbb{R}$ für alle $\omega' \in \Omega'$ definiert als

$$g(\omega') := \mathbb{E}[Z|\{X = \omega'\}] := \mathbb{E}[Z|X = \omega'] \stackrel{(A.79)}{=} \begin{cases} \frac{\mathbb{E}[\mathbb{1}_{\{X = \omega'\}} Z]}{\mathbb{P}(X = \omega')} & \text{f. } \mathbb{P}(X = \omega') > 0, \\ 0 & \text{sonst.} \end{cases} \quad (A.80)$$

Dann wird die Abbildung $\mathbb{E}[Z|X] : \Omega \rightarrow \mathbb{R}$, gegeben durch

$$\mathbb{E}[Z|X] := g \circ X, \quad (A.81)$$

als die *bedingte Erwartung von Z gegeben X* bezeichnet.

Lemma A.67 (vgl. [30, Lemma 8.10, S. 176] oder [36, 14.1, S. 225 ff.]; ohne Beweis). *Sei alles wie in der Definition A.66. Für die dort definierte ZV $\mathbb{E}[Z|X]$, also die bedingte Erwartung von Z gegeben einer diskreter ZV X , gilt*

- (i) $\mathbb{E}[Z|X]$ ist $\sigma(X)$ -messbar,
- (ii) $\mathbb{E}[\mathbb{E}[Z|X] \mathbb{1}_{\{X \in B'\}}] = \mathbb{E}[Z \mathbb{1}_{\{X \in B'\}}]$ für alle $B' \in \Sigma'$ und
- (iii) $\mathbb{E}[Z|X]$ ist (quasi-)integrierbar.

Bemerkungen zur Definition A.66 bzw. dem Lemma A.67.

- (i) Das Lemma zeigt also, dass die Zufallsvariable $\mathbb{E}[Z|X]$ der Definition A.66 die Anforderungen der allgemeinen Definition A.59 erfüllt und somit *eine* denkbare Variante der *bedingten Erwartung von Z gegeben $\sigma(X)$ bzw. X* ist. Damit gelten für sie auch alle im Satz A.61 präsentierten Eigenschaften.
- (ii) Eine andere mögliche 'Definitionsvariante', welche laut der Bem. zur Def. A.59 (e) bzw. dem Satz A.60 fast sicher gleich ist, wäre etwa die folgende. Würde statt (A.80) ein $h : \Omega' \rightarrow \mathbb{R}$ gewählt werden, welches g in den Fällen $\mathbb{P}(X = \omega') > 0$ gleicht, für $\mathbb{P}(X = \omega') = 0$ aber $h(\omega') = c$ mit $c \in \mathbb{R}$ ist, würde das Lemma A.67 analog und somit auch $g \circ X = h \circ X$ f.s. gelten (h unterscheidet sich von g nur auf einer Nullmenge). Die Festlegung auf 0 ist daher willkürlich und unerheblich (ähnlich wie bei der elementaren bedingten WK – siehe die Bem. zur Def. A.56 (i)). $\triangle$

Diskrete Zufallsvariable bedingt auf eine andere diskrete Zufallsvariable

Ist im Setting der Definitionen A.65 und A.66 die Zufallsvariable Z selbst ebenfalls diskret, so können die bedingten Erwartungen auf Ereignisse oder eine andere diskrete Zufallsvariable direkt folgendermaßen berechnet werden.

Lemma A.68. *Seien $Z : \Omega \rightarrow \Omega'$ eine diskrete Zufallsvariable (also mit abzählbarem Träger $T_Z = \{z_i : i \in I\}$, $I \subseteq \mathbb{N}$), $h : \Omega' \rightarrow \mathbb{R}_+$ und $B \in \Sigma$ ein Ereignis. Dann lässt sich der bedingte Erwartungswert der diskreten, nichtnegativen ZV $h(Z)$ gegeben B durch*

$$\mathbb{E}[h(Z)|B] \stackrel{(A.79)}{=} \mathbb{E}_{\mathbb{P}^B} h(Z) = \sum_{i \in I} h(z_i) \mathbb{P}^B(Z = z_i) \quad \text{bestimmen.} \quad (A.82)$$

Beweis. Im Fall $\mathbb{P}(B) = 0$ gilt sowohl $\mathbb{E}_{\mathbb{P}^B} h(Z) = 0$ (Lemma A.64) als auch $\mathbb{P}^B(Z = z_i) = 0$ für alle $i \in I$ (Def. A.56). Für $\mathbb{P}(B) > 0$ liefert (A.57) die Aussage (mit $\mathbb{P}^B$ statt $\mathbb{P}$). $\square$

Sind Z und h wie im Lemma und zudem $X : \Omega \rightarrow \Omega'_X$ eine weitere diskrete ZV, kann somit $g : \Omega'_X \rightarrow \mathbb{R}$ der Def. A.66 explizit durch eine Reihe beschrieben werden, nämlich

$$\begin{aligned} g(\omega') &\stackrel{(A.80)}{=} \mathbb{E}[h(Z)|X = \omega'] \stackrel{(A.82)}{=} \sum_{i \in I} h(z_i) \mathbb{P}^{\{X = \omega'\}}(Z = z_i) \\ &\stackrel{(A.69)}{=} \sum_{i \in I} h(z_i) \mathbb{P}(Z = z_i | X = \omega') \end{aligned} \quad (A.83)$$

und der ebendort definierte *bedingte EW von $h(Z)$ gegeben X* , $\mathbb{E}[h(Z)|X] = g \circ X$, erfüllt weiterhin die Punkte (i) bis (iii) der Def. A.59 (bzgl. $\sigma(X)$) – siehe das Lemma A.67).

A.2.4 Markov-Ketten

Sei für alles in diesem Unterabschnitt I eine (höchstens) abzählbare Indexmenge, welche *Zustandsraum* genannt wird. Eine Markov-Kette ist ein spezifischer stochastischer Prozess, der üblicherweise in diskreter Zeit $\mathbb{N}_0$ definiert wird, also z.B. $\{X_n\}_{n \geq 0}$.

Definition A.69 (Stochastische Matrix – vgl. z.B. [6, S. 112])

Eine Matrix $\mathbf{P} := (p_{ij})_{i,j \in I}$ heißt *stochastisch*, wenn für alle $i, j \in I$ die Einträge $p_{ij} \geq 0$ sind und $\sum_{j \in I} p_{ij} = 1$ für alle $i \in I$ gilt.

Definition A.70 (Markov-Kette – angelehnt an [6, (8.2), S. 111] u. [41, 1.1, S. 1 f.])

Seien $\lambda := \{\lambda_i\}_{i \in I}$ eine Verteilung, d.h. $\sum_{i \in I} \lambda_i = 1$, und $\mathbf{P}$ eine stochastische Matrix.

Eine Folge von ZV $\{X_n\}_{n \geq 0}$ mit Werten in I heißt *Markov-Kette* mit *Übergangsmatrix* $\mathbf{P}$ und *Anfangsverteilung* λ , kurz *Markov*($\lambda, \mathbf{P}$), falls für alle $n \geq 0$ und

- (i) für alle $i \in I$ gilt, dass $\mathbb{P}(X_0 = i) = \mathbb{P}_{X_0}(\{i\}) = \lambda_i$ ist und
- (ii) für alle $j \in I$ und alle $i_0, \dots, i_n \in I$ mit $\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) > 0$ gilt, dass

$$\mathbb{P}(X_{n+1} = j | X_0 = i_0, \dots, X_n = i_n) = \mathbb{P}(X_{n+1} = j | X_n = i_n) = p_{i_n j} \quad \text{ist.} \quad (\text{A.84})$$

Die Wahrscheinlichkeit dafür, dass sich die Kette in einem Schritt vom Zustand $i \in I$ nach $j \in I$ bewegt, heißt *Übergangswahrscheinlichkeit von i nach j* und ist der Eintrag p_{ij} der Übergangsmatrix. Aus (A.84) folgt für alle $n \geq 0$ und $i \in I$ mit $\mathbb{P}(X_n = i) > 0$

$$\mathbb{P}(X_{n+1} = j | X_n = i) = p_{ij}.$$

Bemerkung zur Def. A.70. Da die Übergangswahrscheinlichkeiten nicht von n und nur vom vorherigen Zustand abhängen, ist der hier definierte Prozess genauer gesagt eine (zeit-)homogene Markov-Kette erster Ordnung. $\triangle$

Satz A.71 (Markov-Eigenschaft (ME) – vgl. [41, Theorem 1.1.2, S. 3]; ohne Beweis)
Seien $\{X_n\}_{n \geq 0}$ Markov($\lambda, \mathbf{P}$), $m \geq 0$ und $i \in I$. Dann ist, wenn auf $X_m = i$ bedingt wird, $\{X_{m+n}\}_{n \geq 0}$ Markov($\delta_i, \mathbf{P}$) und unabhängig von $X_0, \dots, X_m$.

Die Definition A.70 und der Satz A.71 können so interpretiert werden, dass zu einem gewissen Zeitpunkt die Zustände der vergangenen Zeitpunkte keinen Einfluss mehr auf die Zukunft haben, die Markov-Kette also gewissermaßen gedächtnislos ist. Dies gilt bereits intuitiv auch für die entsprechenden bedingten Erwartungen, was im nächsten Satz aber auch formal gezeigt werden soll.

Satz A.72 (Markov-Eigenschaft (ME) für bedingte Erwartungen)

Seien $\{X_n\}_{n \geq 0}$ Markov($\lambda, \mathbf{P}$) und $h : I \rightarrow \mathbb{R}_+$. Dann gilt für alle $n \geq 0$

- (i) und alle $i_0, \dots, i_n \in I$ mit $\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) > 0$

$$\mathbb{E}[h(X_{n+1}) | X_0 = i_0, \dots, X_n = i_n] = \mathbb{E}[h(X_{n+1}) | X_n = i_n] \quad (\text{A.85})$$

(ii) und für die auf die entsprechenden ZV bedingte Erwartung

$$\mathbb{E}[h(X_{n+1})|X_0, \dots, X_n] = \mathbb{E}[h(X_{n+1})|X_n] \quad f.s. \quad \text{und auch} \quad (\text{A.86})$$

$$\mathbb{E}[h(X_{n+1})|X_k, \dots, X_n] = \mathbb{E}[h(X_{n+1})|X_n] \quad f.s. \quad \text{für } 0 \leq k < n. \quad (\text{A.87})$$

Beweis. Im Sinne der Bem. zum Satz A.41 (i) bestimmt der erste der Ausdrücke in (A.84) eindeutig die bedingte Verteilung (siehe die Bemerkung A.58) von X_{n+1} gegeben $X_0 = i_0, \dots, X_n = i_n$ und der zweite die von X_{n+1} gegeben $X_n = i_n$ und sie sind gleich, d.h. $\mathbb{P}_{X_{n+1}|X_0=i_0, \dots, X_n=i_n} \stackrel{(\text{A.84})}{=} \mathbb{P}_{X_{n+1}|X_n=i_n}$. Dadurch wird (i) bzw. (A.85) bestätigt (vgl. auch (A.46)).

Alternativ, da $h(X_{n+1}) : \Omega \rightarrow h(I)$, $X_n : \Omega \rightarrow I$ und $\mathbf{X} := (X_0, \dots, X_n)^T : \Omega \rightarrow I^n$ jeweils diskrete ZV sind ($\mathbf{X}$ ist ein diskreter Z-Vektor mit abzählbarem Wertebereich – vgl. die Bem. zur Def. A.42 und die Bem. A.43), kann mit dem Lemma A.68

$$\begin{aligned} \mathbb{E}[h(X_{n+1})|\mathbf{X} = (i_0, \dots, i_n)^T] &= \mathbb{E}[h(X_{n+1})|X_0 = i_0, \dots, X_n = i_n] \\ &\stackrel{(\text{A.82})}{=} \sum_{j \in I} h(j) \mathbb{P}(X_{n+1} = j | X_0 = i_0, \dots, X_n = i_n) \\ &\stackrel{(\text{A.84})}{=} \sum_{j \in I} h(j) \mathbb{P}(X_{n+1} = j | X_n = i_n) \stackrel{(\text{A.82})}{=} \mathbb{E}[h(X_{n+1})|X_n = i_n] \end{aligned} \quad (\star)$$

erhalten werden, was ebenso (i) zeigt.

Definiere nun – analog zur Def. A.66 bzw. wie in (A.83) – die Funktionen $g_1 : I^n \rightarrow \mathbb{R}_+$ und $g_2 : I \rightarrow \mathbb{R}_+$ folgendermaßen:

$$g_1(i_0, \dots, i_n) := \mathbb{E}[h(X_{n+1})|\mathbf{X} = (i_0, \dots, i_n)^T] \quad \text{und} \quad g_2(i_n) := \mathbb{E}[h(X_{n+1})|X_n = i_n].$$

Anhand dieser Funktionen können nun, wie in (A.81), die bedingten Erwartungswerte $\mathbb{E}[h(X_{n+1})|\mathbf{X}] := \mathbb{E}[h(X_{n+1})|X_0, \dots, X_n] := g_1(X_0, \dots, X_n)$ und $\mathbb{E}[h(X_{n+1})|X_n] := g_2(X_n)$ definiert werden. Nun gilt wegen (i) bzw. $(\star)$ für jedes $\tilde{\omega} \in \Omega$, für welches $\mathbb{P}(\tilde{\omega}) > 0$ ist, dass $g_1(X_0(\tilde{\omega}), \dots, X_n(\tilde{\omega})) = g_2(X_n(\tilde{\omega}))$ ist (setze etwa $\tilde{i}_0 := X_0(\tilde{\omega}), \dots, \tilde{i}_n := X_n(\tilde{\omega}) \Rightarrow \mathbb{P}(X_0 = \tilde{i}_0, \dots, X_n = \tilde{i}_n) = \mathbb{P}(\omega : X_0(\omega) = \tilde{i}_0, \dots, X_n(\omega) = \tilde{i}_n) \geq \mathbb{P}(\tilde{\omega}) > 0$). Da dies also für jedes $\tilde{\omega}$ mit positiver WK gilt, stimmen die soeben definierten bedingten Erwartungen fast sicher überein. Damit ist der erste Teil von (ii), d.h. (A.86), gezeigt.

Für (A.87) beachte, dass $\sigma(X_k, \dots, X_n) \subseteq \sigma(X_0, \dots, X_n)$ gilt $(\star\star)$ und weiters, dass $\mathbb{E}[h(X_{n+1})|X_n]$ gemäß der Def. der bedingten Erwartung (Def. A.59 (i)) messbar bzgl. $\sigma(X_n)$ und somit auch bzgl. $\sigma(X_k, \dots, X_n)$ ist $(\star\star\star)$. Mit den Eigenschaften des bedingten Erwartungswertes (Satz A.61) folgt

$$\begin{aligned} \mathbb{E}[h(X_{n+1})|X_k, \dots, X_n] &\stackrel{(\star\star)}{\stackrel{(\text{A.76})}{=}} \mathbb{E}[\overbrace{\mathbb{E}[h(X_{n+1})|X_0, \dots, X_n]}^{\stackrel{(\text{A.86})}{= \mathbb{E}[h(X_{n+1})|X_n]}} | X_k, \dots, X_n] \\ &\stackrel{(\star\star\star)}{\stackrel{\text{Satz A.61 (a)}}{=}} \mathbb{E}[h(X_{n+1})|X_n] \quad f.s. \end{aligned} \quad \square$$

Notation A.73 (vgl. [41, S. 4])

Sei $\{X_n\}_{n \geq 0}$ Markov($\lambda, \mathbf{P}$) und seien $i, j \in I$. Es wird die folgende Notation festgelegt:

- **n -faches Matrixprodukt von $\mathbf{P}$:** Für $n \geq 0$ ist $\mathbf{P}^n$ das n -fache Matrixprodukt, wobei $\mathbf{P}^0 = \mathbf{I}$ die Einheitsmatrix ist. Schreibe für den Eintrag in der i -ten Zeile und der j -ten Spalte von $\mathbf{P}^n$ kurz $p_{ij}^{(n)}$, also $\mathbf{P}^n = (p_{ij}^{(n)})_{i,j \in I}$. Dann ergibt sich mit den Regeln für die Matrixmultiplikation, dass

$$\begin{aligned} p_{ij}^{(0)} &= \delta_{ij} \quad (= 1 \text{ für } i = j \text{ und } 0 \text{ sonst}), & p_{ij}^{(1)} &= p_{ij}, \\ p_{ij}^{(2)} &= \sum_{k \in I} p_{ik} p_{kj} \quad \text{und} & p_{ij}^{(n)} &= \sum_{k \in I} p_{ik}^{(n-1)} p_{kj} \quad \text{ist.} \end{aligned} \quad (\text{A.88})$$

- **Bedingte WK für Markov-Ketten:** Schreibe für $A \in \Sigma$ und alle $i \in I$ mit $\lambda_i = \mathbb{P}(X_0 = i) > 0$ die bedingte WK gegeben $X_0 = i$ als $\mathbb{P}_i(A) := \mathbb{P}(A | X_0 = i)$. $\mathbb{P}_i$ ist ein W-Maß (vgl. die Bem. zur Def. A.56 (ii)) unter welchem $\{X_n\}_{n \geq 0}$ Markov($\delta_i, \mathbf{P}$). Dies gilt aufgrund der Markov-Eigenschaft – Satz A.71. $\triangle$

Satz A.74 (Höhere Übergangswahrscheinlichkeiten)

Sei $\{X_n\}_{n \geq 0}$ Markov($\lambda, \mathbf{P}$) und seien $i, j \in I$ bzw. $m, n \geq 0$.

- (i) $p_{ij}^{(n)}$, wie in der Not. A.73 festgelegt, ist die n -te **Übergangswahrscheinlichkeit** (von i nach j), d.h. die WK in n Schritten von i nach j zu kommen, und es gilt

$$\mathbb{P}_i(X_n = j) = \mathbb{P}(X_{n+m} = j | X_m = i) = p_{ij}^{(n)}. \quad (\text{A.89})$$

- (ii) Es gilt die sogenannte **Chapman-Kolmogorov'sche Gleichung**

$$p_{ij}^{(m+n)} = \sum_{k \in I} p_{ik}^{(m)} p_{kj}^{(n)}. \quad (\text{A.90})$$

Ohne exakten Beweis. Der Punkt (i) beruht auf der ME (Satz A.71) – siehe dafür etwa [41, Theorem 1.1.3, S. 4 f.]. Durch Iteration von (A.88) kann (ii) erhalten werden (vgl. z.B. auch [30, (17.10), S. 364]). $\square$

Nun werden noch gewisse Eigenschaften einzelner Zustände beleuchtet.

Definition A.75 (Absorbierender Zustand – vgl. z.B. [30, Def. 17.30, S. 374])

Ein Zustand $i \in I$ einer Markov-Kette heißt *absorbierend*, falls $p_{ii} = 1$ ist. D.h., dass dieser Zustand bei einmaligem Eintritt nie mehr verlassen werden kann.

Definition A.76 (Transienter Zustand – vgl. z.B. [41, 1.5, S. 24])

Ein Zustand $i \in I$ einer Markov-Kette heißt *transient*, falls die Kette von i ausgehend fast sicher nur endlich oft zu diesem Zustand zurückkehrt, d.h.:

$$i \text{ transient} \iff \mathbb{P}_i('X_n = i \text{ für } \infty \text{ viele } n') = \mathbb{P}_i\left(\limsup_{n \rightarrow \infty} \{X_n = i\}\right) = 0. \quad (\text{A.91})$$

Satz A.77 (Bedingung für die Transienz eines Zustands – angelehnt an [41, Theorem 1.5.3 (ii), S. 26])

Seien $\{X_n\}_{n \geq 0}$ Markov(λ, P) und $i \in I$ bzw. $m \geq 0$ so, dass $\mathbb{P}(X_m = i) > 0$ ist. Wenn

$$\mathbb{P}(\exists n \geq 1 : X_{m+n} = i | X_m = i) < 1 \text{ ist,} \quad \text{gilt} \quad \sum_{n \geq 0} p_{ii}^{(n)} < \infty \text{ und } i \text{ ist transient.} \quad (\text{A.92})$$

Dabei ist (die Gleichheit gilt wegen der Markov-Eigenschaft bzw. gemäß (A.89))

$$\mathbb{P}(\exists n \geq 1 : X_{m+n} = i | X_m = i) = \mathbb{P}_i(\exists n \geq 1 : X_n = i) =: f_i \quad (\text{A.93})$$

die Rückkehrwahrscheinlichkeit (nach dem ersten Sprung) von i nach i .

Beweis(-skizze). Sei also die Rückkehrwahrscheinlichkeit $f_i < 1$. Beachte, dass $\mathbb{P}_i$ ein W-Maß ist und betrachte den W-Raum $(\Omega, \Sigma, \mathbb{P}_i)$. Es folgt

$$\begin{aligned} \sum_{n \geq 0} p_{ii}^{(n)} &\stackrel{(\text{A.89})}{=} \sum_{n \geq 0} \mathbb{P}_i(X_n = i) \stackrel{(\text{A.44})}{=} \sum_{n \geq 0} \mathbb{E}_{\mathbb{P}_i} [\mathbb{1}_{\{X_n = i\}}] \stackrel{\text{MON}}{\stackrel{(\text{A.21})}{=}} \mathbb{E}_{\mathbb{P}_i} \left[\sum_{n \geq 0} \mathbb{1}_{\{X_n = i\}} \right] \\ &\stackrel{(2.5)}{=} \sum_{r \geq 0} \mathbb{P}_i \left(\sum_{n \geq 0} \mathbb{1}_{\{X_n = i\}} > r \right) \stackrel{(\star)}{=} \sum_{r \geq 0} f_i^r \stackrel{\text{geom. Reihe}}{\stackrel{(f_i < 1 \text{ lt. Vorauss.})}{=}} \frac{1}{1 - \underbrace{f_i}_{< 1}} < \infty. \end{aligned}$$

- ($\star$) Für alle $r \geq 0$ ist $\mathbb{P}_i(\sum_{n \geq 0} \mathbb{1}_{\{X_n = i\}} > r)$ die WK dafür, ausgehend von $X_0 = i$ für mindestens r weitere $n \geq 1$ das Ereignis $X_n = i$ zu beobachten (also insgesamt mind. $(r + 1)$ -mal $X_n = i$). Es ist plausibel, dass das der WK, von i ausgehend mindestens r -mal zurück nach i zu kommen, entspricht – also genau der r -ten Potenz der Rückkehrwahrscheinlichkeit von i nach i , f_i^r . Etwa unter [41, Lemma 1.5.2, S. 25] findet sich ein exakter Beweis dafür.

Damit ist der erste Teil von (A.92), also $\sum_{n \geq 0} \mathbb{P}_i(X_n = i) < \infty$, gezeigt. Mit dem *ersten Borel-Cantelli-Lemma*, Lemma A.24, folgt daraus $\mathbb{P}_i(\limsup_{n \rightarrow \infty} \{X_n = i\}) = 0$. Das ist genau das, was laut der Definition A.76 die Transienz von i bedeutet. $\square$

Korollar zum Satz A.77 (angelehnt an [12, XV.(5.5), S. 389])

Wie im Satz gelte für ein $i \in I$, dass $\sum_{n \geq 0} p_{ii}^{(n)} < \infty$ ist (der Zustand i ist also transient und $\mathbb{P}_i(\limsup_{n \rightarrow \infty} \{X_n = i\}) = 0$). Dann gilt auch

$$\text{für alle anderen } j \in I, \text{ dass} \quad \sum_{n \geq 1} p_{ji}^{(n)} < \infty \quad (\text{A.94})$$

$$\text{und somit} \quad \mathbb{P} \left(\limsup_{n \rightarrow \infty} \{X_n = i\} \right) = \mathbb{P}(\text{'}X_n = i \text{ für } \infty \text{ viele } n\text{'}) = 0 \quad \text{ist.} \quad (\text{A.95})$$

Beweis. Wegen der Chapman-Kolmogorov'schen Gleichung (A.90) gilt für $n \geq 0$ $\sum_{j \in I} p_{ij} p_{ji}^{(n)} = p_{ii}^{(n+1)}$ und somit für ein beliebiges $j \in I$ auch $p_{ij} p_{ji}^{(n)} \leq p_{ii}^{(n+1)}$ ($\star$).

Es folgt $p_{ij} \sum_{n \geq 1} p_{ji}^{(n)} = \sum_{n \geq 1} p_{ij} p_{ji}^{(n)} \stackrel{(\star)}{\leq} \sum_{n \geq 1} p_{ii}^{(n+1)} = \sum_{n \geq 2} p_{ii}^{(n)} \leq \sum_{n \geq 0} p_{ii}^{(n)} < \infty$ (wobei die letzte Ungleichung nach Voraussetzung gilt) und daraus $\sum_{n \geq 1} p_{ji}^{(n)} < \infty$. Da $j \in I$ beliebig war, ist damit (A.94) gezeigt.

Sei nun $J := \{j \in I : \mathbb{P}(X_0 = j) > 0\}$. Da für alle $j \in J$ wegen des soeben Bewiesenen $\sum_{n \geq 1} \mathbb{P}_j(X_n = i) = \sum_{n \geq 1} p_{ji}^{(n)} < \infty$ gilt, führt – wie im Beweis des Satzes – wieder das *erste Borel-Cantelli-Lemma* zu $\mathbb{P}_j(\limsup_{n \rightarrow \infty} \{X_n = i\}) = 0$ ($\star\star$). Nun sind die Ereignisse $\{X_0 = j\}$ für unterschiedliche j paarw. disj. und somit ist (wegen der σ -Add. und der Def. A.70) $\mathbb{P}(\bigcup_{j \in J} \{X_0 = j\}) = \sum_{j \in J} \lambda_j = 1$ ($\lambda_j = 0 \ \forall j \notin J$). Mit dem Satz A.57 über die totale WK erhält man folglich

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \{X_n = i\}\right) \stackrel{(A.71)}{=} \sum_{j \in J} \overbrace{\mathbb{P}\left(\limsup_{n \rightarrow \infty} \{X_n = i\} \mid X_0 = j\right)}^{=0 \ \forall j \in J \ (\star\star)} \mathbb{P}(X_0 = j) = 0. \quad \square$$

Bemerkungen zum Satz A.77.

- (i) Transienz wird oft auch gleich dadurch formuliert, dass die Rückkehrwahrscheinlichkeit f_i kleiner 1 ist (vgl. z.B. [30, Def. 17.30, S. 374] oder [6, S. 117 f.]). Das ist deshalb möglich, da die Implikation im Satz eigentlich auch in die andere Richtung gilt und somit (A.92) auch mit ' $\iff$ ' geschrieben werden kann (vgl. z.B. [6, Theorem 8.2, S. 118]).
- (ii) Im Lichte von (i) heißt die Transienz eines Zustandes $i \in I$ i.a.W. auch, dass er möglicherweise für immer verlassen wird bzw. mit positiver WK nie mehr dorthin zurückgekehrt wird. Dies kann man sich dadurch überlegen, dass

$$1 > f_i \stackrel{(A.93)}{=} \mathbb{P}_i(\exists n \geq 1 : X_n = i) = \mathbb{P}_i\left(\bigcup_{n \geq 1} \{X_n = i\}\right) = \mathbb{P}_i\left(\bigcup_{n \geq 1} \{X_n \neq i\}^c\right)$$

$$\stackrel{\text{deMorgan-Regeln}}{=} 1 - \mathbb{P}_i\left(\bigcap_{n \geq 1} \{X_n \neq i\}\right) = 1 - \mathbb{P}_i(X_n \neq i \ \forall n \geq 1) \quad \text{gilt,}$$

$$\text{also auch } \mathbb{P}_i(X_n \neq i \ \forall n \geq 1) \stackrel{(A.89)}{=} \mathbb{P}(X_{m+n} \neq i \ \forall n \geq 1 \mid X_m = i) > 0.$$

- (iii) Die zweite Aussage (A.95) des Korollars zum Satz zeigt formal etwas, was auch intuitiv plausibel ist: Wenn die WK von i ausgehend unendlich oft wieder bei i zu landen gleich 0 ist, muss das auch ohne diese Ausgangsbedingung gelten. $\triangle$

Bemerkung A.78 (Rekurrenter Zustand – vgl. z.B. [41, 1.5, S. 24 ff.])

Das Gegenstück zur Transienz ist die Rekurrenz. Ein Zustand $i \in I$ ist *rekurrent*, wenn $\mathbb{P}_i(X_n = i \text{ für } \infty \text{ viele } n) = 1$ ist. Dann ist $f_i = 1$ und $\sum_{n \geq 0} p_{ii}^{(n)} = \infty$. Insbesondere ist demnach jeder absorbierende Zustand rekurrent. $\triangle$

Satz A.79. *Ein Zustand einer Markov-Kette ist immer entweder transient oder rekurrent (ohne Beweis; vgl. dazu etwa erneut [41, Theorem 1.5.3, S. 26]).*

A.2.5 Konvergenzarten

Es werden noch kurz die in der Arbeit verwendeten Konvergenzbegriffe für Zufallsvariablen bzw. deren Schreibweise festgelegt und ausgewählte Zusammenhänge beigelegt.

Definition A.80 (Konvergenzarten von ZV vgl. z.B. [21, Kap. 6, S. 215 ff.]

Seien reelle Zufallsvariablen $\{X_n\}_{n \geq 0}$ und X , eine weitere Zufallsvariable, gegeben.

- (i) Die Folge $\{X_n\}_{n \geq 0}$ *konvergiert* ($\mathbb{P}$ -) *fast sicher* (f.s.) gegen X , sofern

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} X_n = X\right) = \mathbb{P}(X_n \rightarrow X) = 1$$

gilt. Dies wird als $X_n \xrightarrow{\text{f.s.}} X$ geschrieben.

- (ii) Man spricht davon, dass $\{X_n\}_{n \geq 0}$ *nach Wahrscheinlichkeit* gegen X *konvergiert* und schreibt dafür $X_n \xrightarrow{\mathbb{P}} X$, falls für alle $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0 \quad (\text{kurz } \mathbb{P}(|X_n - X| \geq \varepsilon) \rightarrow 0) \quad \text{gilt.}$$

- (iii) Seien $1 \leq p < \infty$, $\mathbb{E}|X|^p < \infty$ sowie $\mathbb{E}|X_n|^p < \infty$ für jedes $n \geq 0$. Gilt

$$\lim_{n \rightarrow \infty} \mathbb{E}|X_n - X|^p = 0 \quad (\text{kurz } \mathbb{E}|X_n - X|^p \rightarrow 0),$$

so *konvergiert* $\{X_n\}_{n \geq 0}$ *im* p -*ten Mittel* gegen X , kurz $X_n \xrightarrow{\mathcal{L}^p} X$.

- (iv) Seien $\{F_n\}_{n \geq 0}$ bzw. F die Verteilungsfunktionen von $\{X_n\}_{n \geq 0}$ bzw. X , d.h., es gilt $F_n(x) = \mathbb{P}(X_n \leq x)$ bzw. $F(x) = \mathbb{P}(X \leq x)$ für alle $x \in \mathbb{R}$ sowie $n \geq 0$ (Def. A.29 (iv)). Wenn die Folge $\{F_n\}_{n \geq 0}$ für jedes $x \in \mathbb{R}$, welches auch ein Stetigkeitspunkt von F ist, gegen F konvergiert, d.h.,

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad (\text{kurz } F_n(x) \rightarrow F(x)), \quad \text{wenn } F \text{ in } x \text{ stetig ist,}$$

so *konvergiert* $\{X_n\}_{n \geq 0}$ *in Verteilung* gegen X , kurz $X_n \xrightarrow{d} X$.

Es gelten die folgenden Zusammenhänge zwischen diesen Konvergenzarten.

Satz A.81 (Implikationen bzgl. der Konvergenz im p -ten Mittel)

- (i) Unter den Voraussetzungen der Def. A.80 (iii) gilt

aus $X_n \xrightarrow{\mathcal{L}^p} X$ folgt $\mathbb{E}|X_n|^p \rightarrow \mathbb{E}|X|^p$ und insbesondere

aus $X_n \xrightarrow{\mathcal{L}^1} X$ folgt $\mathbb{E}X_n \rightarrow \mathbb{E}X$.

- (ii) Ist $q > p \geq 1$, so folgt aus $X_n \xrightarrow{\mathcal{L}^q} X$ auch $X_n \xrightarrow{\mathcal{L}^p} X$. Weiters folgt aus $X_n \xrightarrow{\mathcal{L}^p} X$ auch $X_n \xrightarrow{\mathbb{P}} X$.

Ohne Beweis. Siehe z.B. [4, 15.1 Theorem, S. 80] für (i) und [38, Kor. 2.32, S. 48] für die erste Aussage in (ii). Mit dieser folgt insbesondere $X_n \xrightarrow{\mathcal{L}^1} X$ aus $X_n \xrightarrow{\mathcal{L}^p} X$. Daraus ergibt sich mit der *Markov-Ungleichung* (Satz A.37) auch $X_n \xrightarrow{\mathbb{P}} X$, denn für alle $\varepsilon > 0$ gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| \geq \varepsilon) \stackrel{(A.47)}{\leq} \lim_{n \rightarrow \infty} \frac{\mathbb{E}|X_n - X|}{\varepsilon} = 0 \quad (\text{vgl. (ii) bzw. (iii) der obigen Def.}). \quad \square$$

Satz A.82 (Satz von der stetigen Abbildung und weitere Zusammenhänge)

Seien wieder reelle Zufallsvariablen $\{X_n\}_{n \geq 0}$ und X gegeben.

(i) **Satz von der stetigen Abbildung:** Sei $g : \mathbb{R} \rightarrow \mathbb{R}$ auf $C \subseteq \mathbb{R}$ stetig, wobei C so sei, dass $\mathbb{P}(X \in C) = 1$ gilt (d.h., g ist stetig f.s.). Dann gilt:

$$\begin{aligned} X_n \xrightarrow{d} X &\Rightarrow g(X_n) \xrightarrow{d} g(X), & X_n \xrightarrow{\mathbb{P}} X &\Rightarrow g(X_n) \xrightarrow{\mathbb{P}} g(X) \quad \text{und} \\ X_n \xrightarrow{f.s.} X &\Rightarrow g(X_n) \xrightarrow{f.s.} g(X). \end{aligned}$$

(ii) **Implikationen bzgl. den anderen Konvergenzarten:** Es gilt:

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

Ohne Beweis. Für (i) siehe etwa [43, 2.3 Theorem, S. 7] (dort *continuous mapping theorem* genannt – kurz *CMT*). Die Implikationen in (ii) sind z.B. unter [38, Satz 4.14, S. 97 und Satz 7.2 bzw. Abb. 7.1, S. 174 f.] nachzulesen. $\square$

A.2.6 Laplace-Transformation und charakteristische Funktionen

Im Rahmen des Abschnitts 2.6 bzw. des Unterabschnitts 4.1.2 kommen teilweise die folgenden Funktionen zum Einsatz, weshalb diese nun kurz angerissen werden.

Satz A.83 (Laplace-Transformation nichtnegativer ZV – vgl. [18, A.4.4, S. 287])

Definiere für eine ZV $X : \Omega \rightarrow \mathbb{R}_+$ die sogenannte *Laplace-Transformation* oder **Laplace-Transformierte** von X als die Funktion $\mathcal{L}_X : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ mit

$$\mathcal{L}_X(u) := \mathbb{E}[e^{-uX}] \quad \text{für } u \in \mathbb{R}_+. \quad (\text{A.96})$$

Die Laplace-Transformierte bestimmt in eindeutiger Weise die Verteilung von X . D.h., wenn Y eine andere nichtnegative reelle ZV ist, gilt:

$$\begin{aligned} X \stackrel{d}{=} Y &\iff \mathcal{L}_X = \mathcal{L}_Y. & \text{Weiters gilt} \\ \mathbb{E}X^k &= (-1)^k \lim_{u \searrow 0} \mathcal{L}^{(k)}(u) = (-1)^k \mathcal{L}^{(k)}(0^+). \end{aligned} \quad (\text{A.97})$$

Auf diese Weise können also die Momente von X aus der k -ten Ableitung der Laplace-Transformierten (welche in $(0, \infty)$ beliebig oft differenzierbar ist) 'erzeugt' werden.

Ohne Beweis. Mit dem Transformationssatz für ZV (Lemma A.31) kann man erkennen, dass $\mathcal{L}_X(u) \stackrel{(\text{A.96})}{=} \mathbb{E}[e^{-uX}] \stackrel{(\text{A.43})}{=} \int e^{-uX} d\mathbb{P} \stackrel{(\text{A.40})}{=} \int_{x \in \mathbb{R}_+} e^{-ux} d\mathbb{P}_X =: \mathcal{L}_{\mathbb{P}_X}(u)$ für $u \in \mathbb{R}_+$ gilt. Dies entspricht allgemein der Laplace-Transformierten des W-Maßes $\mathbb{P}_X$ – wie etwa in [30, Satz 15.6, S. 304] definiert. Der Satz in dieser Referenz bestätigt dann die Aussage über die Eindeutigkeit. Für die Momenterzeugung im spezifischen Fall nichtnegativer ZV in (A.97) siehe etwa [30, Bsp. 6.29 bzw. (6.8), S. 146]. $\square$

Bemerkung zum Satz A.83. Es gilt $\mathcal{L}_X(u) \stackrel{(A.96)}{=} \mathbb{E}[e^{-uX}] = \mathbb{E}[(e^{-u})^X] \stackrel{(2.6)}{=} f_X(e^{-u})$, falls X $\mathbb{N}_0$ -wertig ist (wobei f die *wahrscheinlichkeitserzeugende Funktion* von X ist).

Häufig wird statt $\mathcal{L}_X$ die sogenannte *momenterzeugende Funktion* als $M_X := \mathbb{E}[e^{uX}]$ für reelle (oder komplexe) u definiert (vgl. z.B. [21, Def. 2.8.5, S. 89]). Im obigen Fall nichtnegativer ZV ist die auf $u \geq 0$ beschränkt definierte $\mathcal{L}_X(u) (= M_X(-u))$ aber immer wohldefiniert ($\mathbb{E}[e^{-uX}] \leq 1$). Das ist bei momenterzeugenden Funktionen i.A. nicht der Fall und auch die Ableitungen müssen für die Momenterzeugung in einer offenen Umgebung von 0 vorhanden sein. In (A.97) ist die einseitige Betrachtung ausreichend. $\triangle$

Satz A.84 (Charakteristische Funktionen)

Definiere die **charakteristische Funktion** $\varphi_X : \mathbb{R} \rightarrow \mathbb{C}$ einer ZV $X : \Omega \rightarrow \mathbb{R}$ durch

$$\varphi_X(t) := \mathbb{E}[e^{itX}] \quad \text{für } t \in \mathbb{C} \quad (i = \sqrt{-1}). \quad (\text{A.98})$$

- (i) **Eindeutigkeitssatz:** Die Verteilung von X ist durch φ_X eindeutig bestimmt.
- (ii) **Stetigkeitssatz von Lévy:** Seien $\{X_n\}_{n \geq 0}$ weitere reelle ZV mit charakteristischen Funktionen $\{\varphi_{X_n}\}_{n \geq 0}$. Dann gilt:

$$X_n \xrightarrow{d} X \quad \Longleftrightarrow \quad \varphi_{X_n}(t) \xrightarrow{n \rightarrow \infty} \varphi_X(t) \quad \text{für alle } t \in \mathbb{R}.$$

Ohne Beweis. Für die Definition und den Eindeutigkeitssatz siehe etwa [38, Def. 7.7, S. 181 bzw. Satz 7.13, S. 187] oder [21, Def. 2.8.6, S. 90 bzw. Satz 2.8.9, S. 94]. Der Stetigkeitssatz steht z.B. unter [21, Satz 2.8.12, S. 97] oder [17, 5.9 (5), S. 190]. $\square$

Bemerkung zum Satz A.84. Es wird der EW der *komplexen* ZV $Z := e^{itX}$ gebildet. Dieser ist allgemein durch $\mathbb{E}Z := \mathbb{E}[Re(Z)] + i\mathbb{E}[Im(Z)]$ definiert (vgl. z.B. [38, S. 180 f.] – $Re(Z)$ ist dabei der Real- und $Im(Z)$ der Imaginärteil von Z). Deshalb ist Z auch genau dann integrierbar, wenn sowohl $Re(Z)$ als auch $Im(Z)$ integrierbar sind.

Beachte, dass mit der *Eulerschen Formel* (vgl. z.B. [22, (68.5), S. 395])

$$e^{iy} = \cos(y) + i \sin(y) \quad \text{für alle } y \in \mathbb{R} \text{ gilt.} \quad (\text{A.99})$$

Daher sind $|Re(Z)| = |\cos(tX)| \leq 1$, $|Im(Z)| = |\sin(tX)| \leq 1$ und somit beide integrierbar. Die charakteristische Funktion φ_X ist deshalb auf ganz $\mathbb{R}$ wohldefiniert.

Für $k \geq 0$ können die k -ten Momente von X durch $\mathbb{E}X^k = i^{-k}\varphi_X^{(k)}(0)$ gewonnen werden, sofern $\mathbb{E}|X|^k < \infty$ ist (vgl. z.B. [21, (2.84), S. 96]; wird M_X – siehe die Bem. zum Satz A.83 – für komplexe Zahlen definiert, gilt $M_X(-it) = \varphi_X(t)$ für alle $t \in \mathbb{R}$). $\triangle$

Beispiel A.85 ($\mathcal{L}_X$ und φ_X einer exponentialverteilten ZV)

Sei X eine mit Parameter $\lambda \in \mathbb{R}_+$ exponentialverteilte ZV, kurz $X \sim \text{Exp}(\lambda)$ (d.h. mit Dichtefunktion $p : \mathbb{R} \rightarrow [0, 1]$, definiert als $p(x) := \lambda e^{-\lambda x} \mathbb{1}_{\{x > 0\}}$). Dann gilt

$$\mathcal{L}_X(u) = \frac{\lambda}{\lambda + u} \quad \text{für } u \in \mathbb{R}_+ \quad \text{bzw.} \quad \varphi_X(t) = \frac{\lambda}{\lambda - it} \quad \text{für } t \in \mathbb{R}. \quad (\text{A.100})$$

Weiters gilt $\mathbb{E}X = \frac{1}{\lambda}$, $\mathbb{V}X = \frac{1}{\lambda^2}$ ($\Rightarrow \mathbb{E}X^2 = \mathbb{V}X + (\mathbb{E}X)^2 = \frac{2}{\lambda^2}$; vgl. z.B. [20, S. 274]). Dass $\mathcal{L}_X(u) = \mathbb{E}[e^{-uX}] = \int_{\mathbb{R}_+} e^{-ux} \lambda e^{-\lambda x} dx = \lambda \int_{\mathbb{R}_+} e^{-x(\lambda+u)} dx = \lambda \left[\frac{e^{-x(\lambda+u)}}{-(\lambda+u)} \right]_0^\infty = \lambda(0 + \frac{1}{\lambda+u})$ für $u \in \mathbb{R}_+$ ist, zeigt den ersten Teil von (A.100). Für φ_X vgl. etwa [17, 5.8 (3), S. 186]. $\triangle$

A.2.7 Martingale in diskreter Zeit

Im Zuge der Normalisierung der Populationsgrößen eines GWP kommen sogenannte Martingale zur Anwendung (vgl. die Def. 2.36 bzw. den Satz 2.37 sowie den Abschnitt 2.6). Diese sind stochastische Prozesse, welche im Mittel fair sind. Es werden nur solche auf den natürlichen Zahlen $\mathbb{N}_0$ betrachtet. Sei dafür in diesem Unterabschnitt stets $\{X_n\}_{n \geq 0}$ ein stochastischer Prozess mit Werten in (Ω', Σ') , wobei von reellwertigen Zufallsvariablen, also z.B. $(\Omega', \Sigma') \subseteq (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, ausgegangen wird.

Zunächst wird der Filtrationsbegriff eingeführt, anschließend die Martingaldefinition vorgenommen und ein passender Konvergenzsatz für nichtnegative Martingale präsentiert.

Filtrationen und daran adaptierte Prozesse

Als eine *Filtration* (auf dem W-Raum $(\Omega, \Sigma, \mathbb{P})$) wird eine in $n \in \mathbb{N}_0$ wachsende Folge $\{\mathcal{F}_n\}_{n \geq 0}$ von Teil- σ -Algebren von Σ bezeichnet. D.h., es gilt $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots \subseteq \Sigma$ bzw. $\mathcal{F}_m \subseteq \mathcal{F}_n \subseteq \Sigma$ für alle $n \geq m \geq 0$ (vgl. etwa [38, Def. 11.1, S. 301]).

Wenn für jedes $n \geq 0$ gilt, dass X_n ($X_n : \Omega \rightarrow \Omega'$) $\mathcal{F}_n - \Sigma'$ messbar ist, so ist der Prozess an die Filtration $\{\mathcal{F}_n\}_{n \geq 0}$ *adaptiert* (vgl. etwa [38, Def. 11.2, S. 302]).

Die sogenannte *natürliche Filtration* eines Prozesses erhält man, wenn jeweils (für jedes $n \geq 0$) die gemäß dem Satz A.11 (ii) kleinste σ -Algebra, bezüglich der die $X_0, \dots, X_n$ messbar sind, $\mathcal{F}_n$ bildet. D.h., wenn $\mathcal{F}_n = \sigma(X_0, \dots, X_n) = \sigma(X_m : m \leq n)$ für jedes $n \geq 0$ ist (siehe etwa [38, Bsp. 11.3, S. 302]). Dementsprechend ist jeder Prozess an seine natürliche Filtration adaptiert.

Martingaldefinition

Für die Definition eines Martingals muss eine Filtration mit angegeben werden. Es wird gefordert, dass ein Martingal an diese Filtration adaptiert ist.

Definition A.86 (Martingal in diskreter Zeit – vgl. z.B. [38, Def. 11.4, S. 303])

Sei $\{\mathcal{F}_n\}_{n \geq 0}$ eine Filtration des Prozesses $\{X_n\}_{n \geq 0}$, wobei der Prozess an diese adaptiert sei. Gilt $\mathbb{E}|X_n| < \infty$ für alle $n \geq 0$ und

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n \text{ f.s.}, \quad (\text{A.101})$$

so wird $\{X_n\}_{n \geq 0}$ als *Martingal bzgl. $\{\mathcal{F}_n\}_{n \geq 0}$* bezeichnet. Sind die X_n für alle $n \geq 0$ $\mathbb{R}_+$ -wertig, so handelt es sich um ein *nichtnegatives Martingal*.

Handelt es sich bei $\{\mathcal{F}_n\}_{n \geq 0}$ um die natürliche Filtration (d.h. $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$ für alle $n \geq 0$), so ist der Prozess, wie erwähnt, immer adaptiert und somit ein *Martingal bzgl. der natürlichen Filtration*. Dann kann (A.101) auch wie folgt geschrieben werden:

$$\mathbb{E}[X_{n+1} | X_0, \dots, X_n] \stackrel{(\text{A.77})}{=} \mathbb{E}[X_{n+1} | \mathcal{F}_n] \stackrel{(\text{A.101})}{=} X_n \text{ f.s.} \quad (\text{A.102})$$

Bemerkungen zur Definition A.86.

- (i) Die Eigenschaft (A.101) gilt auch für größere Schritte. Da $\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \dots \subseteq \mathcal{F}_{m-1}$ für $m > n \geq 0$ gilt, erhält man mit der Turmeigenschaft für die bedingte Erwartung, Satz A.61 (d), iterativ $\stackrel{=X_{m-1} \text{ f.s. (A.101)}}{=}$

$$\begin{aligned} \mathbb{E}[X_m | \mathcal{F}_n] &\stackrel{(A.76)}{=} \mathbb{E}[\overbrace{\mathbb{E}[X_m | \mathcal{F}_{m-1}]}^{\text{wie zuvor}} | \mathcal{F}_n] = \mathbb{E}[X_{m-1} | \mathcal{F}_n] \\ &\stackrel{\text{wie zuvor}}{=} \mathbb{E}[X_{m-2} | \mathcal{F}_n] = \dots = \mathbb{E}[X_{n+1} | \mathcal{F}_n] \stackrel{(A.101)}{=} X_n \text{ f.s.} \end{aligned} \quad (A.103)$$

- (ii) Die Erwartungswerte aller Folgenglieder eines Martingals sind gleich, denn mit (i) und dem Satz A.61 (b) über den totalen EW ergibt sich

$$\mathbb{E}X_n \stackrel{(A.74)}{=} \underbrace{\mathbb{E}[\mathbb{E}[X_n | \mathcal{F}_0]]}_{=X_0 \text{ f.s. mit (A.103)}} = \mathbb{E}X_0 \quad \text{für alle } n \geq 0. \quad (A.104)$$

Deshalb ist es bereits ausreichend, anstelle von $\mathbb{E}|X_n| < \infty$ für alle $n \geq 0$ nur etwa $\mathbb{E}|X_0| < \infty$ zu fordern. $\triangle$

Ein Martingalkonvergenzsatz

Satz A.87 (Konvergenzsatz für nichtnegative Martingale)

Sei $\{X_n\}_{n \geq 0}$ ein nichtnegatives Martingal.

- (i) Dann existiert eine integrierbare, nichtnegative ZV X (demnach ist $0 \leq \mathbb{E}X < \infty$), sodass X_n fast sicher gegen X konvergiert, d.h. $X_n \xrightarrow{f.s.} X$ (siehe die Def. A.80 (i)).
- (ii) Gilt zudem, dass auch $\sup_{n \geq 0} \mathbb{E}X_n^2 < \infty$ ist, so gilt sogar $\mathbb{E}X^2 < \infty$ und die Konvergenz auch im quadratischen Mittel, also $X_n \xrightarrow{\mathcal{L}^2} X$ (siehe die Def. A.80 (iii)).

Ohne Beweis. Es handelt sich um spezifische Resultate von meist allgemeiner gehaltenen Martingalkonvergenzsätzen, wie sie u.a. in der angegebenen Literatur gefunden werden können. Dass jedes nichtnegative Martingal f.s. konvergiert, besagt etwa [40, Theorem II-2-9, S. 26]. Beachte weiters, dass der Prozess, um ein Martingal gemäß der Def. A.86 zu sein (siehe die Bem. (ii) dazu), $0 \leq \sup_{n \geq 0} \mathbb{E}X_n \stackrel{(A.104)}{=} \mathbb{E}X_n = \mathbb{E}X_0 < \infty$ erfüllen muss. Für den Punkt (i) kann dann bspw. [38, Theorem 11.34, S. 323 oder Korollar 11.35, S. 325] herangezogen werden und [38, Satz 11.42, S. 329] liefert (ii). $\square$

Korollar zum Satz A.87. Unter den Vorauss. des Punktes (ii) des Satzes gilt zudem

$$\mathbb{E}X_n \rightarrow \mathbb{E}X < \infty \quad \text{sowie} \quad \mathbb{V}X_n \rightarrow \mathbb{V}X < \infty. \quad (A.105)$$

Beweis. Ist $\sup_{n \geq 0} \mathbb{E}X_n^2 < \infty$ erfüllt, folgt (wegen (ii) des Satzes) $X_n \xrightarrow{\mathcal{L}^2} X$ und mit dem Satz A.81 (ii) auch $X_n \xrightarrow{\mathcal{L}^1} X$. Daraus lässt sich gemäß dem Satz A.81 (i) auch direkt $\mathbb{E}X_n^2 \rightarrow \mathbb{E}X^2 < \infty$ bzw. $\mathbb{E}X_n \rightarrow \mathbb{E}X < \infty$ schließen. Weiters gilt deshalb (mit dem Varianz-Verschiebungssatz) $\mathbb{V}X_n = \mathbb{E}X_n^2 - (\mathbb{E}X_n)^2 \rightarrow \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{V}X < \infty$. $\square$

A.3 Resultate und Begrifflichkeiten – Analysis

Nun werden noch ergänzend einige Definitionen und Resultate über die Eigenschaften von Funktionen, Folgen und Reihen (insbesondere von Doppel- und Potenzreihen), welche in dieser Arbeit angewendet werden, angeführt. Abschließend (im Abschnitt A.3.3) wird noch auf die Doppelfakultät und sogenannte *Catalan-Zahlen* eingegangen.

A.3.1 Funktionen

Auf folgende Basisresultate bezüglich reeller Funktionen wird im Text referenziert.

Monotonie und Konvexität

Satz A.88 (Monotonie von Funktionen über die erste Ableitung – vgl. z.B. [22, Satz 49.5, S. 281]; ohne Beweis). *Ist eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ auf einem Intervall I stetig und in seinem Inneren $I_\circ$ differenzierbar, so ist sie auf I*

- *monoton wachsend (isoton), wenn $f' \geq 0$ auf $I_\circ$ ist (strikt isoton mit $>$).*
- *monoton fallend (antiton), wenn $f' \leq 0$ auf $I_\circ$ ist (strikt antiton mit $<$).*

Satz A.89 (Konvexe und konkave Funktionen über die zweite Ableitung – vgl. z.B. [22, (49.2) und Satz 49.8, S. 282 ff.]; ohne Beweis). *Eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ ist auf einem Intervall I konvex, wenn für alle $x_0, x_1 \in I$ mit $x_0 \neq x_1$ sowie für alle $\lambda \in (0, 1)$ gilt, dass $f(x_0\lambda + x_1(1 - \lambda)) \leq f(x_0)\lambda + f(x_1)(1 - \lambda)$ ist.*

Gilt die Ungleichung umgekehrt ($\geq$), dann ist f konkav. Sind die Ungleichungen strikt, dann ist f strikt konvex ($<$) bzw. strikt konkav ($>$).

Ist f auf I stetig und in seinem Inneren $I_\circ$ zweimal differenzierbar, so ist sie

- *konvex auf I , wenn $f'' \geq 0$ auf $I_\circ$ ist (strikt konvex mit $>$).*
- *konkav auf I , wenn $f'' \leq 0$ auf $I_\circ$ ist (strikt konkav mit $<$).*

Zwischen-, Mittelwert- und Nullstellensatz

Sei $[a, b]$ mit $a, b \in \mathbb{R}$ und $a < b$ ein abgeschlossenes Intervall und sei $f : [a, b] \rightarrow \mathbb{R}$ stetig.

Satz A.90 (Nullstellensatz (v. Bolzano) – vgl. [22, Satz 35.5, S. 223]; ohne Beweis) *Gilt entweder $f(a) > 0$ und $f(b) < 0$ oder $f(a) < 0$ und $f(b) > 0$, so besitzt f mindestens eine Nullstelle im offenen Intervall (a, b) .*

Satz A.91 (Zwischenwertsatz (von Bolzano) – vgl. [22, Satz 35.6, S. 223]; ohne Beweis). *Die Funktion f nimmt jeden Wert im offenen Intervall $(f(a), f(b))$ an.*

Satz A.92 (Mittelwertsatz der Differentialrechnung – vgl. [22, Satz 49.1, S. 279]; ohne Beweis). *Ist f zumindest auf (a, b) differenzierbar, dann gilt*

$$f'(c) = \frac{f(b) - f(a)}{b - a} \quad \text{für mindestens ein } c \in (a, b). \quad (\text{A.106})$$

A.3.2 Reihen und Potenzreihen

Konvergenz bzw. Divergenz von Reihen und deren Umordnungen

Das Konzept der 'normalen' Konvergenz von Folgen und Reihen wird als bekannt vorausgesetzt und andernfalls etwa auf [13], [22] oder [31] verwiesen. Zwei fundamentale, auch hier verwendete Konvergenzprinzipien für Folgen werden dennoch ausformuliert.

Satz A.93 (Monotonieprinzip für Folgen – vgl. z.B. [22, Satz 23.1, S. 155]; ohne Beweis). *Eine Folge $\{a_n\}_{n \geq 0}$, die monoton und beschränkt ist, konvergiert. Insbesondere konvergiert jede monotone Folge genau dann, wenn sie beschränkt ist. Es gilt:*

- wenn sie monoton wachsend (isoton) ist, $\lim_{n \rightarrow \infty} a_n = \sup_{n \geq 0} a_n$.
- wenn sie monoton fallend (antiton) ist, $\lim_{n \rightarrow \infty} a_n = \inf_{n \geq 0} a_n$.

Satz A.94 (Konvergenzprinzip für Cauchy-Folgen – vgl. z.B. [22, Satz 23.3, S. 157]; ohne Beweis). *Sei $\{a_n\}_{n \geq 0}$ eine Folge. Wenn es zu jedem $\varepsilon > 0$ ein $n_0 \geq 0$ gibt, sodass $|a_m - a_n| < \varepsilon$ für alle $n, m > n_0$ gilt, so handelt es sich um eine sogenannte **Cauchy-Folge**.*

Es gilt, dass eine Folge genau dann konvergiert, wenn sie eine Cauchy-Folge ist.

Man darf alle *konvergenten* Reihen gliedweise addieren, subtrahieren und mit einer Konstanten multiplizieren und unter entsprechenden Voraussetzungen ist dies auch mit *bestimmt divergenten* Reihen (siehe die Def. A.95 (iii)), also solchen mit einem uneigentlichen Grenzwert, erlaubt.

Es gibt aber keine allgemeine Entsprechung eines 'Kommutativ-' oder eines 'Assoziativgesetzes' für Reihen. Um zu sehen, unter welchen Bedingungen Umordnungen erlaubt sind, werden im Folgenden ein paar diesbezügliche Definitionen und Resultate angeführt.

Definition A.95 (Absolute bzw. unbedingte Konvergenz und bestimmte Divergenz von Reihen). Eine Reihe $\sum_{n \geq 0} a_n$ heißt

- (i) *absolut konvergent*, wenn die Reihe $\sum_{n \geq 0} |a_n|$ konvergiert.
- (ii) *unbedingt konvergent*, wenn sie konvergent ist und jede *Umordnung* wieder, und zwar wieder gegen den selben Wert, konvergiert (vgl. z.B. [22, S. 197]). Eine Umordnung der Reihe ist dabei der Gestalt $\sum_{n \geq 0} a_{i_n}$, wobei $n \mapsto i_n$ eine Bijektion von $\mathbb{N}_0$ auf $\mathbb{N}_0$ ist. Konvergiert die Reihe zwar, ist aber nicht unbedingt konvergent, so heißt sie *bedingt konvergent*.
- (iii) *bestimmt divergent gegen $+\infty$* , wenn es für alle $G \in \mathbb{R}$ ein $n_0 \geq 0$ gibt, sodass $\sum_{m=0}^n a_m > G$ für alle $n \geq n_0$ gilt. Sie heißt *bestimmt divergent gegen $-\infty$* , wenn $-\sum_{n \geq 0} a_n$ bestimmt gegen $+\infty$ divergiert (vgl. z.B. [13, §4, S. 40]). Man schreibt $\sum_{n \geq 0} a_n = \pm\infty$ und sagt auch, dass die Reihe *uneigentlich konvergent* ist. Eine Reihe die zwar divergiert, aber nicht bestimmt divergiert, nennt man *unbestimmt divergent*.

Satz A.96 (vgl. [22, Satz 31.4, S. 192, bzw. Satz 32.3, S. 197]; ohne Beweis)

- (i) Aus der absoluten Konvergenz einer Reihe $\sum_{n \geq 0} a_n$ folgt ihre Konvergenz und es gilt für sie die sogenannte allgemeine Dreiecksungleichung $|\sum_{n \geq 0} a_n| \leq \sum_{n \geq 0} |a_n|$.
- (ii) **Umordnungssatz:** Absolut konvergente Reihen – und nur diese – sind auch unbedingt konvergent. D.h., dass (nur) bei absolut konvergenten Reihen die Reihenfolge der Summation uneingeschränkt keine Rolle spielt.

Insbesondere sind Reihen mit *nichtnegativen* Gliedern, falls sie konvergieren, klarerweise immer auch absolut konvergent. Dass für solche Reihen Umordnungen auch bei Divergenz erlaubt sind, besagt das nächste Lemma.

Satz A.97 (Konvergenz bzw. Divergenz von Reihen mit nichtnegativen Gliedern – Umordnungssatz)

Sei $\sum_{n \geq 0} a_n$ eine Reihe mit nichtnegativen Gliedern, d.h. $a_n \geq 0$ für alle $n \geq 0$. Dann ist sie entweder absolut konvergent oder sie divergiert bestimmt gegen $+\infty$ (im Sinne der Def. A.95 (iii) konvergiert die Reihe also immer zumindest uneigentlich).

In beiden Fällen ist die Reihe invariant gegenüber Umordnungen – d.h., bei nichtnegativen Reihengliedern spielt die Summationsreihenfolge keine Rolle.

Beweis. Aufgrund ihrer Nichtnegativität gilt $a_n = |a_n|$ für alle Glieder. Deshalb ist die Reihe bei Konvergenz auch absolut konvergent und die zweite Aussage ist in diesem Fall eine Konsequenz aus (ii) des vorherigen Satzes A.96.

Sei die Reihe also nun divergent. Dann kann sie, erneut da ihre Glieder nichtnegativ sind, nur über alle Schranken wachsen und muss bestimmt gegen $+\infty$ divergieren. Angenommen, es würde eine konvergente Umordnung dieser Reihe geben. Dann wäre diese Konvergenz auch absolut (siehe oben), was wiederum wegen des Satzes A.96 (ii) auch für jede weitere Umordnung gelten müsste. Dies stellt aber einen Widerspruch zur Divergenz der Ausgangsreihe dar, weshalb jede Umordnung bestimmt divergieren muss. Ein etwas formaleres Argument, welches sich der Definition A.95 (iii) der bestimmten Divergenz bedient, findet sich in der NR B.14. $\square$

Gemäß dem Satz A.96 (ii) sind Umordnungen also bei absoluter Konvergenz erlaubt. Im spezifischeren Fall von nichtnegativen Reihengliedern sind diese laut dem Satz A.97 sogar bei (in diesem Fall bestimmter) Divergenz durchführbar.

Faltung von Folgen und Cauchy-Produkt von Reihen

Zunächst wird allgemein der Begriff der diskreten Faltung von Folgen eingeführt und dieser anschließend gleich verwendet, um eine Aussage über das Produkt von Reihen zu tätigen.

Definition A.98 (Faltung von Folgen)

Seien $a = \{a_n\}_{n \geq 0}$ und $b = \{b_n\}_{n \geq 0}$ reellwertige Folgen.

Dann wird die Folge $c := \{c_n\}_{n \geq 0}$, deren Glieder für alle $n \geq 0$ durch $c_n := \sum_{i=0}^n a_i b_{n-i} \left(= a_0 b_n + a_1 b_{n-1} + \dots + a_{n-1} b_1 + a_n b_0 = \sum_{i+j=n} a_i b_j \right)$ definiert sind, als (diskrete) *Faltung* von a und b bezeichnet und man schreibt $c = a * b$.

Satz A.99 (Cauchy-Produktformel – vgl. z.B. [13, Satz 3, S. 77] oder [22, Satz 32.6, S. 201]). *Seien $\sum_{n \geq 0} a_n$ und $\sum_{n \geq 0} b_n$ absolut konvergente Reihen und die Folge $\{c_n\}_{n \geq 0}$ bestehe für jedes $n \geq 0$ aus den Folgengliedern $c_n := \sum_{i=0}^n a_i b_{n-i}$ (also so wie in der vorangehenden Definition A.98).*

Dann ist die Reihe $\sum_{n \geq 0} c_n$ auch eine absolut konvergente Reihe und es gilt

$$\sum_{n \geq 0} c_n = \left(\sum_{n \geq 0} a_n \right) \left(\sum_{n \geq 0} b_n \right). \quad (\text{A.107})$$

Die entstandene Reihe heißt **Cauchy-Produkt** der beiden ursprünglichen Reihen.

Beweisskizze. Die Reihe $\sum_{n \geq 0} c_n$ ist ebenfalls absolut konvergent, was – sehr grob gesprochen – u.a. deshalb der Fall ist, da $|c_n| \leq \sum_{i=0}^n |a_i| |b_{n-i}|$ aufgrund der Dreiecksungleichung gilt (siehe den Satz A.96 (i)). Einen exakten Beweis dazu findet sich beispielsweise unter den oben angegebenen Referenzen. Es gilt

$$\begin{aligned} \sum_{n \geq 0} c_n &\stackrel{\text{Def. A.98}}{=} \sum_{n \geq 0} \left(\sum_{i=0}^n a_i b_{n-i} \right) = \overbrace{(a_0 b_0)}^{=c_0} + \overbrace{(a_0 b_1 + a_1 b_0)}^{=c_1} + \overbrace{(a_0 b_2 + a_1 b_1 + a_2 b_0)}^{=c_2} + \dots \\ &\quad + \underbrace{(a_0 b_n + a_1 b_{n-1} + \dots + a_k b_{n-k} + \dots + a_{n-1} b_1 + a_n b_0)}_{=c_n} + \dots \\ &\stackrel{(\star)}{=} (a_0 b_0 + a_0 b_1 + a_0 b_2 + \dots) \\ &\quad + (a_1 b_0 + a_1 b_1 + a_1 b_2 + \dots) + (a_2 b_0 + a_2 b_1 + a_2 b_2 + \dots) + \dots \\ &= a_0 \sum_{n \geq 0} b_n + a_1 \sum_{n \geq 0} b_n + a_2 \sum_{n \geq 0} b_n + \dots = \left(\sum_{n \geq 0} a_n \right) \left(\sum_{n \geq 0} b_n \right). \end{aligned}$$

($\star$) Insbesondere diese Umordnung ist – gemäß dem Satz A.96 (ii) – aufgrund der *absoluten* Konvergenz der Reihe erlaubt und wird nach dem sogenannten Cauchyschen Diagonalverfahren (siehe z.B. [22, S. 138]) durchgeführt. $\square$

Bemerkung zum Satz A.99. Allgemein wird eine Reihe, welche unabhängig von der Anordnung alle Produkte $a_i b_j$ (alle Kombinationen von $i \geq 0$ mit $j \geq 0$) aufsummiert,

als *Produktreihe* bezeichnet. Das Cauchy-Produkt ist nur eine mögliche Anordnung und somit eine spezielle Produktreihe. Man kann zeigen, dass, wenn die Reihen $\sum_{n \geq 0} a_n$ und $\sum_{n \geq 0} b_n$ absolut konvergent sind, jede ihrer Produktreihen absolut gegen das Produkt $(\sum_{n \geq 0} a_n)(\sum_{n \geq 0} b_n)$ konvergiert (vgl. z.B. [22, Satz 32.5, S. 200]). $\triangle$

Doppel- und Mehrfachreihen

Oft können Reihen so geschrieben werden, dass ihre Glieder von zwei oder mehreren unterschiedlichen Indizes abhängen. Im ersteren Fall bilden diese Glieder eine sogenannte *Doppelfolge*, beispielsweise $\{a_{j,k}\}_{j,k \geq 0}$. Ähnlich wie die 'normale' Konvergenz wird die der Doppelfolge, also $\lim_{j,k \rightarrow \infty} a_{j,k} = a$, aufgefasst. D.h, dann existiert für jedes $\varepsilon > 0$ ein $n_0 \geq 0$, sodass $|a_{j,k} - a| < \varepsilon$ für alle $j, k \geq n_0$ ist. Analoges gilt für Mehrfachfolgen – so ist für $k \geq 1$ etwa $\{a_{i_1, \dots, i_k}\}_{i_1, \dots, i_k \geq 0}$ eine k -fache Folge.

Definition A.100 (Doppelreihe und iterierte Reihe – vgl. z.B. [22, Kapitel 45, S. 256 ff.]

Sei $\{a_{j,k}\}_{j,k \geq 0}$ eine Doppelfolge. Der Ausdruck $\sum_{j,k \geq 0} a_{j,k} := \lim_{m,n \rightarrow \infty} \left(\sum_{j=0}^m \sum_{k=0}^n a_{j,k} \right)$ wird als *Doppelreihe* bezeichnet. Man spricht von einer konvergenten Doppelreihe, wenn ihr Grenzwert existiert. Existiert er auch für die Reihe mit den Gliedern $|a_{j,k}|$, so ist diese absolut konvergent.

Die Ausdrücke $\sum_{j \geq 0} \left(\sum_{k \geq 0} a_{j,k} \right)$ bzw. $\sum_{k \geq 0} \left(\sum_{j \geq 0} a_{j,k} \right)$ werden *iterierte Reihen* genannt.

Bemerkung zur Definition A.100. Man lässt, salopp formuliert, bei einer Doppelreihe also beide Indizes gewissermaßen 'gleich schnell' laufen (die Summen in der Klammer sind endlich und können somit beliebig vertauscht werden). Bei einer iterierten Reihe hingegen werden zunächst die inneren Summen bestimmt – jeweils bei fixem äußeren Laufindex – und diese anschließend aufsummiert. $\triangle$

Als Nächstes soll beantwortet werden, unter welchen Umständen die Reihenfolge der Summation keine Rolle spielt und die jeweiligen iterierten Reihen und die Doppelreihe äquivalent sind (siehe (A.108)).

Satz A.101 (Doppelreihensatz von Cauchy – vgl. z.B. [22, Satz 45.2, S. 258] oder [31, Doppelreihensatz, S. 71]; ohne Beweis). *Seien Doppelreihen und iterierte Reihen wie in der Definition A.100 gegeben. Wenn eine der beiden iterierten Reihen absolut konvergent ist, dann sind es die andere iterierte Reihe und auch die Doppelreihe selbst ebenso. Umgekehrt folgt aus der absoluten Konvergenz der Doppelreihe auch die beider iterierten Reihen und in beiden Fällen gilt*

$$\sum_{j,k \geq 0} a_{j,k} = \sum_{j \geq 0} \left(\sum_{k \geq 0} a_{j,k} \right) = \sum_{k \geq 0} \left(\sum_{j \geq 0} a_{j,k} \right). \quad (\text{A.108})$$

Bemerkungen zum Doppelreihensatz A.101 von Cauchy.

- (i) Etwa mit der in der Bem. zur Not. A.1 erwähnten Bijektion $\pi: \mathbb{N}_0^2 \rightarrow \mathbb{N}_0$, $\pi(j, k) := l$, können die Glieder $a_{j,k}$ zu abzählbar vielen $b_l = b_{\pi(j,k)} := a_{j,k}$ umgeschrieben werden (für $l, j, k \geq 0$). Im Fall der absoluten Konvergenz spielen gemäß dem Satz A.96 (ii) Umordnungen bei einfachen Reihen keine Rolle, weshalb sich dann die Doppelreihe als $\sum_{j,k \geq 0} a_{j,k} = \sum_{j,k \geq 0} b_{\pi(j,k)} = \sum_{l \geq 0} b_l$ darstellen lässt.
- (ii) **Nichtnegative Reihenglieder:** Analog zum Satz A.97 (und unter Berücksichtigung von (i)) begründbar, sind bei Doppelreihen mit nichtnegativen Gliedern nur absolute Konvergenz oder bestimmte Divergenz gegen $+\infty$ möglich. In beiden Fällen gilt (A.108) (im divergenten Fall in der Art, dass alle Ausdrücke $+\infty$ sind). Zusammengenommen sind Doppelreihen mit nichtnegativen Gliedern jedenfalls invariant gegenüber Umordnungen.
- (iii) **Mehrfachreihen:** Mit (i) und (ii) lässt sich argumentieren, dass der Doppelreihensatz auch auf Mehrfachreihen angewendet werden kann. Etwa für $k \geq 2$ liefert eine $(k-1)$ -fache Anwendung (vorausgesetzt die Reihenglieder sind nichtnegativ oder die iterierte Reihe bzw. die k -fache Reihe ist absolut konvergent)

$$\sum_{i_1, \dots, i_k \geq 0} a_{i_1, \dots, i_k} = \sum_{i_1 \geq 0} \left(\dots \left(\sum_{i_k \geq 0} a_{i_1, \dots, i_k} \right) \dots \right). \quad (\text{A.109})$$

Dabei ist die Reihenfolge der Indizes in der iterierten Reihe beliebig. $\triangle$

Potenzreihen

In der vorliegenden Arbeit kommen wiederholt sogenannte Potenzreihen zur Anwendung – insbesondere in Form von wahrscheinlichkeitserzeugenden Funktionen (siehe den Unterabschnitt 2.2.2). In diesem Teilabschnitt werden deshalb einige ihrer wesentlichen Eigenschaften angeführt. Abschließend wird eine Verallgemeinerung des Binomialkoeffizienten und ein Resultat zur Binomialreihe – einer speziellen Potenzreihe – angeführt.

Definition A.102 (Potenzreihe, Konvergenzradius)

Gegeben einer Folge $\{a_n\}_{n \geq 0}$, wobei alle $a_n \in \mathbb{R}$ sind, bezeichnet man die Reihe der Form $\sum_{n \geq 0} a_n x^n = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots$ als *Potenzreihe* (mit dem Mittelpunkt 0 – vgl. [22, (63.1), S. 362]). Der *Konvergenzradius* einer Potenzreihe sei durch

$$R := \frac{1}{\limsup_{n \rightarrow \infty} \sqrt[n]{|a_n|}} \quad \text{mit } \frac{1}{0} := \infty \text{ und } \frac{1}{\infty} := 0 \quad (\text{A.110})$$

definiert (vgl. [22, (63.3), S. 363]).

Satz A.103 (Eigenschaften von Potenzreihen; ohne Beweis)

Eine Potenzreihe bzw. deren Summenfunktion $f : \mathbb{R} \rightarrow \mathbb{R}$ mit $f(x) := \sum_{n \geq 0} a_n x^n$ hat die folgenden Eigenschaften (Auswahl):

(a) **Konvergenzsatz und -intervall** (vgl. [22, Satz 63.1, S. 362]):

- Falls $R = 0$ ist, konvergiert sie nur für $x = 0$.
- Falls $R > 0$ ist, ist sie absolut konvergent für $x \in (-R, R)$ und divergent für $|x| > R$. Für die Randpunkte $x = \pm R$ ist keine allgemeine Aussage über das Konvergenzverhalten möglich.

Das Intervall $(-R, R) = \{x : |x| < R\}$, also das, auf dem die Potenzreihe jedenfalls konvergiert, heißt Konvergenzintervall.

Die beiden folgenden Sätze (b) und (c) gelten für Konvergenzradien $R \neq 0$, für (d) und (e) wird zusätzlich $R < \infty$ gefordert.

(b) **Stetigkeitssatz** (vgl. [22, Satz 64.1, S. 368]): Die Summe f der Potenzreihe ist auf ihrem gesamten Konvergenzintervall $(-R, R)$ stetig.

(c) **Differenzierbarkeitssatz** (vgl. [22, Satz 64.2, S. 368]): Die Summe f der Potenzreihe ist auf dem gesamten Konvergenzintervall beliebig oft differenzierbar und ihre Ableitungen können durch gliedweise Differentiation erhalten werden. Es gilt somit für jedes $x \in (-R, R)$

- $f'(x) = \sum_{n \geq 1} n a_n x^{n-1}$,
- $f''(x) = \sum_{n \geq 2} n(n-1) a_n x^{n-2}$ und allgemein für $k \geq 1$
- $f^{(k)}(x) = \sum_{n \geq k} n(n-1) \cdots (n-k+1) a_n x^{n-k} = \sum_{n \geq k} \frac{n!}{(n-k)!} a_n x^{n-k}$.

Der Konvergenzradius ändert sich dabei nicht und bleibt für alle Ableitungen gleich dem der ursprünglichen Potenzreihe. Da aus der Differenzierbarkeit einer Funktion deren Stetigkeit folgt, sind f und alle ihre Ableitungen auf dem Konvergenzintervall $(-R, R)$ stetig und f kann dort beliebig oft stetig differenzierbar genannt werden.

(d) **Grenzwertsatz von Abel** (vgl. [22, Satz 65.1, S. 379]): Konvergiert die Potenzreihe im rechten Endpunkt R ihres Konvergenzintervalls, so ist ihre Summenfunktion im Punkt R linksseitig stetig. Es gilt also $\lim_{x \nearrow R} f(x) = \sum_{n \geq 0} a_n R^n$.

(e) **Differenzierbarkeit im rechten Endpunkt R** (vgl. [22, A 65.5, S. 384]): Konvergiert die Potenzreihe in R und zudem auch die Reihe $\sum_{n \geq 1} n a_n R^{n-1}$, so ist ihre Summenfunktion im Punkt R (linksseitig) differenzierbar, wobei ihre Ableitung dort $f'(R) = \sum_{n \geq 1} n a_n R^{n-1}$ ist.

Korollar zum Satz A.103 (d) (Ergänzung zum Grenzwertsatz von Abel für nichtnegative Koeffizienten – vgl. [22, A 65.7, S. 385]). Eine Potenzreihe $\sum_{n \geq 0} a_n x^n$ mit Konvergenzradius $0 < R < \infty$ und nichtnegativen Koeffizienten a_n divergiere im rechten Endpunkt R ihres Konvergenzintervalls, d.h., sie divergiert (bestimmt) gegen $+\infty$ bzw. $\sum_{n \geq 0} a_n R^n = +\infty$.

Dann divergiert die Reihe auch für $x \nearrow R$. Es gilt also $\lim_{x \nearrow R} \sum_{n \geq 0} a_n x^n = +\infty$.

Beweis (angelehnt an [22, S. 616]). Sei $G > 0$ beliebig. Da die Reihe im Punkt R bestimmt divergiert, gibt es ein $n_0 \geq 0$ mit $\lim_{x \nearrow R} \sum_{n=0}^{n_0} a_n x^n \stackrel{\text{endl. Summe}}{=} \sum_{n=0}^{n_0} a_n R^n > G + 1$. Daher gibt es auch ein $\delta > 0$, sodass $\sum_{n=0}^{n_0} a_n x^n > G$ für alle $x \in (R - \delta, R)$ gilt – d.h., für alle x die nur nahe genug bei R (linksseitig) sind. Für diese x gilt dann (aufgrund der nicht-negativen Koeffizienten) erst recht $\sum_{n \geq 0} a_n x^n \geq \sum_{n=0}^{n_0} a_n x^n > G$. Da G beliebig groß gewählt werden kann, divergiert die Reihe also, wenn x von unten gegen R strebt. $\square$

Bemerkung zum Satz A.103 (e). Zumindest für $R = 1$ und *nichtnegative* Koeffizienten a_n (für alle $n \geq 0$) gilt diese Aussage auch bzgl. höherer Ableitungen. Konvergiert also die Reihe $\sum_{n \geq k} n(n-1) \cdots (n-k+1) a_n$, so ist f auch im Punkt 1 k -mal (linksseitig) differenzierbar und $f^{(k)}(1)$ ist genau diese Reihe. Dies soll kurz für $k = 2$ skizziert werden, für höhere $k > 2$ folgt die Behauptung iterativ.

Sei also $\infty > \sum_{n \geq 2} n(n-1) a_n \geq \sum_{n \geq 2} n a_n \geq \sum_{n \geq 2} a_n$. Da sowohl a_0 als auch a_1 endlich sind, konvergieren demnach auch die Reihen $\sum_{n \geq 1} n a_n$ und $\sum_{n \geq 0} a_n = f(1)$. Laut (e) gilt deshalb für die erste Ableitung $f'(1) = \sum_{n \geq 1} n a_n$. Mit $m = n - 1$ und $b_m = a_{m+1}(m+1)$ können die Reihen zu $\sum_{n \geq 2} n(n-1) a_n = \sum_{m \geq 1} m b_m < \infty$ und $\sum_{n \geq 1} n a_n = \sum_{m \geq 0} b_m < \infty$ umgeschrieben werden. Die erneute Anwendung von (e) auf die von $\{b_m\}_{m \geq 0}$ erzeugte Potenzreihe, also $\sum_{m \geq 0} b_m x^m (= f'(x) \text{ auf } [0, 1])$, führt dann zu $f''(1) = \sum_{m \geq 1} m b_m 1^{m-1} = \sum_{m \geq 1} m(m+1) a_{m+1} = \sum_{n \geq 2} n(n-1) a_n$. $\triangle$

Definition A.104 (Binomialkoeffizient – Verallgemeinerung – vgl. z.B. [22, S. 55])
Definiere für $\alpha \in \mathbb{R}$ und $n \in \mathbb{N}_0$ den *verallgemeinerten Binomialkoeffizienten* als

$$\binom{\alpha}{n} := \begin{cases} \frac{\alpha(\alpha-1)(\alpha-2) \cdots (\alpha-n+1)}{n!} & \text{für } n \geq 1, \\ 1 & \text{für } n = 0. \end{cases} \quad (\text{A.111})$$

Satz A.105 (Binomialreihe – vgl. z.B. [22, (62.14) S. 360 bzw. 3. S. 370]; ohne Beweis)
Sei $\alpha \in \mathbb{R}$ und $-1 < x < 1$. Dann konvergiert die sogenannte *Binomialreihe* wie folgt:

$$\sum_{n \geq 0} \binom{\alpha}{n} x^n = (1+x)^\alpha. \quad (\text{A.112})$$

Falls $\alpha > -1$ ist, gilt (A.112) auch für $x = 1$ und falls $\alpha > 0$ ist, auch für $x = -1$ (vgl. [22, (65.9), S. 383]). Falls $\alpha \in \mathbb{N}_0$ ist, gilt (A.112) sogar für alle $x \in \mathbb{R}$ (da dann $\binom{\alpha}{n} = 0$ für $n > \alpha$ gilt, handelt es sich diesfalls um eine endliche Summe und die Formel folgt direkt aus dem binomischen Lehrsatz).

A.3.3 Doppelfakultät und Catalan-Zahlen

Es wird noch der Begriff der *Doppelfakultät* und eine spezielle Zahlenfolge – die der *Catalan-Zahlen* – eingeführt. Beides kommt bzgl. der Gesamtpopulation eines einfachen geometrischen GWP zur Anwendung (siehe den Abschnitt 4.1.4). Beachte, dass in diesem Teilabschnitt die Darstellungen oft nur für hinreichend große n sinnvoll sind.

Notation A.106 (Doppelfakultät – vgl. z.B. [16])

Sei $n \in \{-1\} \cup \mathbb{N}_0$. Die *Doppelfakultät* wird definiert als

$$n!! := \begin{cases} n \cdot (n-2) \cdot (n-4) \cdots 3 \cdot 1 & \text{für ungerade } n > 0, \\ n \cdot (n-2) \cdot (n-4) \cdots 4 \cdot 2 & \text{für gerade } n > 0, \\ 1 & \text{für } n = 0 \text{ und } n = -1. \end{cases} \quad (\text{A.113})$$

Somit gilt bspw. $9!! = 9 \cdot 7 \cdot 5 \cdot 3 \cdot 1 = 945$. Ungerade $n \geq -1$ lassen sich immer als $n = 2k - 1$ mit $k \in \mathbb{N}_0$ darstellen. Somit gilt für ungerade $n \geq -1$ und $k = \frac{n+1}{2}$ ($0! = 1$)

$$\begin{aligned} n!! &= (2k-1)!! \stackrel{(\text{A.113})}{=} (2k-1) \cdot (2k-3) \cdots 3 \cdot 1 \cdot \frac{2k \cdot (2k-2) \cdots 4 \cdot 2}{2k \cdot (2k-2) \cdots 4 \cdot 2} \\ &= \frac{2k \cdot (2k-1) \cdot (2k-2) \cdot (2k-3) \cdots 4 \cdot 3 \cdot 2 \cdot 1}{2(k) \cdot 2(k-1) \cdots 2 \cdot (2) \cdot 2 \cdot (1)} = \frac{(2k)!}{2^k k!}. \quad \triangle \end{aligned} \quad (\text{A.114})$$

Definition A.107 (Catalan-Zahlen – vgl. z.B. [33, S. 106])

Die Folge $\{c_n\}_{n \geq 0}$ der sogenannten *Catalan-Zahlen* ist definiert als

$$c_n = \frac{1}{n+1} \binom{2n}{n} \quad \text{für alle } n \geq 0, \quad (\text{A.115})$$

d.h. $\{c_n\}_{n \geq 0} = 1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786, \dots$

Catalan-Zahlen sind immer ganzzahlig, denn für jedes $n \geq 0$ ist

$$\begin{aligned} c_n &= \binom{2n}{n} \frac{1}{n+1} = \binom{2n}{n} \frac{1+n-n}{n+1} = \binom{2n}{n} \frac{n+1}{n+1} - \frac{(2n)!}{n!n!} \frac{n}{n+1} \\ &= \binom{2n}{n} - \frac{(2n)!}{(n-1)!(n+1)!} = \binom{2n}{n} - \binom{2n}{n-1} \in \mathbb{N}. \end{aligned}$$

Lemma A.108 Für die Glieder der Catalan-Zahlen-Folge gilt

$$c_n \stackrel{(\text{A.115})}{=} \frac{1}{n+1} \binom{2n}{n} = (-1)^n 2^{2n+1} \binom{\frac{1}{2}}{n+1} \quad \text{für alle } n \geq 0. \quad (\text{A.116})$$

Beweis.

$$\begin{aligned} \binom{\frac{1}{2}}{n+1} &\stackrel{(\text{A.111})}{=} \frac{\overbrace{\frac{1}{2}}^{\frac{1}{2}} \cdot \overbrace{\left(\frac{1}{2}-1\right)}^{-\frac{1}{2}} \cdot \overbrace{\left(\frac{1}{2}-2\right)}^{-\frac{3}{2}} \cdot \overbrace{\left(\frac{1}{2}-3\right)}^{-\frac{5}{2}} \cdots \overbrace{\left(\frac{1}{2}-(n+1)+1\right)}^{\frac{1}{2}-n=-\frac{1}{2}(2n-1)}}{(n+1)!} \quad (n+1 \text{ Faktoren}) \\ &= \frac{1}{2^{n+1}(n+1)!} \cdot 1 \cdot \overbrace{(-1) \cdot (-3) \cdot (-5) \cdots (-(2n-1))}^{(n \text{ Faktoren})} \\ &\stackrel{(\text{A.113})}{=} \frac{(-1)^n}{2^{n+1}(n+1)!} \cdot (2n-1)!! \stackrel{(\text{A.114})}{=} \frac{(-1)^n}{2^{n+1}(n+1)!} \cdot \frac{(2n)!}{2^n n!} \\ &= \frac{(-1)^n}{2^{2n+1}} \cdot \frac{1}{n+1} \cdot \frac{(2n)!}{n!n!} = \frac{1}{(-1)^n 2^{2n+1}} \cdot \frac{1}{n+1} \cdot \binom{2n}{n}. \quad \square \end{aligned}$$

B Appendix

B.1 Nebenrechnungen

Zu den Nebenrechnungen wird das Setting – also die Voraussetzungen und Definitionen – nicht immer extra mit angegeben, sondern diese sind so wie an der Stelle, an welchen auf die Nebenrechnung (kurz NR) verwiesen wird. Gleichungen, welche nur innerhalb einer Nebenrechnung benötigt werden, werden dort geschlossen durchnummeriert.

Nebenrechnung B.1 (EW einer $\mathbb{N}_0$ -wertigen ZV – für (2.5)). Ist eine ZV X $\mathbb{N}_0$ -wertig, so kann ihr Erwartungswert auch durch

$$\begin{aligned} \mathbb{E}X &\stackrel{(2.4)}{=} \sum_{n \geq 0} n \mathbb{P}(X = n) \stackrel{(\star)}{=} \sum_{n \geq 0} n (\mathbb{P}(X > n-1) - \mathbb{P}(X > n)) \\ &= \sum_{n \geq 0} n \mathbb{P}(X > n-1) - \sum_{n \geq 0} n \mathbb{P}(X > n) \\ &\stackrel{(\star\star)}{=} \sum_{n \geq 0} (n+1) \mathbb{P}(X > n) - \sum_{n \geq 0} n \mathbb{P}(X > n) = \sum_{n \geq 0} \mathbb{P}(X > n) \end{aligned}$$

berechnet werden (die Operation in der zweiten Zeile und auch die allerletzte sind dabei aufgrund der Nichtnegativität der Reihenglieder erlaubt – vgl. etwa A.3.2).

($\star$) $\{X = n\} = \{X > n-1\} \cap \{X \leq n\} = \{X \leq n-1\}^c \cap \{X > n\}^c$, woraus mit deMorgan $\{X = n\} = (\{X \leq n-1\} \cup \{X > n\})^c$ folgt, wobei die beiden Mengen in der Klammer disjunkt sind. Deshalb gilt (wegen der endl. Add. und (A.29))
 $\mathbb{P}(X = n) = 1 - (\underbrace{\mathbb{P}(X \leq n-1)}_{=1 - \mathbb{P}(X > n-1)} + \mathbb{P}(X > n)) = \mathbb{P}(X > n-1) - \mathbb{P}(X > n).$

$$(\star\star) \sum_{n \geq 0} n \mathbb{P}(X > n-1) \stackrel{n_{\text{neu}} = n_{\text{alt}} - 1}{=} \sum_{n \geq -1} \underbrace{(n+1) \mathbb{P}(X > n)}_{=0 \text{ für } n=-1} = \sum_{n \geq 0} (n+1) \mathbb{P}(X > n).$$

Nebenrechnung B.2 ($\mathbb{E}[h(S_N)|N = n] = \mathbb{E}h(S_n)$). Es soll (im Setting der Definition 2.11 bzw. des Lemmas 2.13) die Gleichung (2.25) auf eine alternative Weise gezeigt werden. Der Ansatz ähnelt dem im Beweis des Lemmas – es muss hier aber nicht unbedingt vorausgesetzt werden, dass die $\{X_i\}_{i \geq 1}$ $\mathbb{N}_0$ -wertig sind. Die NR gilt auch allgemeiner, wenn sie *i.i.d.*-Kopien einer reellen Zufallsvariablen X sind. Zudem kann auf die Nichtnegativität von h verzichtet werden, sofern $h(S_N)$ integrierbar ist.

Sei also wieder $n \geq 0$ so, dass $\mathbb{P}(N = n) > 0$ ist. Angenommen, $h : \mathbb{R} \rightarrow \mathbb{R}$ ist nichtnegativ oder so, dass $h(S_N)$ integrierbar ist. Es sind $\{X_i\}_{i \geq 1}$ und N unabhängig, deshalb sind auch $S_n = \sum_{i \geq 1} X_i$ und $\mathbb{1}_{\{N=n\}}$ unabhängig, woraus weiters die Unabhängigkeit

B Appendix

von $h(S_n)$ und $\mathbb{1}_{\{N=n\}}$ folgt (siehe den Unterabschnitt A.1.3 über die Unabhängigkeit). Weiters gilt

$$\sum_{n \geq 0} h(S_n) \underbrace{\mathbb{1}_{\{N=n\}}}_{=0, \text{ wenn } n \neq N \text{ ist}} = h(S_N) \quad \text{und deshalb ist} \quad (1)$$

$$\begin{aligned} \mathbb{E}[h(S_N) \mathbb{1}_{\{N=n\}}] &\stackrel{(1)}{=} \mathbb{E} \left[\left(\sum_{m \geq 0} h(S_m) \mathbb{1}_{\{N=m\}} \right) \mathbb{1}_{\{N=n\}} \right] \\ &\stackrel{(A.3)}{=} \mathbb{E} \left[\sum_{m \geq 0} h(S_m) \underbrace{\mathbb{1}_{\{N=m\} \cap \{N=n\}}}_{=0, \text{ wenn } m \neq n \text{ ist}} \right] = \mathbb{E}[h(S_n) \mathbb{1}_{\{N=n\}}] \\ &\stackrel{h(S_n), \mathbb{1}_{\{N=n\}}}{=} \mathbb{E}h(S_n) \mathbb{E} \mathbb{1}_{\{N=n\}} \stackrel{\text{uabh. (A.66) } (\star)}{=} \mathbb{E}h(S_n) \mathbb{P}(N=n) \stackrel{(A.44)}{=} \mathbb{E}h(S_n) \mathbb{P}(N=n). \end{aligned} \quad (2)$$

Daraus folgt (mit (A.80) aus der Definition A.66)

$$\mathbb{E}[h(S_N) | N=n] \stackrel{(A.80)}{=} \frac{\mathbb{E}[h(S_N) \mathbb{1}_{\{N=n\}}]}{\mathbb{P}(N=n)} \stackrel{(2)}{=} \frac{\mathbb{E}h(S_n) \mathbb{P}(N=n)}{\mathbb{P}(N=n)} = \mathbb{E}h(S_n).$$

($\star$) Für den in diesem Schritt angewendeten Produktsatz A.52 wird genau genommen die Integrierbarkeit von $h(S_n)$ vorausgesetzt. Sofern diese nicht gegeben ist, ist h hier aber zumindest nichtnegativ, somit $h(S_n)$ quasiintegrierbar und dann sind beide Seiten $+\infty$ (da $\mathbb{1}_{\{N=n\}} > 0$ ist; für die diesbzgl. analog durchführbare Rechnung in der Beweis(-skizze) des Produktsatzes beachte auch die Bem. zum Satz A.101 (ii)).

Nebenrechnung B.3 ($f'(1^-) > 1 \Rightarrow f(s) < s$ nahe bei 1). Es soll ein analytisches Argument für die Aussage des Lemmas 2.16 (iii) formuliert werden.

Es gelte für die weF f , dass $f'(1^-) > 1$ ist. Laut dem Punkt (i) des selben Lemmas ist somit $p_0 + p_1 < 1$ und deshalb f strikt konvex sowie f' strikt isoton.

Somit ist auch $f'(0) \stackrel{(2.10)}{=} p_1 < 1$ und zudem gilt

- falls $f'(1^-) < \infty$ ist, $f'(1) = f'(1^-) > 1$ und dass f' stetig auf ganz $[0, 1]$ ist (siehe das Lemma 2.7 bzw. die Bem. dazu). Deshalb garantiert der Zwischenwertsatz A.91, dass es ein $t_0 \in (0, 1)$ mit $f'(t_0) = 1$ gibt.
- falls $f'(1^-) = \infty$ ist, dass es für alle $G \in \mathbb{R}$ ein $\delta > 0$ gibt, sodass $f'(1 - \delta) > G$ ist. Also gibt es insbesondere auch ein $\delta > 0$, sodass $f'(1 - \delta) > 1$ ist. Wie zuvor liefert demnach der Zwischenwertsatz, dass es ein $t_0 \in (0, 1 - \delta)$ mit $f'(t_0) = 1$ gibt (f' ist auf $[0, 1 - \delta] \subset [0, 1)$ stetig – Satz 2.5 (b)).

Da f' strikt isoton ist, gilt in beiden Fällen $f'(s) > f'(t_0) = 1$ für alle $s \in (t_0, 1)$ ($\star$).

Sei nun $s \in (t_0, 1)$. Die weF f ist auf $[0, 1]$ stetig, also auch auf $[s, 1]$. Zudem ist sie zumindest auf $[s, 1)$ differenzierbar. Deshalb besagt der Mittelwertsatz A.92, dass es ein $c \in (s, 1) \subset (t_0, 1)$ gibt mit $1 < f'(c) \stackrel{(\star)}{=} \frac{f(1) - f(s)}{1 - s} \stackrel{(A.106)}{=} \frac{1 - f(s)}{1 - s} \stackrel{(2.8)}{=} \frac{1 - f(s)}{1 - s} \Rightarrow 1 - s < 1 - f(s),$

woraus wiederum $f(s) < s$ folgt. Falls $f'(1^-) > 1$ ist, kann somit ausgesagt werden, dass es immer ein $t_0 \in (0, 1)$ gibt, sodass $f(s) < s$ für alle $s \in (t_0, 1)$ gilt – also solange s nur hinreichend nahe bei 1 ist.

Nebenrechnung B.4 ($\mathbb{P}(\limsup_{n \rightarrow \infty} \{Z_n = k\}) = 0 \Rightarrow \mathbb{P}(\lim_{n \rightarrow \infty} Z_n = k) = 0$).

Die oben stehende Implikation soll für die letzte Aussage des Satzes 2.26 gezeigt werden.

Beachte zunächst (siehe (A.7)), dass $\limsup_{n \rightarrow \infty} \{Z_n = k\} \supseteq \liminf_{n \rightarrow \infty} \{Z_n = k\}$ und deshalb auch $0 = \mathbb{P}(\limsup_{n \rightarrow \infty} \{Z_n = k\}) \geq \mathbb{P}(\liminf_{n \rightarrow \infty} \{Z_n = k\})$ gilt. Dementsprechend muss auch $\mathbb{P}(\liminf_{n \rightarrow \infty} \{Z_n = k\}) = 0$ sein. Nun gilt

$$\begin{aligned}
 \omega \in \liminf_{n \rightarrow \infty} \{Z_n = k\} &\stackrel{(A.4)}{\iff} \omega \in \bigcup_{n \geq 0} \bigcap_{m \geq n} \{Z_m = k\} \iff \exists n \geq 0 : \omega \in \bigcap_{m \geq n} \{Z_m = k\} \\
 &\iff \exists n \geq 0 : Z_m(\omega) = k \quad \forall m \geq n \iff \exists n \geq 0 : Z_m(\omega) - k = 0 \quad \forall m \geq n \\
 &\stackrel{(\star)}{\iff} \forall \varepsilon > 0 \text{ gilt } \exists \tilde{n}_\varepsilon \geq 0 : |Z_n(\omega) - k| < \varepsilon \quad \forall n \geq \tilde{n}_\varepsilon \iff \lim_{n \rightarrow \infty} Z_n(\omega) = k \\
 &\iff \omega \in \left\{ \lim_{n \rightarrow \infty} Z_n = k \right\},
 \end{aligned} \tag{B.1}$$

wobei im vorletzten Schritt die Definition des Grenzwertes der Folge $\{Z_n(\omega)\}_{n \geq 0}$ verwendet wurde. Daraus folgt $\mathbb{P}(\lim_{n \rightarrow \infty} Z_n = k) = \mathbb{P}(\liminf_{n \rightarrow \infty} \{Z_n = k\}) = 0$.

($\star$) Die ' $\Rightarrow$ '-Implikation dieser Äquivalenz ist allgemein klar. Die umgekehrte Richtung ' $\Leftarrow$ ' gilt, da die Z_n (und auch k) ganzzahlig sind (wähle z.B. $\varepsilon = 1/2$).

Nebenrechnung B.5 ($f_{Z_{n+1}} = f_{Z_n} \circ f$). Wie in der Bem. zum Korollar 2.31 (ii) angekündigt, soll der Induktionsschritt im Beweis zum Satz 2.28 alternativ mithilfe der *Chapman-Kolmogorov'schen Gleichung* (CKG – vgl. den Satz A.74 (ii)) und angelehnt an [3, S. 2] durchgeführt werden.

Seien dafür $s \in [0, 1]$ und $n, m \geq 0$. Betrachte die höheren Übergangs-WK, wie sie im Satz A.74 (i) bzw. (A.89) formuliert wurden, d.h. $p_{ij}^{(n)} = \mathbb{P}(Z_{n+m} = j | Z_n = i)$. Insbesondere gilt $p_{1k}^{(n)} = \mathbb{P}(Z_n = k | Z_0 = 1) \stackrel{Z_0 \equiv 1}{=} \mathbb{P}(Z_n = k)$ für $k \geq 1$, woraus folgt, dass $f_{Z_n}(s) = \sum_{k \geq 0} \mathbb{P}(Z_n = k) s^k = \sum_{k \geq 0} p_{1k}^{(n)} s^k$ ist ($\star$). Weiters gilt – wegen des zweiten Teils von (2.52) – $\sum_{k \geq 0} p_{lk} s^k = (f(s))^l$ ($\star\star$). Mit der CKG erhält man

$$\begin{aligned}
 f_{Z_{n+1}}(s) &\stackrel{(\star)}{=} \sum_{k \geq 0} p_{1k}^{(n+1)} s^k \stackrel{\text{CKG (A.90)}}{=} \sum_{k \geq 0} \left(\sum_{l \geq 0} p_{1l}^{(n)} p_{lk} \right) s^k \stackrel{\text{abs. konv.}}{=} \sum_{l \geq 0} p_{1l}^{(n)} \underbrace{\left(\sum_{k \geq 0} p_{lk} s^k \right)}_{=(f(s))^l \text{ } (\star\star)} \\
 &= \sum_{l \geq 0} p_{1l}^{(n)} (f(s))^l \stackrel{(\star)}{=} f_{Z_n}(f(s)).
 \end{aligned}$$

Nebenrechnung B.6 (Beweis von (2.57) aus dem Satz 2.33 mit weF). Der Varianzbeweis für $\mathbb{V}Z_n$ des Satzes 2.33 soll alternativ mit den weF durchgeführt werden.

Laut Voraussetzung sind $\mathbb{E}X = \mu < \infty$ sowie $\mathbb{V}X = \sigma^2 < \infty$ ($\iff \mathbb{E}X^2 < \infty$). Weiters wurde die erste Ableitung von f_n für $n \geq 1$ und $s \in [0, 1)$ bereits vor dem Satz als $f'_n(s) = f'(f_{n-1}(s))f'_{n-1}(s)$ bestimmt. Erneutes Ableiten führt zu $f''_n(s) = f''(f_{n-1}(s))[f'_{n-1}(s)]^2 + f'(f_{n-1}(s))f''_{n-1}(s)$.

Mit $f_{n-1}(1^-) = f_{n-1}(1) = 1$, der Stetigkeit aller vorkommenden Funktionen zumindest auf $[0, 1)$ und $f''(1^-) = f''(1) < \infty$ folgt (da $\mathbb{E}X^2 < \infty$ ist – siehe das Lemma 2.7)

$$f''_n(1^-) = f''(1)[f'_{n-1}(1^-)]^2 + f'(1^-)f''_{n-1}(1^-) \stackrel{(2.53)}{\stackrel{(2.55)}}{=} f''(1)(\mu^{n-1})^2 + \mu f''_{n-1}(1^-).$$

B Appendix

Da $f_0''(1^-) = 0$ ist, folgt daraus weiters induktiv, dass $f_n''(1^-) < \infty$ für alle $n \geq 1$ gilt. Demnach kann immer $f_n''(1)$ anstelle von $f_n''(1^-)$ geschrieben werden, also

$$f_n''(1) = f''(1)\mu^{2n-2} + \mu f_{n-1}''(1). \quad (1)$$

Die Anfangswerte dieser Rekursion sind bereits bekannt ($f_0''(1) = 0, f_1''(1) = f''(1)$). Nun sollen ein paar weitere Werte bestimmt werden, um eine Gesetzmäßigkeit zu finden.

$$\begin{aligned} f_2''(1) &= f''(1)\mu^2 + \mu f''(1) = f''(1)(\mu^2 + \mu^1) \\ f_3''(1) &= f''(1)\mu^4 + \mu [f''(1)(\mu^2 + \mu^1)] = f''(1)(\mu^4 + \mu^3 + \mu^2) \\ f_4''(1) &= f''(1)\mu^6 + \mu [f''(1)(\mu^4 + \mu^3 + \mu^2)] = f''(1)(\mu^6 + \mu^5 + \mu^4 + \mu^3) \\ &\vdots \quad \Rightarrow \text{Vermutung: } f_n''(1) = f''(1) \sum_{i=n-1}^{2n-2} \mu^i \quad \text{für } n \geq 1. \end{aligned} \quad (2)$$

Mit $\mu f_{n-1}''(1) \stackrel{(2)}{=} \mu f''(1) \sum_{i=n-2}^{2n-4} \mu^i = f''(1) \sum_{i=n-2}^{2n-4} \mu^{i+1} = f''(1) \sum_{i=n-1}^{2n-3} \mu^i$ (*) wird durch

$$f_n''(1) \stackrel{(1)}{=} f''(1)\mu^{2n-2} + \mu f_{n-1}''(1) \stackrel{(*)}{=} f''(1)\mu^{2n-2} + f''(1) \sum_{i=n-1}^{2n-3} \mu^i = f''(1) \sum_{i=n-1}^{2n-2} \mu^i$$

die aufgestellte Vermutung für $n \geq 1$ induktiv bestätigt (wenn sie für $n-1$ korrekt ist, so gilt das auch für n). Die Summe in (2) kann nun mit einer Indexverschiebung umgeschrieben werden zu $\sum_{i=n-1}^{2n-2} \mu^i = \sum_{i=0}^{n-1} \mu^{i+(n-1)} = \mu^{n-1} \sum_{i=0}^{n-1} \mu^i$, wobei $\sum_{i=0}^{n-1} \mu^i$ die $(n-1)$ -te Partialsumme einer geometrischen Reihe ist. Diese wird zu $\frac{1-\mu^n}{1-\mu}$ für $\mu \neq 1$ bzw. zu $n\mu$ für $\mu = 1$. Zusammenfassend gilt also

$$f_n''(1) = f''(1)\mu^{n-1} \sum_{i=0}^{n-1} \mu^i = \begin{cases} f''(1)\mu^{n-1} \frac{1-\mu^n}{1-\mu} = f''(1) \frac{\mu^n(\mu^n-1)}{\mu(\mu-1)}, & \text{wenn } \mu \neq 1 \text{ ist,} \\ f''(1)\mu^{n-1} n\mu = n f''(1), & \text{wenn } \mu = 1 \text{ ist.} \end{cases} \quad (3)$$

Nun gilt (mit (2.54) aus der Def. 2.32), dass $\sigma^2 = f''(1) + \mu - \mu^2$ und deshalb

$$f''(1) = \sigma^2 + \mu^2 - \mu = \sigma^2 + \mu(\mu - 1) \quad \text{ist.} \quad \text{Zudem folgt aus (2.14) für } Z_n \quad (4)$$

$$\mathbb{V}Z_n \stackrel{(2.46)}{=} f_n''(1) - [f_n'(1)]^2 + f_n'(1) \stackrel{(2.55)}{=} f_n''(1) - (\mu^n)^2 + \mu^n = f_n''(1) - \mu^n(\mu^n - 1). \quad (5)$$

Für $\mu = 1$ ergibt sich somit, dass $\mathbb{V}Z_n \stackrel{(3)/(5)}{=} n f''(1) - 1 \cdot 0 \stackrel{(4)}{=} n(\sigma^2 + 1 \cdot 0) = n\sigma^2$ ist – also wie in (2.57) behauptet. Dies trifft auch auf den Fall $\mu \neq 1$ zu, denn aus

$$\begin{aligned} f_n''(1) &\stackrel{(3)}{\stackrel{(4)}}{=} (\sigma^2 + \mu(\mu - 1)) \frac{\mu^n(\mu^n - 1)}{\mu(\mu - 1)} = \frac{\sigma^2 \mu^n(\mu^n - 1)}{\mu(\mu - 1)} + \mu^n(\mu^n - 1) \quad \text{folgt} \\ \mathbb{V}Z_n &\stackrel{(5)}{=} \frac{\sigma^2 \mu^n(\mu^n - 1)}{\mu(\mu - 1)} + \mu^n(\mu^n - 1) - \mu^n(\mu^n - 1) = \frac{\sigma^2 \mu^n(\mu^n - 1)}{\mu(\mu - 1)}. \end{aligned}$$

Bemerkung. Bei (1) handelt es sich ebenso (also wie im Beweis des Satzes 2.33) um eine Differenzengleichung. Ein entsprechendes Lösungsverfahren würde genauso zu (3) führen.

Nebenrechnung B.7 ($\mathbb{P}(\{Z_n \rightarrow 0\} \cup \{Z_n \rightarrow \infty\}) = 1$ – angelehnt an [9, Theorem 1.2, S. 252]). Die oben stehende Wahrscheinlichkeit wurde im Beweis des Satzes 2.48 verbal argumentiert und soll nun kurz formal hergeleitet werden.

Wenn für ein $\omega \in \Omega$ die $\mathbb{N}_0$ -wertige Folge $Z_n(\omega)$ *nicht* gegen 0 konvergiert oder ∞ strebt, müssen für mindestens ein $k \geq 1$ unendlich viele Glieder dieser Folge gleich k sein. Deshalb gilt $\{\lim_{n \rightarrow \infty} Z_n \in \{0, \infty\}\}^c \subseteq \bigcup_{k \geq 1} \{Z_n = k \text{ für } \infty \text{ viele } n\}$.

Für jedes $\omega \in \Omega$, für welches $Z_n(\omega) \rightarrow 0$ oder $Z_n(\omega) \rightarrow \infty$ gilt, darf umgekehrt $Z_n(\omega) = k$ für *kein* $k \geq 1$ unendlich oft vorkommen $\Rightarrow \{\lim_{n \rightarrow \infty} Z_n \in \{0, \infty\}\} \subseteq \bigcap_{k \geq 1} \{Z_n = k \text{ für } \infty \text{ viele } n\}^c = \left(\bigcup_{k \geq 1} \{Z_n = k \text{ für } \infty \text{ viele } n\}\right)^c$ (mit deMorgan) $\Rightarrow \bigcup_{k \geq 1} \{Z_n = k \text{ für } \infty \text{ viele } n\} \subseteq \{\lim_{n \rightarrow \infty} Z_n \in \{0, \infty\}\}^c$ (da $A \subseteq B \iff B^c \subseteq A^c$ für beliebige Mengen A, B gilt).

Zusammengenommen gilt $\{\lim_{n \rightarrow \infty} Z_n \in \{0, \infty\}\}^c = \bigcup_{k \geq 1} \{Z_n = k \text{ für } \infty \text{ viele } n\}$ und somit

$$\begin{aligned} \mathbb{P}\left(\left\{\lim_{n \rightarrow \infty} Z_n \in \{0, \infty\}\right\}^c\right) &= \mathbb{P}\left(\bigcup_{k \geq 1} \{Z_n = k \text{ für } \infty \text{ viele } n\}\right) \\ &\stackrel{(A.30)}{\leq} \sum_{k \geq 1} \underbrace{\mathbb{P}(\{Z_n = k \text{ für } \infty \text{ viele } n\})}_{=0 \text{ für alle } k \geq 1 \text{ mit (2.40)}} = 0 \\ \Rightarrow \mathbb{P}(\{Z_n \rightarrow 0\} \cup \{Z_n \rightarrow \infty\}) &= \mathbb{P}\left(\lim_{n \rightarrow \infty} Z_n \in \{0, \infty\}\right) \\ &\stackrel{(A.29)}{=} 1 - \mathbb{P}\left(\left\{\lim_{n \rightarrow \infty} Z_n \in \{0, \infty\}\right\}^c\right) \stackrel{(1)}{=} 1 - 0 = 1. \end{aligned} \tag{1}$$

Nebenrechnung B.8 ($\sigma(Z_n) = \sigma(W_n)$ bzw. $\sigma(Z_0, \dots, Z_n) = \sigma(W_0, \dots, W_n)$). Es soll ein formales Argument für (2.63) im Beweis des Satzes 2.37 gefunden werden.

Es wurde $W_n := \frac{Z_n}{\mu^n}$ gesetzt und die ZV Z_n ist eine Abbildung von $(\Omega, \Sigma) \rightarrow (\mathbb{N}_0, \mathcal{P}(\mathbb{N}_0))$. Mit $\Omega' := \{\frac{m}{\mu^n} : m \in \mathbb{N}_0\}$ ist somit W_n eine Abbildung von $(\Omega, \Sigma) \rightarrow (\Omega', \mathcal{P}(\Omega'))$. Setze nun für ein beliebiges $A' \in \mathcal{P}(\mathbb{N}_0)$ die Menge $B' := \{\frac{m}{\mu^n} : m \in A'\}$. Dann ist $B' \in \mathcal{P}(\Omega')$ und weiters gilt $Z_n \in A'$ genau dann, wenn $\frac{Z_n}{\mu^n} \in B'$ gilt $(\star)$. Damit folgt

$$\begin{aligned} \sigma(Z_n) &\stackrel{(A.38)}{=} \{\{Z_n \in A'\} : A' \in \mathcal{P}(\mathbb{N}_0)\} \stackrel{(\star)}{=} \left\{\left\{\frac{Z_n}{\mu^n} \in B'\right\} : B' \in \mathcal{P}(\Omega')\right\} \\ &= \{\{W_n \in B'\} : B' \in \mathcal{P}(\Omega')\} \stackrel{(A.38)}{=} \sigma(W_n). \end{aligned} \tag{1}$$

Da dies auf alle $n \geq 0$ zutrifft, gilt deshalb auch

$$\sigma(Z_0, \dots, Z_n) \stackrel{(A.15)}{=} \sigma\left(\bigcup_{i=0}^n \sigma(Z_i)\right) \stackrel{(1)}{=} \sigma\left(\bigcup_{i=0}^n \sigma(W_i)\right) \stackrel{(A.15)}{=} \sigma(W_0, \dots, W_n).$$

Nebenrechnung B.9 (Aus $0 \leq \mathbb{E}Z_{nl}^{(j)} < \infty$ für alle $n \geq m_0$ folgt $\mathbb{P}(Z_{nl}^{(j)} \rightarrow \infty) = 0$). Es soll obige Implikation formal für den Beweis des Satzes 3.50 gezeigt werden.

Sei dafür alles wie an der Stelle im Beweis, an welcher auf diese NR verwiesen wird. Seien zudem $1 \leq j, l \leq k$ beliebig. Im Beweis wurde die Beschränktheit aller Erwartungswerte $\mathbb{E}Z_{nl}^{(j)}$ für $n \geq m_0$ gezeigt. Aber auch für die nur endlich vielen mit $n < m_0$ lässt

B Appendix

sich eine Schranke finden (vgl. etwa (3.54) in Kombination mit (3.67)). Jeder dieser EW lässt sich als $\mathbb{E}Z_{nl}^{(j)} \stackrel{(2.5)}{=} \sum_{i \geq 0} \mathbb{P}(Z_{nl}^{(j)} > i)$, also eine beschränkte Reihe mit nichtnegativen Gliedern, darstellen. Demnach müssen diese Glieder eine Nullfolge (in i) bilden (vgl. z.B. den Satz A.93 bzw. [22, Sätze 31.2 und 31.3, S. 191 f.]).

Für jedes $n \geq 1$ gilt also $\lim_{i \rightarrow \infty} \mathbb{P}(Z_{nl}^{(j)} > i) = 0$ ($\star$). Insbesondere folgt daraus

$$\begin{aligned} \mathbb{P}(\exists n_0 \geq 1 : Z_{nl}^{(j)} > i \text{ für alle } n \geq n_0) &\leq \mathbb{P}(\exists n_0 \geq 1 : Z_{n_0 l}^{(j)} > i) \\ &= \mathbb{P}\left(\bigcup_{n_0 \geq 1} \{Z_{n_0 l}^{(j)} > i\}\right) \stackrel{(A.30)}{\leq} \sum_{n_0 \geq 1} \mathbb{P}(Z_{n_0 l}^{(j)} > i) \stackrel{(\star)}{\xrightarrow{i \rightarrow \infty}} 0. \end{aligned} \quad (1)$$

Mit der Stetigkeit von oben (Satz A.23 (v)) erhält man

$$\begin{aligned} 0 &\stackrel{(1)}{=} \lim_{i \rightarrow \infty} \mathbb{P}(\exists n_0 \geq 1 : Z_{nl}^{(j)} > i \ \forall n \geq n_0) \stackrel{(A.32)}{=} \mathbb{P}\left(\bigcap_{i \geq 1} \{\exists n_0 \geq 1 : Z_{nl}^{(j)} > i \ \forall n \geq n_0\}\right) \\ &= \mathbb{P}(\forall i \geq 1 \text{ gilt } \exists n_0 \geq 1 : Z_{nl}^{(j)} > i \ \forall n \geq n_0) \\ &\geq \mathbb{P}(\forall G \in \mathbb{R}_+ \text{ gilt } \exists n_0 \geq 1 : Z_{nl}^{(j)} > G \ \forall n \geq n_0) = \mathbb{P}\left(\lim_{n \rightarrow \infty} Z_{nl}^{(j)} = \infty\right), \end{aligned}$$

wobei im letzte Schritt die Def. der bestimmten Divergenz gegen $+\infty$ verwendet wurde.

Nebenrechnung B.10 Seien X eine nichtnegative ZV, $a \in \mathbb{R}_+$ und $b \in \mathbb{R}$ (die Rechnung gilt auch für $b \in \mathbb{C}$ – vgl. etwa die Bem. zum Satz A.84 bzgl. des Erwartungswertes komplexer ZV). Dann folgt aus

$$\begin{aligned} \mathbb{E}[a^{bX} \mathbb{1}_{\{X>0\}}] &\stackrel{(A.79)}{=} \mathbb{P}(X > 0) \mathbb{E}[a^{bX} | X > 0] \quad \text{auch} \quad (1) \\ &\stackrel{=\mathbb{1}_\Omega=1 \text{ (da } X \text{ nichtneg. ist)}}{=} \mathbb{E}[a^{bX} (\mathbb{1}_{\{X=0\}} + \mathbb{1}_{\{X>0\}})] \stackrel{=\mathbb{P}(X=0) \text{ mit (A.44)}}{=} \mathbb{E}[\underbrace{a^{bX} \mathbb{1}_{\{X=0\}}}_{=1 \text{ für } X=0}] + \mathbb{E}[a^{bX} \mathbb{1}_{\{X>0\}}] \\ &\stackrel{(1)}{=} \mathbb{P}(X=0) + \mathbb{P}(X>0) \mathbb{E}[a^{bX} | X>0]. \end{aligned} \quad (B.4)$$

Wenn $\mathbb{P}(X > 0) > 0$ ist, gilt deshalb weiters

$$\mathbb{E}[a^{bX} | X > 0] \stackrel{(B.4)}{=} \frac{\mathbb{E}[a^{bX}] - \mathbb{P}(X=0)}{\mathbb{P}(X > 0)} \stackrel{(A.29)}{=} \frac{\mathbb{E}[a^{bX}] - \mathbb{P}(X=0)}{1 - \mathbb{P}(X=0)}. \quad (B.5)$$

Nebenrechnung B.11 (Grenzwertbildungen für den Unterabschnitt 4.1.2 – superkritische geometrische GWP). Seien $u \in \mathbb{R}_+$ und $t \in \mathbb{R}$. Berechne zunächst die Grenzwerte von $\mu^n(1 - e^{-u\mu^{-n}})$ bzw. $\mu^n(1 - e^{it\mu^{-n}})$ für $n \rightarrow \infty$ unter Zuhilfenahme der *Regel von de L'Hospital* (vgl. z.B. [22, Satz 50.1, S. 287]).

Für $x \in \mathbb{R}$ besagt diese Regel bzgl. einer Funktion $x \mapsto x(1 - e^{-\frac{u}{x}})$

$$\lim_{x \rightarrow \infty} x(1 - \underbrace{e^{-\frac{u}{x}}}_{\rightarrow 1}) = \lim_{x \rightarrow \infty} \frac{1 - e^{-\frac{u}{x}}}{\frac{1}{x}} \stackrel{(\frac{0}{0})}{=} \lim_{x \rightarrow \infty} \frac{-e^{-\frac{u}{x}} \frac{u}{x^2}}{-\frac{1}{x^2}} = u \lim_{x \rightarrow \infty} \underbrace{e^{-\frac{u}{x}}}_{\rightarrow 1} = u. \quad (1)$$

Analog kann für eine Funktion $x \mapsto x(1 - e^{\frac{it}{x}})$ vorgegangen werden. Alternativ erhält man aber auch mit kartesischen Koordinaten wie folgt (ebenfalls sehr ähnlich wie in (1)):

$$\begin{aligned} \lim_{x \rightarrow \infty} x(1 - e^{\frac{it}{x}}) &\stackrel{(A.99)}{=} \lim_{x \rightarrow \infty} \frac{1 - \cos\left(\frac{t}{x}\right) - i \sin\left(\frac{t}{x}\right)}{x^{-1}} \stackrel{(\frac{0}{0})}{=} \lim_{x \rightarrow \infty} \frac{-\frac{t}{x^2}(\sin\left(\frac{t}{x}\right) - i \cos\left(\frac{t}{x}\right))}{-\frac{1}{x^2}} \\ &= \lim_{x \rightarrow \infty} \left(\underbrace{\left(\sin\left(\frac{t}{x}\right)t - i \cos\left(\frac{t}{x}\right)t\right)}_{\rightarrow 0} \right) = -it. \end{aligned} \quad (2)$$

Da $\mu > 1$ ist, strebt $m := \mu^n$ gegen ∞ , wenn n gegen ∞ strebt. Deshalb folgt

$$\lim_{n \rightarrow \infty} \mu^n(1 - e^{-u\mu^{-n}}) = \lim_{m \rightarrow \infty} m(1 - e^{-\frac{u}{m}}) = u. \quad \lim_{n \rightarrow \infty} \mu^n(1 - e^{it\mu^{-n}}) \stackrel{\text{analog}}{=} -it. \quad (\text{B.7})$$

Es wird noch kurz die Konvergenz der C-Funktionen $\{\varphi_n\}_{n \geq 0}$ skizziert (ähnlich dem Beweis des Satzes 4.6 bzgl. $\{\mathcal{L}_n\}_{n \geq 0}$). Genau genommen müssten dafür die weF $\{f_n\}_{n \geq 0}$ jedoch für $s \in \mathbb{C}$ mit $|s| \leq 1$ def. sein (vgl. die Bem. zum Satz 2.56 (ii)). Analog zu (2.79) gilt dann für $n \geq 0$ und $t \in \mathbb{R}$

$$\varphi_n(t) \stackrel{(2.76)}{=} f_n(e^{it\mu^{-n}}) \stackrel{(4.16)}{=} \frac{\mu^n(1 - e^{it\mu^{-n}}) + \mu e^{it\mu^{-n}} - 1}{\underbrace{\mu \mu^n(1 - e^{it\mu^{-n}})}_{\rightarrow -it \text{ (B.7)}} + \underbrace{\mu e^{it\mu^{-n}} - 1}_{\rightarrow 1}} \xrightarrow{n \rightarrow \infty} \frac{-it + \mu - 1}{\underbrace{-it\mu + \mu - 1}_{= \varphi(t) \text{ wegen (2.77)}}}. \quad (\text{B.8})$$

Nebenrechnung B.12 ($|h_{n+1}(s) - h_n(s)| \leq \mu|h_n(s) - h_{n-1}(s)|$).

Diese Ungleichung soll im Fall geometrischer GWP für (4.57) gezeigt werden.

Sei zunächst $n \geq 0$ beliebig. Da $Z_0 \equiv 1$ und h_n eine weF ist, gilt definitionsgemäß (siehe (4.23)) immer $T_n > 0$ und somit $h_n(0) \stackrel{(2.9)}{=} \mathbb{P}(T_n = 0) = 0$. Zudem gilt $h_n(1) \stackrel{(2.8)}{=} 1$, weshalb die Ungleichung für $s \in \{0, 1\}$ und $n \geq 1$ trivialerweise erfüllt ist. Beachte weiters, dass $h_{n+1}(s) < h_n(s)$ für alle $s \in (0, 1)$ gilt (siehe (4.30)).

Zusammengenommen genügt es zu zeigen, dass $(h_n - h_{n+1}) \leq \mu(h_{n-1} - h_n)$ auf $(0, 1)$ und für alle $n \geq 1$ gilt. Das ist gleichbedeutend damit, dass für alle $s \in (0, 1)$

$$g_n(s) := \frac{h_n(s) - h_{n+1}(s)}{h_{n-1}(s) - h_n(s)} \leq \mu \quad \text{für alle } n \geq 1 \text{ gilt,} \quad \text{wobei } g_n : (0, 1) \rightarrow \mathbb{R}_+ \text{ ist.} \quad (1)$$

Es soll induktiv vorgegangen werden. Bestimme für den Anfang $f(s) \stackrel{(4.5)}{=} \frac{1}{1+\mu-\mu s}$ sowie $f(sf(s)) = \frac{1}{1+\mu-\frac{\mu s}{1+\mu-\mu s}} = \frac{1+\mu-\mu s}{(1+\mu)^2-\mu s(\mu+2)}$ und verwende das Lemma 4.15 (ii) um die

B Appendix

Beziehungen $h_0(s) = s$, $h_1(s) = sf(s)$ bzw. $h_2(s) = sf(sf(s))$ zu erhalten. Daraus folgt mit $n=1$ für alle $s \in (0, 1)$

$$\begin{aligned} g_1(s) &\stackrel{(1)}{=} \frac{h_1(s) - h_2(s)}{h_0(s) - h_1(s)} \stackrel{s \neq 0}{=} \frac{f(s) - f(sf(s))}{1 - f(s)} = \frac{\frac{1}{1+\mu-\mu s} - \frac{1+\mu-\mu s}{(1+\mu)^2 - \mu s(\mu+2)}}{1 - \frac{1}{1+\mu-\mu s}} \\ &= \frac{(1+\mu)^2 - \mu s(\mu+2) - (1+\mu)^2 + 2(1+\mu)\mu s - (\mu s)^2}{((1+\mu)^2 - \mu s(\mu+2))(\mu - \mu s)} \\ &= \frac{\mu^2 s - \mu^2 s^2}{((1+\mu)^2 - \mu s(\mu+2))\mu(1-s)} \stackrel{s \neq 1}{=} \frac{\mu s}{(1+\mu)^2 - \mu s(\mu+2)} =: \frac{u(s)}{v(s)}. \end{aligned}$$

Nun ist g_1 auf ganz $(0, 1)$ stetig und mit der Quotientenregel ist $g'_1 = \frac{u'v - v'u}{v^2} > 0$ genau dann, wenn $u'v - v'u > 0$ ist ($v^2 > 0$ auf $(0, 1)$). Stelle also fest, dass für alle $s \in (0, 1)$ $u'(s)v(s) - v'(s)u(s) = \mu(1+\mu)^2 - \mu^2 s(\mu+2) + \mu^2 s(\mu+2) = \mu(1+\mu)^2 > 0$ und somit g_1 auf $(0, 1)$ strikt isoton ist (siehe den Satz A.88). Deshalb gilt

$$g_1(s) \leq \lim_{s \nearrow 1} g_1(s) = \frac{\mu}{(1+\mu)^2 - \mu(\mu+2)} = \mu \quad \text{für alle } s \in (0, 1).$$

Nehme nun für $n \geq 2$ an, dass $g_{n-1} \leq \mu$ gilt (Ind.-Ann.). Sei wieder $s \in (0, 1)$. Aus

$$\begin{aligned} g_n(s) &\stackrel{(1)}{\stackrel{(4.29)}{=}} \frac{sf(h_{n-1}(s)) - sf(h_n(s))}{sf(h_{n-2}(s)) - sf(h_{n-1}(s))} \stackrel{s \neq 0}{=} \frac{(f \circ h_{n-1})(s) - (f \circ h_n)(s)}{(f \circ h_{n-2})(s) - (f \circ h_{n-1})(s)} \quad \text{folgt} \\ g_n &\stackrel{(4.5)}{=} \frac{\frac{1}{1+\mu-\mu h_{n-1}} - \frac{1}{1+\mu-\mu h_n}}{\frac{1}{1+\mu-\mu h_{n-2}} - \frac{1}{1+\mu-\mu h_{n-1}}} = \frac{\mu h_{n-1} - \mu h_n}{\mu h_{n-2} - \mu h_{n-1}} \cdot \underbrace{\frac{1+\mu-\mu h_{n-2}}{1+\mu-\mu h_n}}_{\leq \frac{1+\mu-\mu h_n}{1+\mu-\mu h_n} = 1 \text{ (da } h_{n-2} \geq h_n \text{ ist (4.30))}} \\ &\leq \frac{h_{n-1} - h_n}{h_{n-2} - h_{n-1}} \stackrel{(1)}{=} g_{n-1} \stackrel{\text{Ind.-Ann.}}{\leq} \mu. \end{aligned}$$

Nebenrechnung B.13 (Der Prozess $\{\tilde{Z}_n\}_{n \geq 0}$ ist ein GWP). Es soll bewiesen werden, dass es sich beim auf das Aussterben bedingten Prozess $\{\tilde{Z}_n\}_{n \geq 0}$ auch im Fall $\mu > 1$ tatsächlich um einen GW-Prozess handelt. Die Vorgehensweise ist angelehnt an den Beweis in [3, Theorem I.8.3, S. 53 f.]. Sei also $\mu > 1 \Rightarrow \eta < 1 \Rightarrow \mathbb{P}(T < \infty) > 0$ (mit (4.28)).

Stelle zunächst als Hilfsbeobachtung fest, dass die Aussage des Lemmas 4.27 problemlos zu $\mathbb{P}(Z_n = k, (Z_{n-1}, \dots, Z_0)^T \in A, T < \infty) = \mathbb{P}(Z_n = k, (Z_{n-1}, \dots, Z_0)^T \in A) \eta^k$ für $k, n \geq 0$ und $A \subseteq \mathbb{N}_0^{n-1}$ erweitert werden kann (beachte im entsprechenden Beweisschritt, dass die $\{Z_{m,i}\}_{m \geq 0}$ für jedes $i \geq 1$ i.i.d.-Kopien von $\{Z_n\}_{n \geq 0}$ und somit auch unabhängig von allen $\{Z_l\}_{l \leq n}$ sind). Demnach gilt insbesondere auch für alle $j, k, n, j_0, \dots, j_n \geq 0$

$$\begin{aligned} \mathbb{P}(Z_{n+1} = k, Z_n = j, T < \infty) &= \mathbb{P}(Z_{n+1} = k, Z_n = j) \eta^k \\ \mathbb{P}(Z_{n+1} = k, Z_n = j_n, \dots, Z_0 = j_0, T < \infty) &= \mathbb{P}(Z_{n+1} = k, Z_n = j_n, \dots, Z_0 = j_0) \eta^k, \\ \mathbb{P}(Z_n = j_n, Z_{n-1} = j_{n-1}, \dots, Z_0 = j_0, T < \infty) &= \mathbb{P}(Z_n = j_n, Z_{n-1} = j_{n-1}, \dots, Z_0 = j_0) \eta^{j_n}. \end{aligned} \tag{1}$$

Bemerkung. Für den restlichen Teil dieser NR wird immer davon ausgegangen, dass die Ereignisse, auf welche bedingt wird, eine positive Wahrscheinlichkeit haben.

Nun gilt für die Übergangswahrscheinlichkeiten von $\{\tilde{Z}_n\}_{n \geq 0}$ für alle $j, k, n \geq 0$

$$\begin{aligned} \mathbb{P}(\tilde{Z}_{n+1}=k | \tilde{Z}_n=j) &\stackrel{(4.59)}{=} \mathbb{P}(Z_{n+1}=k | Z_n=j, T < \infty) = \frac{\mathbb{P}(Z_{n+1}=k, Z_n=j, T < \infty)}{\mathbb{P}(Z_n=j, T < \infty)} \\ &\stackrel{(1) \text{ und Lemma 4.27}}{=} \frac{\mathbb{P}(Z_{n+1}=k, Z_n=j)\eta^k}{\mathbb{P}(Z_n=j)\eta^j} = \mathbb{P}(Z_{n+1}=k | Z_n=j) \frac{\eta^k}{\eta^j}. \end{aligned} \quad (2)$$

Bemerkung. Die Übergangs-WK summieren sich tatsächlich zu 1 auf. Für $j \geq 0$ gilt:

$$\sum_{k \geq 0} \mathbb{P}(Z_{n+1}=k | Z_n=j) \frac{\eta^k}{\eta^j} \stackrel{(2.38)}{=} \frac{1}{\eta^j} \sum_{k \geq 0} p_{jk} \eta^k \stackrel{(2.52)}{=} \frac{1}{\eta^j} (f(\eta))^j \stackrel{(2.68)}{=} \frac{1}{\eta^j} \eta^j = 1.$$

Als Nächstes soll gezeigt werden, dass der Prozess $\{\tilde{Z}_n\}_{n \geq 0}$ (wie $\{Z_n\}_{n \geq 0}$) ebenso eine Markov-Kette ist. Dies sieht man daran (vgl. den Beweis des Satzes 2.25), dass für alle passenden $k, n, j_0, \dots, j_n \geq 0$

$$\begin{aligned} \mathbb{P}(\tilde{Z}_{n+1}=k | \tilde{Z}_n=j_n, \dots, \tilde{Z}_0=j_0) &\stackrel{(4.59)}{=} \mathbb{P}(Z_{n+1}=k | Z_n=j_n, \dots, Z_0=j_0, T < \infty) \\ &\stackrel{(1) \text{ und (2.39) (ME)}}{=} \mathbb{P}(Z_{n+1}=k | Z_n=j_n) \frac{\eta^k}{\eta^j} \stackrel{(2)}{=} \mathbb{P}(\tilde{Z}_{n+1}=k | \tilde{Z}_n=j) \quad \text{gilt.} \end{aligned}$$

Als Vergleichsobjekt sei nun $\{Z_n^*\}_{n \geq 0}$ mit $\{\tilde{p}_k\}_{k \geq 0} \stackrel{(4.61)}{=} \{p_k \eta^{k-1}\}_{k \geq 0}$ als Nachkommensverteilung ein GWP. Es soll gezeigt werden, dass $\{Z_n^*\}_{n \geq 0}$ und $\{\tilde{Z}_n\}_{n \geq 0}$ die gleiche Verteilung besitzen – ist dies der Fall, bestätigt sich die eingangs getätigte Aussage.

Es sei $f^*(s) \stackrel{(2.6)}{=} \sum_{k \geq 0} \tilde{p}_k s^k \stackrel{(4.61)}{=} \frac{1}{\eta} \sum_{k \geq 0} p_k (s\eta)^k \stackrel{(2.45)}{=} \frac{1}{\eta} f(s\eta)$ (Δ) für alle $s \in [0, 1]$ die weF der zu $\{Z_n^*\}_{n \geq 0}$ gehörigen Nachkommenszahl X^* . Weiters sei für jedes $n \geq 0$ die weF der Populationsgröße $f_n^* := f_{Z_n^*}$. Induktiv folgt dann $f_n^* = \tilde{f}_n$, denn für alle $s \in [0, 1]$ gilt

$$\begin{aligned} f_0^*(s) &\stackrel{(2.30)}{=} s \stackrel{(2.30)}{\underset{(4.62)}{=}} \frac{1}{\eta} s \eta \stackrel{(2.30)}{=} \tilde{f}_0(s), \quad f_1^*(s) \stackrel{(2.30)}{=} f^*(s) \stackrel{(\Delta)}{=} \frac{1}{\eta} f(s\eta) \stackrel{(2.30)}{\underset{(4.62)}{=}} \tilde{f}_1(s) \quad \text{und} \\ f_{n+1}^*(s) &\stackrel{(2.30)}{=} f_1^*(f_n^*(s)) \stackrel{\text{Ind.-Ann.}}{\underset{(4.62)}{=}} \frac{1}{\eta} f_1\left(\frac{1}{\eta} f_n(s\eta)\right) = \frac{1}{\eta} f_1(f_n(\eta s)) \\ &\stackrel{(2.30)}{=} \frac{1}{\eta} f_{n+1}(\eta s) \stackrel{(4.62)}{=} \tilde{f}_{n+1}(s) \end{aligned} \quad \text{für } n \geq 1.$$

Mit dem Satz 2.6 (bzw. (2.11)) lässt sich deshalb $Z_n^* \stackrel{d}{=} \tilde{Z}_n$ für jedes $n \geq 0$ schließen.

Das bedeutet aber noch nicht zwangsläufig, dass auch die *gemeinsamen* Verteilungen der Populationsgrößen aller Generationen (der beiden Prozesse) übereinstimmen. Betrachte also schließlich für $j, k, n \geq 0$ und $s \in [0, 1]$

$$\begin{aligned} \sum_{k \geq 0} \mathbb{P}(Z_{n+1}^*=k | Z_n^*=j) s^k &\stackrel{(2.38)}{\underset{(2.52)}{=}} (f^*(s))^j \stackrel{(\Delta)}{=} \left(\frac{1}{\eta} f(s\eta)\right)^j \\ &\stackrel{(2.38)}{\underset{(2.52)}{=}} \frac{1}{\eta^j} \sum_{k \geq 0} \mathbb{P}(Z_{n+1}=k | Z_n=j) (s\eta)^k = \sum_{k \geq 0} \mathbb{P}(Z_{n+1}=k | Z_n=j) \frac{\eta^k}{\eta^j} s^k \\ &\stackrel{(2)}{=} \sum_{k \geq 0} \mathbb{P}(\tilde{Z}_{n+1}=k | \tilde{Z}_n=j) s^k. \end{aligned}$$

Es sind also die Übergangswahrscheinlichkeiten beider Prozesse gleichermaßen die Koeffizienten ein und derselben Reihe (und zwar der j -ten Potenz von f^* – ungeachtet

des Wertes von s). Zusammenfassend handelt es sich sowohl bei $\{Z_n^*\}_{n \geq 0}$ als auch bei $\{\tilde{Z}_n\}_{n \geq 0}$ um Markov-Ketten mit identischer Anfangsverteilung und gleicher 'Übergangsmechanik' (der Ausdruck stammt aus [3]). Dementsprechend muss $\{Z_n^*\}_{n \geq 0} \stackrel{d}{=} \{\tilde{Z}_n\}_{n \geq 0}$ gelten (was etwa induktiv aus $Z_0^* \equiv \tilde{Z}_0 \equiv 1$ erhalten werden kann). Also ist auch der auf das Aussterben bedingte Prozess ein GWP mit Nachkommensverteilung $\{\tilde{p}_k\}_{k \geq 0}$.

Nebenrechnung B.14 (Invarianz gegenüber Umordnungen divergenter Reihen mit nichtnegativen Gliedern). Es soll ein alternatives Argument für jenes am Schluss des Beweises des Satzes A.97 gefunden werden.

Angenommen, für eine Reihe mit nichtnegativen Gliedern gilt $\sum_{n \geq 0} a_n = \infty$ – sie divergiert also bestimmt gegen $+\infty$. Sei $G > 0$ (beliebig groß). Gemäß der Definition A.95 (iii) gibt es ein $n_0 \geq 0$, sodass $\sum_{n=0}^{n_0} a_n > G$ ist $(\star)$. Definiere (wie im Punkt (ii) der genannten Def.) die Bijektion $\phi : \mathbb{N}_0 \rightarrow \mathbb{N}_0$ für alle $n \geq 0$ als $\phi(n) := i_n \Rightarrow a_{i_n} = a_{\phi(n)}$. Mit der Umkehrabbildung ϕ^{-1} gilt auch $a_n = a_{i_{\phi^{-1}(n)}}$ $(\star\star)$. Sei weiters $m_0 := \max(\phi^{-1}(0), \phi^{-1}(1), \dots, \phi^{-1}(n_0))$.

Dann gilt $m_0 \geq n_0$ und $\{\phi^{-1}(0), \dots, \phi^{-1}(n_0)\} \subseteq \{0, \dots, m_0\}$ $(\star\star\star)$ und weiter

$$G \stackrel{(\star)}{<} \sum_{n=0}^{n_0} a_n \stackrel{(\star\star)}{=} \sum_{n=0}^{n_0} a_{i_{\phi^{-1}(n)}} \stackrel{a_{i_n} \geq 0}{\leq} \sum_{n=0}^{m_0} a_{i_n} \leq \sum_{n=0}^m a_{i_n} \quad \text{für alle } m \geq m_0.$$

Da G beliebig war, existiert also für alle $G > 0$ ein $m_0 \geq 0$, sodass $\sum_{n=0}^m a_{i_n} > G$ für alle $m \geq m_0$ ist. Das bedeutet definitionsgemäß $\sum_{n \geq 0} a_{i_n} = +\infty$.

Nebenrechnung B.15 (Bestimmung von ρ , u und v für das numerische Bsp. im Unterabschnitt 4.2.2). Für die Berechnung der Eigenwerte von $\mathbf{M}$ kann bekanntermaßen das *charakteristische Polynom* von M bzw. die Determinante von $\mathbf{M} - \rho \mathbf{I}_k$ bestimmt und auf 0 gesetzt werden (sodass $\mathbf{M} - \rho \mathbf{I}_k$ singulär ist – vgl. ggf. etwa [34, S. 123 ff.]). Betrachte also

$$\begin{aligned} \det(\mathbf{M} - \rho \mathbf{I}_k) &\stackrel{(4.69)}{=} \det \left(\begin{pmatrix} \frac{8}{7} - \rho & \frac{3}{7} & 0 \\ \frac{16}{35} & \frac{36}{35} - \rho & \frac{16}{35} \\ 0 & \frac{3}{7} & \frac{8}{7} - \rho \end{pmatrix} \right) \\ &= \left(\frac{8}{7} - \rho \right) \left(\left(\frac{36}{35} - \rho \right) \left(\frac{8}{7} - \rho \right) - \frac{16 \cdot 3}{35 \cdot 7} \right) - \frac{16}{35} \left(\frac{3}{7} \left(\frac{8}{7} - \rho \right) \right) \end{aligned}$$

(wobei die Determinante mit dem *Entwicklungssatz von Laplace*, entwickelt nach der ersten Spalte, ermittelt wurde – vgl. etwa [34, Satz 5.4.9, S. 114]). Daran erkennt man direkt $\rho = \frac{8}{7}$ als eine Nullstelle. Für die anderen beiden kann der letzte Ausdruck durch $(\frac{8}{7} - \rho)$ dividiert werden, womit $\rho^2 - \frac{76}{35}\rho + \frac{192}{145} = 0$ zu lösen bleibt. Insgesamt erhält man die Menge der Nullstellen des Polynoms als $\{\frac{12}{7}, \frac{8}{7}, \frac{16}{35}\}$, wobei $\rho = \frac{12}{7}$ die größte ist.

Wird die Eigenwertgleichung $\mathbf{uM} = \rho \mathbf{u}$ (für linke Eigenvektoren – vgl. (3.82)) gelöst, ergeben sich die drei Gleichungen $u_1 \frac{8}{7} + u_2 \frac{16}{35} = u_1 \frac{12}{7}$, $u_1 \frac{3}{7} + u_2 \frac{36}{35} + u_3 \frac{3}{7} = u_2 \frac{12}{7}$ und $u_2 \frac{16}{35} + u_3 \frac{8}{7} = u_3 \frac{12}{7}$. Daraus erhält man $\{(u_1, u_2, u_3) > \mathbf{0} : u_1 = u_3, u_2 = \frac{5}{4}u_3\}$ als Kandidatenmenge für $\mathbf{u}$. Der daraus gemäß dem Satz 3.39 zu $\mathbf{u} \cdot \mathbf{1} = 1$ normalisierte Vektor (siehe (3.83)) ist der gesuchte linke Eigenvektor $\mathbf{u} = (\frac{4}{13}, \frac{5}{13}, \frac{4}{13})^T$.

Völlig analog ergibt sich die Lösungsmenge $\{(v_1, v_2, v_3) > \mathbf{0} : v_1 = v_3, v_2 = \frac{4}{3}v_3\}$ aus der Gleichung $\mathbf{Mv} = \rho \mathbf{v}$. Der einzige Vektor daraus, der $\mathbf{uv} = 1$ (siehe wieder (3.83)) erfüllt, ist $\mathbf{v} = (\frac{39}{44}, \frac{13}{11}, \frac{39}{44})^T$.

Nebenrechnung B.16 (Berechnungen mit `yac_symbol`-Objekten – Quellcode unter B.2.1). Es soll kurz anhand des numerischen Beispiels in 4.2.2 vorgeführt werden, wie mit den expliziten Ausdrücken für $\mathbf{f}_n$ verfahren werden kann. Dafür steht die R-Funktion `fn.sym.calc` zur Verfügung. Diese Routine liefert Objekte der Klasse `yac_symbol`, welche mit YACAS-Befehlen (vgl. etwa [54]) manipulierbar sind.

Mit dem Befehl `simplify()` des 'Ryacac'-Paketes wird etwa erreicht, dass die Koeffizienten bei $\mathbf{s}^r$ für $r \in \mathbb{N}_0^3$ abgelesen werden können. Weiters sorgt der YACAS-Befehl "Rationalize" (in R anwendbar mit `y_fn()`) dafür, dass Kommazahlen vermieden und exaktere Ergebnisse ausgegeben werden. Zusammengenommen liefert etwa das Kommando `f2.sym <- y_fn(simplify(fn.sym.calc(f.str,n=2,time=T)),"Rationalize")` für $\mathbf{s} \in [0, 1]^3$ (ergänzt um die Zeile `unlist(apply(f2.sym,tex))`) für eine L^AT_EX-Darstellung)

$$\mathbf{f}_2(\mathbf{s}) = \left(\begin{array}{l} \frac{9142857143s_1^4}{1000000000} + \frac{137142857s_2s_1^2}{10000000} + \frac{6857142857s_3s_1}{35000000000} + \frac{558s_2^2}{1715} + \frac{1071428571}{87500000000} \\ \frac{1097142857s_2s_1^2}{9800000000} + \frac{11008s_3^2s_1^2}{42875} + \frac{1481142857s_3s_2^2s_1}{5000000} + \frac{329142857s_3s_1}{20000000} + \frac{1666285713s_2^4}{12250000000} \\ + \frac{4248s_2^2}{42875} + \frac{1097142857s_3^2s_2}{9800000000} + \frac{355142857}{12250000000} \\ \frac{9142857143s_3^4}{1000000000} + \frac{137142857s_2s_3^2}{10000000} + \frac{6857142857s_1s_3}{35000000000} + \frac{558s_2^2}{1715} + \frac{1071428571}{87500000000} \end{array} \right)$$

für die zweite Generation (mit marginalen arithmetischen Ungenauigkeiten: z.B. die Konstante in $\mathbf{f}_{21}$ wäre $\frac{3}{245}$, was aber erst ab der 12. Nachkommastelle abweicht). Daran erkennt man, dass diese Berechnungen bereits nach einer Iteration unhandlich werden (und aus Kapazitätsgründen generell nur für wenige Generationen durchführbar sind).

Dennoch kann diese Funktionalität genutzt werden, um die Populationsgrößen $\mathbf{Z}_n$ bzw. $\mathbf{Z}_n^{(j)}$ zu analysieren. Dazu wird `f2.sym` zuerst wieder in den `character`-Vektor `f2.str` umgerechnet, um die Funktionen aus 4.2.1 nutzen zu können (siehe `f.sym.to.f.str` in der Zeile 3 der R-Helferfunktionen unter B.2.2).

In den Zeilen 1-4 der unten stehenden R-Ausgabe werden etwa die Wahrscheinlichkeiten $\mathbb{P}(\mathbf{Z}_2^{(1)} = \mathbf{r})$ mit `wfkt.calc` berechnet (mit geringfügigen Adaptionen wäre dies analog für beliebige $\mathbf{Z}_0$ über (3.45) bzw. $\mathbf{f}_{\mathbf{Z}_2}$ durchführbar).

Die expliziten Ausdrücke für $\mathbf{f}_n$ bieten aber auch eine alternative Möglichkeit zur Bestimmung der Momente. Dies wird abschließend kurz für $\mathbb{E}\mathbf{Z}_2$ mit $\mathbf{Z}_0 = (3, 2, 5)^T$ gezeigt. Betrachte dafür einen MGWP $\{\tilde{\mathbf{Z}}_n\}_{n \geq 0}$ mit $\tilde{\mathbf{Z}}_0 = \mathbf{Z}_0$ und $\mathbf{f}_2$ als weF der Reproduktionszahlen. Dann gilt $\mathbb{E}\mathbf{Z}_2^T \stackrel{(3.65)}{=} \mathbf{Z}_0^T \mathbf{D} \mathbf{f}_2(\mathbf{1}) \stackrel{(3.53)}{=} \tilde{\mathbf{Z}}_0^T \tilde{\mathbf{M}} \stackrel{(3.65)}{=} \mathbb{E}\tilde{\mathbf{Z}}_1^T$, wobei $\tilde{\mathbf{M}}$ die EW-Matrix (Def. 3.21) des 'neuen' Prozesses sei. Diese Gleichheit erkennt man an den Zeilen 7-8 der Ausgabe (jeweils berechnet mit der R-Funktion `EZn.calc`).

```

1 > f2.str <- f.sym.to.f.str(f2.sym)
2 > wfkt.calc(f2.str)$pr[wfkt.calc(f2.str)$pr[,1] != 0,1]
3 r=(0,0,0) r=(4,0,0) r=(2,1,0) r=(0,2,0) r=(1,0,1)
4 0.0122449 0.1865889 0.2798834 0.3253644 0.1959184
5 > rbind(EZn.calc(f.str,n=2,Z0=c(3,2,5)), EZn.calc(f2.str,n=1,Z0=c(3,2,5)))
6      EZ2_1      EZ2_2      EZ2_3
7 [1,] 7.47102 10.34449 10.08327
8 [2,] 7.47102 10.34449 10.08327

```

B.2 Quellcodes

B.2.1 R-Quellcode für den Abschnitt 4.2 – MGWP mit R

Nun folgen die direkt ausführbaren Quellcodes der im Abschnitt 4.2 (und in der NR B.16) durchgeführten Analysen.

Laden der Abhängigkeiten und R-Funktionen

Vor der Verarbeitung der Quellcodes in den nächsten Teilabschnitten müssen die entsprechenden Quelldateien wie im Unterabschnitt B.2.2 beschrieben zur Verfügung gestellt werden. Anschließend können die benötigten Pakete und Funktionen mit dem gleich folgenden Code geladen werden. Das Verzeichnis, in welchem die Quelldateien abgelegt sind, sollte in der `character`-Variable `sourcepath` stehen (mit `'/'` am Ende).

```
1 source(file = paste0(sourcepath, "MGWP_dependencies.R"))
2 source(file = paste0(sourcepath, "MGWP_functions.R"))
3 source(file = paste0(sourcepath, "MGWP_helper.R"))
```

R-Quellcode für den Unterabschnitt 4.2.2

Die Zeilen, welche nur dem L^AT_EX-Export dienen (Zeilen 25, 40 und 56), wurden auskommentiert (die Erstellung der entsprechenden Funktionen erfolgte nur für diesen Zweck, weshalb deren Code ausgespart wird).

Weiters sind Berechnungen oder Simulationen mit einer langen Laufzeit deaktiviert. Beim ersten Durchlauf müssen darum die Auskommentierungen in den Zeilen 29-33, 43 und 67-68 aufgehoben werden. Im `datapath` (siehe die unter B.2.2 beschriebenen Vorbereitungen) sollte ein Verzeichnis `"Sim_numex"` angelegt (oder der Code in der Zeile 42 angepasst) werden.

```
1  ### Numerisches Beispiel Unterabschnitt 4.2.2: ###
2  # Eingabe weF f((4.67)) und W-Funktion berechnen: ###
3  ### R-Ausgabe 4.1 #####
4  f.str <- c("4/7*s1^2+3/7*s2", "1/35+16/35*s1*s3+18/35*s2^2", "3/7*s2+4/7*s3^2")
5  invisible(wfkt.calc(f.str,time=T,loop=F))
6  wfkt <- wfkt.calc(f.str,time=T,loop=T)
7  (pr <- wfkt$pr)

8  # M und Eigenwerte aus weF berechnen:###
9  ### R-Ausgabe 4.2 #####
10 M.calc(f.str,check.PR=T)[c("M", "rho", "Klasse")]

11 # Überprüfung M^2:
12 Mn.calc(M.calc(f.str)$M,2)

13 # f_n Rekursionen:###
14 ### R-Ausgabe 4.3 #####
15 fn.sym.calc(f.str,n=10,all=T,eval=T,time=T)[c(1,5,10),]
16 invisible(fns.calc(f.str,n=10,time=T))
17 (fns.500 <- fns.calc(f.str,n=500,all=T,time=T))[c(1,5,10,100,500),]
18 diff(fns.500)[c(15:17),]
```

```

19 ##### Chunk Bemerkung 4.32 #####
20 fns.calc(f.str,n=500,s=c(0.0001,0.8,0.5))

21 # eta.calc Fixpunkt:####
22 ##### R-Ausgabe 4.4 #####
23 (eta <- eta.calc(f.str))
24 all.equal(fns.500[500,],eta,check.attributes=F)

25 #xtable.m(as.matrix(eta,nrow=length(f.str)),digits=8) # LATEX-Export (4.70)

26 # eta.sim Simulation:####
27 ##### R-Ausgabe 4.5 #####
28 max.gen <- 10; iter.vec <- c(5e4,5e5,5e6)
29 # (eta.hat <- set_rownames( # Ausgeblendet, lange Berechnungsdauer
30 # t(sapply(iter.vec,MGWP.sim,dists.sim.create(f.str),max.gen,only.eta=T)),
31 # paste0("iter=",iter.vec)))

32 # eta.hat speichern/laden:
33 # list.save(list(eta.hat),file=paste0(datapath,"eta.hat.RData"))
34 eta.hat <- list.load(paste0(datapath,"eta.hat.RData"))[[1]]

35 ##### Chunk Bemerkung 4.33 (i) Rest-AWK #####
36 unname((eta-fns.500[max.gen,]) / (1-fns.500[max.gen,]))

37 # Für xtable - Abweichungen:
38 report.m <- rbind(eta,fns.500[20,]-eta,
39                  eta.hat - matrix(rep(eta,3),nrow=3,byrow=T))
40 # xtable.eta(report.m,iter.vec) # LATEX-Export Tabelle 4.7

41 # Trajektorie simulieren / Momente bestimmen: ####
42 Z0 <- c(0,1,0); numexpath <- paste0(datapath,"Sim_numex/")
43 # MGWP.sim(iter.vec[2],dists.sim.create(f.str),max.gen,Z0=Z0,only.eta=F,
44   ↪ savepath=numexpath) # Ausgeblendet, lange Berechnungsdauer
45 traj <- MGWP.sim.load(iter.vec[2],max.gen,Z0,numexpath)$traj
46 MGWP.sim.load(iter.vec[2],max.gen,Z0,numexpath)$timer

47 ##### R-Ausgabe 4.6 #####
48 est.list <- MGWP.sim.est(traj,max.gen)
49 ### Schätzer für EW/Cov/AWK:
50 est.list[c(1,2,4)]
51 ### Vergleich: Berechnete Größen
52 EZn.calc(f.str,max.gen,Z0)
53 VZn.calc(f.str,max.gen,Z0)
54 eta.calc(f.str,Z0=Z0)$AWK

55 # Für Bemerkung 4.35 Vj:
56 Vj.list <- Vj.list.calc(f.str)
57 # invisible(sapply(Vj.list,xtable.m,2)) # LATEX-Export
58 # Alternativ sind die Vj berechenbar durch:
59 # apply(diag(3),1, function(i) VZn.calc(f.str,n=1,Z0=i),simplify = F)

60 # für Bemerkung 4.36 größere Z0:
61 ##### Chunk Bemerkung 4.36 (i) 'Rest-AWK' #####
62 c(eta.calc(f.str,Z0=c(1,0,1))$AWK, eta.calc(f.str,Z0=c(1,1,1))$AWK)

63 # für Bemerkung 4.36 (ii):
64 Zn.sample <- est.list[[3]]
65 # z10 für noch lebende Pfade ermitteln:
66 zn.surv <- Zn.sample[apply(Zn.sample,1,sum)!=0,,drop=F]; dim(zn.surv)
67 # 'Rest'-AWK der noch lebenden Pfade:
68 # AWK.zn.surv <- apply(zn.surv,1,function(i) eta.calc(f.str,Z0=i)$AWK) #
69   ↪ Ausgeblendet, längere Berechnungsdauer
70 # list.save(list(AWK.zn.surv),file=paste0(numexpath,"AWK.zn.surv.RData"))
71 AWK.zn.surv <- list.load(paste0(numexpath,"AWK.zn.surv.RData"))[[1]] # Laden
72 # die größte (Rest-)AWK davon:
73 i.max <- which(AWK.zn.surv==max(AWK.zn.surv))

```

B Appendix

```

72 # Gesamtpopulation der noch lebenden Relisierungen bei n = 10:
73 zn.surv.gesamtpop <- apply(zn.surv,1,sum)
74 # die kleinste Gesamtpopulation davon:
75 j.max <- which(zn.surv.gesamtpop==min(zn.surv.gesamtpop))
76 # sind i.max und j.max eindeutig und betreffen den gleichen Pfad? -> ja
77 i.max==j.max; length(i.max); length(j.max)
78 zn.surv[j.max, ]; AWK.zn.surv[i.max]           # Population und 'Rest'-AWK ausgeben
79 eta.calc(f.str,zn.surv[j.max,])$AWK           # Gegencheck

80 # für Bsp. 4.37 relative Häufigkeiten:
81 ##### Chunk 1 Bsp. 4.37 #####
82 n <- c(1,2,3,5,10); prop <- NULL
83 for(i in n){ mean.i <- MGWP.sim.est(traj,i)[[1]]
84               prop <- rbind(prop,mean.i / sum(mean.i)) }
85 set_rownames(prop,paste0("n=",n))

86 ##### Chunk 2 Bsp. 4.37 #####
87 M.calc(f.str)$u

88 ##### Chunk 3 Bsp. 4.37 #####
89 est.list.new <- MGWP.sim.est(
90   MGWP.sim(iter=1,dists.sim.create(f.str),max.gen=20,Z0=c(3,1,1000)),n=20)
91 est.list.new[[1]] / sum(est.list.new[[1]])

92 all.equal(eta.calc(f.str,Z0=c(3,1,1000))$AWK,0)
93 eta.calc(f.str,Z0=c(3,1,1000))$AWK==0

```

R-Quellcode für den Unterabschnitt 4.2.3

```

1 ##### Weitere Multitype GWP Unterabschnitt 4.2.2: #####
2 ##### R-Ausgabe 4.7 #####
3 f.str.A <- c("7/8+1/16*s1+1/16*s2^2", "15/17+2/17*s1*s2")
4 c(format(M.calc(f.str.A,check.PR=T)$rho,digits=6),M.calc(f.str.A)$Klasse)
5 f.str.B <- function(p) c(paste0(p,"s1+",1-p,"s2"),
6   paste0("1/2*",p,"s1+1/2*s2+1/2*",1-p,"s3"),
7   paste0(p,"s2+",1-p,"s3"))
8 c(M.calc(f.str.B(0.4),check.PR=T)$rho,M.calc(f.str.B(0.4))$Klasse)

9 ##### R-Ausgabe 4.8 #####
10 mu <- 1.1; (f.str.C <- paste0("1 / (1+",mu,"-s1*",mu,")"))
11 sapply(1:5,function(n) VZn.calc(f.str.C,Z0=1,n))

12 MGWP.sim(iter=100,dists=dists.geom.create(mu),max.gen=20,Z0 = 1,type="independent")

```

R-Quellcode für die Nebenrechnung B.16

```

1 ##### Nebenrechnung B.16: #####
2 # Eingabe f (4.67)
3 f.str <- c("4/7*s1^2+3/7*s2", "1/35+16/35*s1*s3+18/35*s2^2", "3/7*s2+4/7*s3^2")
4 # f2 als yac_symbol (exaktes Ergebnis mit "Rationalize"):
5 f2.sym <- y_fn(simplify(fn.sym.calc(f.str,n=2,time=T)),"Rationalize")
6 unlist(sapply(f2.sym,tex))           #  $\text{TeX-Output } f_2(s)$ 

7 ##### Chunk NR B.16: #####
8 # yac_symbol in string umwandeln:
9 f2.str <- f.sym.to.f.str(f2.sym)
10 # W-Funktion von  $Z_2^{(2)}$  (also mit  $Z_0 = (0,1,0)^T$ )
11 wfkt.calc(f2.str)$pr[wfkt.calc(f2.str)$pr[,1]!=0,1]
12 rbind(EZn.calc(f.str,n=2,Z0=c(3,2,5)), EZn.calc(f2.str,n=1,Z0=c(3,2,5)))

```

B.2.2 R-Quellcode für verwendete Funktionen und Abhängigkeiten

Vor der Ausführung der Quellcodes unter B.2.1 müssen die folgenden drei Codeteile wie jeweils angemerkt als .R-Quelldateien gespeichert werden. Anhand der beiden letzteren können die Funktionsweisen der dokumentierten R-Funktionen und jene der Hilfsfunktionen nachvollzogen werden.

R-Quellcode für Abhängigkeiten und Verzeichnisse

Dieser Quellcode sollte als MGWP_dependencies.R im sourcepath (siehe B.2.1) abgelegt werden (falls nötig, müssen zudem die Pakete zunächst mit `install.packages` installiert werden). In der Zeile 8 kann ein passendes Datenverzeichnis angegeben werden (welches manuell angelegt werden muss).

```

1 require(Branching); require(Ryacas); require(FixedPoint) # siehe 4.2
2 require(tictoc)      # für timer-Funktionen
3 require(rlang)       # für environment-handling
4 require(rlist)       # für list.save
5 require(magrittr)    # u.a. für set_colnames, pipe %>% (Letzteres nicht verwendet)
6 datapath <- paste0("D:/Data/") # hier Datenverzeichnis eingeben ('/' am Ende)

```

Es sollen noch die beiden R-Pakete 'tikzDevice' (Version 0.12.3.1 – vgl. [56]) und 'xtable' (Version 1.8.4 – vgl. [51]) gesondert erwähnt werden. Mit ersterem wurden fast alle Abbildungen (bis auf die Abb. 3.1) und mit dem zweiten alle Tabellen in dieser Arbeit generiert. Die Erstellung von Rohfassungen der Funktionsdokumentationen im Abschnitt 4.2.1 bewerkstelligte die Funktion `Rd2latex` aus 'tools'.

Im Unterabschnitt 4.2.1 dokumentierte R-Funktionen

Dieser Quellcode sollte als MGWP_functions.R im sourcepath abgelegt werden.

```

1 ##### Funktionen für die Analyse der weF #####
2 # Funktion(en) aus f-string erzeugen:
3 wef.create <- function(f.str,save=F){
4   k <- length(f.str)
5   # string mit Vektorinput:
6   f.str.vec <- f.str
7   for(i in 1:length(f.str.vec)){ f.str.vec <- gsub(paste0("s",i),
8     paste0("s[",i,"]"),f.str.vec) }
9   # 'Normale' Funktionen:
10  # f_j Liste (Eingabe s_1,...,s_k):
11  s.str <- paste0(rep("s",k),1:k,collapse=",")
12  fj.list <- sapply(1:k, function(i)
13    eval(parse(text = paste0("function(", s.str , ")" , f.str[i]))))
14  # f_j (Eingabe j und s_1,...,s_k):
15  eval(parse(text=paste0("fj <- function(",s.str,"j=k) fj.list[[j]](",s.str,")"))))
16  # f (Eingabe s_1,...,s_k):
17  eval(parse(text=paste0("f <- function(",s.str,") fj.list[[j]](",s.str,")"))))
18  eval(parse(text=paste0("f <- function(",s.str,")",
19    paste0("c(",paste(f.str,collapse=","),")"))))
20  # Vektorfunktionen:
21  # f_j Liste (Eingabe s):
22  fj.vec.list <- sapply(1:k,

```

B Appendix

```

23         function(i) eval(parse(text = paste0("function(s)", f.str.vec[i])))
24 # fj (Eingabe j und s):
25 fj.vec <- function(s,j=k){fj.vec.list[[j]](s)}
26 # f (Eingabe s):
27 eval(parse(text=paste0("f.vec <- function(s)",
28                         paste0("c(",paste(f.str.vec,collapse=", "),")")"))))
29 if(!isTRUE(all.equal(f.vec(rep(1,k)),rep(1,k)))) warning("f(1)=1 ist nicht erfüllt!")
30 if(save){ CE <- caller_env()
31   assign("fj.list",fj.list,envir=CE); assign("fj",fj,envir=CE)
32   assign("f",f,envir=CE); assign("fj.vec.list",fj.vec.list,envir=CE)
33   assign("fj.vec",fj.vec,envir=CE); assign("f.vec",f.vec,envir=CE)
34   assign("k",k,envir=CE); assign("f.str.vec",f.str.vec,envir=CE)
35 }else{ list(fj.list=fj.list, fj=fj, f=f,fj.vec.list=fj.vec.list, fj.vec=fj.vec,
36           f.vec=f.vec,f.str.vec=f.str.vec, k=k) }}

37 # p(r)-Berechnung:
38 wfkt.calc <- function(f.str,max.order = 10,red = TRUE, time = FALSE, time.break=20,
39                       loop = TRUE, tol.loop = 0.0001){
40   if(time) tic("pr-Berechnung")
41   k <- length(f.str); tol.break <- F
42   dims <- asplit(matrix(rep(0:max.order,k),ncol=k),2)
43   r.matrix <- expand.grid(dims); colnames(r.matrix) <- 1:k
44   if(!loop){ pr.matrix <- t(apply(r.matrix,1,function(r)
45     1/prod(factorial(r)) * dfr.e.calc(f.str,l=r,eval=T)))
46     if(k==1) pr.matrix <- t(pr.matrix)
47     tol.break <- isTRUE(all.equal(1,min(colSums(pr.matrix,na.rm=T)),
48                                   tolerance=tol.loop))
49   }else{ pr.matrix <- NULL; timer <- 0
50     for(i in 1:nrow(r.matrix)){ tic(); r <- as.numeric(r.matrix[i,])
51       pr <- 1/prod(factorial(r)) * dfr.e.calc(f.str,l=r,eval=T)
52       pr.matrix <- rbind(pr.matrix,pr)
53       loop.end <- toc(quiet=T); timer <- timer+loop.end$toc-loop.end$tic
54       if(timer>time.break){ # Abbruch, wenn es zu lange dauert
55         message(paste0("timeout: ",time.break," Sekunden")); break }
56       # Abbruch, wenn exakt 1:
57       if(isTRUE(all.equal(1,min(colSums(pr.matrix,na.rm=T))))) break
58       tol.break <- isTRUE(all.equal(1,min(colSums(pr.matrix,na.rm=T)),
59                                     tolerance=tol.loop))
60       if(tol.break) break }} # Abbruch, wenn fast 1
61   colnames(pr.matrix) <- paste0("p",1:k)
62   rownames(pr.matrix) <- paste0("r=",apply(r.matrix[1:nrow(pr.matrix),,drop=F],
63     1,paste0,collapse=",",""))
64   r.matrix <- r.matrix[1:nrow(pr.matrix),,drop=F]
65   if(red) { red.idx <- rowSums(pr.matrix)!=0 # reduzieren auf p(r) > 0?
66     r.matrix <- r.matrix[red.idx,,drop=F]
67     pr.matrix <- pr.matrix[red.idx,,drop=F] }
68   rownames(r.matrix) <- paste0("r",1:nrow(r.matrix))
69   if(!isTRUE(all.equal(1,min(colSums(pr.matrix,na.rm=T))))) {
70     warning(paste0("Mind. eine Verteilung summiert sich nicht auf 1 (",
71                   ifelse(tol.break,"innerhalb","AUSSERHALB"),
72                   " der angegebenen Toleranz)."),call.=F) }
73   if(sum(is.na(pr.matrix))>0) warning("NAs produziert.",call.=F)
74   if(time) toc(); list(pr=pr.matrix,r.matrix=r.matrix) }

75 # EW-Matrix M aus wef berechnen:
76 M.calc <- function(f.str,check.PR = FALSE, n.check.PR = 10){
77   k <- length(f.str)
78   M <- apply(diag(k),1,function(l) dfr.e.calc(f.str,l,s=rep(1,k)))
79   # Eigenwert- und -vektorberechnung
80   eigM <- eigen(M)
81   rho <- max(eigM$values)
82   # rechter EV v und linker EV u:
83   v <- setNames(eigM$vectors[,which(eigM$values==rho)],paste0("v",1:k))
84   u <- setNames(t(eigen(t(M))$vectors)[which(eigM$values==rho),],paste0("u",1:k))
85   # Überprüfung der Eigenwert-Gleichungen:
86   if(isFALSE(all.equal(t(t(rho*v)),M%*%v))) warning("EW-Gleichung für v nicht erfüllt")

```

```

87 if(isFALSE(all.equal(rho*t(u),u/%M))) warning("EW-Gleichung für u nicht erfüllt")
88 # u normieren und v so skalieren, dass u/%v=1 ist:
89 (u<-u/sum(u)); v <- (v/as.numeric(u/%v))
90 # ggf. Überprüfung der positiven Regularität:
91 if(check.PR) ifelse(Reduce("|",sapply(Mn.calc(M.calc(f.str)$M,n.check.PR,T),
92                                     function(i) prod(i>0))),
93                     print("M ist garantiert positiv regulär."),
94                     print(paste0("Positive Regularität von M zumindest bis M^",
95                                   n.check.PR," nicht feststellbar.")))
96 list(M=M,rho=rho,u=u,v=v,
97      Klasse=ifelse(rho<1,"subkritisch",ifelse(rho==1,"kritisch","superkritisch"))) }

98 # f_n-Rekursionen - explizite Funktionen:
99 fn.sym.calc <- function(f.str,n=1,eval=FALSE,s=rep(0,length(f.str)),all=FALSE,
100                      include.f0 = FALSE, time = FALSE){
101   if(time) ifelse(eval,tic("Gesamt"),tic("Verkettung"))
102   if(time & eval) tic("Verkettung (ohne Evaluierung)")
103   k <- length(f.str)
104   if(k==1) f.vec <- wef.create(f.str)$f
105   else f.vec <- wef.create(f.str)$f.vec
106   ### Funktion für Rekursion:
107   fn.sym.rekursion <- function(f.sym,n,fn.sym=f.sym,all=all,m=1){
108     if(n==1){ if(all) return(setNames(list(fn.sym),paste0("f",m)))
109               else return(fn.sym) }
110     if(n>1){ rekursion <- fn.sym.rekursion(f.sym,n-1,f.vec(fn.sym),all=all,m=m+1)
111             if(all) c(setNames(list(fn.sym),paste0("f",m)),rekursion)
112             else rekursion }
113   fn.sym <- fn.sym.rekursion(ysym(f.str),n,all=all) # eigentliche Berechnung
114   if(time & eval) toc()
115   if(eval){ if(time) tic("Evaluierung")
116             eval.fn.exp <- function(fn.exp){eval(parse(text=paste0("s",1:k,"<-s[",1:k,"]")))}
117             eval(fn.exp) }
118   if(all){ fn.sym <- t(simplify2array(lapply(lapply(fn.sym,as_r),
119                                               function(i) eval.fn.exp(i))))
120           if(include.f0) fn.sym <- rbind(s,fn.sym)
121           rownames(fn.sym) <- paste0("f",
122                                     ifelse(include.f0,0,1):n,"(",paste0(s,collapse=","),")")
123   }else fn.sym <- setNames(eval.fn.exp(as_r(fn.sym)),
124                           paste0("f_",n,"(",1:k,"")(",paste0(s,collapse=","),")"))
125   if(time){ toc();toc() }
126   }else if(time) toc(); return(fn.sym) }

127 # f_n-Rekursionen - direkt:
128 fns.calc <- function(f.str, n = 1, s = rep(0,length(f.str)), all = FALSE,
129                    include.f0 = FALSE, time = FALSE){
130   if(time) tic("fn Rekursion")
131   f.vec <- wef.create(f.str)$f.vec; k <- length(f.str)
132   if(length(s)!=k) stop("Dimension von s stimmt nicht mit der von f überein.")
133   ### Funktion für Rekursion:
134   fns.rekursion <- function(f.vec,n=1,s=rep(0,k),all=T){
135     if(n==1) return(f.vec(s))
136     if(n>1){ if(all) return(rbind(f.vec(s),fns.rekursion(f.vec,n-1,f.vec(s),all=all)))
137             else return(fns.rekursion(f.vec,n-1,f.vec(s),all=all)) }
138   fns <- fns.rekursion(f.vec,n,s,all) # eigentliche Berechnung
139   if(all){ if(include.f0) fns <- rbind(s,fns)
140           rownames(fns) <- paste0("f",ifelse(include.f0,0,1):n,
141                                     "(",paste0(s,collapse=","),")")
142   }else names(fns) <- paste0("f",n,"(",1:k,"")(",paste0(s,collapse=","),")")
143   if(time) toc(); return(fns) }

144 # η und AWK-Berechnung:
145 eta.calc <- function(f.str = NULL, Z0 = NULL,
146                    f.vec = wef.create(f.str)$f.vec, k=length(f.str)){
147   if(k==0) stop("Wenn f.vec statt f.str angegeben wird, bitte k mitgeben.")
148   eta <- setNames(FixedPoint(f.vec,Inputs=rep(0,k),Method="Aitken")$FixedPoint,
149                  paste0("eta(",1:k,""))

```

B Appendix

```

150   if(is.null(ZO)) eta
151   else list(eta=eta,AWK=prod(sapply(1:k,function(i) eta[i]^ZO[i]))) }

152   ##### Funktionen für die Bestimmung der Momente #####
153   # EW-Vektoren  $\mathbb{E}Z_n$  berechnen:
154   EZn.calc <- function(f.str=NULL,n,ZO,M=M.calc(f.str)$M,all=FALSE,
155                        include.EZO=FALSE|(n==0),simplify=all){
156     if(!(is.vector(ZO) & ifelse(length(ZO)==1,T,length(ZO)==ncol(M))))
157       stop("Länge/Format von ZO passt nicht.")
158     if(all & n==0 & !include.EZO) stop("Das ist nicht möglich.")
159     if(all){ outlist <- setNames(lapply(Mn.calc(M,n,all,includeM0=include.EZO),
160                                       function(i) as.vector(t(i)%*%ZO)),
161                                paste0("EZ", (1-include.EZO*1):n))
162       if(simplify) simplify2array(outlist)
163     } else outlist
164   } else setNames(as.vector(t(Mn.calc(M,n,all))%*%ZO),paste0("EZ",n,"_",1:length(ZO))) }

165   # Kovarianzmatrizen  $\mathbb{V}Z_n$  berechnen:
166   VZn.calc <- function(f.str=NULL,n,ZO,M=M.calc(f.str)$M){
167     if(!(is.vector(ZO) & ifelse(length(ZO)==1,T,length(ZO)==ncol(M))))
168       stop("Länge/Format von ZO passt nicht.")
169     Vj.list <- Vj.list.calc(f.str)
170     #  $\mathbb{E}Z_0, \mathbb{E}Z_1, \dots, \mathbb{E}Z_{n-1}$ :
171     EZnm1 <- EZn.calc(n=n-1,M=M,ZO=ZO,all = T,include.EZO = T,simplify = F)
172     # innere Summe in (3.75):
173     innersum <- function(i,M,ZO,Vj.list,k=length(ZO)){
174       Reduce("+",lapply(1:k, function(j) EZnm1[[i]][j]*Vj.list[[j]])) }
175     M.nm1.list <- Mn.calc(M,n,list=T,includeM0 = T) #  $M_0, M_1, \dots, M_n$ 
176     # äußere Summe in (3.75)
177     Reduce("+",lapply(1:n, function(i)
178       t(M.nm1.list[[n-i+1]])%*%innersum(i,M,ZO,Vj.list)%*%M.nm1.list[[n-i+1]])) }

179   ##### Funktionen für Simulationen und Schätzungen #####
180   # Verteilung für 'Branching' erstellen:
181   dists.sim.create <- function(f.str = NULL, wfkt = wfkt.calc(f.str)){
182     pr = wfkt$pr; r.matrix = wfkt$r.matrix; k = ncol(pr)
183     size <- setNames(as.list(colSums(pr!=0,na.rm = T)),paste0("dist",1:ncol(pr)))
184     prob <- setNames(lapply(1:k,function(i) subset(pr[,i],pr[,i]!=0)),
185                      paste0("dist",1:ncol(pr)))
186     vect <- setNames(lapply(1:k,function(i) subset(r.matrix,pr[,i]!=0)),
187                      paste0("dist",1:ncol(pr)))
188     list(sizes=size,probs=prob,vects=vect) }

189   # MGWP simulieren (mit rBGM aus 'Branching'):
190   MGWP.sim <- function(iter,dists,max.gen=10,ZO=NULL,only.eta=FALSE,
191                        savepath=NULL,type="general",time=TRUE){
192     k <- length(dists[[1]])
193     if(only.eta){
194       if(time) tic(paste0("Simulation eta.hat (" ,k,"x",iter,""))
195       schwellen <- FA.schwellen(iter*k,time*100); est.eta <- NULL
196       for(j in 1:k){ ZO <- diag(k)[j,]
197         AST <- sapply(1:iter,function(i){FA.check(i+(j-1)*iter,schwellen)
198           sum((rBGM(dists=dists,type=type,d=k,n=max.gen,z0=ZO)$o.tt.s)[max.gen,])==0})
199         est.eta[j] <- sum(AST)/iter }; if(time) toc()
200       setNames(est.eta,paste0("eta.hat(",1:k,""))
201     } else{ if(is.null(ZO)) stop("Bitte ZO eingeben oder only.eta=TRUE setzen.")
202       if(time) tic(paste0("Simulation Trajektorien (" ,iter,""))
203       schwellen <- FA.schwellen(iter,time*100); traj <- NULL
204       for(i in 1:iter){
205         traj[[i]] <- rbind(ZO,rBGM(dists=dists,type=type,d=k,n=max.gen,z0=ZO)$o.tt.s)
206         rownames(traj[[i]]) <- paste0("Z",0:max.gen)
207         FA.check(i,schwellen) }
208       if(time) timerstop <- toc()
209       if(!is.null(savepath)){
210         if(!dir.exists(savepath)) dir.create(savepath)
211         list.save(list(traj=traj,
```



```

212         timer=ifelse(time,timerstop$toc - timerstop$tic,"deaktiviert")),
213         file=paste0(savepath,"iter",iter,"_gen",max.gen,
214                     "_Z0(",paste0(Z0,collapse=","),").RData"))
215     }else traj }}

216 # Simulationsergebnisse laden:
217 MGWP.sim.load <- function(iter,max.gen=10,Z0,loadpath){
218   if(!dir.exists(loadpath)) stop("Verzeichnis existiert nicht.")
219   loadfile <- paste0(loadpath,"iter",iter,"_gen",max.gen,
220                       "_Z0(",paste0(Z0,collapse=","),").RData")
221   if(!file.exists(loadfile))
222     stop("Datei nicht vorhanden - bitte zuerst Simulation durchführen.")
223   list.load(loadfile) }

224 ### Simulationsergebnisse - Kenngrößen schätzen:
225 MGWP.sim.est <- function(traj,n=NULL){
226   k <- ncol(traj)[[1]]
227   if(is.null(n)){ n <- nrow(traj)[[1]]-1; only.eta <- T
228   }else only.eta <- F
229   Zn.sample <- t(apply(traj,function(i) i[n+1,]))
230   if(k==1){ Zn.sample <- t(Zn.sample); AWK <- sum(Zn.sample==0)/length(Zn.sample)
231   }else AWK <- sum(apply(Zn.sample,1,sum)==0)/nrow(Zn.sample)
232   if(!only.eta) setNames(list(apply(Zn.sample,2,mean),
233                                 cov(Zn.sample), Zn.sample, AWK),
234                           c(paste0("Z",n,".",c("mean","cov","sample")), "AWK"))
235   else AWK }

```

R-Helferfunktionen

Dieser Quellcode sollte als MGWP_helper.R im sourcepath abgelegt werden.

```

1  ##### Allgemeine Helferfunktionen: #####
2  # Helferfunktion yac_symbol in string umwandeln (für die NR B.16):
3  f.sym.to.f.str <- function(f.sym) unlist(strsplit(gsub("^|.\\$", "", yac(f.sym)), ","))

4  # Forstschrittsanzeige (implizit aufgerufen):
5  FA.schwellen <- function(iter,F.Anzeige=100){
6    ifelse(!(F.Anzeige==F),
7    return(floor(seq(iter/F.Anzeige,iter,length.out=F.Anzeige))),return(NULL)) }
8  FA.check <- function(i,schwellen){
9    if(i %in% schwellen){
10      cat(paste0("\r",format(schwellen[which(schwellen==i)]/tail(schwellen,1)*100,
11                            digits=2,nsml=2,decimal.mark=","),
12          "% fertiggestellt",ifelse(i %in% tail(schwellen,1),"n","")) }
13    flush.console() }

14 # Verteilung für rBGWM aus 'Branching' aufbereiten (siehe das Bsp. 4.38)
15 dists.geom.create <- function(mu=NULL,p=NULL){
16   if((is.null(p)&is.null(mu)) | (!is.null(p)&!is.null(mu)))
17     stop("Bitte mu ODER p eingeben.")
18   if(is.null(p)) p <- 1/(1+mu)
19   data.frame( name="geom", param1=p, stringsAsFactors=F) }

20 ##### Helferfunktionen Momente (implizit aufgerufen): #####
21 # M^n Rekursion (n-fache Matrixmultiplikation) - einzeln oder Liste
22 # (includeM0 nur mit list relevant; Achtung: dann ist der Index um 1 verschoben)
23 Mn.calc <- function(M,n,list=F,includeM0=F,Mn=M,m=1){
24   k <- max(1,ncol(M))
25   if(n==0){ if(list) return(setNames(list(diag(k)), "M0"))
26   else return(diag(k)) }
27   if(n==1){ if(list){if(m>1|!includeM0) return(setNames(list(Mn),paste0("M",m)))
28   else return(setNames(list(diag(k),Mn),paste0("M",c(0,1))))}
29   }else return(Mn) }

```

B Appendix

```

30 if(n>1){ rekursion <- Mn.calc(M,n-1,list,Mn=Mn/%M,m=m+1)
31 if(list) c(setNames(list(Mn),paste0("M",m)),rekursion)
32 if(list){ outlist <- c(setNames(list(Mn),paste0("M",m)),rekursion)
33 if(includeM0) c(setNames(list(diag(k),Mn),paste0("M",c(m-1,m))),rekursion)
34 else c(setNames(list(Mn),paste0("M",m)),rekursion)
35 }else rekursion }}

36 # Kovarianzen  $Cov[X_i^{(j)}, X_l^{(j)}]$  aus weF berechnen:
37 covXj_il.calc <- function(f.str,j,i,l){
38 k <- length(f.str); s <- rep(1,k); fstr <- f.str[[j]]
39 dfjr.e.calc(fstr,k,diag(k)[i,]+diag(k)[l,],s,T) -
40 dfjr.e.calc(fstr,k,diag(k)[i,],s,T)*dfjr.e.calc(fstr,k,diag(k)[l,],s,T) +
41 (i==l)*dfjr.e.calc(fstr,k,diag(k)[i,],s,T) }

42 # Kovarianzmatrizen  $V_j = \mathbb{V}X^{(j)}$  aus weF berechnen
43 Vj.list.calc <- function(f.str){
44 Vj.list <- NULL; k <- length(f.str)
45 for(j in 1:k){ Vj <- matrix(NA,nrow=k,ncol=k)
46 for(i in 1:k){ for(l in 1:k) Vj[i,l] <- covXj_il.calc(f.str,j,i,l) }
47 Vj.list[[j]] <- Vj }
48 names(Vj.list) <- paste0("V",1:k); Vj.list }

49 ##### Ableitungen (implizit aufgerufen): #####
50 # Rekursion für höhere Ableitungen l-ter Ordnung:
51 dl <- function(expr, name, l=1) {
52 if(l < 1) return(expr)
53 if(l == 1) D(expr, name)
54 else dl(D(expr, name), name, l-1) }

55 # partielle Ableitung  $f^{(j)}$  nach  $s_1^{l_1}, \dots, s_k^{l_k}$ :
56 dfjr.e.calc <- function(fj.str,k=3,l=rep(0,k),s=rep(0,k),eval=F){
57 d.fjr.e <- parse(text=fj.str)
58 for(j in 1:k){ d.fjr.e <- as.expression(dl(d.fjr.e, paste0("s",j), l = 1[j])) }
59 if(eval){ # s als Variable:
60 for(j in 1:k) eval(parse(text = paste0("s",j,"<=",s[j])))
61 return(eval(d.fjr.e))
62 }else return(d.fjr.e) }

63 # partielle Ableitung f nach  $s_1^{l_1}, \dots, s_k^{l_k}$  (dfjr.e.calc wird dafür benötigt)
64 dfr.e.calc <- function(f.str,l=rep(0,length(f.str)),s=rep(0,length(f.str)),eval=T){
65 k <- length(f.str); d.fr.e <- NULL
66 for(i in 1:k){ d.fjr.e <- dfjr.e.calc(f.str[[i]],k=k,l=l,s=s,eval=eval)
67 if(eval) d.fr.e <- c(d.fr.e,d.fjr.e)
68 else d.fr.e <- c(d.fr.e,as.expression(d.fjr.e)) }
69 d.fr.e }

```

Abkürzungsverzeichnis

Add.	Additivität (oder tw. <i>Additive</i>)
Ann.	Annahme(n)
AWK	Aussterbewahrscheinlichkeit(en)
CAS	Computeralgebrasystem
C-Funktion	charakteristische Funktion
deMorgan	de Morgansche Regeln
Distr.	Distributiv(-gesetz)
EW	Erwartungswert(e)
f.s.	fast sicher
f.ü.	fast überall
GW	Galton-Watson
GWP	Galton-Watson-(Verzweigungs-)prozess(e)
i.A.	im Allgemeinen
i.a.W.	in anderen Worten
<i>i.i.d.</i>	independent and identically distributed
Ind.-Ann.	Induktions-Annahme
ME	Markov-Eigenschaft
MGWP	Multitype Galton-Watson-(Verzweigungs-)prozess(e)
MON	Monotone Konvergenz
MW	Mittelwert
NR	Nebenrechnung
o.B.d.A.	ohne Beschränkung der Allgemeinheit
paarw. disj.	paarweise disjunkt
s^2	(korrigierte) Stichprobenvarianz
tot. EW	totaler Erwartungswert
uabh.	unabhängig
weF	wahrscheinlichkeitserzeugende Funktion(en)
W-Funktion	Wahrscheinlichkeitsfunktion
WK	Wahrscheinlichkeit(en)
W-Maß	Wahrscheinlichkeitsmaß
W-Raum	Wahrscheinlichkeitsraum
ZV	Zufallsvariable(n)
Z-Vektor	Zufallsvektor

Abbildungsverzeichnis

2.1	WeF für GWP mit geometrischen Nachkommensverteilungen.	44
3.1	Schematische denkbare Realisierung eines MGWP mit $k = 3$	68
4.1	Konvergenz der Folge $\{f_n\}_{n \geq 0}$ – geometrische Fälle.	105
4.2	Beispielhafte Pfade eines superkritischen geometrischen GWP.	110
4.3	Histogramm der noch lebenden Realisierungen von W_{40} – superkritischer geometrischer Fall.	111

Tabellenverzeichnis

4.1	W-Funktion von Z_5 im superkritischen geometrischen Fall.	103
4.2	Momente von Z_n im superkritischen geometrischen Fall.	103
4.3	Konvergenz der Folge $\{f_n\}_{n \geq 0}$ – geometrische Fälle.	105
4.4	WK für den Aussterbezeitpunkt τ – superkritischer geometrischer Fall. . .	115
4.5	Verteilung bzw. W-Funktion von T – geometrische Fälle.	125
4.6	Verteilungsfunktion von T – geometrische Fälle.	125
4.7	Vergleich der ermittelten Werte bzw. Schätzer für η	145

Verzeichnis dokumentierter R-Funktionen

<code>wef.create</code>		131
<code>wfkt.calc</code>		131
<code>M.calc</code>		132
<code>fn.sym.calc</code> , <code>fns.calc</code>		133
<code>eta.calc</code>		134
<code>EZn.calc</code> , <code>VZn.calc</code>		135
<code>dists.sim.create</code>		136
<code>MGWP.sim</code> , <code>MGWP.sim.load</code>		136
<code>MGWP.sim.est</code>		138

Verzeichnis der R-Ausgaben

4.1	Numerisches Bsp.: Berechnung der Reproduktionsverteilungen.	140
4.2	Numerisches Bsp.: Analyse von $\mathbf{M}$	140
4.3	Numerisches Bsp.: Bestimmung von η durch f_n -Iterationen.	142
4.4	Numerisches Bsp.: Bestimmung von η mit <code>eta.calc</code>	143
4.5	Numerisches Bsp.: Schätzungen $\hat{\eta}$ für η	143
4.6	Numerisches Bsp.: Schätzungen mit <code>MGWP.sim.est</code> / Berechnungen mit <code>eta.calc</code> , <code>EZn.calc</code> und <code>VZn.calc</code>	146
4.7	Weitere MGWP: Numerische Beispiele.	148
4.8	Weitere MGWP: Varianzberechnung für den geometrischen Fall.	149

Literaturverzeichnis

- [1] M. M. Andersen und S. Højsgaard. „Ryacas: A Computer Algebra System in R“. In: *Journal of Open Source Software* 4.42 (Okt. 2019). URL: <https://doi.org/10.21105/joss.01763>.
- [2] D. Atanasov, V. Stoimenova und N. M. Yanev. „Branching Process Modelling of COVID-19 Pandemic Including Immunity and Vaccination“. In: *Stochastics and Quality Control* 36.2 (Dez. 2021), S. 157–164. URL: <https://doi.org/10.1515/eqc-2021-0040>.
- [3] K. B. Athreya und P. E. Ney. *Branching Processes*. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete 196. Berlin: Springer-Verlag, 1972.
- [4] H. Bauer. *Measure and Integration Theory*. De Gruyter Studies in Mathematics 26. Berlin: Walter de Gruyter, 2001.
- [5] E. Behrends. *Elementare Stochastik: Ein Lernbuch - von Studierenden mitentwickelt*. Wiesbaden: Springer Spektrum, 2013.
- [6] P. Billingsley. *Probability and Measure*. 3rd ed. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons, 1995.
- [7] K. G. Binmore. *Mathematical Analysis: A Straightforward Approach*. 2nd ed. Cambridge: Cambridge University Press, 1982.
- [8] O. Deiser. *Einführung in die Mengenlehre: Die Mengenlehre Georg Cantors und ihre Axiomatisierung durch Ernst Zermelo*. 2., verb. und erw. Aufl. Berlin: Springer-Verlag, 2004.
- [9] L. Devroye. „Branching Processes and Their Applications in the Analysis of Tree Structures and Tree Algorithms“. In: *Probabilistic Methods for Algorithmic Discrete Mathematics*. Hrsg. von M. Habib, C. McDiarmid, J. Ramirez-Alfonsin und B. Reed. Algorithms and Combinatorics 16. Berlin: Springer-Verlag, 1998, S. 249–314.
- [10] M. Drmota, B. Gittenberger, G. Karigl und A. Panholzer. *Mathematik für Informatik*. 4., erw. Aufl. Berliner Studienreihe zur Mathematik Bd. 17. Lemgo: Heldermann Verlag, 2007.
- [11] J. Elstrodt. *Maß- und Integrationstheorie*. 6., korrigierte Aufl. Berlin: Springer-Verlag, 2009.
- [12] W. Feller. *An Introduction to Probability Theory and Its Applications: Volume I*. 3. ed., rev. print. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons, 1968.

- [13] O. Forster. *Analysis 1: Differential- und Integralrechnung einer Veränderlichen*. 9., überarb. Aufl. Wiesbaden: Friedr. Vieweg & Sohn Verlag, 2008.
- [14] G. Geenens und C. Wang. „Local-Likelihood Transformation Kernel Density Estimation for Positive Random Variables“. In: *Journal of Computational and Graphical Statistics* 27.4 (Juni 2018), S. 822–835. URL: <https://doi.org/10.1080/10618600.2018.1424636>.
- [15] H.-O. Georgii. *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. 2., bearb. Aufl. Berlin: Walter De Gruyter, 2004.
- [16] H. Gould und J. Quaintance. „Double Fun with Double Factorials“. In: *Mathematics Magazine* 85.3 (Juni 2012), S. 177–192. URL: <https://doi.org/10.4169/math.mag.85.3.177>.
- [17] G. Grimmett und D. Stirzaker. *Probability and Random Processes*. 3rd ed. Oxford: Oxford University Press, 2001.
- [18] P. Haccou, P. Jagers und V. A. Vatutin. *Branching Processes: Variation, Growth, and Extinction of Populations*. Cambridge Studies in Adaptive Dynamics 5. Cambridge: Cambridge University Press, 2005.
- [19] T. E. Harris. *The Theory of Branching Processes*. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete 119. Berlin: Springer-Verlag, 1963.
- [20] N. Henze. *Stochastik für Einsteiger: Eine Einführung in die faszinierende Welt des Zufalls*. 7., überarb. u. erw. Aufl. Wiesbaden: Friedr. Vieweg & Sohn Verlag, 2008.
- [21] C. Hesse. *Angewandte Wahrscheinlichkeitstheorie: Eine fundierte Einführung mit über 500 realitätsnahen Beispielen und Aufgaben*. Braunschweig: Friedr. Vieweg & Sohn Verlag, 2003.
- [22] H. Heuser. *Lehrbuch der Analysis: Teil 1*. 15., durchges. Aufl. Mathematische Leitfäden. Stuttgart: B. G. Teubner, 2003.
- [23] H. Heuser. *Lehrbuch der Analysis: Teil 2*. 12., durchges. Aufl. Mathematische Leitfäden. Stuttgart: B. G. Teubner, 2002.
- [24] R. v. d. Hofstad. *Random Graphs and Complex Networks: Volume I*. Cambridge Series in Statistical and Probabilistic Mathematics 43. Cambridge: Cambridge University Press, 2017.
- [25] P. Jagers. *Branching Processes with Biological Applications*. Wiley Series in Probability and Mathematical Statistics. London: John Wiley & Sons, 1975.
- [26] O. Jones, R. Maillardet und A. Robinson. *Introduction to Scientific Programming and Simulation Using R*. 2nd ed. The R Series. Boca Raton: Taylor & Francis Group, 2014.
- [27] S. Karlin und H. M. Taylor. *A First Course in Stochastic Processes*. 2nd ed. San Diego: Academic Press, 1975.

- [28] H. Kesten und B. P. Stigum. „Additional Limit Theorems for Indecomposable Multidimensional Galton-Watson Processes“. In: *The Annals of Mathematical Statistics* 37.6 (Dez. 1966), S. 1463–1481.
- [29] M. Kimmel und D. E. Axelrod. *Branching Processes in Biology*. 2nd ed. Interdisciplinary Applied Mathematics 19. New York: Springer Science & Business Media, 2015.
- [30] A. Klenke. *Wahrscheinlichkeitstheorie*. 3., überarb. u. erg. Aufl. Berlin: Springer Spektrum, 2013.
- [31] K. Königsberger. *Analysis 1*. 5., neu bearb. Aufl. Berlin: Springer-Verlag, 2001.
- [32] K. Königsberger. *Analysis 2*. 3., überarb. Aufl. Berlin: Springer-Verlag, 2000.
- [33] T. Koshy. *Catalan Numbers with Applications*. Oxford: Oxford University Press, 2009.
- [34] H.-J. Kowalsky und G. O. Michler. *Lineare Algebra*. 12., überarb. Aufl. Berlin: Walter de Gruyter, 2003.
- [35] U. Krengel. *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. 8., erw. Aufl. Wiesbaden: Friedr. Vieweg & Sohn Verlag, 2005.
- [36] N. Kusolitsch. *Maß- und Wahrscheinlichkeitstheorie: Eine Einführung*. 2., überarb. u. erw. Aufl. Berlin: Springer Spektrum, 2014.
- [37] M. Loève. *Probability Theory I*. 4th ed. Graduate Texts in Mathematics 45. New York: Springer Science & Business Media, 1977.
- [38] D. Meintrup und S. Schäffler. *Stochastik: Theorie und Anwendungen*. Statistik und ihre Anwendungen. Berlin: Springer-Verlag, 2005.
- [39] C. J. Mode. *Multitype Branching Processes: Theory and Applications*. Modern Analytic and Computational Methods in Science and Mathematics 34. New York: American Elsevier Publishing Company, 1971.
- [40] J. Neveu. *Discrete-Parameter Martingales*. Amsterdam: North-Holland Publishing Company, 1975.
- [41] J. R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics 2. Cambridge: Cambridge University Press, 1998.
- [42] K. D. Schmidt. *Maß und Wahrscheinlichkeit*. 2., durchges. Aufl. Heidelberg: Springer-Verlag, 2011.
- [43] A. W. v. d. Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics 3. Cambridge: Cambridge University Press, 1998.
- [44] M. G. Velasco, I. Garcia und G. Yanev. *Controlled Branching Processes*. Branching Processes, Branching Random Walks and Branching Particle Fields Set Vol. 2. London: ISTE/John Wiley & Sons, 2018.
- [45] H. W. Watson und F. Galton. „On the Probability of the Extinction of Families.“ In: *The Journal of the Anthropological Institute of Great Britain and Ireland* 4 (1875), S. 138–144.

- [46] H. S. Wilf. *Generatingfunctionology*. 3rd ed. Wellesley: A K Peters, 2006.
- [47] D. Williams. *Probability with Martingales*. 12th print. Cambridge: Cambridge University Press, 2010.
- [48] K.-U. Witt und M. E. Müller. *Algorithmische Informationstheorie: Berechenbarkeit und Komplexität verstehen*. Berlin: Springer Spektrum, 2020.

Software- und Paketressourcen

- [49] M. M. Andersen u. a. *Ryacas: R Interface to the 'Yacas' Computer Algebra System*. 16. Jan. 2023. URL: <https://CRAN.R-project.org/package=Ryacas>.
- [50] S. Baumann und M. Klymak. *FixedPoint: Algorithms for Finding Fixed Point Vectors of Functions*. 19. Dez. 2022. URL: <https://CRAN.R-project.org/package=FixedPoint>.
- [51] D. B. Dahl, D. Scott, C. Roosen, A. Magnusson und J. Swinton. *xtable: Export Tables to LaTeX or HTML*. 21. Apr. 2019. URL: <https://CRAN.R-project.org/package=xtable>.
- [52] G. Mazur. *Yacas. Yet Another Computer Algebra System*. URL: <http://www.yacas.org/index.html> (zuletzt besucht am 23. Feb. 2023).
- [53] H. D. Nguyen, A. T. Jones und G. J. McLachlan. *logKDE: Computing Log-Transformed Kernel Density Estimates for Positive Data*. 9. Aug. 2018. URL: <https://CRAN.R-project.org/package=logKDE>.
- [54] A. Pinkus, S. Winnitzky und M. Grzegorz. *Yacas. Reference Manual*. URL: https://yacas.readthedocs.io/en/latest/reference_manual/index.html (zuletzt besucht am 23. Feb. 2023).
- [55] R Core Team. *R: A Language and Environment for Statistical Computing*. Wien: R Foundation for Statistical Computing, 2022. URL: <https://www.R-project.org/>.
- [56] C. Sharpsteen und C. Bracken. *tikzDevice: R Graphics Output in LaTeX Format*. 12. Jan. 2020. URL: <https://CRAN.R-project.org/package=tikzDevice>.
- [57] C. J. Torres-Jimenez. *Branching: Simulation and Estimation for Branching Processes*. 6. Dez. 2016. URL: <https://CRAN.R-project.org/package=Branching>.