

Einführung in die mathematische Modellierung

Claus Peter Ortlieb

Sommersemester 2009

Inhaltsverzeichnis

1 Exakte Naturwissenschaft und Modellbegriff	1
1.1 Die Anfänge: Galileis Fallgesetze	2
1.1.1 Der empiristische Mythos und die Legende von Pisa	2
1.1.2 Beweis des Ersten Fallgesetzes	5
1.1.3 Beweis des Zweiten Fallgesetzes	6
1.1.4 Experimente: Die Produktion von Tatsachen	9
1.1.5 Zur Methodenabhängigkeit der Erkenntnisse	10
1.2 Newtons Instrumente: Infinitesimalrechnung und Differentialgleichungen	11
1.2.1 Das Gesetz der Veränderung als neuer mathematischer Gegenstand	11
1.2.2 Gleichförmig beschleunigte Bewegung	13
1.2.3 Im Verhältnis zur Geschwindigkeit gebremste Bewegung	14
1.2.4 Beschleunigung gegen den Widerstand eines Mediums: Linearer Fall	15
1.2.5 Beschleunigung gegen den Widerstand eines Mediums: Quadratischer Fall	16
1.2.6 Beschleunigung gegen den Widerstand eines Mediums: Allgemeiner Fall	17
1.2.7 Galileis Zeitmessungen	18
1.3 Trennung von Mathematik und Physik und Verallgemeinerung ihrer Methode	20
1.3.1 Nichteuklidische Geometrie	20
1.3.2 Axiomatische Methode und Angewandte Mathematik	23
1.3.3 Der Modellbegriff	24
1.3.4 Zur Gültigkeit von Modellen	27
1.4 Vom realen Problem zum mathematischen Modell	28
1.4.1 Der Modellierungsprozess	29
1.4.2 Modellierungs-„Rezepte“	29
1.4.3 Simulationswerkzeuge	31
1.4.4 Klassifikation von Modellen	31
1.5 Literatur	33
2 Die Dynamik von Wachstum und Zerfall	35
2.1 Exponentielle Zerfallsprozesse: Vier Modelltypen	35
2.1.1 Kontinuierliche Zustände, diskrete Zeit: Iterationsprozess	35
2.1.2 Kontinuierliche Zustände, kontinuierliche Zeit: Differentialgleichung	36
2.1.3 Zusammenhang von diskreter und kontinuierlicher Lösung	37
2.1.4 Halbwertszeit	37
2.1.5 Diskrete Zustände, diskrete Zeit: Markov-Prozess	37
2.1.6 Diskrete Zustände, kontinuierliche Zeit: Kontinuierlicher Markov-Prozess	41
2.1.7 Zusammenfassung	44
2.1.8 Methodischer Exkurs: Erzeugende Funktionen	45
2.2 Exponentielle Zerfallsprozesse: Modelle und Daten	47
2.2.1 Stochastische Simulationen	48
2.2.2 Anpassung der Modellparameter an Daten	48
2.2.3 Ein nichtlineares Ausgleichsproblem	49
2.2.4 Logarithmierte Daten: Lineares Ausgleichsproblem	49

2.2.5	Zur Wahl des Ausgleichsproblems	50
2.2.6	Ein Ansatz für kleine Bestände bei exakten Messungen	50
2.3	Eindimensionale Wachstumsmodelle	52
2.3.1	Absolute, relative und exponentielle Wachstumsrate	52
2.3.2	Deterministische Modelle exponentiellen Wachstums	53
2.3.3	Stochastische Modelle exponentiellen Wachstums	53
2.3.4	Die logistische Differentialgleichung	56
2.3.5	Logistisches Wachstum in diskreter Zeit	59
2.3.6	Stochastische Modelle logistischen Wachstums	63
2.3.7	Dichteabhängiges Wachstum: Modelle und Daten	69
2.3.8	Ein allgemeiner Ansatz für dichteabhängiges Wachstum	73
2.4	Ein allgemeiner Rahmen zur Modellierung dynamischer Prozesse	76
2.4.1	Der Zustandsraum	76
2.4.2	Diskrete und kontinuierliche Zeit	77
2.4.3	Deterministische und stochastische Entwicklungsgesetze	77
2.4.4	Automaten	78
2.4.5	Deterministische dynamische Systeme	78
2.4.6	Existenz und Eindeutigkeit bei gewöhnlichen Differentialgleichungen	80
2.4.7	Positive Invarianz	81
2.5	Literatur	82
3	Modellierung mit linearen Gleichungen	83
3.1	Drei lineare Modelle	83
3.1.1	Populationswachstum mit altersabhängiger Fruchtbarkeit und Sterblichkeit	83
3.1.2	Ökonomische Verflechtung	88
3.1.3	Meinungsbildung in Gruppen. Das Abelson-Modell	92
3.2	Die Zusammenhangsstruktur mehrdimensionaler Modelle	94
3.2.1	Zusammenhangskomponenten und Kondensations-Digraph	96
3.2.2	Meinungsbildung in Gruppen	98
3.2.3	Der Digraph einer nichtnegativen Matrix	98
3.2.4	Positive Potenzen nichtnegativer Matrizen	99
3.3	Der Lösungsraum linearer dynamischer Systeme	101
3.3.1	Existenz und Eindeutigkeit bei Anfangswertaufgaben	101
3.3.2	Die lineare Struktur des Lösungsraums	102
3.3.3	Die Rolle der Eigenwerte der Systemmatrix	102
3.3.4	Komplexe Lösungen	103
3.3.5	Komplexe Fundamentalsysteme: Diagonalisierbare Systemmatrizen	104
3.3.6	Exkurs: Jordan'sche Normalform und Jordan-Sequenzen	104
3.3.7	Komplexe Fundamentalsysteme: Allgemeiner Fall	105
3.3.8	Stabilität	106
3.3.9	Inhomogene lineare dynamische Systeme mit konstanten Einträgen	106
3.4	Nichtnegative Matrizen. Der Satz von Perron-Frobenius	107
3.4.1	Der Perron-Frobenius-Eigenwert	108
3.4.2	Positive Matrizen	108
3.4.3	Allgemeiner Fall	109
3.4.4	Irreduzible Matrizen	109
3.4.5	Langzeitverhalten diskreter dynamischer Systeme	110
3.4.6	Stabilität und Produktivität nichtnegativer Matrizen	110
3.4.7	Input-Output-Tabellen und Verflechtungsmatrix	111

3.5	Analyse der Altersklassen-Modelle	112
3.5.1	Einfacher Leslie-Prozess. Langzeitverhalten	112
3.5.2	Populationen mit zwei Geschlechtern. Langzeitverhalten	113
3.5.3	Wachstum oder Zerfall?	113
3.6	Kooperative Matrizen	115
3.6.1	Der Satz von Perron-Frobenius für kooperative Matrizen	115
3.6.2	Langzeitverhalten linearer Differentialgleichungen	116
3.6.3	Stabilität kooperativer Matrizen	116
3.6.4	Strenge Monotonie des Perron-Frobenius-Eigenwerts	116
3.7	Analyse des Abelson-Modells	117
3.7.1	Stark zusammenhängende Gruppen	117
3.7.2	Gruppen mit genau einer Führungsclique	118
3.7.3	Gruppen mit mindestens zwei autonomen Untergruppen	118
3.7.4	Beispiel	118
3.8	Markov-Prozesse	119
3.8.1	Klassifikation von Markov-Prozessen und ihren Zuständen	120
3.8.2	Eine deterministische Interpretation von Markov-Prozessen	121
3.8.3	Beispiele	122
3.8.4	Kontinuierliche Markov-Prozesse	125
3.8.5	Reguläre Markov-Prozesse	127
3.8.6	Ein irregulärer Markov-Prozess	128
3.8.7	Absorbierende Markov-Prozesse	130
3.8.8	Lineare Kompartiment-Modelle mit externen Einträgen	134
3.9	Literatur	136

1 Exakte Naturwissenschaft und Modellbegriff

Der *Begriff* des mathematischen Modells ist erst gegen Ende des 19. Jahrhunderts aufgekommen. Nichtsdestoweniger ist mathematische Modellbildung aus heutiger Sicht das, was die neuzeitliche mathematische Naturwissenschaft seit Galilei betreibt. Das Kant'sche Diktum (KANT 1786 / 1996, Vorrede),

daß in jeder besonderen Naturlehre nur so viel eigentliche Wissenschaft angetroffen werden könne, als darin Mathematik anzutreffen ist,

besagt in moderner Terminologie: Es ist der Gebrauch *mathematischer Modelle*, der die *exakten Wissenschaften* „exakt“ bzw. die *hard sciences* „hart“ macht.

Nun ist weder „eigentlich“ noch „exakt“ noch „hart“ ein in diesem Zusammenhang präzise geklärter Begriff.¹ Insofern handelt es sich bei der hier vorliegenden Aussage nicht um einen Tatsachenbehauptung, sondern sie ist schlicht definitorisch. Auf der anderen Seite sind die darin verwendeten Worte positiv besetzt, und in der Tat haben die exakten Naturwissenschaften ihrer sichtbaren technischen Erfolge wegen für andere Wissenschaften die Rolle eines Vorbilds übernommen, dem es nachzueifern gelte. Mathematische Modelle werden daher seit Ende des 19. Jahrhunderts zunehmend auch in „weichen“ Wissenschaften eingesetzt, und sei es auch nur aus Gründen des wissenschaftlichen Prestiges.

Auch und gerade wenn es um den Einsatz, die Rolle, die Sinnhaftigkeit und die Grenzen mathematischer Modellbildung etwa in den Lebens- und Sozialwissenschaften gehen soll, ist es nützlich, sich ihrer Herkunft aus der mathematisch-naturwissenschaftlichen Methode klar zu werden. Schließlich ist es nicht ausgemacht, dass ein Instrument, welches sich in einem bestimmten Umfeld (hier der Physik) bewährt hat, auch in jedem beliebigen anderen erfolgreich eingesetzt werden kann. Zu fragen ist daher

- nach der Rolle der Mathematik für die neuzeitliche Physik, mit der zusammen sie sich entwickelt hat,
- nach den Bedingungen der Möglichkeit dieses offenbar erfolgreichen Zusammenwirkens und
- nach dem Charakter und ggf. der Begrenztheit dieses Erfolges.

Das soll hier an einigen Beispielen aus den Anfängen geschehen, wobei der Schwerpunkt weniger auf den Ergebnissen als vielmehr auf der Methode liegt, mit der sie gewonnen wurden. Mein erstes Ziel ist es dabei, der irrigen empiristischen Auffassung entgegen zu treten, Ausgangspunkt und zentrales Werkzeug aller Naturwissenschaft sei die Erfahrung. Tatsächlich wird sich die Rolle der Mathematik als schwerer wiegend und damit zugleich problematischer erweisen als vielfach auch von Naturwissenschaftlern angenommen. Das eigentlich Neue an der neuzeitlichen exakten Naturwissenschaft ist die Vorstellung von *mathematischen Gesetzen* hinter den vielfältigen Erscheinungen des Naturgeschehens. Dieses Weltbild erst impliziert auch die besondere Rolle des Experiments: Nur hier, in der bewussten Herstellung einer künstlichen Situation, und nicht in der Beobachtung mit „freiem Auge“, lassen sich die mathematischen Idealwelten mit der Erfahrung zumindest näherungsweise zur Deckung bringen. Die Konstitution der mathematisch-naturwissenschaftlichen Methode, mit der die Namen Galileis und Newtons verbunden sind, wird in den Abschnitten 1 und 2 analysiert.

Bis ins 19. Jahrhundert hinein bilden Mathematik und Physik eine Einheit, gelten mathematische Sätze einfach als wahre Aussagen über die Natur. Dann aber beginnt diese Einheit zu bröckeln. Es wird deutlich

¹ „Wissenschaft“ ist es übrigens ebenso wenig.

und am Ende eines Jahrhunderts des Übergangs auch klar ausgesprochen, dass unsere Bilder von der Natur Schöpfungen unseres Geistes sind, die aus der Wirklichkeit nicht einfach herausgelesen werden. Im Zuge dieser wissenschaftlichen Revolution der Moderne wird die Mathematik zu einem eigenen Fach, und es entwickelt sich der Modellbegriff. Davon handelt Abschnitt 3.

Die Tätigkeit des Modellierens, der Modellierungsprozess, ist Gegenstand von Abschnitt 4.

1.1 Die Anfänge: Galileis Fallgesetze

Die Gesetze des freien Falls schwerer Körper stehen am Beginn der neuzeitlichen Physik. Sie besagen:

G1 Alle Körper fallen gleich schnell.

G2 Bei einem Fall aus der Ruhelage verhalten sich die zurückgelegten Wege wie die Quadrate der Zeiten.

Mit diesen Gesetzen geriet Galileo Galilei (1564 - 1642) in Widerspruch zu der in seiner Zeit vorherrschenden aristotelischen Wissenschaft, deren Lehre besagte:

Ar Jeder Körper hat das Bestreben, den ihm zukommenden Platz einzunehmen. Leichte Körper bewegen sich nach oben, schwere fallen nach unten. Je schwerer der Körper, desto schneller fällt er.

Tatsächlich handelt es sich hier um einen der seltenen Fälle, in denen sich die neuzeitliche Physik direkt mit mittelalterlichen Vorstellungen konfrontieren lässt, denn in der Regel behandelt sie Fragen, die sich Menschen in anderen oder früheren Gesellschaften gar nicht stellten. Umso interessanter ist, wie sich Galileis Fallgesetze durchsetzten.

1.1.1 Der empiristische Mythos und die Legende von Pisa

Ein fester Bestandteil des Bildes, das die Moderne im Allgemeinen und die westliche Wissenschaft im Besonderen von sich selber hat, ist die Vorstellung, sie orientiere sich an Tatsachen, während vergangene Kulturen doch eher ihren Mythen und anderen Hirngespinnsten gefolgt und daher folgerichtig und völlig zu Recht inzwischen vergangen seien. Als ein Paradigma dafür dient bis heute Galileis Auseinandersetzung mit der Autorität der aristotelischen Wissenschaft und der katholischen Kirche, obwohl doch die auf Galilei und Newton zurückgehende Mechanik ihren allgemeinen Geltungsanspruch schon längst hat aufgeben müssen. Noch Bertolt Brechts um 1945 entstandenes Theaterstück „Leben des Galilei“ lebt von dem aufklärerischen Pathos dieses Kampfes des die Tatsachen aufdeckenden „kalten Auges der Wissenschaft“ gegen den „tausendjährigen Perlmutterdunst von Aberglauben und alten Wörtern“, durch den allein die Herrschaft „selbstsüchtiger Machthaber“ weiterhin aufrecht erhalten werden kann. Die vor dem Hintergrund des Abwurfs der ersten Atombombe unumgängliche Kritik Brechts² wird denn auch ausschließlich auf der moralischen Ebene vorgetragen, dass nämlich Galilei sich habe einschüchtern lassen und sein Wissen den Machthabern überliefert habe, „es zu gebrauchen, es nicht zu gebrauchen, es zu mißbrauchen, ganz, wie es ihren Zwecken diene“. Jeder brave Naturwissenschaftler kann dem zu Recht entgegenhalten, dass Galileis Lehre trotz seines Widerrufs schließlich zum Allgemeingut geworden, die Wahrheit eben nicht aufzuhalten sei, auch wenn das der Menschheit wenig genützt zu haben scheint.

Die mit den Namen Galileis und Newtons verbundene und heute als „klassisch“ bezeichnete Mechanik spielte von Beginn der Neuzeit bis ins 19. Jahrhundert hinein die Rolle einer Leitwissenschaft. In gewisser Hinsicht ist sie es noch heute, auch wenn ihre Ergebnisse durch die „moderne“ Physik des 20.

²im Stück als Selbstkritik Galileis formuliert

Jahrhunderts ihren universellen Anspruch verloren haben. Denn die an ihr entwickelte und mit durchschlagendem Erfolg angewandte mathematisch-naturwissenschaftliche Methode hat im letzten Jahrhundert an Bedeutung weiter gewonnen und eine Vorbildfunktion für die westliche Wissenschaft aller Fakultäten eingenommen, zumindest ihrer jeweiligen Mainstreams, sodass selbst noch die Kritiker ihrer Übertragung etwa in die Sozialwissenschaften sich mit ihr auseinandersetzen müssen. So richtig deren Argument ist, dass eine Methode sich an ihren Gegenstand anzupassen habe und „Gesellschaft“ eben nicht dasselbe sei wie „Natur“, so sehr leiden derartige Diskussionen oft darunter, dass der Empirismus, also die „Tatsachen-Fraktion“, die Interpretations-Hegemonie darüber gewonnen hat, was diese Methode eigentlich leiste und welcher Art die mit ihr zu erzielenden Ergebnisse seien. Die Behauptung, es handele sich dabei um objektive, für jedermann überprüfbare Tatsachen, wird gar nicht mehr in Frage gestellt.

Ein Beispiel für dieses Phänomen ist die folgende Geschichte, die die Wissenschaftsgeschichtsschreibung über fast drei Jahrhunderte hinweg als gesichertes Wissen anzubieten hatte. Sie betrifft den freien Fall schwerer Körper, den ersten Teil des galileischen Fallgesetzes, und figurierte als der „Schlag, von dem sich die aristotelische Wissenschaft nie wieder erholte“:

An dieser Stelle müssen wir auf die berühmten Experimente zum Fall der Körper zu sprechen kommen, sind diese doch aufs engste verknüpft mit dem schiefen Turm von Pisa, einem der kuriosesten Baudenkmäler Italiens. Beinahe zweitausend Jahre zuvor hatte Aristoteles behauptet, daß im Falle zweier verschiedener Gewichte gleichen Materials, die aus gleicher Höhe fielen, das schwerere den Erdboden vor dem leichteren erreiche, und dies gemäß dem Verhältnis ihrer jeweiligen Schwere. Das Experiment ist gewiß nicht schwierig; nichtsdestoweniger war niemand auf die Idee gekommen, einen derartigen Beweis zu führen, weshalb diese Behauptung kraft des Machtwortes des Aristoteles unter die Axiome der Wissenschaft von der Bewegung aufgenommen worden war. Galilei forderte nun unter Berufung auf die Sinneswahrnehmung die Autorität des Aristoteles heraus und behauptete, daß die Kugeln in gleicher Zeit fielen, abgesehen von einer unbedeutenden, auf dem unterschiedlichen Luftwiderstand beruhenden Differenz. Die Aristoteliker verspotteten diese Idee und verweigerten ihr das Gehör. Galilei aber ließ sich nicht einschüchtern und beschloß, seine Gegner dazu zu zwingen, gleich ihm der Tatsache ins Auge zu sehen. Daher bestieg er eines Morgens vor der versammelten Universität - Professoren und Studenten - den schiefen Turm, zwei Kugeln mit sich führend, eine zehn- und eine einpfündige. Er legte sie auf den Rand des Turms und ließ sie zugleich fallen. Und sie fielen gemeinsam und schlugen gemeinsam am Boden auf.³

Es ist wohl das Verdienst Alexandre Koyrés⁴, fast 300 Jahre nach dem Tod Galileis der Geschichte von seinen Versuchen zum freien Fall am schiefen Turm von Pisa endgültig den Garaus gemacht zu haben, sodass heute kein Wissenschaftshistoriker, der ernst genommen werden will, sie noch erzählen kann. An der Geschichte ist eigentlich nur wahr, dass Galilei um das Jahr 1590 herum eine schlecht bezahlte und auf drei Jahre befristete Stelle als Professor für Mathematik an der Universität Pisa innehatte. Die Legende kam 60 Jahre nach dem beschriebenen Vorfall erstmals auf und wurde von späteren Wissenschaftshistorikern immer weiter ausgeschmückt. Was einem ohne weitere historische Kenntnis auffällt, ist ihre Inkonsistenz: Was hätte die aristotelischen Professoren, denen hier ihr Dogmatismus vorgehalten wird, wohl dazu veranlassen sollen, zusammen zu laufen, wenn einer ihrer unbedeutendsten Kollegen ein irrsinniges Experiment veranstaltet? Die Geschichte widerspricht allen Gebräuchen an Universitäten dieser Zeit und wohl auch noch heutiger Universitäten. Sie wurde von Galilei selbst nie erwähnt,⁵ und schließlich: Die Experimente wären schief gegangen, bzw. sie wurden gemacht (1640, 1645, 1650), mit

³J.J. Fahie. Galilei, His Life and Work, London 1903, 24 f., zitiert nach KOYRÉ 1998, 124

⁴Galiléé et l'expérience de Pise: À propos d'une légende, Annales de l'Université de Paris 1937, KOYRÉ 1998, 123 - 134

⁵In einem Traktat Galileis aus demselben Jahr 1590 findet sich sogar der gegenteilige Hinweis: Wenn man Kugeln aus Holz und Blei von einem hohen Turm fallen lässt, bewegt sich das Blei weit voraus, s. FÖLSING 1996, 85

großen und kleinen Eisenkugeln, mit gleich großen Tonkugeln, eine massiv, die andere hohl, mit Kugeln aus verschiedenen Materialien, und sie sind (im Sinne der Legende) allesamt schief gegangen.⁶

Das eigentlich spannende an diesem modernen Märchen ist, dass es 300 Jahre lang zum allgemeinen Bildungsgut gehörte, gewissermaßen zum gesicherten Bestand unseres naturwissenschaftlichen Wissens. Wie alle Märchen transportiert auch dieses eine Botschaft, nämlich die von der neuzeitlichen Rationalität, die unvoreingenommen die Tatsachen sprechen lasse, während das finstere Mittelalter sich nur auf Autoritäten berufe und Lehrbuchwissen tradiere. Der spät geführte Nachweis, dass es sich hierbei um einen Mythos handelt, den Mythos des Empirismus, ändert nichts an dessen Wirksamkeit. 350 Jahre nach Galilei ist dieses Weltbild so selbstverständlich geworden, dass es keiner Begründung mehr bedarf. Und wie ein Blick in ein Standardlehrbuch der Experimentalphysik zeigt, ist auch das mit ihm verbundene Märchen zu schön, um einfach weggelassen zu werden, nur weil es ein Märchen ist:

*Zunächst sei untersucht, ob die Fallbewegung von der Art des fallenden Körpers, z. B. von seiner Größe oder seinem Gewicht abhängig ist. Wir machen folgende Versuche: Zwei gleichgroße Kugeln aus Aluminium und Blei, die also sehr verschiedenes Gewicht haben, lassen wir gleichzeitig aus derselben Höhe zu Boden fallen. Wir stellen fest, daß sie zu gleicher Zeit zu Boden aufschlagen, wie bereits Galilei (1590) durch Fallversuche am schiefen Turm von Pisa festgestellt hat. Nehmen wir drei gleiche Kugeln aus demselben Stoff, so kommen diese natürlich zur gleichen Zeit am Boden an. Verbinden wir nun zwei dieser Kugeln fest miteinander (etwa durch einen hindurchgehenden Stift), und lassen wir diese Doppelkugel mit der dritten Einzelkugel gleichzeitig fallen, so schlagen auch diese Körper von verschiedener Größe und verschiedenem Gewicht gleichzeitig am Boden auf. Der hieraus zu ziehenden Folgerung, daß alle Körper, unabhängig von Gestalt, Art und Gewicht, gleich schnell fallen, scheint aber folgender Versuch zu widersprechen: Lassen wir eine Münze und ein gleich großes Stück Papier fallen, so beobachten wir, daß die Münze wesentlich früher unten ankommt, als das zur gleichen Zeit aus derselben Höhe fallende Papierstückchen; letzteres flattert in unregelmäßiger Bewegung zu Boden und benötigt zum Durchfallen eine größere Zeit. Der Gegensatz ist indessen nur scheinbar. Bei diesem letzten Versuch macht sich nämlich der Widerstand der Luft störend bemerkbar. Die beim Fall an dem Körper vorbeiströmende Luft hemmt die Fallbewegung, und zwar umso stärker, je größer die Angriffsfläche der Luft an dem betreffenden Körper ist. Ballen wir das Papierstück zu einer kleinen Kugel zusammen, so fällt es ebenso rasch wie die Münze. Der störende Einfluß des Luftwiderstandes auf den freien Fall läßt sich noch durch einen von Newton angegebenen Versuch anschaulich zeigen. Ein etwa 2 m langes, mehrere Zentimeter weites Glasrohr, das an beiden Enden zugeschmolzen ist, enthält eine Bleikugel, ein Stück Kork und eine Flaumfeder. Befinden sich die drei Körper am Boden der Röhre und dreht man diese rasch um 180°, so beobachtet man, wie zuerst die Bleikugel, dann das Korkstück und schließlich die Flaumfeder unten ankommen. Pumpt man aber die Luft aus der Röhre und wiederholt man den Versuch, so erkennt man, daß nunmehr die drei Körper im gleichen Augenblick auf dem Boden des Rohres aufschlagen. Wir dürfen also das Erfahrungsgesetz aussprechen: **Im luftleeren Raum fallen alle Körper gleich schnell.**⁷*

Wieso eigentlich nur im luftleeren Raum, in Pisa hat es doch schließlich auch funktioniert? Die Schlussfolgerung bleibt ebenso undurchsichtig wie die Begründung. Der Grund liegt darin, dass hier Aussagen mit völlig verschiedenem methodischem Status wild durcheinandergeworfen werden:

- Der Text enthält falsche und richtige Behauptungen über alltägliche Beobachtungen, wobei die

⁶s. KOYRÉ 1998, 129 - 132

⁷Bergmann-Schäfer: Lehrbuch der Experimentalphysik, Band I, Mechanik, Akustik, Wärme, 9. verbesserte Auflage, Berlin 1974, S. 40

richtigen gerade diejenigen sind, die zum galileischen Fallgesetz in Widerspruch stehen. Sie werden unter Hinweis auf den „störenden“ Luftwiderstand einfach weginterpretiert.

- Es wird ein Gedankenexperiment durchgeführt (Kugel und Doppelkugel), aus welchem das Fallgesetz logisch zwingend, aber ohne Rückgriff auf irgendeine Beobachtung sich ergibt.
- Es wird schließlich ein Experiment beschrieben, das durchzuführen einen hohen technischen Aufwand erfordert (Leerpumpen der Röhre). Erst in der so hergestellten künstlichen Situation lässt sich das behauptete Gesetz auch beobachten.

Das Ganze dann als „Erfahrungsgesetz“ zu bezeichnen, ist schon stark und setzt in der Tat die Verwirrung voraus, die zuvor erst gestiftet werden musste. Von dieser Verwirrung lebt der Empirismus.

Der Text ist ein Beispiel dafür, wie wenig die meisten Naturwissenschaftler von der Geschichte und Methode der Wissenschaft wissen, die sie selbst betreiben.⁸ Das war keineswegs immer so, sondern was sich hier konstatieren lässt, hat vielmehr den Charakter einer Verfallserscheinung. Galilei selbst jedenfalls war sich seines Vorgehens, anders als die meisten seiner Epigonen, durchaus bewusst. Es lohnt daher, zu den Quellen zurückzugehen.

Was brachte das galileische Fallgesetz in die Welt, wenn es denn die Erfahrung nicht sein konnte, weder die unmittelbare Beobachtung, denn die lehrt etwas anderes, noch ein Experiment im luftleeren Raum, das Galilei schon deswegen nicht durchführen konnte, weil ihm dazu die technischen Mittel fehlten? Die schlichte Antwort ist: Das Fallgesetz ergibt sich aus einem logischen Argument, einem mathematischen Beweis oder, wie man heute sagen würde, einem Gedankenexperiment. Das Argument war schon 1585 von dem Mathematiker Benedetti in Venedig veröffentlicht worden und ist auch in dem oben zitierten Text aus dem Physiklehrbuch enthalten, wenn auch dort seines methodischen Stellenwerts völlig beraubt.

1.1.2 Beweis des Ersten Fallgesetzes

Benedetti argumentierte: Zwei gleiche Körper fallen gleich schnell, das jedenfalls scheint unbestritten. Verbindet man sie nun durch einen leichten (im Idealfall masselosen) Stab, so ändert sich an ihrer Ge-

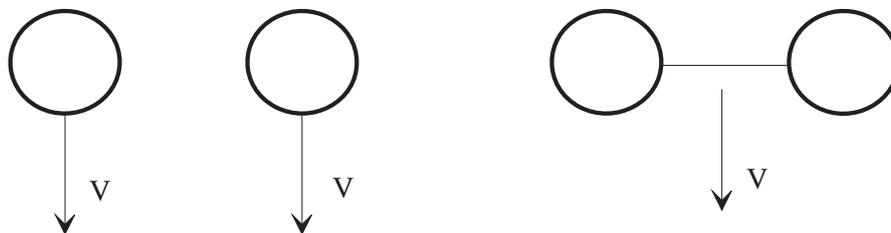


Abbildung 1.1: Benedettis Argument

⁸Die allgemeine Begriffsverwirrung lässt sich noch steigern: „Jeder, der glaubt, dass physikalische Gesetze bloße gesellschaftliche Konventionen seien, ist eingeladen, vom Fenster meiner Wohnung die Überschreitung dieser Grenzen zu versuchen (ich wohne im 22. Stock).“ So der Physiker Alan Sokal in Reaktion auf die Zumutungen der postmodernen Kulturwissenschaft (s. *lingua franca* 6 (4), 62 - 64, 1996). Wenn irgend jemand tatsächlich behauptet haben sollte, physikalische Gesetze seien Konventionen, so wäre das die erste Dummheit. Die zweite ist die, die Sokal selber verzapft, entweder Dummheit oder aber bewusste Irreführung. Die Tatsache, dass ein Sprung aus dem 22. Stock in aller Regel zum Tode führt, hat mit physikalischen Gesetzen im modernen Sinne absolut nichts zu tun. Von der Existenz 22-stöckiger Gebäude einmal abgesehen, dürfte das in jeder Gesellschaft bekannt gewesen sein, z. B. als mögliche Form des Todesstrafe (Tarpejische Felsen). Für ein physikalisches Gesetz reicht das aber hinten und vorne nicht. Der einzige Zusammenhang besteht darin, dass ein physikalisches Gesetz zu dieser Tatsache nicht in Widerspruch geraten sollte, was übrigens auch das aristotelische Fallgesetz **A**r nicht tut.

schwindigkeit nichts, die dann aber einem Körper doppelter Masse zukommt (vgl. Abbildung 1.1). Genauso lässt sich mit drei, sieben oder auch hunderttausend Körpern argumentieren, in jedem Fall ergibt sich dieselbe Geschwindigkeit für Körper beliebig verschiedener Masse.

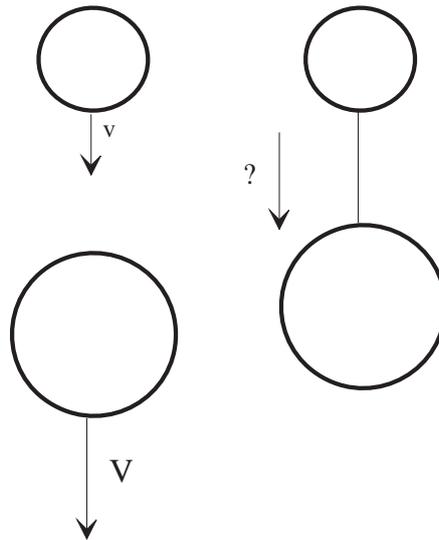


Abbildung 1.2: Galileis Argument

Galilei machte daraus einen Widerspruchsbeweis⁹: Wäre das aristotelische Fallgesetz **A_r** richtig, so müsste ein schwererer Körper einem leichteren vorausziehen. Verbindet man nun beide mit einer Schnur, so müsste der schwerere Körper den leichteren hinter sich herziehen, der leichtere den schwereren aber abbremsen (vgl. Abbildung 1.2). Es ergäbe sich eine kleinere Geschwindigkeit als die des ursprünglichen schwereren Körpers, allerdings für einen insgesamt noch schwereren, ein Widerspruch also.

Beide Beweise des Ersten Fallgesetzes **G₁** abstrahieren von der Gestalt der Körper, beziehen sich also nur auf ihre Masse. Gezeigt wurde also: Hängt die Fallgeschwindigkeit von der Gestalt der Körper, ihrer Massenverteilung, nicht ab, so müssen alle Körper gleich schnell fallen. Dieses Ergebnis steht nun aber in offensichtlichem Widerspruch zur Empirie, die Körper fallen schließlich nicht gleich schnell. Würden nun Logik und Empirie gleichermaßen berücksichtigt, so wäre daraus der Schluss zu ziehen, dass von der Gestalt der Körper eben nicht abstrahiert werden darf. Diesen Schluss zieht Galilei nun aber gerade *nicht*, und genau hier liegt das revolutionär Neuartige seiner Naturbetrachtung: Er *entscheidet* sich für die Logik und Mathematik und gegen die unmittelbare Empirie und damit für eine Naturauffassung, die Antike oder Mittelalter nur als verrückt hätten ansehen können.

1.1.3 Beweis des Zweiten Fallgesetzes

Die hier zum Ausdruck kommende Naturauffassung ist heute ohne Alternative, wir haben uns so an sie gewöhnt, dass wir die Welt nur noch in ihrem Lichte sehen und deshalb meinen, sie aus Erfahrung und Beobachtung gewonnen zu haben. Die Folge sind derart wirre und von Verständnislosigkeit für die eigene Methode geprägte Ausführungen wie die oben aus einem physikalischen Standardlehrbuch zitierten. Wer wissen will, wie Physik funktioniert, halte sich daher besser an die Originale. Galilei selbst hat in einem reichhaltigen Schrifttum die von ihm entwickelten und verwendeten Methoden sehr genau beschrieben. Sein Alterswerk, die *Discorsi* von 1638 werden 300 Jahre später in der Jubiläumsausgabe von Max v. Laue als „erstes Lehrbuch der Physik“ bezeichnet. Das typische Vorgehen wird am Dritten Tag¹⁰ am

⁹GALILEI 1638/1995, 57/58

¹⁰Die *Discorsi* sind als ein über mehrere Tage sich hinziehendes Gespräch dargestellt.

Beispiel des Zweiten Fallgesetzes besonders deutlich.

Auch hier sind es nicht irgendwelche sinnlichen Wahrnehmungen, die den Ausgangspunkt bilden. Das wäre allerdings auch sehr verwunderlich bei einem Gesetz, welches sich mit bloßem Auge schwerlich beobachten lässt. Tatsächlich handelt es sich um einen klassischen mathematischen Text in der Form Definition - Satz - Beweis. Zunächst wird die Definition der gleichförmigen Bewegung eingeführt:

Definition 1.1.1 (GALILEI 1638/1995, 141)

Ich nenne diejenige Bewegung gleichförmig, bei welcher die in irgendwelchen gleichen Zeiten vom Körper zurückgelegten Strecken unter einander gleich sind.

über die dann zunächst einige Theoreme aufgestellt und mit geometrischen Mitteln bewiesen werden. Dann erfolgt der Übergang zur beschleunigten Bewegung (GALILEI 1638/1995, 146):

Bisher war die gleichförmige Bewegung behandelt worden, jetzt gehen wir zur beschleunigten Bewegung über. Zunächst muss eine der natürlichen Erscheinung genau entsprechende Definition gesucht und erläutert werden. Obgleich es durchaus gestattet ist, irgend eine Art der Bewegung beliebig zu ersinnen und die damit zusammenhängenden Ereignisse zu betrachten ..., so haben wir uns dennoch entschlossen, diejenigen Erscheinungen zu betrachten, die bei den frei fallende Körpern in der Natur vorkommen, und lassen die Definition der beschleunigten Bewegung zusammenfallen mit dem Wesen einer natürlich beschleunigten Bewegung. ...

Wenn ich daher bemerke, dass ein aus der Ruhelage von bedeutender Höhe herabfallender Stein nach und nach neue Zuwüchse an Geschwindigkeit erlangt, warum soll ich nicht glauben, dass solche Zuwüchse in allereinfachster, Jedermann plausibler Weise zu Stande kommen? Wenn wir genau aufmerken, werden wir keinen Zuwachs einfacher finden, als denjenigen, der in immer gleicher Weise hinzutritt. Das erkennen wir leicht, wenn wir an die Verwandtschaft der Begriffe der Zeit und der Bewegung denken: denn wie die Gleichförmigkeit der Bewegung durch die Gleichheit der Zeiten und Räume bestimmt und erfasst wird ..., so können wir durch ebensolche Gleichheit der Zeittheile die Geschwindigkeitszunahmen als einfach zu Stande gekommen erfassen: mit dem Geiste erkennen wir diese Bewegung als einförmig und in gleicher Weise stetig beschleunigt, da zu irgend welchen gleichen Zeiten gleiche Geschwindigkeitszunahmen sich addiren.

Die alltägliche Erfahrung kommt bei diesen Überlegungen insofern ins Spiel, als festgehalten wird, dass ein losgelassener Stein fällt und dabei immer schneller wird. Doch das ist alles, genauer wird die Bewegung nicht beschrieben. Vielmehr kommt jetzt das *Prinzip der Einfachheit* ins Spiel: Unter allen Bewegungen mit den genannten Eigenschaften wird eine möglichst einfache ausgewählt. Dieses Prinzip und nicht irgend eine vorausgegangene Messung oder genaue Beobachtung führt schließlich zu

Definition 1.1.2 (GALILEI 1638/1995, 148)

Gleichförmig oder einförmig beschleunigte Bewegung nenne ich diejenige, die von Anfang an in gleichen Zeiten gleiche Geschwindigkeitszunahmen ertheilt.

Die Empirie kann an dieser Stelle schon deswegen gar nicht zum Zuge kommen, weil sich die Geschwindigkeit eines fallenden Steines (zumindest zu Zeiten Galileis) nicht direkt messen lässt. Unmittelbar messen lassen sich Strecken und Zeitspannen.¹¹ Ob also die Definition 1.1.2 der gleichförmig beschleunigten Bewegung bei den frei fallenden Körpern in der Natur vorkommt, wie Galilei behauptet, lässt sich daher nicht so ohne weiteres feststellen. Galilei verbleibt denn auch auf der mathematischen Ebene und beweist mit geometrischen Mitteln zwei auf einander aufbauende Sätze:

¹¹Die Zeitmessung stellte allerdings mangels präziser und für die Erfassung kurzer Zeitspannen geeigneter Uhren ebenfalls ein schwerwiegendes Problem dar.

Satz 1.1.3 (GALILEI 1638/1995, 158)

Die Zeit, welche irgendeine Strecke von einem Körper von der Ruhelage aus mittelst einer gleichförmig beschleunigten Bewegung zurückgelegt wird, ist gleich der Zeit, in welcher dieselbe Strecke von demselben Körper zurückgelegt würde mittelst einer gleichförmigen Bewegung, deren Geschwindigkeit gleich wäre dem halben Betrage des höchsten und letzten Geschwindigkeitswerthes bei jener ersten gleichförmig beschleunigten Bewegung.

Satz 1.1.4 (GALILEI 1638/1995, 159)

Wenn ein Körper von der Ruhelage aus gleichförmig beschleunigt fällt, so verhalten sich die in gewissen Zeiten zurückgelegten Strecken wie die Quadrate der Zeiten.

Zum Beweis:

In Satz 1.1.3 werden zwei verschiedene Bewegungen einander gegenüber gestellt: Eine gleichförmig beschleunigte, bei der im Zeitintervall $[0, t_1]$ die Geschwindigkeit gleichmäßig von 0 auf einen maximalen Wert v_1 zunimmt, und eine gleichförmige mit der konstanten Geschwindigkeit $v_1/2$. Offenbar sind die

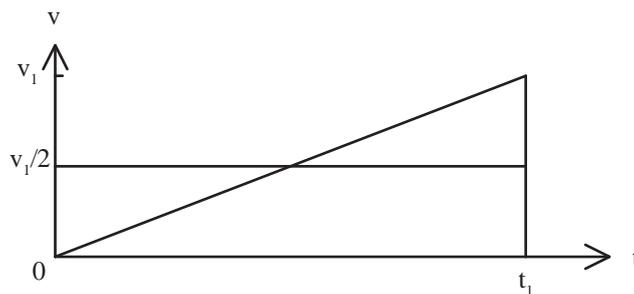


Abbildung 1.3: Zum Beweis von Satz 1.1.3

Flächen unter den beiden verschiedenen Geschwindigkeitsprofilen gleich (vgl. Abbildung 1.3). Nun ist aber, in moderner Terminologie, bei einer gegebenen Geschwindigkeitsfunktion $v(t)$

$$s = \int_0^{t_1} v(t) dt$$

der im Zeitintervall $[0, t_1]$ durchlaufene Weg, und dieses Integral beschreibt ja gerade die Fläche unter dem Graphen von $v(t)$, woraus sich die Aussage von Satz 1.1.3 ergibt. Galilei selbst steht wohl die Geometrie zur Verfügung, mit der die Gleichheit der beiden Flächen gezeigt werden kann, nicht aber der Integralbegriff, der sich in seinen Formulierungen allerdings bereits andeutet, in denen von sich addierenden *Geschwindigkeits-* oder *Bewegungsmomenten* die Rede ist.

Ist nun bei einer gleichförmig beschleunigten Bewegung zum Zeitpunkt t_1 die Geschwindigkeit v_1 und zum Zeitpunkt t_2 die Geschwindigkeit v_2 erreicht, so gilt nach Satz 1.1.3 für die jeweils zurückgelegten Wege

$$s_1 = \frac{1}{2} v_1 t_1 \text{ und } s_2 = \frac{1}{2} v_2 t_2 ,$$

andererseits wegen der gleichförmigen Beschleunigung

$$\frac{v_1}{v_2} = \frac{t_1}{t_2} ,$$

woraus

$$\frac{s_1}{s_2} = \frac{t_1^2}{t_2^2}$$

und damit Satz 1.1.4 folgt.

1.1.4 Experimente: Die Produktion von Tatsachen

Aus Satz 1.1.4 wird das Fallgesetz **G2**, wenn man in der Voraussetzung den mathematischen Begriff der gleichförmig beschleunigten Bewegung durch den physikalischen des freien Falls ersetzt. Ob diese Gleichsetzung berechtigt ist, lässt sich, wie bereits bemerkt, nicht direkt überprüfen, wohl aber in der Form des jetzt vorliegenden Satzes, in dem nur noch Wege und Zeiten auftreten, aber keine Geschwindigkeiten. Die mathematische Herleitung von Satz 1.1.4 erweist sich also als eine Voraussetzung dafür, dass überhaupt ein Experiment durchgeführt werden kann, und der Satz selbst gibt Hinweise darauf, was zu messen ist.

Die von Galilei beschriebenen Experimente fanden mit auf schiefen Ebenen abrollenden Kugeln statt. Tatsächlich befassen sich große Teile der *Discorsi* mit weiteren mathematischen Überlegungen, schiefe Ebenen betreffend, auf die ich nicht näher eingehe. Es zeigt aber, dass der vor dem allerersten Experiment getriebene theoretische Aufwand noch erheblich höher war, als in meiner Kurzfassung hier dargestellt.

GALILEI (1638/1995, 162) beschreibt seine Experimente:

Auf einem Lineale, oder sagen wir auf einem Holzbrette von 12 Ellen Länge, bei einer halben Elle Breite und drei Zoll Dicke, war auf dieser letzten schmalen Seite eine Rinne von etwas mehr als einem Zoll Breite eingegraben. Dieselbe war sehr gerade gezogen, und um die Fläche recht glatt zu haben, war inwendig ein sehr glattes und reines Pergament aufgeklebt; in dieser Rinne liess man eine sehr harte, völlig runde und glattpolirte Messingkugel laufen.

Nach Aufstellung des Brettes wurde dasselbe einerseits gehoben, bald eine, bald zwei Ellen hoch; dann liess man die Kugel durch den Kanal fallen und verzeichnete in sogleich zu beschreibender Weise die Fallzeit für die ganze Strecke: häufig wiederholten wir den einzelnen Versuch, zur genaueren Ermittlung der Zeit, und fanden gar keine Unterschiede, auch nicht einmal von einem Zehnthel eines Pulsschlages. Darauf liessen wir die Kugel nur durch ein Viertel der Strecke laufen, und fanden stets genau die halbe Fallzeit gegen früher. Dann wählten wir andere Strecken, und verglichen die gemessene Fallzeit mit der zuletzt erhaltenen und mit denen von $\frac{2}{3}$ oder $\frac{3}{4}$ oder irgend anderen Bruchtheilen; bei wohl hundertfacher Wiederholung fanden wir stets, dass die Strecken sich verhielten wie die Quadrate der Zeiten: und dieses zwar für jedwede Neigung der Ebene, d. h. des Kanales, in dem die Kugel lief. ...

Zur Ausmessung der Zeit stellten wir einen Eimer voll Wasser auf, in dessen Boden ein enger Kanal angebracht war, durch den ein feiner Wasserstrahl sich ergoss, der mit einem kleinen Becher aufgefangen wurde, während einer jeder beobachteten Fallzeit: das dieser Art aufgesammelte Wasser wurde auf einer sehr genauen Waage gewogen; aus den Differenzen der Wägungen erhielten wir die Verhältnisse der Gewichte und die Verhältnisse der Zeiten, und zwar mit solcher Genauigkeit, dass die zahlreichen Beobachtungen niemals merklich von einander abwichen.

Einer unbefangenen, nicht durch entsprechende Schulversuche voreingenommenen Leserin der Beschreibung Galileis dürfte vor allem die absolute Künstlichkeit des ganzen Geschehens ins Auge springen. Hier wird nicht einfach beobachtet, wie schwere Gegenstände fallen, sondern es wird mit großem Aufwand eine besondere Situation bewusst *hergestellt*, die herbei zu führen einem Laien niemals in den Sinn käme. Aus der Sicht der mathematischen Naturwissenschaft kann das gar nicht anders sein, geht es doch darum, den im mathematischen Satz unterstellten Idealbedingungen möglichst nahe zu kommen und das heißt, störende Einflüsse auszuschalten. Das trifft auch für das Erste Fallgesetz zu: Es gilt nur dort, wo die Gestalt der Körper und damit der Luftwiderstand keine Rolle spielt, also im Vakuum, das Galilei selbst noch gar nicht herstellen konnte.

Aus der Sicht Außenstehender freilich muss ein solches Vorgehen, das die Tatsachen, um die es geht, erst mit hohem Aufwand produziert, die Aussagekraft der damit gewonnenen Ergebnisse erheblich beein-

trächtigen. Das Experiment (nicht die Beobachtung und der Rückgriff auf Erfahrung) ist eine Erfindung der neuzeitlichen Naturwissenschaft, und niemand ist gezwungen, es als Erkenntnisinstrument zu akzeptieren. Ob die Natur sich außerhalb der künstlichen experimentellen Situation so verhält, wie die aus mathematischen Abstraktionen und experimentellen Überprüfungen gewonnenen Gesetze es besagen, darf schließlich mit Fug und Recht bezweifelt werden. Bei der Beurteilung der Auseinandersetzungen etwa Galileis mit den mittelalterlichen, aristotelischen Vorstellungen seiner Zeit sollte man sich diese Diskrepanz in den Erkenntnis*methoden* immer vor Augen halten. Ich komme darauf zurück.

Die tatsächliche Ausführung von Experimenten stößt in Galileis Zeiten auf größere Schwierigkeiten, als das aus der vorliegenden Textstelle deutlich wird, weil nämlich die technischen Voraussetzungen erbärmlich sind, verglichen etwa mit denen, die heute die Physiksammlung einer normalen Schule bietet. Galileis Versuche zur Bestimmung der konstanten Beschleunigung im freien Fall sind völlig wertlos, er vermeidet in seinen Schriften so weit wie möglich die Angabe eines konkreten Zahlenwerts, und wenn er es doch tut, liegt er jedes Mal völlig falsch, nämlich etwa bei der Hälfte des heute gültigen. Galilei bediente sich denn auch zur Stützung seiner Auffassung hin und wieder des „Kunstgriffs“, Experimente, die er sich bloß erdacht hatte, als wirklich durchgeführte zu beschreiben (KOYRÉ 1998, 129). Auch bei dem hier zitierten Text stellt sich die Frage, ob Versuchsergebnisse nicht einfach schöngeredet wurden.

Das von Galilei beschriebene Vorgehen zur Messung der Zeit weist auf ein weiteres, mit Experimenten verbundenes Problem hin. Nicht nur die Anlage von Experimenten ist theoriegeleitet, sondern bereits die Erhebung von Messdaten innerhalb eines Experiments ist es. Jeder Messung liegt eine *Theorie des Messinstruments* zu Grunde, im vorliegenden Text ist es diese: *Die aus dem Eimer fließende und anschließend gewogene Wassermenge ist der Zeit, in der sie ausfließt, proportional*; hinzu kommt noch die *Theorie der benutzten Waage*. Die jeweilige Theorie des Messinstruments wird in einem Experiment aber nicht überprüft, sondern immer vorausgesetzt.

Durch die Verfeinerung der Messinstrumente im Laufe der Entwicklung der Experimentalphysik wird dieses Problem nicht etwa behoben, sondern vielmehr verstärkt. Präzisionsuhren etwa beruhen in weit höherem Maße auf physikalischen Theorien als Galileis einfache Konstruktionen. Denkbar ist daher, dass in Experimenten Messinstrumente eingesetzt werden, die auf eben jenen Theorien beruhen, die im Experiment überprüft werden sollen. Dieser Problemstrang der mangelnden Aussagekraft von Experimenten wird im vorliegenden Text nicht weiter verfolgt, vgl. dazu COLLINS/PINCH 1999.

1.1.5 Zur Methodenabhängigkeit der Erkenntnisse

Der Unterschied zwischen Beobachtung und Experiment kann gar nicht scharf genug hervorgehoben werden, seine Nichtbeachtung hat schon viele in die Irre führt, so etwa den braven Übersetzer Emil Strauss von Galileis *Dialog* ins Deutsche 1890 in seiner Einleitung, der

... die falsche, ja thörichte aristotelische Behauptung ..., daß die Fallgeschwindigkeit proportional zur Schwere und umgekehrt proportional der Dichtigkeit des Mediums sei ...

als Beleg für die Überlegenheit der neuzeitlichen Naturwissenschaft gegenüber mittelalterlichen und anderen Denkweisen anführt. Dieser Ausspruch ist ein besonders schönes Beispiel für ein unkritisches und geschichtsblindes Denken, das die eigene Erkenntnisform für die einzig mögliche und Vertreter anderer Kulturen, die zu anderen Ergebnissen kommen, einfach nur für dumm oder verblendet hält. Dabei hat Aristoteles so Unrecht ja nicht, solange man sich auf alltägliche Beobachtungen bezieht. Oder anders gesagt: Galilei wäre zu einem ähnlichen Ergebnis gekommen, wenn er so vorgegangen wäre, wie die Legende vom Schiefen Turm es erzählt. Galileis ganz anderes, als Fallgesetz formuliertes Ergebnis beruht auf einer ganz anderen Methode, die u. a. gerade darin besteht, von der „Dichtigkeit des Mediums“ zu abstrahieren. Seine Überprüfung im Experiment setzt voraus, dass Versuchsbedingungen hergestellt werden, bei denen die Dichte vernachlässigt werden kann. Galilei selbst ist sich seines Vorgehens jedenfalls

bewusster als sein Übersetzer 250 Jahre später. Im Zusammenhang mit der Wurfbewegung¹² schreibt er (GALILEI 1638/1995, 224):

In Betreff des Widerstandes des Mediums gestehe ich zu, dass dessen störender Einfluss bemerklicher sein wird, und wegen seiner mannigfach verschiedenen Beschaffenheit kaum unter feste Regeln gebracht werden kann; so lange wir auch nur den Widerstand der Luft berücksichtigen, so wird dieser alle Bewegungen stören, auf unendlich verschiedene Weise, da unendlich verschieden Gestalt, Gewicht und Geschwindigkeit der geworfenen Körper sich ändern könnten. Wenn z. B. die Geschwindigkeit größer ist, so wird auch der Einfluss der Luft wachsen, und das zwar um so mehr, je leichter die Körper sind, sodass, obwohl die Strecken bei senkrechtem Fall sich wie die Quadrate der Zeiten verhalten sollten, dennoch selbst die allerschwersten Körper von bedeutender Höhe herab solchen Widerstand von der Luft erfahren, dass die Beschleunigung gänzlich aufhört und die Bewegung eine gleichförmige wird; letzteres tritt um so früher ein, und von um so geringeren Höhen, je leichter die Körper sind. . . . Ueber alle die unendlich verschiedenen Möglichkeiten hinsichtlich der Schwere, der Geschwindigkeit und der Gestalt kann keine Theorie gegeben werden.

Letztlich stimmt Galilei der „thörichten aristotelischen Behauptung“ sogar zu, doch ist seine Fragestellung eine ganz andere. Die „Dichtigkeit des Mediums“ ist für ihn nicht Untersuchungsgegenstand, sondern ein Störfaktor, den es auszuschalten gilt. Das Interesse der neuzeitlichen Physik gilt nicht der Vielfalt der Erscheinungen, denn darüber „kann keine Theorie gegeben werden“, sondern den hinter den Erscheinungen wirkenden Gesetzen, von denen Galilei annimmt, dass sie mathematische seien.¹³ Aristoteles hat die mathematisch-experimentelle Methode nicht gekannt, mit seinen auf unmittelbarer Beobachtung beruhenden Methoden musste er zu anderen Ergebnissen kommen.

Uns als modernen Menschen sind die grundlegenden Auffassungen und Prinzipien der neuzeitlichen Naturwissenschaft so vertraut, dass wir ihre Ergebnisse schlicht für Tatsachen halten. „Uns entgeht die Waghalsigkeit Galileis, mit der er beschließt, die Mechanik als Zweig der Mathematik zu behandeln, also die wirkliche Welt der täglichen Erfahrung durch eine bloß vorgestellte Welt der Geometrie zu ersetzen“ (KOYRÉ 1998, 73), und Aussagen über die Natur gegen alle Empirie aus mathematischen Begriffen wie Zeit, Raum und Bewegung abzuleiten. Die hieraus entwickelte Naturauffassung, die uns so evident erscheint, wäre in der griechischen Antike und im Mittelalter als offenkundig falsch oder gar absurd eingestuft worden (vgl. KOYRÉ 1998, 70 - 87).

Es kann heutzutage natürlich nicht darum gehen, zu „vormodernen“ Auffassungen zurückzukehren. Für das Verständnis der modernen Welt wichtig ist aber nicht nur die Kenntnis grundlegender naturwissenschaftlicher Ergebnisse (gar in Form unerschütterlicher „Tatsachen“), sondern ebenso die Kenntnis und Reflexion der Methoden, mit denen sie zu Stande kommen.

1.2 Newtons Instrumente: Infinitesimalrechnung und Differentialgleichungen

1.2.1 Das Gesetz der Veränderung als neuer mathematischer Gegenstand

Die neue Methode der Naturbetrachtung führt rasch auf mathematische Probleme, die mit den bis dahin zur Verfügung stehenden Begriffen und Methoden nicht zu lösen sind. An die Stelle geometrischer Objekte in bekannter Gestalt (Punkte, Geraden, Kreise, Kegelschnitte) treten beliebige Kurven, deren Form es zu bestimmen gilt. Isaac Newton (1642 - 1727) formuliert die beiden Grundaufgaben der Infinitesimalrechnung direkt in der Sprache der Mechanik (PEIFFER / DAHAN-DALMEDICO, 236):

¹²Ein gleichförmig horizontaler und zugleich gleichförmig beschleunigter Bewegung unterworfenen Körper beschreibt eine Halbparabel. (GALILEI 1638/1995, 218)

¹³In dieser Hinsicht steht Galilei und mit ihm die neuzeitliche Physik Platon näher als Aristoteles, vgl. KOYRÉ 1998, 88 - 122.

1. Die Länge der durchlaufenen Strecke sei kontinuierlich angegeben. Man soll die Geschwindigkeit der Bewegung zu einem beliebigen aber festen Zeitpunkt bestimmen.
2. Die Geschwindigkeit der Bewegung sei kontinuierlich angegeben. Zu bestimmen ist die Länge der durchlaufenen Strecke zu einem beliebigen aber festen Zeitpunkt.

Die bis in das 19. Jahrhundert hinein gehende Entwicklung und Präzisierung der Methoden und Begriffe (zu denen auch der der *Funktion* gehört) will ich hier nicht nachvollziehen, sondern vielmehr die Fragestellungen in moderner Darstellung erläutern.

Das erste Grundproblem lautet dann: Gegeben sei der Ort (die durchlaufene Strecke) $x(t)$ zu jedem Zeitpunkt t . Gesucht ist die Geschwindigkeit

$$v(t) = \dot{x}(t)$$

zu jedem Zeitpunkt t . Nach den üblichen physikalischen Konventionen bezeichne dabei

$$\dot{x} = \frac{dx}{dt}$$

die Ableitung nach der Zeit. Gesucht ist also die *Ableitungsfunktion* einer gegebenen (differenzierbaren) Funktion. Für die üblicher Weise betrachteten, so genannten elementaren Funktionen (Polynome, trigonometrische Funktionen, Exponentialfunktion, Logarithmus und daraus zusammengesetzte Funktionen) ist dieses Problem bekanntlich durch den Differentialkalkül seit langer Zeit vollständig gelöst und lässt sich heute automatisch (durch Computeralgebra-Systeme wie *Mathematica* oder *Maple*) exakt bearbeiten.

Anders sieht es mit dem zweiten Grundproblem aus. Gegeben ist hier im einfachsten Fall die Geschwindigkeit $v(t)$ als Funktion der Zeit t . Gesucht ist die Länge der durchlaufenen Strecke zu jedem Zeitpunkt, also eine Funktion $x(t)$ des Ortes von der Zeit, für die

$$\dot{x}(t) = v(t) .$$

Nach dem Hauptsatz der Differential- und Integralrechnung ist das nach Wahl eines beliebigen Zeitpunkts t_0 äquivalent zu

$$x(t) = x(t_0) + \int_{t_0}^t v(s) ds .$$

Zu bestimmen ist also das *Integral* der gegebenen Geschwindigkeitsfunktion $v(t)$. Dieses existiert zwar für eine sehr viel größere Klasse von Funktionen als es die Ableitung tut (etwa für alle stetigen Funktionen), aber selbst für elementare Funktionen lässt es sich nur in Sonderfällen explizit angeben (d. h. wiederum durch elementare Funktionen ausdrücken). Für physikalische Zwecke kommen daher *numerische Näherungsverfahren* ins Spiel, genauer gesagt Verfahren der *numerischen Integration*, wie z. B. die *Newton-Cotes-Formeln*, die das Integral durch endlich viele Auswertungen der zu integrierenden Funktion approximieren. Entsprechende Verfahren gehören heute zu jeder numerischen Standardsoftware, setzten aber den Möglichkeiten präziser physikalischer Berechnungen vor der Erfindung des Computers enge Kapazitätsgrenzen.

In allgemeinerer Form, wie sie in der Mechanik die Regel ist, lautet das zweite Grundproblem: Gegeben ist die Geschwindigkeit $v = v(t, x)$ als Funktion der Zeit t und des Ortes x . Gesucht ist eine Funktion $x(t)$ des Ortes von der Zeit, für die

$$\dot{x}(t) = v(t, x(t)) .$$

In dieser Form handelt es sich bei der zweiten Grundaufgabe der Infinitesimalrechnung um das Problem der Lösung oder *Integration* einer *Differentialgleichung*. Was schon für die Integration von Funktionen

gesagt wurde, gilt auch hier: Erst im Computerzeitalter ist es möglich geworden, beliebige Differentialgleichungen beliebig genau numerisch zu lösen, was die exakte Lösung in Einzelfällen allerdings nicht ausschließt:

Newton legte in seinem Hauptwerk, den *Principia* (NEWTON 1687/1988) eine mathematisch-deduktive und vereinheitlichende Begründung der von Johannes Kepler (1571 - 1630) entdeckten Gesetzmäßigkeiten der Himmelsbewegungen und der wesentlich auf Galilei zurückgehenden sublunaren Physik vor¹⁴ und machte damit aus der Mechanik ein naturwissenschaftliches und philosophisches Denkgebäude, das erst knapp 200 Jahre später zu bröckeln begann. Zugleich sind die *Principia* so etwas wie ein „erstes Lehrbuch der Differentialgleichungen“, indem dort die Bewegung von Körpern auf abstrakte mathematische Weise, eben in Form von Differentialgleichungen abgehandelt werden. Der verwendete Formalismus und die Methoden sind allerdings noch ganz andere, als sie uns heute zur Verfügung stehen. Newton muss den sowohl in der Ableitung als auch im Integral steckenden Grenzwert für jedes einzelne Beispiel „neu erfinden“, ohne ihn dabei auf den Begriff zu bringen. Umso bemerkenswerter sind die Erfolge der von ihm verwendeten Methoden, auf die hier nicht weiter eingegangen wird. Ich stelle im Folgenden einige Beispiele der klassischen Mechanik vor, verwende dabei aber den heute gebräuchlichen formalen Apparat. In den meisten Fällen handelt es sich dabei übrigens nicht um Differentialgleichungen für die Orts-, sondern um solche für die Geschwindigkeitsfunktion.

1.2.2 Gleichförmig beschleunigte Bewegung

Ist $v(t)$ die Geschwindigkeitsfunktion einer gleichförmig beschleunigten Bewegung, so darf bei gegebenem Zeitintervall der Länge $\Delta t > 0$ der Geschwindigkeitszuwachs $v(t + \Delta t) - v(t)$ nur von Δt abhängen:

$$v(t + \Delta t) - v(t) = f(\Delta t)$$

mit einer zunächst beliebigen reellen Funktion f . Die Existenz der gebildeten Grenzwerte vorausgesetzt, ist dann

$$\dot{v}(t) = \lim_{\Delta t \rightarrow 0} \frac{v(t + \Delta t) - v(t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{f(\Delta t)}{\Delta t} =: b$$

mit einer reellen Zahl b , der *konstanten Beschleunigung*. Für einen beliebig gewählten Anfangszeitpunkt t_0 und zugehöriger Anfangsgeschwindigkeit $v(t_0) = v_0$ ergibt sich durch Integration

$$v(t) = v_0 + \int_{t_0}^t b \, ds = v_0 + b(t - t_0).$$

Bezeichnet $x(t)$ den Ort der Bewegung zur Zeit t , so ist

$$\dot{x}(t) = v(t),$$

und daher liefert nochmalige Integration mit dem Anfangsort $x(t_0) = x_0$

$$x(t) = x_0 + v_0(t - t_0) + \frac{b}{2}(t - t_0)^2. \quad (1.1)$$

Mit der speziellen Wahl $t_0 = 0$, $x_0 = 0$ und $v_0 = 0$ ergibt sich hieraus

$$x(t) = \frac{b}{2} t^2$$

das Fallgesetz entsprechend der Formulierung in Satz 1.1.4.

An diesem einfachsten Beispiel einer Differentialgleichung lässt sich hinsichtlich der adäquaten Problemstellung erkennen, was für Differentialgleichungen allgemein gilt: Um eine *eindeutige* Lösung zu

¹⁴In der aristotelischen Physik herrschten in den himmlischen Sphären noch ganz andere Gesetzmäßigkeiten als auf der Erde

erhalten, ist außer der Differentialgleichung ein *Anfangswert* vorzugeben, d. h. die in diesem Sinne ad-äquate Problemstellung für die gesuchte Funktion v hat die Form einer *Anfangswertaufgabe*

$$\dot{v} = b, v(t_0) = v_0.$$

Das Problem der Bestimmung der gesuchten Ortsfunktion x lässt sich auch als *Differentialgleichung 2. Ordnung* formulieren, in der die zweite Ableitung von x auftritt, der dann aber für die eindeutige Lösbarkeit zwei Anfangsbedingungen zugeordnet werden müssen:

$$\ddot{x} = b, x(t_0) = x_0, \dot{x}(t_0) = v_0.$$

Wie jede Differentialgleichung 2. Ordnung lässt sich auch diese als ein *System von Differentialgleichungen 1. Ordnung* in zwei unbekannt Funktionen (x und v) formulieren, was hier der ursprünglich gestellten Aufgabe eher entspricht:

$$\begin{aligned} \dot{x} &= v, & x(t_0) &= x_0 \\ \dot{v} &= b, & v(t_0) &= v_0 \end{aligned}$$

1.2.3 Im Verhältnis zur Geschwindigkeit gebremste Bewegung

Satz 1.2.1 (NEWTON 1687/1988, 123)

Die aufgrund des Widerstands verlorene Bewegung eines Körpers, dem im Verhältnis der Geschwindigkeit Widerstand geleistet wird, verhält sich wie die Wegstrecke, die er in seiner Bewegung zurückgelegt hat.

Der kurze, bündige Beweis, den Newton hierfür gibt, lautet: *Denn weil die in einzelnen gleichen Zeitteilchen verlorene Bewegung wie die Geschwindigkeit ist, d. h. wie die Teilchen des zurückgelegten Weges, so wird, durch Aneinanderreihen, die in der ganzen Zeit verlorene Bewegung sein wie der ganze Weg. Q.e.d.*

Das „Aneinanderreihen der Zeitteilchen“ entspricht im modernen Formalismus der Integration. Newtons Vorgehen besteht also darin, die zu Grunde liegende Differentialgleichung

$$\dot{v} = -c v \text{ (mit einer Konstanten } c > 0)$$

zu integrieren:

$$v(t) = v(t_0) - c \int_{t_0}^t v(s) ds = v(t_0) - c (x(t) - x(t_0)),$$

woraus sich für die „verlorene Bewegung“

$$v(t_0) - v(t) = c (x(t) - x(t_0))$$

ergibt, was der Aussage des Satzes entspricht.

Satz 1.2.2 (NEWTON 1687/1988, 124)

Wenn einem Körper im Verhältnis der Geschwindigkeit Widerstand geleistet wird und derselbe sich allein aufgrund der eingepflanzten Kraft durch ein homogenes Medium bewegen soll, dabei aber gleiche Zeiten genommen werden sollen, so stehen die Geschwindigkeiten an den Anfängen der einzelnen Zeiten in geometrischer Progression, und die in den einzelnen Zeiten beschriebenen Wegstrecken verhalten sich wie die Geschwindigkeiten.

Der von Newton gegebene Beweis ist nicht so kurz wie der von Satz 1.2.1, ich gehe daher sofort zur modernen Darstellung über:

Die gesuchte Geschwindigkeitsfunktion v genüge der Anfangswertaufgabe

$$\dot{v} = -c v, v(0) = v_0 \quad (1.2)$$

mit einer positiven Konstanten c und einer Anfangsgeschwindigkeit $v_0 > 0$. Der Trick zur Lösung der Differentialgleichung besteht darin, alle Terme, in denen v vorkommt, auf die linke Seite zu bringen. Für alle Zeiten t , für die $v(t) > 0$, ist dann mit dem natürlichen Logarithmus \log

$$\frac{d}{dt} \log v(t) = \frac{\dot{v}(t)}{v(t)} = -c$$

und daher nach Integration

$$\log v(t) - \log v_0 = -c t .$$

Wendet man hierauf die (Newton noch nicht bekannte) Exponentialfunktion an, so ergibt sich die Lösung

$$v(t) = v_0 e^{-c t} . \quad (1.3)$$

In Kenntnis dieser Lösung von (1.2) ist deren Herleitung noch einfacher und allgemeiner (ohne die Voraussetzung $v(t) > 0$) zu haben: Für jede Lösung von (1.2) gilt nämlich

$$\frac{d}{dt} [e^{c t} v(t)] = e^{c t} [\dot{v}(t) + c v(t)] = 0 ,$$

woraus die Konstanz des Ausdrucks in eckigen Klammern auf der linken Seite folgt, der aber für $t = 0$ den Wert v_0 hat. Daraus ergibt sich die bereits gefundene Lösung (1.3) ohne jede Voraussetzung an das Vorzeichen von v_0 und $v(t)$.

Integriert man (1.3), so ergibt sich mit der Anfangsbedingung $x(0) = 0$

$$x(t) = \int_0^t v(s) ds = \frac{v_0}{c} (1 - e^{-c t}) . \quad (1.4)$$

Aufgabe 1.1

Zeigen Sie, dass die Lösung (1.3), (1.4) die von Newton in Satz 1.2.2 angegebenen Eigenschaften hat.

Aufgabe 1.2

In einem Brief aus dem Jahre 1604 (zitiert in STENGERS 1998, 418) hat Galilei die Annahme vertreten, „daß der schwere Körper im natürlichen Fall beständig seine Geschwindigkeit vergrößert in dem Maße, wie die Entfernung von seinem Ausgangspunkt wächst“. Danach wäre also die Geschwindigkeit nicht proportional zur Zeit, sondern proportional zum zurückgelegten Weg. Wie würde unter dieser Annahme der freie Fall aus der Ruhelage aussehen? Lässt sich, ohne präzise Messungen durchzuführen, begründen, warum Galilei diese Annahme wieder fallen gelassen und durch die der gleichförmig beschleunigten Bewegung entsprechend Definition 1.1.2 ersetzt hat?

1.2.4 Beschleunigung gegen den Widerstand eines Mediums: Linearer Fall

Im Anschluss an Satz 1.2.2 betrachtet NEWTON (1687/1988, 125) die Bewegung eines Körpers, der einer gleichförmigen Schwere, also einer konstanten Beschleunigung unterliegt, dem aber zugleich durch das Medium, in dem er sich bewegt, Widerstand *im Verhältnis seiner Geschwindigkeit* geleistet wird. In moderner Form ist also für die Geschwindigkeit eine Differentialgleichung

$$\dot{v} = b - c v \quad (1.5)$$

mit positiven Konstanten b und c zu untersuchen. Eine *spezielle* Lösung von (1.5) ist die konstante Funktion

$$\bar{v}(t) = \frac{b}{c} \quad (1.6)$$

Aus ihr und der Kenntnis der Lösungen der zugehörigen *homogenen* Gleichung

$$\dot{v} = -c v \quad (1.7)$$

lassen sich aber sofort *alle* Lösungen von (1.5) bestimmen, und zwar auf Grund des folgenden, sofort nachzurechnenden Ergebnisses:

Satz 1.2.3

v ist genau dann eine Lösung von (1.5), wenn $v - \bar{v}$ eine Lösung von (1.7) ist.

Nun sind aber die Lösungen von (1.7) alle Funktionen $a e^{-ct}$ mit einer reellen Zahl a . Alle Lösungen von (1.5) sind daher von der Form

$$v(t) = \frac{b}{c} + a e^{-ct} \text{ mit } a \in \mathbb{R}.$$

Für die Lösung der Anfangswertaufgabe

$$\dot{v} = b - c v, v(0) = v_0 \quad (1.8)$$

muss

$$v_0 = v(0) = \frac{b}{c} + a$$

und daher

$$a = v_0 - \frac{b}{c}$$

sein, woraus sich als eindeutige Lösung von (1.8)

$$v(t) = \frac{b}{c} + (v_0 - \frac{b}{c}) e^{-ct} = \frac{b}{c} (1 - e^{-ct}) + v_0 e^{-ct} \quad (1.9)$$

ergibt. Wegen $e^{-ct} \rightarrow 0$ gilt $v(t) \rightarrow \frac{b}{c}$ für $t \rightarrow \infty$, langfristig wird die Bewegung gleichförmig.

1.2.5 Beschleunigung gegen den Widerstand eines Mediums: Quadratischer Fall

Im letzten Beispiel wurde angenommen, dass der vom Medium gegen die Bewegung geleistete Widerstand der Geschwindigkeit proportional sei. Zu dieser Voraussetzung bemerkt NEWTON (1687/1988, 134):

Übrigens ist die Hypothese, daß der Widerstand der Körper sich wie die Geschwindigkeit verhalte, eher eine mathematische als eine natürliche. In Medien, die frei von aller Starrheit (Zähigkeit) sind, stehen die Widerstände der Körper im quadratischen Verhältnis der Geschwindigkeiten. Denn durch die Einwirkung eines schnelleren Körpers wird derselben Menge des Mediums in kürzerer Zeit eine größere Bewegung proportional zur größeren Geschwindigkeit mitgeteilt; und daher wird in gleicher Zeit, wegen der größeren Menge des aufgestörten Mediums, eine im quadratischen Verhältnis größere Bewegung mitgeteilt.

Die Argumentation beruht darauf, dass die „Menge des aufgestörten Mediums“, der die Geschwindigkeit v je Zeiteinheit „mitgeteilt“ wird, wiederum zu v proportional ist. Betrachtet wird daher ein Bewegungsgesetz der Form

$$\dot{v} = b - c v^2, v(0) = v_0$$

mit positiven Konstanten b und c und einer beliebigen Anfangsgeschwindigkeit $v_0 \geq 0$. Mit

$$\bar{v} = \sqrt{\frac{b}{c}}$$

lässt es sich in die Form

$$\dot{v} = c (\bar{v}^2 - v^2), \quad v(0) = v_0 \quad (1.10)$$

bringen. Ist $v_0 = \bar{v}$, so ist offenbar die konstante Funktion $v = \bar{v}$ Lösung von (1.10). Ist dagegen $v_0 \neq \bar{v}$, so besteht der Trick zur Lösung der Differentialgleichung wiederum darin, alle Terme, in denen v auftritt, auf die linke Seite zu bringen. Für $v(t) \neq \bar{v}$ ist

$$\frac{\dot{v}}{2\bar{v}} \left(\frac{1}{\bar{v}+v} + \frac{1}{\bar{v}-v} \right) = \frac{\dot{v}}{\bar{v}^2 - v^2} = c.$$

Die linke Seite wurde hier (durch Partialbruchzerlegung) in eine Form gebracht, für die sich leicht eine Stammfunktion angeben lässt:

$$\frac{d}{dt} (\log|\bar{v}+v(t)| - \log|\bar{v}-v(t)|) = \dot{v}(t) \left(\frac{1}{\bar{v}+v(t)} + \frac{1}{\bar{v}-v(t)} \right) = 2c\bar{v}.$$

Unter Berücksichtigung der Rechenregeln für Logarithmus und Exponentialfunktion folgt daraus

$$\frac{\bar{v}-v(t)}{\bar{v}+v(t)} = A e^{-2c\bar{v}t}$$

mit einer Konstanten A , die sich wegen der Anfangsbedingung zu

$$A = \frac{\bar{v}-v_0}{\bar{v}+v_0}$$

ergibt. Umstellen nach $v(t)$ und anschließendes Einsetzen von A liefert als Lösung von (1.10)

$$v(t) = \frac{\bar{v}+v_0 - (\bar{v}-v_0) e^{-2c\bar{v}t}}{\bar{v}+v_0 + (\bar{v}-v_0) e^{-2c\bar{v}t}} \bar{v} = \frac{\bar{v} \sinh(c\bar{v}t) + v_0 \cosh(c\bar{v}t)}{\bar{v} \cosh(c\bar{v}t) + v_0 \sinh(c\bar{v}t)} \bar{v}. \quad (1.11)$$

Auch hier gilt $v(t) \rightarrow \bar{v}$ für $t \rightarrow \infty$, die Bewegung wird gleichförmig.

1.2.6 Beschleunigung gegen den Widerstand eines Mediums: Allgemeiner Fall

Auch wenn Newton versucht, durch theoretische Überlegungen plausibel zu machen, dass der der Bewegung eines Körpers vom Medium geleistete Widerstand quadratisch von seiner Geschwindigkeit abhängen muss, so lässt sich ohne empirische Überprüfungen nicht sagen, dass wirkliche Körper sich so verhalten. Die schwächste Annahme, die sich im Rahmen des Newton'schen Ansatzes machen lässt, dass nämlich der Widerstand eine Funktion der Geschwindigkeit sei, besteht darin, diese Funktion $f(v)$ als stetig und streng monoton über alle Grenzen wachsend vorzusetzen, für die zudem $f(0) = 0$. Ist b die konstante Beschleunigung, so existiert wegen der Eigenschaften von f eine eindeutig bestimmte positive Geschwindigkeit \bar{v} mit $f(\bar{v}) = b$, was zu dem Bewegungsgesetz

$$\dot{v} = f(\bar{v}) - f(v), \quad v(0) = v_0 \text{ mit einem } v_0 \geq 0 \quad (1.12)$$

führt. Auch wenn sich hier, mit einer nicht näher spezifizierten Funktion f , die Lösungen nicht explizit angeben lassen, so ist es doch möglich, ihr Verhalten qualitativ zu bestimmen (vgl. Abbildung 1.4). Wegen

$$\frac{d}{dt} (v(t) - \bar{v})^2 = 2(v(t) - \bar{v}) \dot{v}(t) = 2(v(t) - \bar{v})(f(\bar{v}) - f(v(t))) \leq 0$$

nähert sich jede Lösung $v(t)$ mit wachsendem t dem Wert \bar{v} auf monotone Weise von oben oder unten. Daher existiert $\lim_{t \rightarrow \infty} v(t) = v_1$, und wegen der Monotonie von v ist dann $\lim_{t \rightarrow \infty} \dot{v}(t) = 0$. Der selbe Grenzübergang liefert dann in der Differentialgleichung (1.12)

$$0 = \lim_{t \rightarrow \infty} \dot{v}(t) = f(\bar{v}) - \lim_{t \rightarrow \infty} f(v(t)) = f(\bar{v}) - f(v_1),$$

woraus wegen der strengen Monotonie von f folgt, dass $v_1 = \bar{v}$. Es gilt somit

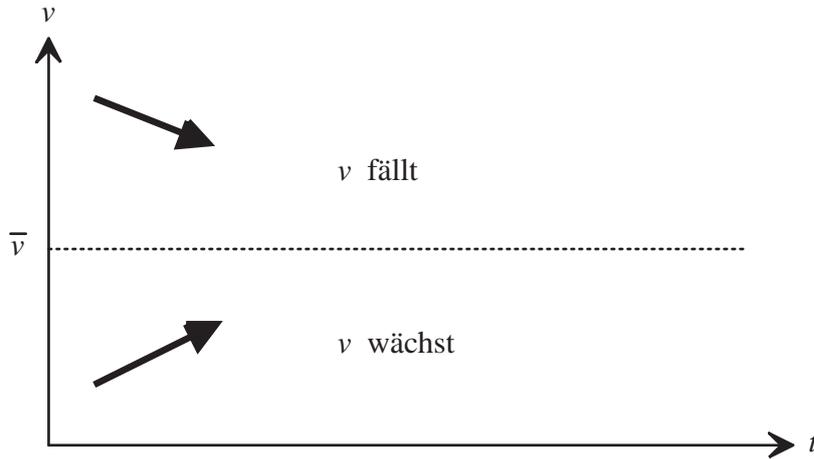


Abbildung 1.4: Beschleunigung gegen Widerstand

Satz 1.2.4

Sei v Lösung von (1.12). Ist $v_0 = \bar{v}$, so ist

$$v(t) = \bar{v} \text{ für alle } t \geq 0.$$

Ist $v_0 < \bar{v}$, so ist

$$v \text{ monoton wachsend, } v(t) \leq \bar{v} \text{ für alle } t \geq 0, \text{ und } \lim_{t \rightarrow \infty} v(t) = \bar{v}.$$

Ist $v_0 > \bar{v}$, so ist

$$v \text{ monoton fallend, } v(t) \geq \bar{v} \text{ für alle } t \geq 0, \text{ und } \lim_{t \rightarrow \infty} v(t) = \bar{v}.$$

In jedem Fall gilt also auch hier: $v(t) \rightarrow \bar{v}$ für $t \rightarrow \infty$, die Bewegung wird gleichförmig.

1.2.7 Galileis Zeitmessungen

Wie oben zitiert, berichtet Galilei, dass er bei seinen Experimenten die Zeit gemessen habe, indem er Wasser aus einem kleinen Loch in einem Eimer auffing und anschließend genau wog. Die Frage ist, ob die gewogene Wassermenge tatsächlich proportional zur verflossenen Zeit ist. In einem zylindrischen Gefäß mit der Grundfläche a befindet sich unten ein kleines Loch mit dem Querschnitt b . Im Gefäß befindet sich Wasser, welches im Laufe der Zeit aus dem Loch austritt. Sei $h = h(t)$ die Höhe des Wasserspiegels über dem Loch zum Zeitpunkt t (vgl. Abbildung 1.5). Evangelista Torricelli (1608 - 1647), ein Schüler Galileis, fand im Jahre 1643 durch theoretische Überlegungen, auf die hier nicht genauer eingegangen wird, heraus,¹⁵ dass die Geschwindigkeit, mit der das Wasser aus dem Loch tritt, ebenso groß sein muss wie die Geschwindigkeit, die ein Gegenstand erreicht, der aus der Ruhelage um die Höhe h frei herunterfällt, also

$$v = \sqrt{2gh}$$

mit der Erdbeschleunigung g . Nun ist aber bv gerade das Wasservolumen, welches pro Zeiteinheit den Eimer durch das Loch verlässt. Für das gesamte, oberhalb des Loches befindliche Wasservolumen

$$V = ah$$

¹⁵Aus dem Energierhaltungssatz lässt sich das Ergebnis schnell herleiten: Bei einer kleinen Volumenänderung ΔV nimmt die potentielle Energie um den Betrag $\rho \Delta V gh$ ab, der in den gleichen Betrag $\frac{1}{2}\rho \Delta V v^2$ kinetischer Energie der ausströmenden Flüssigkeit umgesetzt wird. ρ bezeichne deren Dichte.

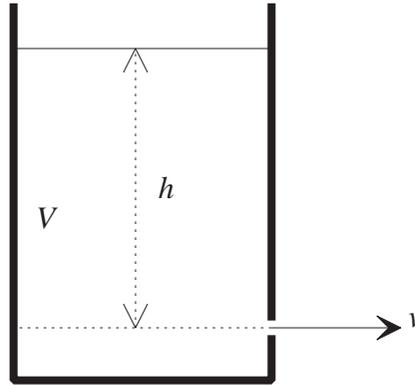


Abbildung 1.5: Loch im Wassereimer

gilt daher

$$a \dot{h} = \dot{V} = -b v = -b \sqrt{2gh}.$$

Ist $h_0 > 0$ die Höhe des Wasserspiegels zum Beginn des Experiments (Zeitpunkt 0), so genügt also h der Anfangswertaufgabe

$$\dot{h} = -c \sqrt{h}, \quad h(0) = h_0, \quad \text{wobei } c = \frac{b}{a} \sqrt{2g} \quad (1.13)$$

Der Trick zur Lösung dieser Differentialgleichung besteht wieder darin, alle h enthaltenden Terme auf die linke Seite zu bringen und eine Stammfunktion für den dort entstehenden Ausdruck zu finden: Solange $h(t) > 0$, gilt

$$\frac{d}{dt} (2 \sqrt{h(t)}) = \frac{\dot{h}(t)}{\sqrt{h(t)}} = -c,$$

woraus

$$\sqrt{h_0} - \sqrt{h(t)} = \frac{c}{2} t$$

folgt. Lösung dieser Gleichung nach $h(t)$ liefert

$$h(t) = \left(\sqrt{h_0} - \frac{c}{2} t \right)^2.$$

Diese Herleitung stand unter der Voraussetzung $h(t) > 0$, was offenbar der Fall ist, solange

$$t < \frac{2}{c} \sqrt{h_0} =: \bar{t}.$$

Dagegen ist für die gefundene Lösung $h(\bar{t}) = 0$, d. h. zum Zeitpunkt \bar{t} ist das Gefäß leer (jedenfalls oberhalb des Loches) und muss es dann aus physikalischen Gründen auch zukünftig bleiben. Als Lösung von (1.13) ergibt sich daher

$$h(t) = \begin{cases} \left(\sqrt{h_0} - \frac{c}{2} t \right)^2 & \text{für } t < \frac{2}{c} \sqrt{h_0} \\ 0 & \text{für } t \geq \frac{2}{c} \sqrt{h_0} \end{cases}. \quad (1.14)$$

Die im Zeitintervall $[0, t]$ mit $t < \bar{t}$ ausgelaufene Wassermenge ist nun aber proportional zu

$$h_0 - h(t) = c \sqrt{h_0} t - \frac{c^2}{4} t^2$$

und daher *nicht* proportional zu t . Für kleine Zeiten t fällt allerdings die in t quadratische Abweichung von der Proportionalität nicht ins Gewicht.

Zu beachten ist, dass es sich hier um eine *Theorie* des Wasseraustritts aus dem Gefäß handelt, für Galileis Fallversuche also eine *Theorie des Messinstruments*. Ob wirkliches Wasser in wirklichen Gefäßen sich so verhält, wäre allerdings erst noch experimentell zu überprüfen.

1.3 Trennung von Mathematik und Physik und Verallgemeinerung ihrer Methode

Einer auf Galilei zurückgehenden Metapher zufolge ist Mathematik „die Sprache, in der das Buch der Natur geschrieben“ sei, und Newtons Hauptwerk enthält bereits im Titel eine ähnliche Aussage. Damit wird der Mathematik eine gewisse Eigenständigkeit zwar nicht abgesprochen, schließlich lassen sich in ein und derselben Sprache verschiedene Bücher schreiben.¹⁶ Die *Neuentwicklung* mathematischer Begriffe und Konzepte in der Geometrie und insbesondere der Analysis, bei Rechenverfahren und in der Datenanalyse geht jedoch Hand in Hand mit den damit zu lösenden Problemen der neuen Naturwissenschaft. Bis ins 19. Jahrhundert hinein bilden Mathematik und Physik eine Einheit. In dieser Sicht sind mathematische Sätze einfach wahre Aussagen über die Natur, und seien sie auch größtenteils nur hypothetischer Art, wenn sich die Voraussetzungen im Experiment nicht so ohne weiteres herstellen lassen.

Im Laufe des 19. Jahrhunderts zerfällt diese Einheit, weil in verschiedener Hinsicht Mehrdeutigkeiten auftreten. Zum einen entwickeln sich konsistente mathematische Theorien, die miteinander nicht vereinbar sind und daher nicht alle zugleich wahre Aussagen über die Natur enthalten können. Zum anderen zeigt sich, dass es verschiedene, zu einander konkurrierende mathematische Theorien zur Erklärung derselben Naturvorgänge geben kann, wodurch erst deutlich wird, dass die mathematische Erklärung nicht der Natur entspringt, sondern auch anderen Gesichtspunkten, wie etwa dem der Zweckmäßigkeit folgt. Am Ende der davon ausgelösten Entwicklung steht die Ablösung der Mathematik von der Physik, ein radikal anderer mathematischer Wahrheitsbegriff und der Begriff des mathematischen Modells in seiner modernen Gestalt.

1.3.1 Nichteuklidische Geometrie

Die um 325 vor unserer Zeitrechnung entstandenen *Elemente* des Euklid sind das erfolgreichste Buch der Mathematikgeschichte, das mehr als zweitausend Jahre lang als Grundlage aller mathematischen Studien gedient hat. Ausgehend von wenigen Axiomen (evidenten Wahrheiten) wird darin die bis dahin bekannte Mathematik systematisch zusammengefasst und hergeleitet. Für die neuzeitliche Naturwissenschaft ist die klassische euklidische Geometrie von Anfang an die adäquate Beschreibung des physikalischen Vorstellung des Raumes. Für Kant zählt sie, ebenso wie der Begriff der Zeit, zu den Prinzipien der Vernunft, die nicht aus der Empirie gewonnen werden, sondern umgekehrt deren systematische Wahrnehmung erst ermöglichen (KANT 1781/1787/1990, 67):

Der Raum ist eine notwendige Vorstellung a priori, die allen äußeren Anschauungen zum Grunde liegt. Man kann sich niemals eine Vorstellung davon machen, daß kein Raum sei, ob man sich gleich ganz wohl denken kann, daß keine Gegenstände darin angetroffen werden. Er wird also als die Bedingung der Möglichkeit der Erscheinungen, und nicht als eine von ihnen abhängende Bestimmung angesehen, und ist eine Vorstellung a priori, die notwendigerweise äußeren Erscheinungen zum Grunde liegt.

¹⁶In den bürgerlichen Zentren, in denen im 17. Jahrhundert die Naturwissenschaft blühte, waren mathematische Fertigkeiten vor allem auch im Rechnungswesen gefragt.

Geometrische Sätze sind demgemäß „synthetische Urteile a priori“, wahre Aussagen über die Natur, die sich gleichwohl unter Absehung von aller Empirie entdecken lassen, und die Geometrie ist die Wissenschaft von dem einen, einzigen Raum, der ohne Alternative ist. Zu Beginn des 19. Jahrhunderts kommen daran Zweifel auf, aufgeworfen durch die Lösung eines alten, den *Elementen* immanenten Problems: Die ersten fünf der euklidischen Axiome lauten (DAVIS/HERSH 1985, 224).

1. *Zwischen zwei beliebigen Punkten kann eine Strecke gezogen werden.*
2. *Jede Strecke kann unbegrenzt gerade verlängert werden.*
3. *Ein Kreis kann mit jedem vorgegebenen Punkt als Mittelpunkt und jedem vorgegebenen Radius gezogen werden.*
4. *Alle rechten Winkel sind gleich.*
5. *Wenn zwei Geraden, die in einer Ebene liegen, sich mit einer anderen Geraden schneiden und wenn die Summe der Innenwinkel auf der einen Seite weniger als diejenige von zwei rechten Winkeln ausmacht, dann werden sich die Geraden schneiden, wenn sie auf der Seite, auf der die Winkelsumme weniger ist als zwei rechte Winkel, ausreichend verlängert werden.*

Während die ersten vier Axiome einfach und evident erscheinen, ist es das fünfte offenbar nicht. Es ist viel komplizierter zu formulieren, und auch wenn man sich durch eine Zeichnung seiner Gültigkeit versichern kann, so ist diese doch von derselben Qualität wie die vieler geometrischer Sätze, die Euklid erst beweist. Die Frage liegt daher nahe, ob dieses Axiom sich nicht vielleicht aus den anderen herleiten lasse. Es wird als *Parallelenaxiom* bezeichnet, obwohl das Wort „parallel“ darin nicht vorkommt, und von Euklid in der Tat erst später eingeführt wird:

Parallele Geraden sind Geraden, die in derselben Ebene liegen, unbeschränkt in beide Richtungen weitergeführt werden und sich weder in der einen noch in der anderen Richtung schneiden.

Mit dieser Definition ist das fünfte Axiom (unter Voraussetzung der ersten vier) zur folgenden Aussage äquivalent:

6. *In einer Ebene seien eine Gerade g und ein Punkt P nicht auf g gegeben. Dann gibt es genau eine Gerade durch P , die zu g parallel ist.*

was heute die gebräuchlichste Formulierung des Parallelenaxioms ist.

Bis Ende des 18. Jahrhunderts bemühten sich Mathematiker, dieses Axiom aus den ersten vier zu beweisen, es gelangen aber immer nur neue äquivalente Formulierungen, etwa (Legendre): *Die Winkelsumme im Dreieck beträgt zwei rechte Winkel*. Ebenso wenig führten Versuche eines indirekten Beweises zum Erfolg, die Annahme der Gültigkeit der ersten vier Axiome und der Ungültigkeit des fünften zu einem Widerspruch zu bringen.

Carl Friedrich Gauß (1777 - 1855) ist wohl der erste, dem es gelingt, unter der Annahme der ersten vier Axiome und der Ersetzung des fünften durch ein anderes, nämlich

7. *Zu einer Geraden g und einem nicht auf ihr liegenden Punkt P gibt es unendlich viele Geraden durch P , die g nicht schneiden.*

eine neue Geometrie zu konstruieren, die er *nichteuklidisch* nennt. Gauß selber hat in Erwartung mangelnder Akzeptanz seine Überlegungen nie veröffentlicht. Sie werden allerdings in einem Briefwechsel erwähnt (s. MEHRTENS 1990, 46). Auch den unabhängig von Gauß und voneinander veröffentlichten Arbeiten von Janos Bolyai (1802 - 1860) und Nikolai Iwanowitsch Lobatschewskij (1793 - 1856) bleibt

zunächst wenig Resonanz beschieden. In der auf Gauß/Bolyai/Lobatschowskij zurückgehenden, heute als *hyperbolisch* bezeichneten Geometrie ist *die Winkelsumme im Dreieck stets kleiner als zwei rechte Winkel*.

Gauß scheint sich des Spaltes, der sich hier zwischen Mathematik und Physik aufzutun beginnt, bewusst zu sein. Zumindest die Zurechnung geometrischer Konzepte zum Kant'schen Apriori ist ihm fraglich geworden, wenn er schreibt (zitiert nach MEHRTENS 1990, 46)

*Wir müssen in Demuth zugeben, dass, wenn die Zahl bloss unseres Geistes Product ist, der Raum auch ausser unserem Geiste eine Realität hat, der wir a priori unsere Gesetze nicht vollständig vorschreiben können.*¹⁷

Geodät, der er auch ist, versucht Gauß empirisch festzustellen, welche Geometrie der Realität am besten angepasst ist, indem er die Winkelsumme im Dreieck ausmisst, das von drei Berggipfeln im Harz gebildet wird. Das Ergebnis übertrifft die zwei rechten Winkel um weniger als 15 Bogenminuten und bleibt damit im Rahmen der Messgenauigkeit auf dem Boden der euklidischen Geometrie.

Bernhard Riemann (1826 - 1866) schließlich stellt in seinem berühmten Habilitationsvortrag von 1854 (veröffentlicht 1868) *über die Hypothesen, welcher der Geometrie zu Grunde liegen eine Geometrie vor, die elliptische*, in der das fünfte euklidische Axiom ersetzt ist durch

8. *Zu einer Geraden g und einem nicht auf ihr liegenden Punkt P gibt es keine zu g parallele Gerade durch P , m.a.W. alle Geraden durch P schneiden g .*

In dieser Geometrie ist die *Winkelsumme im Dreieck größer als zwei rechte Winkel*. Auch Riemann ist sich der Tragweite des Problems für das Verhältnis von Mathematik und Erfahrung bewusst, wenn er schreibt (zitiert nach PEIFFER / DAHAN-DALMEDICO 1994, 163)

Ich habe mir daher zunächst die Aufgabe gestellt, den Begriff einer mehrfach ausgedehnten Größe aus allgemeinen Größenbegriffen zu konstruieren. Es wird daraus hervorgehen, daß eine mehrfach ausgedehnte Größe verschiedener Maßverhältnisse fähig ist und der Raum also nur einen besonderen Fall einer dreifach ausgedehnten Mannigfaltigkeit bildet. Hiervon aber ist eine notwendige Folge, daß sich die Sätze der Geometrie nicht aus allgemeinen Größenbegriffen ableiten lassen, sondern daß diejenigen Eigenschaften, durch welche sich der Raum von anderen denkbaren dreifach ausgedehnten Größen unterscheidet, nur aus der Erfahrung entnommen werden können.

Erst jetzt dringt die Ungeheuerlichkeit ins gemeinsame mathematische Bewußtsein, die in der Existenz konkurrierender mathematischer Theorien besteht, stellt doch diese Tatsache den Begriff der mathematischen Wahrheit in Frage:

Niemand kann zwei Herren dienen. Man kann nicht der Wahrheit dienen und der Unwahrheit. Wenn die euklidische Geometrie wahr ist, so ist die nichteuklidische Geometrie falsch, und wenn die nichteuklidische wahr ist, so ist die euklidische Geometrie falsch.

Gottlob Frege (1848 - 1925) hält hier noch einmal die traditionelle Position hoch (zitiert nach MEHRTENS 1990, 117), bereits in Reaktion auf die sich abzeichnende Relativierung des mathematischen Wahrheitsbegriffs. Freges Position hätte aber zur notwendigen Folge, wie Riemann und bereits Gauß bemerkt haben, die Erfahrung zum Schiedsrichter über die Richtigkeit mathematischer Aussagen zu machen, was nach allgemeiner Auffassung die Stringenz der mathematischen Argumentation doch stark beeinträchtigen würde.

¹⁷Das Aufkommen der nichteuklidischen Geometrie gilt in interessierten philosophischen Kreisen zuweilen als Widerlegung von Kant. Richtig ist das allenfalls in Bezug auf die Zurechnung bestimmter geometrischer Ideen zu den Vorstellungen a priori. Es ändert aber nichts daran, dass derartige Vorstellungen die mathematische Naturwissenschaft erst konstituieren, sondern zeigt nur, dass ihre Inhalte geschichtlichen Veränderungen unterworfen sind.

1.3.2 Axiomatische Methode und Angewandte Mathematik

In seinem berühmten, die Mathematik des 20. Jahrhunderts prägenden Vortrag *Mathematische Probleme*, gehalten auf dem zweiten internationalen Mathematikerkongress, bringt David Hilbert (1862 -1943) die sich abzeichnende radikale Abkehr von der traditionellen Auffassung auf den Punkt (vgl. MEHRTENS 1990, 108 ff). Ein Jahr zuvor hat er in seinen *Grundlagen der Geometrie* bereits am Beispiel vorgemacht, was als Programm für die gesamte Mathematik gedacht ist (HILBERT 1899/1968, 2):

Wir denken drei verschiedene Systeme von Dingen: die Dinge des ersten Systems nennen wir Punkte und bezeichnen sie mit A, B, C, \dots ; die Dinge des zweiten Systems nennen wir Geraden und bezeichnen sie mit a, b, c, \dots ; die Dinge des dritten Systems nennen wir Ebenen und bezeichnen sie mit $\alpha, \beta, \gamma, \dots$; ... Wir denken die Punkte, Geraden, Ebenen in gewissen gegenseitigen Beziehungen und bezeichnen diese Beziehungen durch Worte wie „liegen“, „zwischen“, „kongruent“; die genaue und für mathematische Zwecke vollständige Beschreibung dieser Beziehungen erfolgt durch die Axiome der Geometrie.

Die hier verwendeten Worte sind zunächst leere Hülsen ohne jede Bedeutung, die tatsächliche Wortwahl ist nur ein (allerdings nicht ganz unwichtiges) Zugeständnis an die mathematische Konvention. So etwas wie Bedeutung erhalten die Worte erst durch die dann folgenden Axiome (z. B.: *Zu zwei Punkten A, B gibt es stets genau eine Gerade a , die mit jedem der beiden Punkte A, B zusammengehört*). Die Axiome sind nicht wie bei Euklid evidente Wahrheiten, können es gar nicht sein, weil die in ihnen vorkommenden Worte keine eigenständigen Sinn haben. Es handelt sich vielmehr um *Setzungen*, die die verwendeten Worte erst in gewisser Weise definieren. Mathematische Sätze sind dann nicht mehr an sich wahr oder falsch, sondern immer nur aus einem klar definierten Axiomensystem ableitbar oder nicht. Aufgabe der Mathematik ist es, ableitbare Sätze abzuleiten und für nicht ableitbare die Unlösbarkeit des Ableitungsproblems nachzuweisen (wie etwa beim Parallelenaxiom in Bezug auf die ersten vier euklidischen Axiome geschehen). Im Jahre 1900 jedenfalls zeigt sich Hilbert von der Lösbarkeit (in diesem Sinne) aller mathematischen Probleme überzeugt (s. MEHRTENS 1990, 110).

In diesem Programm haben sowohl die euklidische Geometrie als auch die nichteuklidischen ihren Platz in der Mathematik, sie widersprechen sich nicht einmal mehr, sondern beruhen einfach auf verschiedenen Axiomensystemen. Insofern herrscht grenzenlose Freiheit, allerdings mit einer kleinen Einschränkung: Jedes der untersuchten Axiomensysteme muss widerspruchsfrei sein, es darf aus ihm nicht eine Aussage und ihr logisches Gegenteil zugleich ableitbar sein, da andernfalls (nach einem Satz der Aussagenlogik) alle überhaupt formulierbaren Sätze auch ableitbar wären. Das Problem der Widerspruchsfreiheit führt ebenso wie das der Entscheidbarkeit (ist ein bestimmter mathematischer Satz aus einem Axiomensystem ableitbar oder nicht?) in die Abgründe der mathematischen Logik, in die ich mich hier nicht begeben möchte, weil es mir im vorliegenden Kontext eher um die Beziehungen zwischen der Mathematik und anderen Wissenschaften geht. Die allerdings sind für die moderne Mathematik, die dem Hilbert'schen Programm folgt, ausgesprochen schillernd:

Auf der einen Seite konstituiert Hilbert die Mathematik als eigenständiges Fach, welches sich nicht über seine Inhalte, sondern ausschließlich über die *Form* definiert, in die diese zu bringen seien. Auch für die weitere Entwicklung betont Hilbert diese Eigenständigkeit (HILBERT 1900, 26/27):

Bei der Weiterentwicklung einer mathematischen Disziplin wird sich jedoch der menschliche Geist, ermutigt durch das Gelingen der Lösungen, seiner Selbständigkeit bewußt; er schafft aus sich selbst heraus oft ohne erkennbare äußere Anregung, allein durch logisches Kombinieren, durch Verallgemeinern, Spezialisieren, durch Trennen und Sammeln der Begriffe in glücklichster Weise neue und fruchtbare Probleme und tritt selbst als der eigentliche Frager in den Vordergrund ... und so entstehen überhaupt fast alle feineren Fragen der modernen Zahlen- und Funktionentheorie.

Auf der anderen Seite erweitert das Absehen von allen konkreten Inhalten und die damit verbundene Loslösung der Mathematik von der Physik die Anwendungsfelder. Und die Orientierung an ihnen erweist sich mehr als je zuvor als erforderlich, soll Mathematik eine Wissenschaft bleiben und nicht zu einem reinen Glasperlenspiel verkommen. Die Widerspruchsfreiheit allein gibt kein hinreichendes Kriterium für eine Antwort auf die Frage, welche der unabsehbar vielen möglichen Axiomensysteme es zu behandeln lohnt und welche nicht. Hier müssen Aspekte hinzutreten, die (zumindest im Rahmen des Hilbert'schen Programms) außermathematischer Art sind. So nennt Hilbert selbst beispielsweise in der Einleitung seiner *Grundlagen der Geometrie* als deren Aufgabe die *logische Analyse unserer räumlichen Anschauung*, und ohne diesen Hintergrund wären seine darauf folgenden Abhandlungen zwar logisch und mathematisch korrekt, doch völlig substanzlos.

Was schließlich die Möglichkeiten der Angewandten Mathematik, den Umfang und die Anzahl ihrer Betätigungsfelder angeht, so gehen hier die Ansichten auseinander. Hilbert vertritt in dieser Frage, wie es wohl oft in der Anfangseuphorie einer neuen Entwicklung passiert, einen durchaus imperialistischen Standpunkt, wenn er im Jahre 1918 bemerkt (zitiert nach MEHRTENS 1990, 132)

Ich glaube: Alles was Gegenstand des wissenschaftlichen Denkens überhaupt sein kann, verfällt, sobald es zu Bildung einer Theorie reif ist, der axiomatischen Methode und damit unmittelbar der Mathematik. ... In dem Zeichen der axiomatischen Methode erscheint die Mathematik berufen zu einer führenden Rolle in der Wissenschaft.

und vier Jahre später (zitiert nach MEHRTENS 1990, 132)

Ich bemerkte einmal, dass die Frage, was angewandte Mathematik sei, mit der Gegenfrage beantwortet werden könnte: Was ist nicht angewandte Mathematik? In der Tat, was wir auch für Begebenheiten oder Erscheinungen in der Natur oder im praktischen Leben antreffen, überall wird der mathematische Gesinnte und Eingestellte einen mathematischen Kern finden.

Der erste Satz impliziert, dass hier alles, was nicht mathematisierbar ist, als nicht wissenschaftlich oder zumindest noch nicht reif zur Bildung einer Theorie deklariert wird. Nun ist der Gedanke keineswegs neu, dass alle Dinge, die überhaupt menschlicher Erkenntnis zugänglich sind, sich durch mathematische Ableitungen erfassen ließen. Er geht bereits auf René Descartes (1596 - 1650) zurück. Ihn ausgerechnet im Zusammenhang mit der axiomatischen Methode wieder aufzuwärmen, ist zumindest ziemlich vorwitzig angesichts des gerade durch die von Hilbert vollendete Entwicklung sich erstmals im vollen Ernst stellenden Problems der Beziehungen zwischen Mathematik und Realität. Die Frage, ob der angeblich überall zu findende „mathematische Kern“ einen Sachverhalt angemessen beschreibt, wird hier nämlich einfach übergangen.

Gerade diese Kernfrage aber nach dem Zusammenhang zwischen der nunmehr abgekoppelten Mathematik und allen anderen Wissensgebieten, nach den Bedingungen der Möglichkeit, informelle Inhalte mathematisch zu formalisieren, ist selbst nicht mathematisierbar. Denn jeder entsprechende Versuch müsste einen nicht formalisierten Gegenstand zunächst einer Operation gerade der Art unterziehen, die zu problematisieren ist.¹⁸

1.3.3 Der Modellbegriff

Heinrich Hertz (1857 - 1894) ist dem breiteren Publikum vor allem durch seine experimentellen Studien zu Radio- oder allgemeiner elektromagnetischen Wellen bekannt, weniger durch seine Untersuchungen,

¹⁸Die mathematische Logik etwa zeigt in einem etwas anderen, aber ähnlich gelagerten Bereich die Vergeblichkeit eines solchen Bemühens: Als inhaltliche mathematische Theorie ist sie durchaus von Interesse, als eine „Metatheorie“ mathematischen Denkens, die sie angeblich auch sein soll, aber taugt sie nicht, weil die Frage der *Interpretation*, also der Beziehung zwischen ihren formalen Zeichenketten und „konkreten“ mathematischen Objekten (wie etwa den natürlichen Zahlen) sich nicht wiederum als formale Zeichenkette darstellen lässt.

die die gesamte theoretische Physik seiner Zeit betreffen. In den letzten drei Jahren seines Lebens konzentriert er sich auf die begriffliche Überarbeitung der Grundlagen der klassischen Mechanik. Die ein halbes Jahr nach seinem Tod erschienene Schrift *Die Prinzipien der Mechanik, in neuem Zusammenhange dargestellt* ist der Versuch einer „Mechanik ohne Kraftbegriff“, genauer gesagt, einer Mechanik, in der die Kraft kein Grundbegriff, sondern aus anderen Begriffen abgeleitet ist, nämlich der Bewegung verborgener Massen des Äthers. Die Ätherhypothese ist von der modernen Physik kurz darauf verworfen worden, dagegen gibt es ähnliche Erklärungsmuster für die Bewegungen von Körpern in der allgemeinen Relativitätstheorie (durch Gravitationsfelder gekrümmter Raum).

Berühmt geworden ist vor allem die Einleitung zu dieser Abhandlung, in der Hertz sich ein Instrumentarium schafft, seinen eigenen Aufbau der Mechanik und die beiden bis dahin bekannten¹⁹ hinsichtlich der Vor- und Nachteile gegeneinander abzuwägen. Er verwendet offenbar als Erster das Wort „Modell“, das allerdings nur einmal auftritt und im alltagssprachlich-technischen Sinne als Metapher benutzt wird. Dabei wird der *Begriff* des Modells im modernen Sinne aber herausgearbeitet, auch wenn andere Worte („Bilder“, „Scheinbilder“, „Symbole“) verwendet werden, mit denen er die Arbeitsweise der „bewußten Naturerkenntnis“ charakterisiert (HERTZ 1894/1996, 67, Hervorhebung C.P.O.):

Es ist die nächste und in gewissem Sinne wichtigste Aufgabe unserer bewußten Naturerkenntnis, daß sie uns befähige, zukünftige Erfahrungen vorauszusehen, um nach dieser Voraussicht unser gegenwärtiges Handeln einrichten zu können. Als Grundlage für die Lösung jener Aufgabe der Erkenntnis benutzen wir unter allen Umständen vorangegangene Erfahrungen, gewonnen durch zufällige Beobachtungen oder durch absichtlichen Versuch. Das Verfahren aber, dessen wir uns zur Ableitung des Zukünftigen aus dem Vergangenen und damit zur Erlangung der erstrebten Voraussicht stets bedienen, ist dieses: Wir machen uns innere Scheinbilder oder Symbole der äußeren Gegenstände, und zwar machen wir sie von solcher Art, daß die denknöthigen Folgen der Bilder stets wieder die Bilder seien von den naturnotwendigen Folgen der abgebildeten Gegenstände. Damit diese Forderung überhaupt erfüllbar sei, müssen gewisse Übereinstimmungen vorhanden sein zwischen der Natur und unserem Geiste. Die Erfahrung lehrt uns, daß die Forderung erfüllbar ist und daß also solche Übereinstimmungen in der Tat bestehen. Ist es uns einmal geglückt, aus der angesammelten bisherigen Erfahrung Bilder von der verlangten Beschaffenheit abzuleiten, so können wir an ihnen, wie an Modellen, in kurzer Zeit die Folgen entwickeln, welche in der äußeren Welt erst in längerer Zeit oder als Folgen unseres eigenen Eingreifens auftreten werden; wir vermögen so den Tatsachen voranzueilen und können nach der gewonnenen Einsicht unsere gegenwärtigen Entschlüsse richten. Die Bilder, von welchen wir reden, sind unsere Vorstellungen von den Dingen; sie haben mit den Dingen die eine wesentliche Übereinstimmung, welche in der Erfüllung der genannten Forderung liegt, aber es ist für ihren Zweck nicht nötig, daß sie irgend eine weitere übereinstimmung mit den Dingen haben. In der Tat wissen wir auch nicht, und haben auch kein Mittel zu erfahren, ob unsere Vorstellungen von den Dingen mit jenen in irgend etwas anderem übereinstimmen, als allein in eben jener einen fundamentalen Beziehung

Hertz' Sprachgebrauch markiert erkenntnistheoretisch eine Distanz zu Galilei oder Newton, die ähnlich groß sein dürfte wie die zwischen der modernen, axiomatischen Mathematik und der klassischen mit ihrem absoluten Wahrheitsbegriff. Und auch hier liegt der Grund in den mit der neuzeitlichen Wissenschaft inzwischen gemachten Erfahrungen: Wenn die „Modelle“ oder „inneren Scheinbilder“ von „der Natur“ auseinander zu fallen beginnen, dann lassen sie sich nicht mehr als naturgegeben denken, sondern

¹⁹Neben dem Newton'schen Zugang mit *Kraft* als Grundbegriff ist das der im Laufe des 19. Jahrhunderts aufkommende und auf Variationsprinzipien beruhende, in dem *Energie* als Grundbegriff an die Stelle der Kraft tritt. Hertz versucht in seinem Zugang, mit den (allen drei Zugängen gemeinsamen) drei Grundbegriffen *Zeit*, *Raum* und *Masse* auszukommen.

nur noch als von uns selbst hervorgebracht, als ein spezifisches Instrument, mit spezifischen Möglichkeiten und Grenzen.

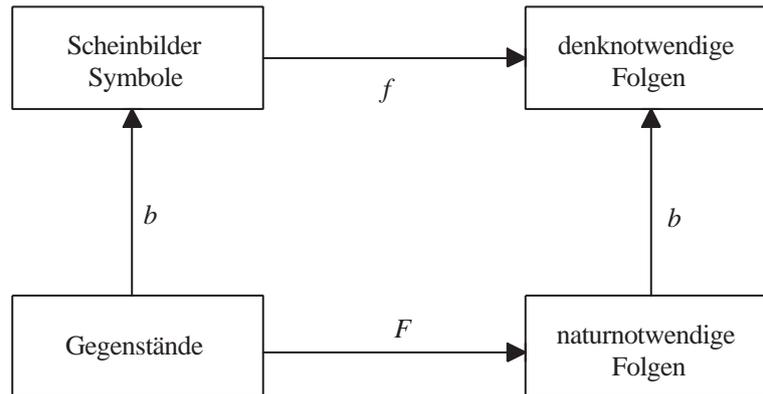


Abbildung 1.6: Zum Hertz'schen Modellbegriff

Die von den „inneren Scheinbildern“ zu erfüllende Bedingung erinnert an ein kommutierendes Diagramm, wie es in der modernen Algebra gebräuchlich ist (vgl. Abbildung 1.6).²⁰ Zu erfüllen ist die Forderung

$$f \circ b = b \circ F .$$

Gegenstand der Naturerkenntnis ist die (unbekannte) Abbildung F , wogegen f sich aus prinzipiell zugänglichen Operationen der Mathematik und Logik zusammensetzt. Die eigentliche Modellbildung besteht in der Wahl der Abbildung b . Für die geeignete Wahl nennt Hertz drei Kriterien: *Zulässigkeit*, *Richtigkeit* und *Zweckmäßigkeit* (HERTZ 1894/1996, 67):

Eindeutig sind die Bilder, welche wir uns von den Dingen machen wollen, noch nicht bestimmt durch die Forderung, daß die Folgen der Bilder wieder die Bilder der Folgen seien. Verschiedene Bilder derselben Gegenstände sind möglich und diese Bilder können sich nach verschiedenen Richtungen unterscheiden. Als unzulässig sollten wir von vornherein solche Bilder bezeichnen, welche schon einen Widerspruch gegen die Gesetze unseres Denkens in sich tragen, und wir fordern also zunächst, daß alle Bilder logisch zulässige oder kurz zulässige seien. Unrichtig nennen wir zulässige Bilder dann, wenn ihre wesentlichen Beziehungen den Beziehungen der äußeren Dinge widersprechen, das heißt wenn sie jener ersten Grundforderung nicht genügen. Wir verlangen demnach zweitens, daß unsere Bilder richtig seien. Aber zwei zulässige und richtige Bilder derselben äußeren Gegenstände können sich noch unterscheiden nach der Zweckmäßigkeit. Von zwei Bildern desselben Gegenstandes wird dasjenige das zweckmäßigere sein, welches mehr wesentliche Beziehungen des Gegenstandes widerspiegelt als das andere; welches, wie wir sagen wollen, das deutlichere ist. Bei gleicher Deutlichkeit wird von zwei Bildern dasjenige zweckmäßiger sein, welches neben den wesentlichen Zügen die geringere Zahl überflüssiger oder leerer Beziehungen enthält, welches also das einfachere ist. Ganz werden sich leere Beziehungen nicht vermeiden lassen, denn sie kommen den Bildern schon deshalb zu, weil es eben nur Bilder und zwar Bilder unseres besonderen Geistes sind und also von den Eigenschaften seiner Abbildungsweise mitbestimmt sein müssen.

²⁰Es wäre allerdings ein Missverständnis, hierin ein „mathematisches Modell“ der Modellbildung zu sehen, da keine der drei enthaltenen Abbildungen hinsichtlich ihrer realen Bedeutung präzise gefasst ist.

Das Kriterium der Richtigkeit, also die Grundforderung, wird in der Physik, an die Hertz hier ausschließlich denkt, in der Regel durch ein Experiment überprüft, also durch die bewusste und theoriegeleitete Herstellung von Versuchsbedingungen, die den Idealvorstellungen des Modells möglichst nahe kommen und an denen sich seine Vorhersagen überprüfen lassen.

Die Verbreitung mathematischer Modelle in anderen Wissenschaftsbereichen bringt eine methodische Schwierigkeit mit sich, die die Physik so nicht kennt, mit Ausnahme vielleicht in der Kosmologie: Klimamodelle, Modelle von Volkswirtschaften oder komplexer biologischer Systeme etwa lassen sich nicht experimentell überprüfen: Der Beobachter kann die im Modell unterstellten Idealbedingungen nicht herstellen, er muss das beobachtete System so nehmen, wie es ist. Hier dürfte das Hauptproblem bei der Übertragung der mathematisch-naturwissenschaftlichen Methode auf die meisten anderen Wissensgebiete liegen.

Dennoch gibt es weitere Anforderungen an eine saubere mathematische Modellbildung, die sich auch außerhalb der Physik erfüllen lassen, aber leider oft nicht erfüllt werden (HERTZ 1894/1996, 68):

Wir haben bisher die Anforderungen aufgezählt, welche wir an die Bilder selbst stellen; etwas ganz anderes sind die Anforderungen, welche wir an eine wissenschaftliche Darlegung solcher Bilder stellen. Wir verlangen von der letzteren, daß sie uns klar zum Bewußtsein führe, welche Eigenschaften den Bildern zugelegt seien um der Zulässigkeit willen, welche um der Richtigkeit willen, welche um der Zweckmäßigkeit willen. Nur so gewinnen wir die Möglichkeit an unseren Bildern zu ändern, zu bessern. Was den Bildern beigelegt wurde um der Zweckmäßigkeit willen, ist enthalten in den Bezeichnungen, Definitionen, Abkürzungen, kurzum in dem, was wir nach Willkür hinzutun oder wegnehmen können. Was den Bildern zukommt um ihrer Richtigkeit willen, ist enthalten in den Erfahrungstatsachen, welche beim Aufbau der Bilder gedient haben. Was den Bildern zukommt, damit sie zulässig seien, ist gegeben durch die Eigenschaften unseres Geistes. Ob ein Bild zulässig ist oder nicht, können wir eindeutig mit ja und nein entscheiden, und zwar mit Gültigkeit unserer Entscheidung für alle Zeiten. Ob ein Bild richtig ist oder nicht, kann ebenfalls eindeutig mit ja und nein entschieden werden, aber nur nach dem Stande unserer gegenwärtigen Erfahrung und unter Zulassung der Berufung an spätere reifere Erfahrung. Ob ein Bild zweckmäßig sei oder nicht, dafür gibt es überhaupt keine eindeutige Entscheidung, sondern es können Meinungsverschiedenheiten bestehen. Das eine Bild kann nach der einen, das andere nach der andern Richtung Vorteile bieten, und nur durch allmähliches Prüfen vieler Bilder werden im Laufe der Zeit schließlich die zweckmäßigsten gewonnen.

Es kommt also darauf an, nicht nur die in ein Modell eingehenden Annahmen deutlich herauszuarbeiten, sondern ebenso die Gründe für diese Annahmen, die sehr verschiedenartig sein können. Zu diesen Gründen zählt immer auch der Zweck des Modells: Welche Fragen soll es beantworten, welche beobachteten Phänomene erklären, wovon darf im jeweiligen Zusammenhang abstrahiert werden?

1.3.4 Zur Gültigkeit von Modellen

Die „Richtigkeit“ von Modellen im Sinne von Hertz lässt sich nicht beweisen, sondern nur an Experimenten überprüfen und bei negativem Ausgang widerlegen. Umgekehrt folgt aus der Übereinstimmung mit experimentellen Beobachtungen allenfalls eine Art „vorläufige Richtigkeit“ bis zum Beweis des Gegenteils im nächsten Experiment.

Problematisch wird die Situation dann, wenn Experimente mit dem durch das Modell beschriebenen System nicht möglich sind oder sich verbieten: Eine aus einem mathematischen Modell gewonnene Aussage hat (wie jeder mathematische Satz) immer die Form: Unter den Voraussetzungen A, B, C, \dots tritt X, Y, Z, \dots ein. Geraten nun die Vorhersagen X, Y, Z zu Beobachtungen in Widerspruch, so kann das u. a. daran liegen,

- dass die vereinfachenden Annahmen A, B, C nur in der konkret beobachteten Situation nicht erfüllt waren,
- dass sie von dem beschriebenen System grundsätzlich nicht erfüllt werden können,
- dass schludrig gearbeitet wurde, entweder in Form von Mehrdeutigkeiten und Inkonsistenzen bei der Interpretation (Abbildung b) oder in Form logischer oder mathematischer Fehlschlüsse,

was ganz verschiedene Konsequenzen hinsichtlich der Beurteilung des Modells nach sich zieht, aber nicht so ohne weiteres auseinander gehalten werden kann.

Logische Fehler und Inkonsistenzen können natürlich auch dazu führen, dass Aussagen gewonnen werden, die das Modell eigentlich gar nicht hergibt, die sich aber empirisch mehr oder weniger bestätigen lassen.²¹ Die Übereinstimmung mit Beobachtungsdaten ist daher weder ein notwendiges noch ein hinreichendes Kriterium für die Richtigkeit eines Modells, sondern nur ein Teilaspekt. BOSSEL (1992, 36) bezeichnet dieses Kriterium als *empirische Gültigkeit*, nennt aber noch drei weitere Kriterien:

- *Verhaltensgültigkeit* ist ein ähnliches Kriterium, wobei allerdings von der numerischen Übereinstimmung mit Daten abgesehen wird. Gefragt ist nur, ob Modell und beschriebenes System ein ähnliches qualitatives Verhalten aufweisen (z. B. Schwingungen, Periodizitäten u. a.).
- *Strukturgültigkeit* zielt auf die Entsprechung der Wirkungsstrukturen von Modell und Original. Ein Modell ist mit Sicherheit dann nicht struktur­gültig, wenn nur das qualitative oder quantitative Systemverhalten durch eine mathematische Struktur nachgebildet wird, ohne auf die inneren Wirkungsmechanismen des Originals Bezug zu nehmen.
- *Anwendungsgültigkeit* fragt danach, ob das Modell seinem Zweck entspricht und Antworten auf die gestellten Fragen gibt. Ein Modell ist mit Sicherheit dann nicht anwendungsgültig, wenn es mathematisch unlösbare Probleme aufwirft und eine Simulation mit den aktuell zur Verfügung stehenden Mitteln nicht möglich ist.

Es dürfte im Allgemeinen schwierig sein, präzise festzustellen, dass eines dieser Kriterien erfüllt ist, schon eher, dass es nicht erfüllt ist. Die Kriterien sind daher nur als Aspekte zu verstehen, unter denen sich mathematische Modelle diskutieren lassen.

1.4 Vom realen Problem zum mathematischen Modell

Wie kommt man bei gegebener realer Fragestellung zu einem mathematischen Modell? Darauf gibt es keine eindeutige Antwort, kein Rezept, das immer zum Ziel führt. Das Problem ist komplex, komplexer als die Mathematik selbst, weil die reale Fragestellung nicht schon in formalisierter Form vorliegt (sonst bräuhete man nicht mehr zu "modellieren") und daher ihre Beziehung zur Mathematik, mit der sie bearbeitet werden soll, selbst nicht mathematisierbar ist. Dennoch lassen sich einige Regeln angeben, die beachtet werden sollten.

²¹Nach meinem Eindruck scheint das ein in weiten Teilen der herrschenden Volkswirtschaftslehre beliebtes Vorgehen zu sein, weshalb denn dort auch die Ansicht häufig anzutreffen ist, es komme auf Erklärungsgehalt und Begründung der verwendeten Modelle gar nicht an, sondern nur darauf, ob zutreffende Voraussagen möglich sind, so etwa FRIEDMAN 1953. Dann stellt sich allerdings die Frage, warum man auf Modelle, denen gar kein Erklärungsanspruch zukommt, nicht ganz verzichtet und sich auf die Extrapolation von Beobachtungsdaten beschränkt.

1.4.1 Der Modellierungsprozess

Der gemeinsame Durchschnitt der wenigen Lehrbücher über Mathematische Modellbildung ist das Diagramm in Abbildung 1.7, deren Ähnlichkeit zu Abbildung 1.6 vielleicht ins Auge springt, aber nicht zur Verwirrung führen darf: Dort geht es um ein Kriterium für die Richtigkeit von Modellen, hier um die Tätigkeiten, die auszuführen sind, um zu einem Modell zu kommen, den *Modellierungsprozess*: Ausgangspunkt ist ein reales Problem oder erklärungsbedürftiges Phänomen, hieraus wird ein mathema-

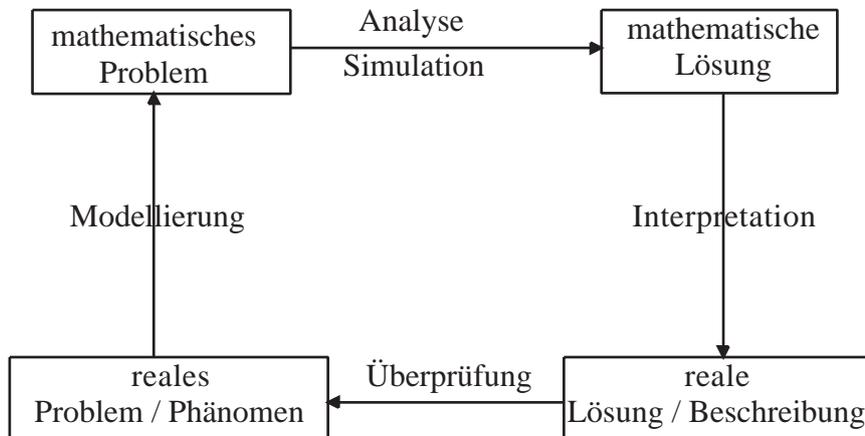


Abbildung 1.7: Der Modellierungsprozess

tisches Problem entwickelt, ein Bild der Wirklichkeit, dieses wird mit mathematischen Methoden gelöst, die mathematische Lösung wird hinsichtlich ihrer realen Bedeutung interpretiert und auf ihre Relevanz für das reale Problem überprüft.

Das Schema kann den irreführenden Eindruck vermitteln, es handele sich um eine Art Algorithmus, den man nur in Gang setzen müsse, um gesicherte Erkenntnisse zu produzieren. So ist es keineswegs, auch wenn das Schema durchaus hilfreich sein kann. Aber es bleibt ein Schema, der tatsächliche Prozess ist meistens komplexer, weil z. B. die Formulierung und die Lösung des mathematischen Problems miteinander in Wechselwirkung stehen. Im Folgenden wird versucht, einige Leitlinien für die mathematische Modellbildung zu formulieren:

1.4.2 Modellierungs-,„Rezepte“

A. Modellentwicklung: Vom realen zum mathematischen Problem

1. **Worin besteht das reale Problem?** Versuche, die reale Fragestellung möglichst präzise herauszuarbeiten! Was ist wesentlich, was unwesentlich? Welche Ziele sollen erreicht werden? Wie genau müssen im Rahmen dieser Ziele die Antworten sein? Oft zeigt sich, dass eine an sich als klar erscheinende Problemstellung alles andere als klar ist. Insbesondere in den "weichen" Wissenschaften ist die Präzisierung der Fragen manchmal schon der wesentliche Nutzen mathematischer Modellierung.
2. **Gesetzmäßigkeiten:** Durch welche Gesetzmäßigkeiten ist das reale Problem bestimmt, und wie lassen sie sich in mathematischer Sprache fassen? Sofern das Problem in Zusammenhang mit einer anderen Wissenschaft steht, ist es in der Regel bereits „theoretisch vorbelastet“. Welche Vorstellungen hat die Substanzwissenschaft von den zu Grunde liegenden Gesetzmäßigkeiten? Wie weit liegen sie bereits in mathematisierter Form vor? Was lässt sich davon nutzbar machen?
3. **Übertragen von Ansätzen aus bekannten Modellen:** Gibt es bereits mathematische Modelle für ähnliche Probleme? Wurde ein Teilproblem bereits anderswo modelliert? Gibt es strukturelle Analogien zu Fragestellungen aus ganz anderen Wissensbereichen? Lässt sich beispielsweise ein physikalisches / mechanisches Modell für das Problem formulieren?

4. **Benötigte und überflüssige Informationen:** Welche Informationen (Daten) werden benötigt? Welche sind vorhanden, welche davon ggf. überflüssig, welche müssen erst noch beschafft werden? Wie sicher sind diese Informationen? „Modelliererinnen“ sollten generell nicht davon ausgehen, dass der „Anwender“ von sich aus weiß, was wichtig ist und was nicht. Modellierungsprozesse haben manchmal nur die Funktion, empirische Untersuchungen anzuregen, auf die ohne den Versuch der Mathematisierung aber niemand gekommen wäre.
5. **Modellvariablen und -parameter:** Durch welche Variablen soll das Modell beschrieben werden? Welche extern vorgegebenen Parameter gehen in das Problem ein? Welche inhaltliche Bedeutung haben sie? Wie genau lassen sie sich bestimmen? Welche Variationsbreite haben sie? Was sind die Maßeinheiten von Variablen und Parametern? Mit den Maßeinheiten lassen sich die in das Modell eingehenden mathematischen Terme einer ersten Konsistenzprüfung unterziehen: Ein Term beispielsweise, der aus einer Summe von Größen mit verschiedenen Maßeinheiten besteht, ist sinnlos.
6. **Keine Lösung ohne eindeutig formuliertes Problem:** Bevor mathematische Methoden angewandt werden, muss das mathematische Problem präzise formuliert sein. Da das aus der realen Fragestellung meist nicht eindeutig hervorgeht, sind hier Entscheidungen zu treffen, oder es können auch mehrere alternative mathematische Problemformulierungen entwickelt werden, die dann aber klar voneinander zu trennen sind. Falsch ist es und führt immer zu unzulässigen, in sich widersprüchlichen Modellen, sich fehlender Kriterien wegen gar nicht zu entscheiden.
7. **So einfach wie möglich !** Vereinfache zu Beginn radikal! Komplizierter kann man ein Modell immer noch machen. Für die reale Fragestellung ist das Modell nur dann von Nutzen, wenn es sich als mathematisches Problem auch mathematisch lösen lässt. An dieser Stelle geht natürlich das eigene mathematische Vorwissen in die Modellierung ebenso ein wie die technische Möglichkeit der Lösung: Die Computerisierung etwa macht heute ganz andere Modelle sinnvoll als noch vor 40 Jahren.
8. **Bewusstheit des eigenen Vorgehens** Gib Rechenschaft über die Überlegungen, die zum Modell führten! Das Bewusstsein für die im Modell vorgenommenen Vereinfachungen und die Gründe für sie (Zulässigkeit, Richtigkeit, Zweckmäßigkeit) darf nicht verloren gehen, sonst lässt sich das Modell nicht mehr verbessern.

B. Analyse und Simulation: Vom mathematischen Problem zu seiner Lösung

1. **Das reale Problem nicht vergessen !** Für den Weg vom mathematischen Problem zu seiner Lösung stehen natürlich alle mathematischen Werkzeuge zur Verfügung, und es scheint sich hier um ein rein innermathematisches Vorgehen zu handeln. Behalte trotzdem das real zu lösende Problem in Erinnerung! Es kann als Leitfaden auch für den mathematisch zu beschreitenden Weg dienen.
2. **Analytische Lösungen und qualitatives Modellverhalten:** Lässt sich das Modellverhalten auf analytischem Wege zumindest qualitativ bestimmen? Gibt es (im regelhaft nicht erreichbaren Idealfall) eine geschlossene Formel für die Lösung?
3. **Spezialfälle und Vereinfachungen, Modellreduktion:** Lässt sich eine Lösung zumindest für spezielle Fälle bestimmen? Gibt es einfachere, aber ähnliche Probleme (z. B. geringerer Dimension), die sich analytisch lösen lassen? Versuche, die Anzahl der Modellparameter reduzieren, z. B. durch passende Wahl der Maßeinheiten und dimensionslose Schreibweise des Problems.
4. **Computersimulationen und Parameterstudien:** Für Computersimulationen ist es nötig, alle Modellparameter mit numerischen Zahlenwerten zu belegen. Sofern nicht bekannt, sind diese zu variieren. Das geht nicht für unbegrenzt viele, hier liegt die Bedeutung der vorausgegangenen Reduktion ihrer Anzahl. Darüber hinaus ist es vielleicht möglich, weitere Parameter auf Grund der realen Gegebenheiten festzulegen. Welche der zu variierenden Parameter sind kritisch, d. h. von welchen hängt das Modellverhalten sensitiv ab?

C. Interpretation und Validierung des Modells

1. **Interpretierbarkeit von Ergebnissen:** Lassen sich die gefundenen mathematischen Ergebnisse und die mathematischen Voraussetzungen, unter denen sie gelten, überhaupt real deuten? Sind Ergebnisse und Voraussetzungen realistisch? Liegen die gefundenen Lösungen im interpretierbaren Bereich, werden

beispielsweise Bestandgrößen nicht negativ? Lässt sich das analytisch oder qua Computersimulation gefundene Modellverhalten in der Sprache der realen Fragestellung ausdrücken?

2. **Visualisierung der Ergebnisse:** Bringe die Ergebnisse in eine Form, die überblickt und interpretiert werden kann! Bei einem ungeordneten Haufen von auch nur 1000 Zahlen ist das unmöglich. Visualisiere die Ergebnisse! Bediene dich dazu der Möglichkeiten moderner Computer!
3. **Vergleich mit Beobachtungsdaten und Experimenten:** Lassen sich die gefundenen Ergebnisse mit Beobachtungsdaten vergleichen? Gibt das Modell Anlass zu Experimenten oder Beobachtungen, die erst noch durchzuführen sind? Stimmen Modellverhalten und Beobachtungsdaten überein, ggf. nach Anpassung der Modellparameter? Im negativen Falle ist zu klären, woher die Diskrepanz kommt und was im Modell daher zu verändern ist.

1.4.3 Simulationswerkzeuge

Mathematische Modelle sind oft aufgrund ihrer Komplexität analytischen Methoden nur schwer zugänglich. Ein wichtiger Bestandteil des Modellierungsprozesses ist daher die Auswertung eines fertigen mathematischen Modells mit Hilfe von Computersimulationen. Hat man es zudem mit einer großen Datenmenge zu tun, müssen Simulationsergebnisse oder auch Messdaten visualisiert werden.

Grundlage einer Computersimulation ist ein numerisches Verfahren oder ein Algorithmus, der das gegebene mathematische Problem approximativ löst. Sind diese Verfahren komplex und erfordern die Bewältigung großer Datenmengen, so spricht man auch vom *Wissenschaftlichen Rechnen* (engl.: Scientific Computing). Hier werden numerische Algorithmen mit Hilfe höherer Programmiersprachen (wie beispielsweise C++ oder JAVA) in Computerprogramme übersetzt und am Computer implementiert. Eingesetzt werden häufig auch kommerzielle Programmbibliotheken, die bereits fertige Routinen für einzelne numerische Verfahren enthalten, oder fertige Simulationspakete (MATLAB, SIMULINK, MATEMATICA...), die Module zu verschiedenen numerischen Verfahren und diverse Visualisierungstechniken beinhalten.

1.4.4 Klassifikation von Modellen

Mathematische Modelle lassen sich unter verschiedenen Gesichtspunkten klassifizieren. Einer davon ist der Gegenstandsbereich, auf den sie sich beziehen, die Klassifikation entspricht dann der der Wissenschaften, in denen mathematische Modelle eingesetzt werden.

Eine Klassifikation nach der eingesetzten Mathematik

Ein zweiter ist die Art der Mathematik, die in den Modellen zum Zuge kommt. RAPOPORT (1980) unterscheidet für die Sozialwissenschaften

- *klassische* Modelle, in denen Differential- und Differenzgleichungen verwendet werden,
- *stochastische* Modelle, die wesentlich mit stochastischen Prozessen arbeiten,
- *strukturelle* Modelle, die u. a. diskrete Mathematik einsetzen.

Differential- und Differenzgleichungen setzen die Kenntnis bzw. Annahme von Entwicklungsgesetzen voraus, denen die betrachteten Systeme dann auf deterministische Weise folgen. Stochastische Prozesse weichen die Determiniertheit der Zukunft aus den Anfangsbedingungen insofern auf, als zufällige Einflüsse mit einbezogen werden. Strukturelle Modelle verzichten dem gegenüber ganz auf Prognosen und weisen ein mehr oder weniger großes Spektrum möglicher Abläufe aus oder verzichten ganz auf Vorhersagen.

Eine differenzierte Erörterung der Klassifikation mathematischer Modelle nach der eingesetzten Mathematik findet sich bei ORTLIEB / V. DRESKY / GASSER / GÜNZEL (2009: 12 ff.).

Eine Klassifikation nach dem Modellzweck

In der Literatur zur mathematischen Modellbildung außerhalb der *hard sciences* hat sich eine weitere Klassifikation von Modellen herausgebildet, die stark am Modellzweck orientiert ist, nämlich die Unterscheidung von deskriptiven (beschreibenden, beobachtenden) Modellen, numerischen Simulationsmodellen und konzeptionellen (analytischen, hypothetischen) Modellen (vgl. GERSHENFELD 1999, WISSEL 1989).

- *Deskriptive Modelle* dienen der Erfassung, Aufbereitung, Komprimierung und Analyse meist großer Datenmengen, ohne dass damit ein Erklärungsanspruch verbunden ist. Nichtsdestoweniger handelt es sich dabei um mathematische Modellbildung, da entschieden werden muss, welche potentiellen Eigenschaften unübersichtlicher Daten zum Vorschein gebracht und welche in den Hintergrund gedrängt werden sollen.
- *Simulationsmodelle* sind der Versuch, einer komplexen Realität dadurch gerecht zu werden, dass möglichst viele bekannte Einzelheiten erfasst werden. Ihres Umfangs wegen lassen sich diese Modelle nur auf dem Computer numerisch durchrechnen. Das setzt voraus, dass jedem (bekannten oder unbekanntem) Parameter im Modell ein numerischer Wert zugewiesen sein muss. Der Zweck dieser Modelle besteht in quantitativen Prognosen, möglicher Weise auch in der Entwicklung von Vorschlägen für Management- oder Politikmaßnahmen. Die ausschließlich numerische Simulation setzt einen engen Zusammenhang zu empirisch erhobenen Daten voraus, die mangelnde oder ungenaue Datenerfassung kann daher zu einem Problem werden. Ein Folgeproblem besteht in der Schwierigkeit, die Auswirkungen von Ungenauigkeiten in den Eingangsdaten abzusehen. Die Komplexität eines Simulationsmodells macht sein Systemverhalten oft undurchsichtig, zugespitzt könnte man sagen: Das Verständnis der komplexen Realität wird nicht gefördert, sondern ersetzt. Prognosen über den empirisch bereits erfassten Rahmen hinaus sind daher immer mit großer Vorsicht zu genießen.
- *Konzeptionelle Modelle* verfahren genau umgekehrt, indem sie bewusst von starken Idealisierungen ausgehen. Sie beschreiben Situationen, die es in der Realität so gar nicht oder nur im Extremfall gibt, etwa im Sinne von Gedankenexperimenten der klassischen Physik. Derartige Modelle müssen schon deswegen möglichst einfach gehalten werden, um einer Analyse zugänglich zu sein. In der Regel operieren sie mit nichtnumerischen Parametern. Das Ziel dieser Modelle ist es, die Wirkungen von einzelnen Komponenten der komplexen Realität genauer zu bestimmen und so Verständnis für Zusammenhänge zu erzeugen. Eine quantitative Übereinstimmung der aus konzeptionellen Modellen gewonnenen Vorhersagen mit Beobachtungsdaten, die nicht in einem entsprechend angelegten Experiment gewonnen wurden, ist nicht zu erwarten.²²

Eine Klassifikation nach der Art der Information über das modellierte System

Die Möglichkeiten, im konkreten Fall ein mathematisches Modell zu entwickeln, hängen in starkem Maße davon ab, welche Art von Informationen ich über das zu modellierende reale System habe bzw. bekommen kann. Eine Klassifikation nach deren Genauigkeit besteht in der Unterscheidung von so genannten Black-, Grey- und White-Modellen. Dieser Sprachgebrauch lehnt sich an den Begriff der *black box* an, also eines Systems, dessen Input-Output-Verhalten sich beobachten lässt, dessen innerer Mechanismus aber (noch) unbekannt ist.

²²Um noch einmal auf die Volkswirtschaftslehre zurück zu kommen: Die dort extensiv gebrauchten mathematischen Modelle sind ihrem Ansatz nach konzeptioneller Art. Einen wie immer mit ihnen verbundenen Erklärungsanspruch einfach fallen zu lassen und statt dessen auf ihre Prognosefähigkeit zu setzen, ist daher ein sinnloses Unterfangen.

- Ein *Black-Modell* ist ein Modell für eine black box. Um zu ihm zu gelangen, müssen willkürliche Annahmen gemacht werden, etwa dass der Zusammenhang zwischen den eingehenden und den ausgehenden Daten linear ist. Anschließend kann dann versucht werden, freie Modellparameter so zu wählen, dass eine möglichst gute Übereinstimmung mit den Beobachtungen erzielt wird. Beispiel: Reaktion eines durchschnittlichen Autofahrers auf bestimmte Verkehrssituationen (Abbremsen des Fahrzeugs vor ihm, Auftauchen eines Hindernisses usw.).
- Ein *White-Modell* ist das andere Extrem, nämlich ein Modell für ein System, bei dem es für die Funktionsweise aller seiner Elemente bereits bewährte mathematische Theorien gibt, die dann „nur“ noch geeignet kombiniert werden müssen. Beispiel: Reaktion eines Autos auf bestimmten Aktionen seines Fahrers (Lenken, Gasgeben, Bremsen).
- Ein *Grey-Modell* liegt irgendwo dazwischen: Gewisse Elemente sind im Sinne eines White-Modells bekannt, bei anderen handelt es sich mehr oder weniger um eine black box. Beispiel: Mensch-Maschine-Systeme Autofahrer-Auto.

Eine differenzierte Erörterung der Klassifikation mathematischer Modelle nach ihrer Durchsichtigkeit findet sich bei ORTLIEB / V. DRESKY / GASSER / GÜNZEL (2009: 8 ff.).

1.5 Literatur

- BOSSSEL, H.: *Modellbildung und Simulation*, Braunschweig 1992
- COLLINS, H. / PINCH, T.: *Der Golem der Forschung. Wie unsere Wissenschaft die Natur erfindet*, Berlin 1999
- DAVIS, P. J. / HERSH, R.: *Erfahrung Mathematik*, Basel 1985
- FÖLSING, A.: *Galileo Galilei. Prozeß ohne Ende. Eine Biographie*, Reinbek 1996
- FRIEDMAN, M.: in *The methodology of positive economics*, in M. Friedman: *Essays in positive economics*, 5 - 43, Chicago 1953
- GALILEI, G.: *Discorsi e dimostrazioni matematiche, intorno a due novo scienze*, 1638, Übersetzung von A. v. Oettingen 1890, Nachdruck, Frankfurt 1995
- GERSHENFELD, N.: *The nature of mathematical modeling*, Cambridge 1999
- HERTZ, H.: *Die Prinzipien der Mechanik in neuem Zusammenhange dargestellt*. Einleitung, 1894, mit *Anmerkungen* von J. KUCZERA, Frankfurt 1996
- HILBERT, D.: *Grundlagen der Geometrie*, 1899, 10. Aufl., Stuttgart 1968
- HILBERT, D.: *Die Hilbertschen Probleme. Vortrag "Mathematische Probleme", gehalten auf dem 2. Internationalen Mathematikerkongreß Paris 1900.*, 4. Aufl., Thun 1998
- KANT, I.: *Kritik der reinen Vernunft*, 1781, 2. Auflage 1787, Nachdruck, Hamburg 1990
- KANT, I.: *Metaphysische Anfangsgründe der Naturwissenschaft*, 1786, Kants Werke auf CD-ROM, Berlin 1996
- KOYRÉ, A.: *Leonardo, Galilei, Pascal. Die Anfänge der neuzeitlichen Naturwissenschaft*, Frankfurt 1998
- MEHRTENS, H.: *Moderne - Sprache - Mathematik*, Frankfurt 1990
- NEWTON, I.: *Philosophia naturalis principia mathematica*, 1687, Übersetzung von E. Dellian, Hamburg 1988
- ORTLIEB, C. P. / V. DRESKY, C. / GASSER, I. / GÜNZEL, S.: *Mathematische Modellierung. Eine Einführung in zwölf Fallstudien*, Wiesbaden 2009
- PEIFFER, J. / DAHAN-DALMEDICO, A.: *Wege und Irrwege - Eine Geschichte der Mathematik*, Basel 1994
- RAPOPORT, A.: *Mathematische Methoden in den Sozialwissenschaften*, Würzburg 1980

SERRES, M. (Hrsg.): *Elemente einer Geschichte der Wissenschaften*, Frankfurt 1998

STENGERS, I.: *Die Galilei-Affären*, in SERRES 1998, 395 - 443

WISSEL, C.: *Theoretische Ökologie*, Berlin 1989

2 Die Dynamik von Wachstum und Zerfall

Veränderungen des Bestandes einer Substanz oder Population sollen in diesem Kapitel als Prototyp für *dynamische Prozesse* dienen, für deren Modellierung außerdem ein allgemeiner Rahmen vorgestellt wird. Die Gestalt der mathematischen Modelle hängt entscheidend davon ab, durch welche mathematischen Strukturen die Zeit und die Zustände des jeweils betrachteten Systems und durch welche Art von Gesetzmäßigkeiten ihre Veränderungen beschrieben werden. Ich beginne mit einer Problemklasse, die trotz ihrer Einfachheit auf ganz verschiedene Art und Weise mathematisch modelliert werden kann.

2.1 Exponentielle Zerfallsprozesse: Vier Modelltypen

Betrachtet wird eine (anorganische, organische oder lebende) Substanz, gemessen durch ihre Masse oder Individuenanzahl, zu der nichts hinzukommt und die im Laufe der Zeit zerfällt oder abstirbt, wie es etwa in den folgenden Beispielen der Fall ist:

- (a) Bäume nehmen Kohlendioxid aus der Luft auf und lagern den darin enthaltenen Kohlenstoff in ihrem Holz an. Ein gewisser Anteil dieses Kohlenstoffs ist radioaktiv (C^{14}). Wird nun ein Baum gefällt, so kommt in das Holz von außen kein Kohlenstoff mehr hinzu, das enthaltene C^{14} zerfällt. Wieviel von ihm ist nach 10, 100, 1000 Jahren noch übrig? Weiß man das, so kann man umgekehrt aus dem gemessenen C^{14} -Gehalt auf das Alter des Holzes schließen (C^{14} -Methode zur Altersbestimmung).
- (b) Ein Gewässer ist durch einen Unglücksfall mit einer toxischen Substanz verunreinigt, die dadurch abgebaut wird, dass ständig frisches Wasser hinzukommt und dieselbe Menge des vorhandenen Wassers abfließt. Wie entwickelt sich die Konzentration des Giftes im Laufe der Zeit?
- (c) In einem abgeschlossenen Habitat (z. B. einem See) wird eine Generation einer bestimmten Art betrachtet (z. B. alle Forellen, die im Jahre 1992 ein Jahr alt waren). Eine solche *Alterskohorte* kann im Laufe der Zeit natürlich nur weniger werden. Wie entwickelt sie sich, wann ist sie vollständig gestorben?
- (d) Für die Fischerei und die Festlegung von Fischfangquoten ist die Frage nicht ganz unerheblich, welche Menge an anderen Fischen etwa ein durchschnittlicher Kabeljau in der Nordsee pro Tag frisst. Das lässt sich, anders als vielleicht auf dem Lande, im Meer aber nicht durch direkte Beobachtung feststellen. Was sich messen lässt, ist der Mageninhalt von gefangenen Fischen. Um hieraus auf die tägliche Menge ihrer Nahrung schließen zu können, muss die Geschwindigkeit ihrer Verdauung, also der Leerung ihres Magens bekannt sein. Das wiederum lässt sich auch unter Laborbedingungen untersuchen. Wie entwickelt sich der Mageninhalt im Laufe der Zeit, wenn keine Nahrung aufgenommen wird?

Es bezeichne im Folgenden t die Zeit und $x(t)$ den Bestand zum Zeitpunkt t . Die Betrachtung beginne zum Zeitpunkt $t = 0$, für den $x(0) = x_0$. Gefragt ist nach $x(t)$ für alle $t \geq 0$.

2.1.1 Kontinuierliche Zustände, diskrete Zeit: Iterationsprozess

Sei $\Delta t > 0$ ein beliebig gewähltes Zeitintervall fester Länge. Das System werde nur zu den Zeitpunkten

$$t = 0, \Delta t, 2\Delta t, 3\Delta t, \dots$$

betrachtet, d. h. die Zeit wird durch ganze Zahlen beschrieben. Die im Weiteren entscheidende **Grundannahme** lautet nun:

In jedem Zeitintervall der Länge Δt zerfällt bzw. stirbt ein fester Anteil

$$p = p(\Delta t) \in [0, 1]$$

des vorhandenen Bestandes, unabhängig vom Zeitpunkt, an dem das passiert. Oder anders formuliert: Die relative Zerfallsrate p ist konstant.

Es ist dann

$$x(t + \Delta t) = x(t) - p x(t) = (1 - p) x(t) = q x(t) \quad (2.1)$$

mit $q = 1 - p$.

Aufgabe 2.1

Diskutieren Sie, für welche der oben aufgeführten Beispiele die hier gemachte Grundannahme sinnvoll ist und für welche sie in welcher Weise modifiziert werden müsste.

Mit $x(0) = x_0$ ergibt sich für die betrachteten Zeitpunkte $t = m \Delta t$ aus (2.1) sofort

$$x(m\Delta t) = x_0 q^m \text{ für alle } m \in \mathbb{N}_0. \quad (2.2)$$

Für den hier betrachteten Modellansatz mit *diskreter Zeit* ist damit die gestellte Aufgabe gelöst.

2.1.2 Kontinuierliche Zustände, kontinuierliche Zeit: Differentialgleichung

Bei manchen realen Fragestellungen wird eine bestimmte Wahl von Δt nahegelegt (z. B. Tag oder Jahr), bei anderen ist dagegen jede Wahl so gut oder schlecht wie jede andere. Um für eine möglichst große Anzahl von Zeitpunkten Aussagen über $x(t)$ zu erhalten, wird man versuchen, Δt möglichst klein zu wählen. Schreibt man nun (2.1) in der Form

$$\frac{x(t + \Delta t) - x(t)}{\Delta t} = - \frac{p(\Delta t)}{\Delta t} x(t),$$

so zeigt sich, was passiert, wenn Δt immer kleiner wird: Auf der linken Seite entsteht die Ableitung $\dot{x}(t)$. Damit auch rechts etwas entsteht, muss

$$\lim_{\Delta t \rightarrow 0} \frac{p(\Delta t)}{\Delta t} =: \alpha$$

existieren, was genau dann der Fall ist, wenn $p(0) = 0$ und p (als Funktion von Δt) an der Stelle 0 differenzierbar ist. Unter diesen Annahmen ergibt sich für $\Delta t \rightarrow 0$ die Differentialgleichung

$$\dot{x}(t) = - \alpha x(t), \quad (2.3)$$

die zusammen mit der Anfangsbedingung $x(0) = x_0$ die eindeutige Lösung

$$x(t) = x_0 e^{-\alpha t} \quad (2.4)$$

hat. Die für den exponentiellen Zerfall charakteristische Größe α heißt auch *exponentielle Zerfallsrate*.

2.1.3 Zusammenhang von diskreter und kontinuierlicher Lösung

Betrachtet man die Lösung (2.4) für die speziellen Zeitpunkte $m \Delta t$ mit $m \in \mathbb{N}_0$ und $\Delta t > 0$ fest, so ergibt sich

$$x(m \Delta t) = x_0 e^{-\alpha m \Delta t} = x_0 (e^{-\alpha \Delta t})^m .$$

Der Vergleich mit der Lösung (2.2) zeigt, dass die beiden Lösungen in den diskreten Zeitpunkten $m \Delta t$ genau dann übereinstimmen, wenn

$$q(\Delta t) = 1 - p(\Delta t) = e^{-\alpha \Delta t} .$$

2.1.4 Halbwertszeit

Die Geschwindigkeit der relativen Abnahme der Lösung (2.4) ist durch die Konstante α gegeben: Je größer α , desto schneller der Zerfall. Eine andere, insbesondere beim radioaktiven Zerfall verwendete Charakterisierung ist die *Halbwertszeit*, d. i. der Zeitraum, in dem sich der Bestand auf die Hälfte verringert, also die Zeit τ , für die

$$x_0 e^{-\alpha \tau} = \frac{x_0}{2} ,$$

woraus sich unabhängig von x_0

$$\tau = \frac{\log 2}{\alpha} \approx \frac{0.6931471}{\alpha}$$

ergibt.

Im Falle von C^{14} beträgt die Halbwertszeit 5568 Jahre.

Aufgabe 2.2

Lebendes und frisch geschlagenes Holz hat eine C^{14} -Zerfallsrate von 6.68 Zerfallsakten pro Minute und Gramm. Im Jahre 1950 wurden in der Lascaux-Höhle in Frankreich (Höhlenmalereien) Holzkohle mit einer C^{14} -Zerfallsrate von 0.97 pro Minute und Gramm gemessen. Wann entstanden vermutlich die Höhlenmalereien?

2.1.5 Diskrete Zustände, diskrete Zeit: Markov-Prozess

Die Modellansätze (2.1) und (2.3) gehen beide davon aus, dass der jeweils vorhandene Bestand als beliebige nichtnegative reelle Zahl ausgedrückt werden kann. Das ist aber unrealistisch, wenn er in Anzahlen von Teilchen oder Individuen gemessen wird. In diesem Falle wird der Zustand des Systems zu jedem Zeitpunkt durch eine nichtnegative ganze Zahl ausgedrückt. Um das auch symbolisch deutlich zu machen, bezeichne $N(t)$ die im System vorhandene Anzahl von Individuen oder Teilchen zum Zeitpunkt t . $N(0) = n$ sei vorgegeben. Sei wiederum $\Delta t > 0$ ein beliebig gewähltes Zeitintervall fester Länge. Ein deterministisches Entwicklungsgesetz der Form

$$N(t + \Delta t) = f(N(t))$$

scheint für einen Zerfallsprozess nicht angemessen, da das Ergebnis eine kleinere ganze Zahl sein müsste¹, so dass die Substanz unabhängig von der Wahl von Δt spätestens nach n Schritten verschwunden wäre. Eine mögliche Lösung besteht darin, das Entwicklungsgesetz als einen stochastischen Prozess aufzufassen: Die Teilchen zerfallen bzw. die Individuen sterben mit einer gewissen Wahrscheinlichkeit. Der oben formulierten Grundannahme, dass immer ein fester Anteil zerfällt, entspricht in diesem Rahmen die folgende:

¹ $N(t + \Delta t) = N(t)$ hätte die Stabilisierung des Bestandes bei diesem Wert zur Folge.

In jedem Zeitintervall der Länge Δt zerfällt jedes Teilchen bzw. stirbt jedes Individuum mit einer festen Wahrscheinlichkeit

$$p = p(\Delta t)$$

unabhängig von den anderen und unabhängig vom aktuellen Zeitpunkt.

Unter dieser Annahme ist

$$p_{jk} := P(N(t + \Delta t) = j \mid N(t) = k) = \binom{k}{j} p^{k-j} (1-p)^j = \binom{k}{j} p^{k-j} q^j \quad (2.5)$$

die Wahrscheinlichkeit dafür, dass während eines Zeitintervalls der Länge Δt von den k vorhandenen Individuen $k - j$ sterben und j überleben (Binomialverteilung, vgl. BARTH / HALLER 1996, S. 231).

Prognosen über die Entwicklung des Systems können jetzt nicht mehr in der Angabe von $N(t)$ für jeden Zeitpunkt $t = m\Delta t$ bestehen, denn schon nach einem einzigen Zeitschritt ist es ebenso möglich, dass alle Individuen gestorben sind, wie, dass sie alle überlebt haben, und alle Zustände dazwischen natürlich auch. Angeben lässt sich nur die Wahrscheinlichkeit für die verschiedenen Zustände. Es bezeichne

$$w_j(t) = P(N(t) = j)$$

die Wahrscheinlichkeit dafür, dass zum Zeitpunkt t noch genau j Individuen leben. Der Vektor

$$w(t) = \begin{pmatrix} w_0(t) \\ w_1(t) \\ \vdots \\ w_n(t) \end{pmatrix} \in \mathbb{R}^{1+n}$$

ist dann eine Wahrscheinlichkeitsverteilung auf der endlichen Menge $\{0, 1, \dots, n\}$. Offenbar ist

$$w(0) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Mit den in (2.5) definierten Übergangswahrscheinlichkeiten p_{jk} dafür, dass ein System mit k Individuen in einem Zeitschritt in ein System mit j Individuen übergeht, ist dann

$$w_j(t + \Delta t) = \sum_{k=j}^n p_{jk} w_k(t) \quad \text{für } j = 0, \dots, n \quad (2.6)$$

oder in vektorieller Darstellung

$$w(t + \Delta t) = P w(t)$$

mit der $(1+n, 1+n)$ -Matrix $P = (p_{jk})$, wobei p_{jk} durch (2.5) definiert ist und $p_{jk} = 0$ für $j > k$. P ist eine *stochastische Matrix*, d. h. die Spalten von P sind Wahrscheinlichkeitsverteilungen auf der Menge $\{0, 1, \dots, n\}$, ihre Einträge sind nichtnegativ, und die Spaltensummen sind alle 1.

Aufgabe 2.3

Zeigen Sie: Mit $w \in \mathbb{R}^{1+n}$ ist auch $P w$ eine Wahrscheinlichkeitsverteilung auf $\{0, 1, \dots, n\}$.

Eine durch eine stochastische Matrix definierte Iteration der Form (2.6) wird *Markov-Prozess* genannt. Die Lösung für die diskreten Zeitpunkte $t = m\Delta t$, $m \in \mathbb{N}_0$ lässt sich formal leicht angeben:

$$w(m\Delta t) = P^m w(0) \text{ für alle } m \in \mathbb{N}_0 .$$

Die Potenzen P^m der stochastischen Matrix P wären allerdings noch zu berechnen, was etwas mühsam, aber hier noch möglich ist. Es gilt: $P^m = (p_{jk}^{(m)})$ mit

$$p_{jk}^{(m)} = \binom{k}{j} q^{mj} (1 - q^m)^{k-j} , \text{ wobei } q = 1 - p ,$$

und daher für die Lösung von (2.6)

$$w_j(m\Delta t) = \binom{n}{j} q^{mj} (1 - q^m)^{n-j} . \quad (2.7)$$

Dieses Ergebnis lässt sich durch elementare stochastische Überlegungen auch direkt gewinnen, also ohne die mühsamen und deshalb hier gar nicht durchgeführten Rechnungen: Für jeden der betrachteten Zeitpunkte $m\Delta t$ ist (2.7) eine Binomialverteilung mit der Grundwahrscheinlichkeit q^m für das Einzelereignis. Das aber besteht gerade darin, dass ein einzelnes Teilchen die m Zeitschritte überlebt. Da nun aber die n Teilchen nach Voraussetzung unabhängig von einander zerfallen oder überleben, muss die Anzahl der überlebenden Teilchen binomial verteilt sein, d. h. es muss sich (2.7) ergeben.

Zahlenbeispiel: Drei Fische im verschmutzten Gewässer

In einem stark verschmutzten Gewässer leben noch 3 Fische. Die Wahrscheinlichkeit für jeden von ihnen, den jeweils nächsten Tag zu erleben, sei $4/5$. Das System hat also 4 mögliche Zustände, beschrieben durch die Anzahl der noch lebenden Fische, und es ist $p = 1/5$. In Abbildung 2.1 sind die möglichen

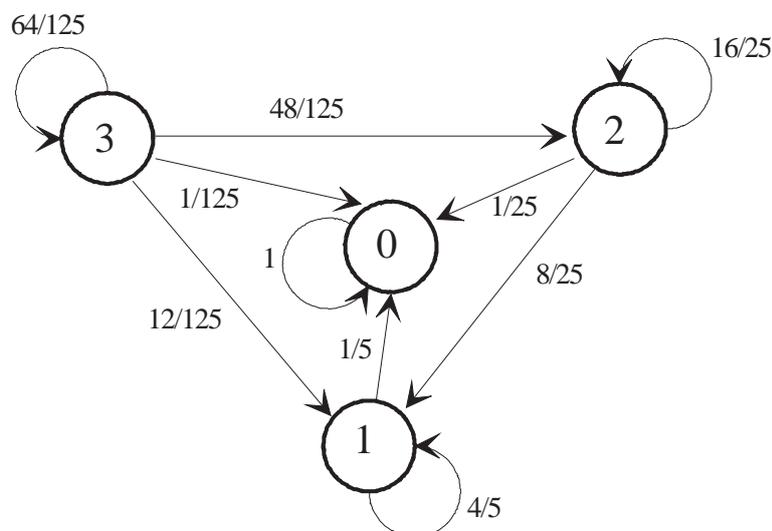


Abbildung 2.1: Ein einfacher Markov-Prozess

Übergänge durch Pfeile gekennzeichnet, an denen die Übergangswahrscheinlichkeiten eingetragen sind. Das System befindet sich zu jedem Zeitpunkt in einem der Zustände und bewegt sich in jedem Zeitschritt

entlang einem der Pfeile mit der an ihm eingetragenen Wahrscheinlichkeit (die Summe der Wahrscheinlichkeiten an den von einem Zustand ausgehenden Pfeilen ist 1). Die Iteration (2.6), hier mit

$$P = \begin{pmatrix} 1 & \frac{1}{5} & \frac{1}{25} & \frac{1}{125} \\ 0 & \frac{4}{5} & \frac{8}{25} & \frac{12}{125} \\ 0 & 0 & \frac{16}{25} & \frac{48}{125} \\ 0 & 0 & 0 & \frac{64}{125} \end{pmatrix},$$

beschreibt die Entwicklung der Wahrscheinlichkeitsverteilung für den vom System eingenommenen Zustand.

Aufgabe 2.4

Berechnen Sie diese Entwicklung und stellen Sie sie grafisch dar.

Es ist intuitiv klar, dass das System irgendwann einmal im Zustand 0 landen muss. Die Frage ist, wann das passiert, genauer: Wie groß ist der Erwartungswert T_k für die Zahl der Zeitschritte, die das im Zustand k befindliche System benötigt, um in den Zustand 0 zu gelangen? Zwischen den Größen T_k besteht der folgende Zusammenhang

$$T_0 = 0, T_k = \sum_{j=0}^n p_{jk}(T_j + 1) = \sum_{j=0}^n p_{jk}T_j + 1 \text{ für } k \neq 0,$$

den man sich folgendermaßen klar macht: Befindet sich das System im Zustand k , so befindet es sich einen Zeitschritt später mit Wahrscheinlichkeit p_{jk} im Zustand j , von wo es im Mittel noch T_j Zeitschritte bis in den Zustand 0 benötigt. Mit Wahrscheinlichkeit p_{jk} werden also von k aus im Mittel $T_j + 1$ Zeitschritte in den Zustand 0 benötigt. Damit ergibt sich ein lineares Gleichungssystem für die T_k , aus denen sie sich berechnen lassen, im hier betrachteten System rekursiv:

$$T_1 = \frac{1}{5}T_0 + \frac{4}{5}T_1 + 1 = \frac{4}{5}T_1 + 1 \text{ ergibt } T_1 = 5,$$

was der mittleren Lebenserwartung eines einzelnen Fisches entspricht.

$$T_2 = \frac{1}{25}T_0 + \frac{8}{25}T_1 + \frac{16}{25}T_2 + 1 = \frac{16}{25}T_2 + \frac{13}{5} \text{ ergibt } T_2 = \frac{65}{9}.$$

$$T_3 = \frac{1}{125}T_0 + \frac{12}{125}T_1 + \frac{48}{125}T_2 + \frac{64}{125}T_3 + 1 = \frac{64}{125}T_3 + \frac{319}{75} \text{ ergibt } T_3 = \frac{1595}{183}.$$

Die drei Fische sind also im Mittel nach

$$\frac{1595}{183} \approx 8.7$$

Tagen alle gestorben.

Eine derartige Aussage lässt sich übrigens mit den deterministischen Modellen (2.1) oder (2.3) nicht machen.

Die Dynamik des Erwartungswerts der Populationsgröße

Auch ohne explizite Kenntnis der Lösung von (2.6) lässt sich die im Mittel zu erwartende Populationsgröße

$$x(t) = \sum_{j=0}^n j w_j(t)$$

leicht berechnen. Aus (2.6) folgt nämlich

$$x(t + \Delta t) = \sum_{j=0}^n j w_j(t + \Delta t) = \sum_{j=0}^n j \sum_{k=j}^n p_{jk} w_k(t) = \sum_{k=0}^n \left(\sum_{j=0}^k j p_{jk} \right) w_k(t).$$

Nun ist aber

$$\sum_{j=0}^k j p_{jk} = k (1 - p)$$

gerade der Erwartungswert der hier vorliegenden Binomialverteilung (BARTH / HALLER 1996, S. 240). Also ist

$$x(t + \Delta t) = (1 - p) x(t),$$

$x(t)$ folgt also der Dynamik (2.1). Im Lichte des hier untersuchten stochastischen Modells lässt sich also sagen, dass das deterministische Modell (2.1) die Veränderung des Erwartungswerts der Größe der betrachteten Substanz oder Population beschreibt.

2.1.6 Diskrete Zustände, kontinuierliche Zeit: Kontinuierlicher Markov-Prozess

Wie beim Übergang von (2.1) zu (2.3) wird jetzt die Abhängigkeit des Markov-Prozesses (2.6) von Δt betrachtet mit der Absicht, Δt beliebig klein werden zu lassen. Ich bringe dazu (2.6) in die Form

$$\frac{w_j(t + \Delta t) - w_j(t)}{\Delta t} = - \frac{1 - p_{jj}(\Delta t)}{\Delta t} w_j(t) + \sum_{k=j+1}^n \frac{p_{jk}(\Delta t)}{\Delta t} w_k(t), \quad (2.8)$$

in der auf der linken Seite beim Grenzübergang $\Delta t \rightarrow 0$ die Ableitung $\dot{w}_j(t)$ entsteht. Andererseits ist

$$p_{jk}(\Delta t) = \binom{k}{j} p(\Delta t)^{k-j} (1 - p(\Delta t))^j$$

mit der Wahrscheinlichkeit $p(\Delta t)$, dass ein einzelnes Teilchen bzw. Individuum im Verlauf des Zeitraum Δt zerfällt bzw. stirbt.

Auch hier werde angenommen, dass $p(0) = 0$ und p an der Stelle $\Delta t = 0$ differenzierbar ist:

$$p(\Delta t) = \alpha \Delta t + o(\Delta t).$$

Dabei bezeichnet $o(\Delta t)$ eine beliebige Funktion von Δt , für die

$$\lim_{\Delta t \rightarrow 0} \frac{o(\Delta t)}{\Delta t} = 0.$$

Es ist dann

$$\begin{aligned} p_{jj}(\Delta t) &= (1 - p(\Delta t))^j = 1 - j \alpha \Delta t + o(\Delta t) \\ p_{j,j+1}(\Delta t) &= (j+1) p(\Delta t) (1 - p(\Delta t))^j = (j+1) \alpha \Delta t + o(\Delta t) \\ p_{jk}(\Delta t) &= o(\Delta t) \quad \text{für } k > j+1 \end{aligned}$$

Einsetzen in (2.8) und der anschließende Grenzübergang $\Delta t \rightarrow 0$ liefern

$$\begin{aligned} \dot{w}_j(t) &= -j \alpha w_j(t) + (j+1) \alpha w_{j+1}(t) \quad \text{für } j = 0, \dots, n-1 \\ \dot{w}_n(t) &= -n \alpha w_n(t) \end{aligned} \quad (2.9)$$

oder in vektorieller Schreibweise

$$\dot{w}(t) = A w(t)$$

mit der $(1+n, 1+n)$ -Matrix

$$A = \alpha \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 2 & 0 & & \vdots \\ 0 & 0 & -2 & 3 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & & & 0 & -n+1 & n \\ 0 & 0 & \cdots & 0 & 0 & -n \end{pmatrix}.$$

Die Spaltensummen von A sind alle 0, und die Außerdiagonalelemente von A sind nichtnegativ. Eine lineares Differentialgleichungssystem mit einer solchen Matrix wird *kontinuierlicher Markov-Prozess* genannt. Ist $w(0)$ eine Wahrscheinlichkeitsverteilung auf $\{0, 1, \dots, n\}$ und w Lösung eines kontinuierlichen Markov-Prozesses, so ist $w(t)$ für alle $t \geq 0$ eine Wahrscheinlichkeitsverteilung auf $\{0, 1, \dots, n\}$.

Zusammen mit einer Anfangsbedingung, im hier betrachteten Fall ist das

$$w_j(0) = 0 \text{ für } j = 0, \dots, n-1, w_n(0) = 1,$$

besitzen derartige Systeme eine eindeutig bestimmte und für alle $t \in \mathbb{R}$ definierte Lösung (HEUSER 1989, S. 462).

Drei Fische im verschmutzten Gewässer

Die oben bereits betrachtete Situation der drei Fische im verschmutzten Gewässer lässt sich im kontinu-

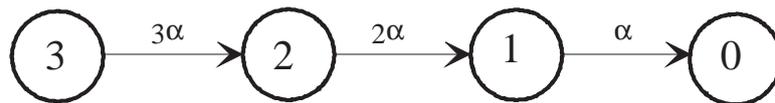


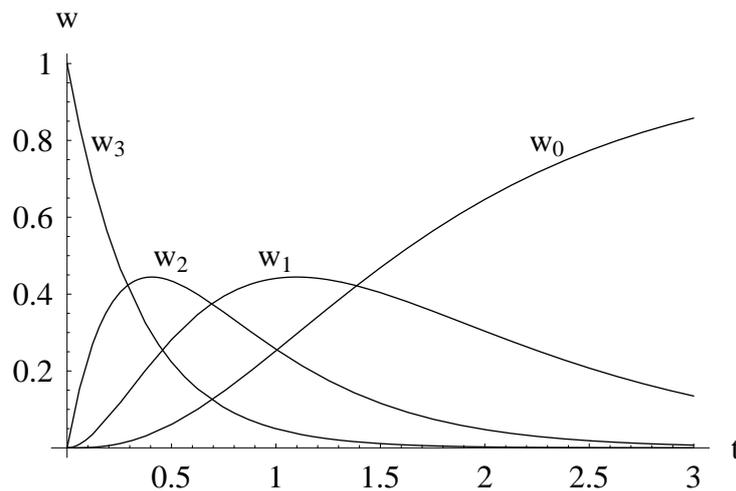
Abbildung 2.2: Kontinuierlicher Markov-Prozess

ierlichen Modell entsprechend Abbildung 2.2 darstellen. Hier gibt es nur Übergänge von n auf $n-1$. In "infinitesimaler Zeit" kann es keinen Sprung von 3 auf 1 geben, dazwischen liegt immer der Zustand 2. Die Zahlenwerte an den Pfeilen in Abbildung 2.2 sind keine Wahrscheinlichkeiten, sondern haben die Dimension "Wahrscheinlichkeit pro Zeit". Sie geben an, welcher Anteil des auf dem Zustand am Anfang eines Pfeils liegenden Gewichts pro Zeiteinheit in den Zustand am Ende des Pfeils hineinfließt.² Das zugehörige Differentialgleichungssystem ist durch das Diagramm (und die Forderung, dass die Spaltensummen der Systemmatrix 0 sein müssen) eindeutig bestimmt und lautete hier:

$$\begin{aligned} \dot{w}_0 &= \alpha w_1 \\ \dot{w}_1 &= -\alpha w_1 + 2\alpha w_2 \\ \dot{w}_2 &= -2\alpha w_2 + 3\alpha w_3 \\ \dot{w}_3 &= -3\alpha w_3 \end{aligned}$$

Die (durch die hinzu kommende Anfangsbedingung) eindeutig bestimmte Lösung lässt sich mit einem

²Entsprechende Modelle lassen sich formulieren, wenn es darum geht, Schadstoffflüsse zwischen verschiedenen *Kompartimenten* zu beschreiben, vgl. HEUSER 1989, S. 483 ff.

Abbildung 2.3: Lösung des kontinuierlichen Markov-Prozesses ($\alpha = 1$)

Computeralgebra-System wie etwa MATHEMATICA exakt bestimmen und lautet (siehe auch Abbildung 2.3)

$$\begin{aligned} w_0(t) &= (1 - e^{-\alpha t})^3 \\ w_1(t) &= 3 e^{-\alpha t} (1 - e^{-\alpha t})^2 \\ w_2(t) &= 3 e^{-2\alpha t} (1 - e^{-\alpha t}) \\ w_3(t) &= e^{-3\alpha t} \end{aligned}$$

Die allgemeine Lösung

Die hier für den Fall $n = 3$ gefundene Lösung lässt vermuten, dass allgemein

$$w_j(t) = \binom{n}{j} e^{-j\alpha t} (1 - e^{-\alpha t})^{n-j} \quad (2.10)$$

die gesuchte Lösung von (2.9) ist.

Aufgabe 2.5

Zeigen Sie, dass das so ist, indem Sie (2.10) in (2.9) einsetzen.

Die Populationsgröße ist also zu jedem Zeitpunkt $t \geq 0$ binomial verteilt, wobei sich mit wachsendem t der Parameter $e^{-\alpha t}$ dieser Binomialverteilung von 1 auf 0 verschiebt. Auch dieses Ergebnis lässt sich – wie oben – mit elementaren stochastischen Überlegungen direkt gewinnen, wenn man berücksichtigt, dass $e^{-\alpha t}$ gerade die Wahrscheinlichkeit für einen einzelnen Fisch ist, zum Zeitpunkt t noch am Leben zu sein.

Zusammenhang von diskreter und kontinuierlicher Lösung

Auch für die Lösungen (2.7) und (2.10) der beiden stochastischen Modelle gilt, wie schon im deterministischen Fall: Sie stimmen in den diskreten Zeitpunkten $m\Delta t$ genau dann überein, wenn

$$q(\Delta t) = 1 - p(\Delta t) = e^{-\alpha \Delta t} .$$

Die Dynamik des Erwartungswerts

Für den Erwartungswert des Bestandes

$$x(t) = \sum_{j=0}^n j w_j(t)$$

ergibt sich aus (2.9)

$$\begin{aligned} \dot{x} &= \sum_{j=0}^{n-1} j (-j \alpha w_j + (j+1) \alpha w_{j+1}) - n^2 \alpha w_n \\ &= \sum_{j=0}^{n-1} j(j+1) \alpha w_{j+1} - \sum_{j=1}^n j^2 \alpha w_j \\ &= - \sum_{j=0}^{n-1} (j+1) \alpha w_{j+1} \\ &= - \alpha x \end{aligned}$$

Der Erwartungswert genügt also der Differentialgleichung (2.3), und es ist daher

$$x(t) = n e^{-\alpha t} .$$

Dieses Ergebnis hätte sich auch direkt aus der allgemeinen Lösung (2.10) und dem Satz über den Erwartungswert der Binomialverteilung (BARTH / HALLER 1996, S. 240) ableiten lassen.

Varianz und Standardabweichung

Für die Varianz der binomial verteilten Zufallsgröße (2.10) gilt (BARTH / HALLER 1996, S. 240)

$$Var w(t) = \sum_{j=0}^n (j - x(t))^2 w_j(t) = n e^{-\alpha t} (1 - e^{-\alpha t}) = x(t) (1 - e^{-\alpha t}) ,$$

für die Standardabweichung somit

$$\sigma(w(t)) = \sqrt{Var w(t)} = \sqrt{x(t) (1 - e^{-\alpha t})} .$$

Die Größe

$$\frac{\sigma(w(t))}{x(t)} = \frac{\sqrt{1 - e^{-\alpha t}}}{\sqrt{x(t)}}$$

ist ein Maß für die mittlere relative Abweichung des Zustands vom Mittelwert, also der relativen Abweichung der Lösung des stochastischen Modells (2.9) von der des deterministischen Modells (2.3). Je größer also die Anzahl unabhängig voneinander zerfallender Teilchen bzw. sterbender Individuen (noch) ist, desto angemessener ist es, das einfachere deterministische Modell an Stelle des komplizierteren stochastischen zu verwenden.

2.1.7 Zusammenfassung

Die hier untersuchten Modelle (2.1), (2.3), (2.6) und (2.9) beschreiben Zerfalls- oder Sterbeprozesse mit gleichartigen, unabhängig voneinander und von der Zeit zerfallenden Teilchen bzw. sterbenden Individuen. Sie hängen eng miteinander zusammen: Die diskreten Modelle stimmen in den Zeitpunkten, für die sie definiert sind, mit dem entsprechenden kontinuierlichen überein. Die deterministischen Modelle beschreiben die (exponentiell fallende) Dynamik des Mittelwerts des entsprechenden stochastischen. Je größer der Bestand der Substanz oder Population, gemessen in der Anzahl ihrer Teilchen oder Mitglieder, desto angemessener die Beschreibung des Zerfalls durch ein deterministisches Modell.

Achtung: Die hier hergeleiteten Ergebnisse sind **nicht prototypisch** für beliebige dynamische Modelle. Sie gelten einzig und allein für die hier betrachtete Problemklasse des exponentiellen Zerfalls. Das betrifft sowohl den Zusammenhang zwischen deterministischen und stochastischen als auch den zwischen diskreten und kontinuierlichen Modellen. Die Wahl der "richtigen" Modellklasse bleibt daher ein Problem, das im Auge zu behalten ist.

2.1.8 Methodischer Exkurs: Erzeugende Funktionen

Ist $w = (w_j : j \in \mathbb{N}_0)$ eine Wahrscheinlichkeitsverteilung auf der Menge der nichtnegativen ganzen Zahlen (also der möglichen Bestände einer Population, gemessen in der Anzahl ihrer Mitglieder), so heißt die durch die Potenzreihe

$$G(z) := \sum_{j=0}^{\infty} w_j z^j \quad (2.11)$$

definierte Funktion die zugehörige *erzeugende Funktion*. Der Konvergenzradius der Potenzreihe (2.11) ist mindestens 1, da sie für $z = 1$ konvergiert:

$$G(1) = \sum_{j=0}^{\infty} w_j = 1.$$

Ist die Wahrscheinlichkeitsverteilung nicht durch eine explizite Formel, sondern, wie in den obigen stochastischen Modellen des Zerfalls, nur implizit durch eine Iterationsvorschrift oder ein System von Differentialgleichungen angegeben, so ist es manchmal leichter, die erzeugende Funktion zu bestimmen, als die Wahrscheinlichkeitsverteilung selbst. Kennt man aber G als Funktionsausdruck in z , so lässt sich daraus wegen

$$w_j = \frac{G^{(j)}(0)}{j!} \text{ für alle } j \in \mathbb{N}_0$$

die Wahrscheinlichkeitsverteilung ermitteln. Und auch ohne explizite Kenntnis aller w_j lassen sich charakteristische Größen bestimmen. So ist

$$\mu_w = \sum_{j=0}^{\infty} j w_j = G'(1)$$

der Mittelwert von w , sofern G an der Stelle 1 tatsächlich gliedweise differenziert werden kann, was für konkrete Funktionen jeweils zu überprüfen wäre, und für die Varianz erhält man

$$\begin{aligned} \sigma_w^2 &= \sum_{j=0}^{\infty} (j - \mu_w)^2 w_j = \sum_{j=0}^{\infty} j^2 w_j - \mu_w^2 = \\ &= \sum_{j=0}^{\infty} j(j-1) w_j + \mu_w - \mu_w^2 = G''(1) + G'(1)(1 - G'(1)). \end{aligned}$$

Zerfallsprozess in diskreter Zeit

Wie oben betrachte ich, der Einfachheit halber mit $\Delta t = 1$, eine Population von $N(0) = n$ Individuen, von denen jedes während einer Zeiteinheit mit der Wahrscheinlichkeit p stirbt, unabhängig von den anderen und unabhängig vom aktuellen Zeitpunkt. Sei $w_j(t)$ die Wahrscheinlichkeit dafür, dass die Population nach t Zeiteinheiten aus genau j Individuen besteht. Da kein Individuum hinzukommt, ist $w_j(t) = 0$ für $j > n$, sodass für jeden Zeitpunkt t eine Wahrscheinlichkeitsverteilung über der Menge $\{0, 1, \dots, n\}$ vorliegt. Gemäß (2.5) und (2.6) ist für alle $t \in \mathbb{N}_0$

$$w_j(t+1) = \sum_{k=0}^n \binom{k}{j} p^{k-j} q^j w_k(t) \text{ für } j = 0, \dots, n \quad (2.12)$$

mit der Überlebenswahrscheinlichkeit $q = 1 - p$. Zusammen mit der Anfangsbedingung

$$w_j(0) = 0 \text{ für } j = 0, \dots, n-1, w_n(0) = 1 \quad (2.13)$$

sind dann die Wahrscheinlichkeitsverteilungen $w(t)$ für alle $t \in \mathbb{N}_0$ eindeutig bestimmt.

Die zugehörigen erzeugenden Funktionen

$$G_t(z) := \sum_{j=0}^n w_j(t) z^j$$

sind dann allesamt Polynome vom Grade n . Aus (2.12) ergibt sich

$$\begin{aligned} G_{t+1}(z) &= \sum_{j=0}^n w_j(t+1) z^j = \sum_{j=0}^n \sum_{k=0}^n \binom{k}{j} p^{k-j} q^j w_k(t) z^j = \\ &= \sum_{k=0}^n w_k(t) \left(\sum_{j=0}^n \binom{k}{j} p^{k-j} q^j z^j \right) = \\ &= \sum_{k=0}^n w_k(t) (p + qz)^k = G_t(p + qz) \end{aligned}$$

und damit aus (2.13) iterativ

$$\begin{aligned} G_0(z) &= z^n \\ G_1(z) &= (p + qz)^n \\ G_2(z) &= (p(1 + q) + q^2 z)^n \\ G_3(z) &= (p(1 + q + q^2) + q^3 z)^n \\ &\vdots \\ G_t(z) &= \left(p \sum_{s=0}^{t-1} q^s + q^t z \right)^n = \\ &= (1 - q^t + q^t z)^n \end{aligned}$$

Aufgabe 2.6

Bestimmen Sie hieraus Mittelwert und Varianz der Wahrscheinlichkeitsverteilungen $w(t)$.

Zerfallsprozess in kontinuierlicher Zeit

Modelliert man den Zerfallsprozess in kontinuierlicher Zeit, so ergibt sich für die nunmehr für alle reellen $t \geq 0$ definierten Wahrscheinlichkeiten $w_j(t)$ das System

$$\begin{aligned} \dot{w}_j(t) &= -j \alpha w_j(t) + (j+1) \alpha w_{j+1}(t) \text{ für } j = 0, \dots, n-1 \\ \dot{w}_n(t) &= -n \alpha w_n(t) \end{aligned} \quad (2.14)$$

von Differentialgleichungen (siehe (2.9)) zusammen mit der Anfangsbedingung (2.13). Die zugehörige erzeugende Funktion

$$G(t, z) := \sum_{j=0}^n w_j(t) z^j$$

ist auch hier für jedes $t \geq 0$ ein Polynom in z vom Grade n . Aus (2.14) folgt

$$\begin{aligned} \frac{\partial G}{\partial t}(t, z) &= \sum_{j=0}^n \dot{w}_j(t) z^j = \\ &= \alpha \sum_{j=0}^{n-1} (-j w_j(t) + (j+1) w_{j+1}(t)) z^j - \alpha n w_n(t) z^n = \\ &= -\alpha z \sum_{j=1}^n j w_j(t) z^{j-1} + \alpha \sum_{j=1}^n j w_j(t) z^{j-1} = \\ &= \alpha (1-z) \frac{\partial G}{\partial z}(t, z). \end{aligned}$$

Zusammen mit (2.13) erhält man also für G die Gleichungen

$$\frac{\partial G}{\partial t}(t, z) + \alpha (z-1) \frac{\partial G}{\partial z}(t, z) = 0, \quad G(0, z) = z^n. \quad (2.15)$$

Es handelt sich hier um eine lineare partielle Differentialgleichung 1. Ordnung zusammen mit einer Anfangsbedingung. Diese Aufgabe lässt sich mit der *Charakteristikenmethode* lösen, auf die ich hier nur rudimentär, auf das spezielle Beispiel bezogen eingehe (siehe auch STRAUSS 1995, 6 - 11): Entlang jeder Kurve $(t, z(t))$ in der tz -Ebene mit einer Lösung z der Differentialgleichung

$$\dot{z} = \alpha (z-1) \quad (2.16)$$

hat wegen

$$\frac{d}{dt} G(t, z(t)) = \frac{\partial G}{\partial t}(t, z(t)) + \frac{\partial G}{\partial z}(t, z(t)) \dot{z}(t) = \frac{\partial G}{\partial t}(t, z(t)) + \alpha (z(t)-1) \frac{\partial G}{\partial z}(t, z(t)) = 0$$

die gesuchte Funktion G einen konstanten Wert. Da die Lösungen von (2.16) gerade die Funktionen der Form

$$z(t) = 1 + c e^{\alpha t}$$

mit einem beliebigen $c \in \mathbb{R}$ sind, heißt das, das für jedes $c \in \mathbb{R}$

$$G(t, 1 + c e^{\alpha t}) = \text{const} = G(0, 1 + c) = (1 + c)^n \text{ für alle } t \geq 0.$$

Da sich nun aber alle $z \in \mathbb{R}$ bei gegebenem t in der Form $z = 1 + c e^{\alpha t}$ ausdrücken lassen, indem nämlich $c = (z-1) e^{-\alpha t}$ gewählt wird, erhält man als eindeutige Lösung von (2.15)

$$G(t, z) = (1 + (z-1) e^{-\alpha t})^n = (1 - e^{-\alpha t} + z e^{-\alpha t})^n.$$

Aufgabe 2.7

Bestimmen Sie hieraus Mittelwert und Varianz der Wahrscheinlichkeitsverteilungen $w(t)$ und vergleichen Sie das Ergebnis mit dem oben bereits festgestellten. Es sollte übereinstimmen.

2.2 Exponentielle Zerfallsprozesse: Modelle und Daten

Es wird weiterhin ein exponentieller Zerfallsprozess betrachtet, in dem der Erwartungswert des Bestandes einem Gesetz der Form

$$x(t) = x_0 e^{-\alpha t}$$

genügt. Die charakteristische, vom Beobachtungszeitpunkt unabhängige Größe eines solchen Prozesses ist die exponentielle Zerfallsrate α bzw. sind bei gegebenem Zeitintervall Δt die davon abhängigen Größen

$$q = q(\Delta t) = e^{-\alpha \Delta t}, \quad p = p(\Delta t) = 1 - q(\Delta t) = 1 - e^{-\alpha \Delta t}.$$

In der Regel sind nicht diese Parameter bekannt, sondern Daten in Gestalt der Größe des Bestandes für endlich viele Zeitpunkte. Wie sehen die Daten aus, die von solchen Prozessen erzeugt werden, und wie lässt sich umgekehrt aus den Daten der charakteristische Parameter eines solchen Prozesses ermitteln?

2.2.1 Stochastische Simulationen

Auch wenn man, wie etwa beim radioaktiven Zerfall, davon ausgehen kann, dass sich der betrachtete Zerfallsprozess der oben formulierten Grundannahme entsprechend verhält, so ist dennoch bei kleinen Bestandsgrößen mit deutlich erkennbaren Abweichungen von dem durch (2.2) bzw. (2.4) beschriebenen exponentiellen Zerfallsgesetz zu rechnen, wie sich bei Simulationen unter Einsatz eines Zufallsprozessors zeigt.³ Eine solche Simulation unter Verwendung von MATHEMATICA⁴ ergibt etwa für $N_0 = 100$

Zeit t	0	1	2	3	4	5	6	7	8	9	10
Bestand N	100	88	78	70	65	61	54	48	44	36	36

Tabelle 2.1: Stochastische Simulation des exponentiellen Zerfalls mit $p = 0.1$

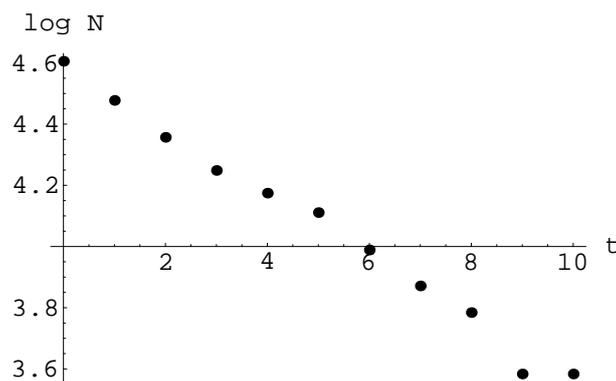


Abbildung 2.4: Logarithmierte Daten aus Tabelle 2.1

und $p = 0.1$ das in Tabelle 2.1 dargestellte Ergebnis. Im Mittel (oder bei deterministischem Zerfallsgesetz exakt) müsste sich der Bestand in jedem Zeitschritt um 10 Prozent vermindern. Die Abweichung vom deterministischen Zerfallsgesetz

$$N = N(t) = N_0 (1 - p)^t = N_0 e^{-\alpha t}$$

lässt sich mit bloßem Auge erkennen, wenn man dieses in logarithmierter Form als

$$\log N = \log N_0 + t \log(1 - p) = \log N_0 - \alpha t$$

mit den entsprechend logarithmierten Daten vergleicht (vgl. Abbildung 2.4): Die Datenpunkte sollten auf einer Geraden liegen, deren Steigung $-\alpha$ ist.

2.2.2 Anpassung der Modellparameter an Daten

Die umgekehrte Situation kommt häufiger vor: Gegeben seien Messdaten eines Zerfallsprozesses etwa in Gestalt von Tabelle 2.1 oder allgemeiner durch $n + 1$ Zahlenpaare

$$(t_0, y_0), (t_1, y_1), \dots, (t_n, y_n)$$

³Mit den meisten Programmiersprachen lassen sich, in der Regel durch eine Funktion namens Random, sogenannte *Zufallszahlen* erzeugen. Korrekter ist es, von *Pseudo-Zufallszahlen* zu sprechen, da sie im Computer immer nur durch deterministische Prozesse erzeugt werden können. Die Erzeugung von Pseudo-Zufallszahlen ist ein Kapitel für sich, auf das hier nicht näher eingegangen wird.

⁴Das MATHEMATICA-Standardpaket Statistics'Discrete Distributions enthält die Funktion BinomialDistribution, mit der sich binomial verteilte Zufallszahlen erzeugen lassen.

mit den Messpunkten $t_0 < t_1 < \dots < t_n$ und den Messwerten y_0, y_1, \dots, y_n . Zum Zeitpunkt t_i wurde jeweils der Bestand y_i gemessen. Unterstellt man einen exponentiellen Zerfallsprozess, so sind also N_0 und α zu bestimmen, sodass

$$N_0 e^{-\alpha t_i} \approx y_i \text{ für } i = 0, 1, \dots, n. \quad (2.17)$$

Ist $n \geq 2$, so ist nicht zu erwarten, dass sich diese Beziehungen mit Gleichheit erfüllen lassen, da es sich um ein überbestimmtes Gleichungssystem ($n + 1$ Gleichungen für die beiden Unbekannten N_0 und α) handelt. (2.17) ist daher noch kein präzise gestelltes mathematisches Problem, ein solches wäre erst noch zu formulieren. Wie immer in solchen Situationen gibt es dazu verschiedene Möglichkeiten.

2.2.3 Ein nichtlineares Ausgleichsproblem

Ein gängiges Verfahren zur Behandlung überbestimmter Gleichungssysteme besteht darin, die Quadrate der in den Gleichungen auftretenden Fehler aufzusummieren und die Unbekannten so zu bestimmen, dass diese Summe möglichst klein wird. Im Falle (2.17) sind also N_0 und α zu bestimmen, sodass

$$\sum_{i=0}^n (N_0 e^{-\alpha t_i} - y_i)^2 = \min ! \quad (2.18)$$

Der Nachteil dieser Neuformulierung von (2.17) ist, dass ein *nichtlineares Ausgleichsproblem*, wie es in Gestalt von (2.18) vorliegt, sich nicht so ohne Weiteres lösen lässt und noch nicht einmal seine eindeutige Lösbarkeit gesichert ist. "Nichtlinear" bedeutet hier, dass die *Residuen*

$$N_0 e^{-\alpha t_i} - y_i, ,$$

deren Quadrate hier aufsummiert werden, von den gesuchten Parameterwerten N_0 und α nichtlinear abhängen.

2.2.4 Logarithmierte Daten: Lineares Ausgleichsproblem

Logarithmiert man (2.17), so ergibt sich das überbestimmte Gleichungssystem

$$\log N_0 - \alpha t_i \approx \log y_i \text{ für } i = 0, 1, \dots, n, \quad (2.19)$$

das in den Unbekannten α und $\log N_0$ linear ist. Das zugehörige *lineare Ausgleichsproblem*

$$\sum_{i=0}^n (\log N_0 - \alpha t_i - \log y_i)^2 = \min ! \quad (2.20)$$

lässt sich mit Standardverfahren der numerischen Mathematik eindeutig lösen, wie sie in jedem einschlägigen Lehrbuch beschrieben sind (siehe etwa OPFER 1994, S. 197 ff).⁵ Im Falle der Daten aus Tabelle 2.1 ergibt sich auf sechs Dezimalstellen genau

$$\log N_0 = 4.58681 \text{ und } \alpha = 0.103101 .$$

Mit diesen Werten wird die *Ausgleichsgerade* durch die Punkte in Abbildung 2.4 bestimmt, die hier in Abbildung 2.5 dargestellt ist. Diesem Ergebnis entspricht

$$N_0 = 98.1809 \text{ und } p = 1 - e^{-\alpha} = 0.0979639 .$$

⁵Bei Verwendung von MATHEMATICA leistet das die eingebaute Funktion Fit.

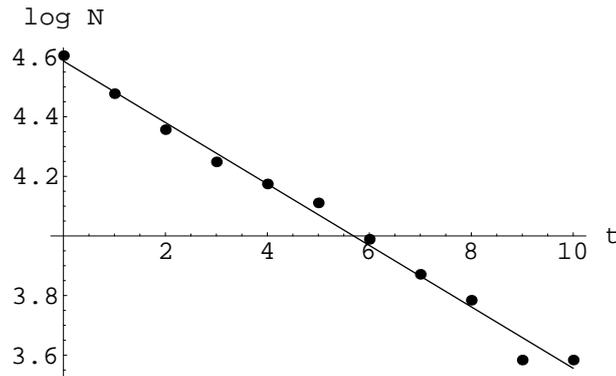


Abbildung 2.5: Logarithmierte Daten und Ausgleichsgerade

2.2.5 Zur Wahl des Ausgleichsproblems

Geht man, wie es im hier betrachteten Beispiel der Fall ist, davon aus, dass die linken und rechten Seiten von (2.17) bzw. (2.19) nur wenig voneinander abweichen, so gilt für die Residuen von (2.19)

$$\log y_i - \log N_0 + \alpha t_i = \log \frac{y_i}{N_0 e^{-\alpha t_i}} = \log \left(1 + \frac{y_i - N_0 e^{-\alpha t_i}}{N_0 e^{-\alpha t_i}} \right) \approx \frac{y_i - N_0 e^{-\alpha t_i}}{N_0 e^{-\alpha t_i}}.$$

Was also in (2.20) minimiert wird, ist die Quadratsumme der *relativen* Fehler, also der relativen Abweichungen der gemessenen Werte y_i vom "wahren" Wert $N_0 e^{-\alpha t_i}$. Dagegen wird in (2.18) die Quadratsumme der *absoluten* Fehler minimiert.

Um einschätzen zu können, welcher der beiden Ansätze angemessener erscheint, muss man sich die Begründung der von Carl Friedrich Gauß eingeführten Fehlerquadratmethode genauer ansehen. Gauß ist von einer Situation ausgegangen, in der die Messwerte y_i mit einem normalverteilten Messfehler *gleicher Varianz* für alle i behaftet sind, und hat gezeigt, dass *in dieser Situation* die Minimierung der Fehlerquadratsumme die beste Schätzung des Mittelwerts liefert.

Für die Wahl des richtigen Ansatzes sind daher Kenntnisse oder Annahmen darüber erforderlich, wie die Fehler, also die Abweichungen von der Gleichheit in (2.17) zu Stande kommen und in welchen Größenverhältnissen sie zueinander stehen. Daten, die (aus messtechnischen oder anderen Gründen) ungenau sind, wird man weniger ernst nehmen als genauere Daten, und die Fehlerquadrate mit entsprechenden Gewichten belegen.

Kann man davon ausgehen, dass der absolute Fehler bei allen Messungen ungefähr derselbe ist, so wird man also Ansatz (2.18) wählen. Ist dagegen der relative Fehler bei allen Messungen gleich einzuschätzen, so ist Ansatz (2.20) vorzuziehen.

2.2.6 Ein Ansatz für kleine Bestände bei exakten Messungen

Die Zahlen in Tabelle 2.1 und ihre Abweichungen vom exponentiellen Zerfallsgesetz sind *nicht* durch Messfehler zu Stande gekommen, sondern dadurch, dass wir es hier mit einem stochastischen Prozess zu tun haben, der nur im Mittel dem exponentiellen Zerfallsgesetz entspricht und desto stärker davon abweicht, je kleiner der Bestand ist. Werden zu den äquidistanten Zeitpunkten $t = 0, 1, \dots, n$ exakte Messungen y_0, y_1, \dots, y_n des (ganzzahligen) Bestands durchgeführt und entsteht der Bestand y_i aus dem Bestand y_{i-1} durch einen binomial verteilten Zufallsprozess mit dem Parameter (der Überlebenswahrscheinlichkeit)

$$q = 1 - p = e^{-\alpha},$$

so ist $y_i \approx q y_{i-1}$, und der dabei auftretende Fehler ist proportional zur Standardabweichung der Binomialverteilung, also proportional zu $\sqrt{y_{i-1}}$. Die gewichteten Fehler

$$\frac{y_i - q y_{i-1}}{\sqrt{y_{i-1}}}$$

sind also von der gleichen Größenordnung. Es liegt daher nahe, den unbekannt Parameter q so zu wählen, dass

$$\sum_{i=1}^n \frac{(y_i - q y_{i-1})^2}{y_{i-1}} = \sum_{i=1}^n \frac{y_i^2}{y_{i-1}} - 2q \sum_{i=1}^n y_i + q^2 \sum_{i=1}^n y_{i-1}$$

minimal wird, was für

$$q = \frac{y_1 + \dots + y_n}{y_0 + \dots + y_{n-1}} \text{ und damit } p = \frac{y_0 - y_n}{y_0 + \dots + y_{n-1}}$$

der Fall ist. Für die Daten aus Tabelle 2.1 ergibt sich damit

$$p = 0.0993789 .$$

Achtung: Es ist im wahrsten Sinne des Wortes Zufall, dass im hier durchgerechneten Zahlenbeispiel die zuletzt genannte Methode das bessere Ergebnis liefert: Das hat nur mit der zufällig erzeugten Tabelle 2.1 zu tun.

Aufgabe 2.8

Führen Sie die Rechnungen dieses Abschnitts, einschließlich der Erzeugung zufälliger Daten, mehrfach durch.

Aufgabe 2.9

Beim radioaktiven Zerfall werden nicht die Bestände (Anzahl der Teilchen), sondern die in einer gewissen Zeiteinheit zerfallenden Teilchen gemessen. Bei einer unbekannt radioaktiven Substanz werde über 10 Stunden die Anzahl der in einer Stunde zerfallenden Teilchen gemessen:

Stunde	1	2	3	4	5	6	7	8	9	10
Anzahl der Zerfallsakte	109	84	75	58	34	38	26	18	28	21

Schätzen Sie die Halbwertszeit der Substanz.

Maximum Likelihood: Eine andere Interpretation desselben Ansatzes

Bei gegebener Zerfallswahrscheinlichkeit $p \in (0, 1)$ und, wie vorausgesetzt, voneinander unabhängigen Zerfallsakten ist die Wahrscheinlichkeit dafür, dass es ausgehend von y_0 gerade zu der Reihe y_0, y_1, \dots, y_n von Beständen kommt,

$$\binom{y_0}{y_1} \binom{y_1}{y_2} \dots \binom{y_{n-1}}{y_n} p^{y_0 - y_n} (1 - p)^{y_1 + \dots + y_n} .$$

Ein in der statistischen Datenanalyse gebräuchlicher Ansatz zur Schätzung von p besteht darin, p so zu wählen, dass die Wahrscheinlichkeit für das tatsächlich beobachtete Ergebnis y_0, y_1, \dots, y_n möglichst groß wird. Dazu ist der Ausdruck

$$p^{y_0 - y_n} (1 - p)^{y_1 + \dots + y_n}$$

zu maximieren, was gerade durch (wie oben)

$$p = \frac{y_0 - y_n}{y_0 + \dots + y_{n-1}}$$

geleistet wird.

2.3 Eindimensionale Wachstumsmodelle

Bestände müssen nicht zerfallen, sie können auch größer werden. Dafür kann es "externe" Gründe geben, wie etwa bei der Schadstoffmenge in einem Gewässer, wenn Bestände von außen zugeführt werden. Bei lebenden Systemen (Populationen) kommen "interne" Gründe hinzu, sie wachsen "aus sich heraus". Davon handeln die Modelle dieses Abschnitts, wobei davon abstrahiert wird, dass auch solche Systeme einer externen Energiezufuhr bedürfen, etwa in Form von Nahrung.

- (a) Laut FISCHER-WELTALMANACH '98 kamen im Jahr 1995 in Argentinien auf je 1000 Einwohner 20 Geburten und 8 Todesfälle. Nimmt man an, dass diese Zahlen über Jahre hinweg konstant bleiben, so wächst also die Bevölkerung um 1.2 Prozent pro Jahr. Wieviele Einwohner hat Argentinien im Jahr 2020, wenn es 1995 ca. 33 Millionen waren?
- (b) Das Wachstum einer Volkswirtschaft bzw. des Bruttoinlandsprodukts (BIP) pro Kopf wird üblicherweise in Prozent pro Jahr angegeben, also als relative Wachstumsrate. Bei MANKIV 1999, S. 559 findet man für das "reale" BIP pro Kopf in den USA die folgenden Zahlen, in Preisen von 1985: Im Jahre 1870 betrug das BIP pro Kopf 2244 \$, im Jahre 1990 dagegen 18258 \$. Hieraus wird als durchschnittliche Wachstumsrate 1.76 Prozent pro Jahr errechnet.

2.3.1 Absolute, relative und exponentielle Wachstumsrate

Es bezeichne wieder $x(t)$ den Bestand zum Zeitpunkt t . Unter der *absoluten Wachstumsrate* w wird die Geschwindigkeit verstanden, mit der sich der Bestand verändert, also

$$w = \dot{x}(t) \text{ oder } w = \frac{x(t + \Delta t) - x(t)}{\Delta t},$$

je nachdem, ob die Zeit kontinuierlich oder in Zeitintervallen einer festen Länge $\Delta t > 0$ gemessen wird. Die Maßeinheit der absoluten Wachstumsrate ist *Bestand pro Zeiteinheit*. Bei Populationen ergibt sie sich als Differenz von Geburten und Todesfällen pro Zeiteinheit. Bei der Modellentwicklung mit einem

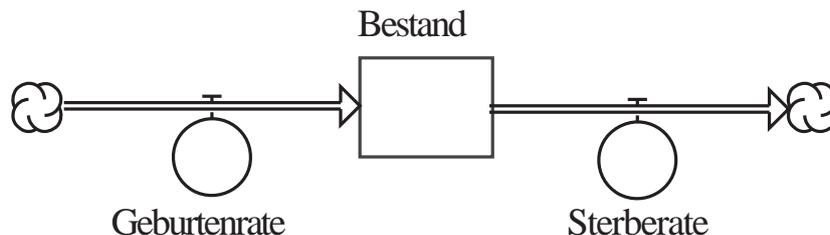


Abbildung 2.6: STELLA-Grundelement dynamischer Modelle

Programm wie STELLA stellt die Beziehung zwischen Bestand und absoluter Wachstumsrate das Grundelement dar (Abbildung 2.6). Mit dem Diagramm ist die Gleichung

$$\text{Bestand}(t + \Delta t) = \text{Bestand}(t) + (\text{Geburtenrate} - \text{Sterberate}) \Delta t$$

fest verbunden. Für die Modellierung ist dann entscheidend, wie Geburten- und Sterberate ihrerseits von anderen Modellgrößen abhängen.

Wenn in der Zeitung von einem 3.3-prozentigen Wachstum des Bruttoinlandsprodukts die Rede ist, so handelt es sich bei $p = 0.033$ um die *relative Wachstumsrate*, die auf den vorhandenen Bestand bezogen ist:

$$p = \frac{\Delta x}{x} = \frac{x(t + \Delta t) - x(t)}{x(t)},$$

für die daher

$$x(t + \Delta t) = x(t) + p x(t) = (1 + p) x(t) .$$

Die relative Wachstumsrate ist eine dimensionslose Zahl, die stets von der Wahl des Zeitintervalls Δt abhängt, weshalb dieses auch immer mit angegeben werden muss.

Lässt man das Zeitintervall immer kleiner werden, so heißt der Grenzwert

$$\alpha = \lim_{\Delta t \rightarrow 0} \frac{p(\Delta t)}{\Delta t} ,$$

sofern existent, die *exponentielle Wachstumsrate*. Es ist dann

$$\dot{x}(t) = \alpha x(t) \text{ oder } \alpha = \frac{\dot{x}}{x} = \frac{d}{dt} \log x .$$

Die Maßeinheit der exponentiellen Wachstumsrate ist 1/Zeiteinheit.

Achtung: In verschiedenen Texten werden die verschiedenen Wachstumsraten w , p und α manchmal einfach als "Wachstumsrate" bezeichnet. Beachten Sie stets, welche gemeint ist.

2.3.2 Deterministische Modelle exponentiellen Wachstums

Ist die exponentielle Wachstumsrate α bzw. (bei fest gewähltem Δt) die relative Wachstumsrate p eine Konstante, so spricht man von *exponentiellem Wachstum*. Bei gegebenem Anfangsbestand $x(0) = x_0$ ist dann

$$x(m \Delta t) = x_0 (1 + p)^m \text{ für alle } m \in \mathbb{N}_0 \quad (2.21)$$

bzw.

$$x(t) = x_0 e^{\alpha t} \text{ für alle } t \geq 0 . \quad (2.22)$$

Für die speziellen Zeitpunkte $t = m \Delta t$ mit $m \in \mathbb{N}_0$ ergibt sich die Übereinstimmung der beiden Lösungen, falls

$$1 + p(\Delta t) = e^{\alpha \Delta t} .$$

Aufgabe 2.10

Lösen Sie die in den oben aufgeführten Beispielen (mit $\Delta t = 1$ Jahr) enthaltenen Problemstellungen, in denen offenbar exponentielles Wachstum unterstellt wird.

2.3.3 Stochastische Modelle exponentiellen Wachstums

Der Bestand einer Population werde als Anzahl der ihr angehörenden Individuen beschrieben. Sei $N(t)$ die Individuenzahl zum Zeitpunkt t . Das Wachstum der Population lässt sich, wie beim reinen Zerfalls- oder Todesprozess, wieder stochastisch modellieren, ohne dass hier allerdings die Beziehung zu den deterministischen Modellen noch so eindeutig wäre wie dort: Für ein stochastisches Modell macht es einen Unterschied, ob eine relative Wachstumsrate von etwa $p = 0.012$ pro Jahr nur durch Geburten oder dadurch zustande kommt, dass sich eine höhere Anzahl von Geburten mit einer geringeren Anzahl von Todesfällen überlagert wie etwa im oben angeführten Beispiel (a). Im ersteren Fall ist die Population nach einem Jahr mit Sicherheit nicht kleiner geworden, im letzteren kann sie dagegen sogar ausgestorben sein. Auch die Anzahl der Nachkommen je Geburt beeinflusst das Modell.

NÖBAUER/TIMISCHL 1979 (S. 58 - 64) untersuchen Modelle in kontinuierlicher Zeit, bei denen jedes Individuum in einem kleinen Zeitintervall $[t, t + \Delta t]$ unabhängig von den anderen mit Wahrscheinlichkeit

$$\begin{array}{ll} \alpha \Delta t + o(\Delta t) & \text{sich verdoppelt,} \\ \beta \Delta t + o(\Delta t) & \text{stirbt,} \\ 1 - (\alpha + \beta) \Delta t + o(\Delta t) & \text{erhalten bleibt ohne Verdopplung.} \end{array}$$

Diese Annahme führt für die Wahrscheinlichkeiten $w_j(t)$, dass die Population zum Zeitpunkt t aus j Individuen besteht, auf ein (unendlichdimensionales) System linearer Differentialgleichungen

$$\dot{w}_j = -(\alpha + \beta)jw_j + \alpha(j-1)w_{j-1} + \beta(j+1)w_{j+1} \quad (j \in \mathbb{N}_0).$$

Mit der Methode der erzeugenden Funktion lässt sich zeigen, dass auch hier der Mittelwert

$$x(t) = \mu_{w(t)} = \sum_{j=0}^{\infty} j w_j(t) = n e^{(\alpha-\beta)t}$$

(mit der Anfangspopulation von n Individuen) dem Wachstumsgesetz des deterministischen Modells genügt, zu dem es dennoch einen sehr entscheidenden Unterschied gibt: Für die Wahrscheinlichkeit, dass die Population irgendwann ausstirbt, gilt

$$\lim_{t \rightarrow \infty} w_0(t) = \begin{cases} (\beta/\alpha)^n & , \text{ falls } \beta < \alpha \\ 1 & , \text{ falls } \beta \geq \alpha \end{cases}.$$

Differenzen ergeben sich insbesondere im Falle $\beta = \alpha$, in dem beim deterministischen Modell die Population die konstante Größe n hat, während sie im stochastischen Modell mit Sicherheit ausstirbt.

Es soll im folgenden ein stochastisches Modell in diskreter Zeit $t \in \mathbb{N}_0$ untersucht werden, das auf der folgenden Grundannahme beruht:

Jedes Mitglied einer Population möge eine jede Periode $[t, t+1]$ (t ganzzahlig) unabhängig von den anderen mit der Wahrscheinlichkeit q überleben und sich in diesem Falle am Ende der Periode verdoppeln.

Man sollte sich vielleicht eine Bakterienkultur vorstellen, deren Mitglieder sich durch Zellteilung vermehren, die in festen Zeitschritten erfolgt. Intuitiv wird man erwarten, dass sich die Population im Mittel in jedem Zeitschritt um den Faktor $2q$ vergrößert. Das wäre allerdings noch nachzuweisen.

Betrachtet man die Population zu den Zeitpunkten $t = 0, 1, 2, \dots$ direkt nach der Verdopplung der Überlebenden, so besteht sie aus einer geraden Anzahl von Individuen, weshalb es genügt, solche Anzahlen zu betrachten. Es sei, mit einer beliebigen natürlichen Zahl n , $N(0) = 2n$ die Größe der Anfangspopulation. Besteht die Population zu irgendeinem Zeitpunkt aus $2k$ Individuen, so ist die Anzahl, die das nächste Zeitintervall überlebt, binomial verteilt, wie oben für den exponentiellen Zerfallsprozess ausgeführt. Anschließend verdoppeln sich die Überlebenden. Daher ist

$$p_{jk} := P(N(t+1) = 2j \mid N(t) = 2k) = \binom{2k}{j} q^j (1-q)^{2k-j}$$

die Übergangswahrscheinlichkeit dafür, dass in einem Zeitschritt aus $2k$ vorhandenen Individuen $2j$ sich entwickeln. Die Wahrscheinlichkeiten $w_j(t)$, dass zum Zeitpunkt t die Population genau $2j$ Mitglieder hat, ergeben sich dann aus

$$w_j(t+1) = \sum_{k=0}^{\infty} p_{jk} w_k(t)$$

zusammen mit der Anfangsbedingung

$$w_n(0) = 1, w_j(t) = 0 \text{ für } j \neq n$$

auf eindeutige Weise. Die unendliche Summation stellt dabei kein Problem dar, da zu jedem Zeitpunkt t nur endlich viele der $w_j(t)$ von 0 verschieden sind, denn da sich in jedem Zeitschritt die maximal mögliche Populationsgröße verdoppelt, ist $w_j(t) = 0$ für $j > n 2^t$. In Tabelle 2.2 ist für $n = 1$ und $q = 0.6$ die Entwicklung der Wahrscheinlichkeiten $w_j(t)$ für die ersten drei Zeitschritte erfasst. Zu erkennen ist das Anwachsen von $w_0(t)$ mit wachsendem t , die Verdopplung der Anzahl der von 0 verschiedenen w_j mit jedem Zeitschritt und dass die Verteilung $w_1(t), \dots, w_{2^t}(t)$ offenbar ein einziges Maximum besitzt, dass sich mit wachsendem t nach hinten verschiebt.

Bestand	$t = 0$	$t = 1$	$t = 2$	$t = 3$
0	0	0.16	0.24602	0.29988
2	1	0.48	0.28570	0.18774
4	0	0.36	0.29722	0.22469
6	0	0	0.12442	0.14289
8	0	0	0.04666	0.08805
10	0	0	0	0.03622
12	0	0	0	0.01556
14	0	0	0	0.00418
16	0	0	0	0.00078

Tabelle 2.2: Entwicklung der $w_j(t)$ für $q = 0.6$ und $n = 1$ **Aufgabe 2.11**

Führen Sie die Berechnung der $w_j(t)$ für einige weitere Zeitschritte durch und überprüfen sie, ob die festgestellten Beobachtungen weiterhin gültig bleiben.

Um der Entwicklung der Wahrscheinlichkeitsverteilungen auch analytisch auf die Spur zu kommen, führe ich wieder die erzeugenden Funktionen

$$G_t(z) := \sum_{j=0}^{\infty} w_j(t) z^j$$

ein, die wegen

$$G_0(z) = z^n$$

und der Verdopplung des Grades in jedem Zeitschritt für jedes $t \in \mathbb{N}_0$ ein Polynom vom Grade $n2^t$ ist. Es ist dann

$$\begin{aligned} G_{t+1}(z) &= \sum_{j=0}^{\infty} w_j(t+1) z^j = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \binom{2k}{j} q^j (1-q)^{2k-j} w_k(t) z^j = \\ &= \sum_{k=0}^{\infty} w_k(t) \sum_{j=0}^{\infty} \binom{2k}{j} q^j (1-q)^{2k-j} z^j = \\ &= \sum_{k=0}^{\infty} w_k(t) (1-q+qz)^{2k} = G_t((1-q+qz)^2), \end{aligned}$$

woraus sich die $G_t(z)$ iterativ berechnen lassen: Mit

$$f(z) := (1-q+qz)^2$$

und der t -ten Iterierten f^t von f (der t -maligen Hintereinanderausführung von f) ist

$$G_t(z) = G_{t-1}(f(z)) = G_{t-2}(f^2(z)) = \dots = G_0(f^t(z)) = f^t(z)^n.$$

Leider lässt sich für $f^t(z)$ keine geschlossene Formel angeben. Trotzdem lassen sich einige relevante Informationen aus den erzeugenden Funktionen herausholen. So gilt etwa für den Mittelwert $x(t) := \mu_{w(t)}$:

$$x(t+1) = G'_{t+1}(1) = 2q G'_t(1) = 2q x(t),$$

d. h. der Mittelwert vergrößert sich in jedem Zeitschritt um den Faktor $2q$, woraus sich wegen $x(0) = n$

$$x(t) = n (2q)^t$$

ergibt. Auch die Frage nach der Wahrscheinlichkeit des langfristigen Aussterbens der Population lässt sich beantworten. Denn wenn auch für

$$w_0(t) = G_t(0) = f^t(0)^n$$

keine geschlossene Formel angegeben werden kann, so lässt sich doch die Konvergenz dieser Folge für $t \rightarrow \infty$ zeigen:

Aufgabe 2.12

Zeigen Sie, dass die Folge $z_t = f^t(0)$, die definiert ist durch

$$z_0 = 0, z_{t+1} = (1 - q + q z_t)^2 \quad (t \geq 0),$$

streng monoton wächst, und

$$\lim_{t \rightarrow \infty} z_t = \min \left\{ 1, \left(\frac{1-q}{q} \right)^2 \right\}$$

Daraus folgt

$$\lim_{t \rightarrow \infty} w_0(t) = \min \left\{ 1, \left(\frac{1-q}{q} \right)^{2n} \right\} = \begin{cases} \left(\frac{1-q}{q} \right)^{2n} & , \text{ falls } q > \frac{1}{2} \\ 1 & , \text{ falls } q \leq \frac{1}{2} \end{cases}$$

Auch im Falle $q = \frac{1}{2}$, in dem der Mittelwert konstant gleich n ist, stirbt also die Population langfristig mit Sicherheit aus.

Anders als beim exponentiellen Zerfall liefern hier die stochastischen Modelle qualitativ andere Ergebnisse als die deterministischen.

2.3.4 Die logistische Differentialgleichung

Biologische und soziale Systeme können in der Regel nicht unbeschränkt wachsen, wie es mit einer konstanten relativen Wachstumsrate unterstellt wird. Um dem Rechnung zu tragen, kann man die relative bzw. exponentielle Wachstumsrate als *streng monoton fallende Funktion des Bestandes* ansetzen, die jenseits einer bestimmten Kapazitätsgrenze negativ wird. Ob das durch eine sinkende Geburten- oder eine wachsende Sterberate verursacht wird, spielt zunächst keine Rolle, jedenfalls bei deterministischen Modellen. Der belgische Mathematiker Verhulst schlug 1838 vor, die exponentielle Sterberate als lineare Funktion des Bestands x anzusetzen und nannte die entstehende Differentialgleichung der Form

$$\dot{x} = a x - b x^2 = a \left(1 - \frac{x}{K} \right) x \quad (2.23)$$

logistisch. Dabei sind a und b positive Konstanten, $K = a/b$. Verhulsts Arbeiten dazu blieben zunächst weitgehend unbeachtet und wurden erst 1921 in einer Arbeit von Pearl zur "Biologie des Todes" wieder aufgegriffen.⁶ Die logistische oder Pearl-Verhulst'sche Differentialgleichung (2.23) dürfte inzwischen die am häufigsten verwendete Gleichung der mathematische Ökologie sein. Die exponentielle Zerfallsrate

$$\alpha(x) = a - b x = a \left(1 - \frac{x}{K} \right)$$

als (affin) lineare Funktion von x anzusetzen, ist die *einfachste* Annahme, die sich in diesem Zusammenhang machen lässt. Ob sie auch *angemessen* ist, muss im konkreten Einzelfall allerdings noch untersucht werden.

⁶s. HEUSER 1989, 22/23. Die zwanziger Jahre des 20. Jahrhunderts gelten als "Goldenes Zeitalter der mathematischen Biologie", in dem manches grundlegende Differentialgleichungs-Modell aus diesem Bereich entstand.

- (c) **Innerspezifische Konkurrenz:** Unter *intra-* oder *innerspezifischer Konkurrenz* wird in der Populationsökologie die Konkurrenz unter den Individuen einer Art verstanden, die dadurch zustande kommt, dass sie sich knappe Ressourcen wie Nahrung oder (bei Pflanzen) Licht gegenseitig streitig machen. Dabei erhöht sich mit wachsender Populationsdichte der Konkurrenzdruck. Bezeichnet x die Populationsdichte, so ist (jedenfalls bei frei herumlaufenden Tieren) x^2 proportional zur Häufigkeit des Aufeinandertreffens zweier Individuen, die sich dabei evtl. gegenseitig Nahrung wegnehmen. Nimmt man an, dass das zu einer entsprechenden Erhöhung der Sterberate führt, so ergibt sich bei ansonsten konstanter Geburten- und "natürlicher" Sterberate die Differentialgleichung

$$\dot{x} = a x - b x^2 .$$

- (d) **Ausbreitung von Gerüchten:** In einer menschlichen Population von N Personen verbreite sich ein Gerücht oder auch eine Information dadurch, dass es weitererzählt wird. Sei x die Anzahl der Personen, die das Gerücht bereits kennen, $N - x$ also die Anzahl von Personen, die es noch nicht kennen. Geht man davon aus, dass jede Person jede andere pro Zeiteinheit mit gleicher Wahrscheinlichkeit trifft, dass diese Treffen voneinander unabhängig sind und dass bei jedem Treffen mit gleicher Wahrscheinlichkeit das Gerücht zur Sprache kommt, so ist seine Ausbreitungsgeschwindigkeit zum Produkt $x(N - x)$ proportional, weshalb mit einem $c > 0$

$$\dot{x} = c x (N - x) = a x \left(1 - \frac{x}{N}\right) \text{ mit } a = c N .$$

Beachten Sie die vielen einschränkenden Voraussetzungen, insbesondere die, dass jede Person zu jeder anderen gleichermaßen Kontakt hat.

- (e) **Ein Epidemie-Modell:** Ersetzt man im letzten Beispiel "Gerücht" durch "Virusinfektion", die sich dadurch ausbreitet, dass ein Infizierter mit einem Gesunden in Kontakt (einer bestimmten Intensität) kommt, so erhält man ein Epidemie-Modell. Die Population, um deren Wachstum es hier geht, ist die der Erkrankten oder auch, wenn man so will, die der Viren, gemessen in der Anzahl der von ihnen befallenen Personen. Unter der weiteren Annahme, dass Erkrankte mit einer konstanten exponentiellen Rate d gesunden, dann aber sofort wieder infizierbar sind, ergibt sich für die Anzahl x der infizierten Personen die Differentialgleichung

$$\dot{x} = c x (N - x) - d x = a x \left(1 - \frac{x}{K}\right) \text{ mit } a = c N - d \text{ und } K = N - \frac{d}{c} ,$$

was (nur) im Falle $d < c N$ wieder auf eine logistische Differentialgleichung (2.23) führt.

Zwei spezielle **Lösungen** von (2.23) sind die konstanten Funktionen

$$x(t) = 0 \text{ bzw. } x(t) = K \text{ für alle } t \in \mathbb{R} .$$

Solche Lösungen einer Differentialgleichung, in denen sich der Zustand nicht ändert, heißen auch *Gleichgewichtslösungen*. Für $x(t) \neq 0$ und $x(t) \neq K$ ist dagegen (2.23) äquivalent zu

$$a = \frac{\dot{x}}{x \left(1 - \frac{x}{K}\right)} = \left(\frac{1}{x} + \frac{1}{K - x}\right) \dot{x} = \frac{d}{dt} (\log|x| - \log|K - x|) ,$$

und das wiederum (nach Integration) zu

$$a t + \text{const} = \log|x| - \log|K - x| = \log \left| \frac{x}{K - x} \right| .$$

Bildet man links und rechts die Exponentialfunktion, so ergibt sich wiederum äquivalent, mit einer reellen Konstanten $C \neq 0$, die Gleichung

$$C e^{at} = \frac{x}{K-x}$$

die sich nunmehr nach $x = x(t)$ auflösen lässt zu

$$x(t) = \frac{K C e^{at}}{1 + C e^{at}}.$$

Für die spezielle Lösung von (2.23), die der Anfangsbedingung $x(0) = x_0$ mit einem $x_0 \geq 0$ genügt, ist

$$x_0 = \frac{K C}{1 + C} \text{ und daher } C = \frac{x_0}{K - x_0}$$

zu wählen, woraus sich

$$x(t) = \frac{K x_0 e^{at}}{K - x_0 + x_0 e^{at}} = \frac{K x_0}{x_0 + (K - x_0) e^{-at}} \quad (2.24)$$

als die eindeutige Lösung von (2.23) mit $x(0) = x_0$ ergibt, was im Übrigen in dieser Form auch in den Fällen $x_0 = 0$ und $x_0 = K$ richtig ist, in denen sich die entsprechenden Gleichgewichtslösungen ergeben.

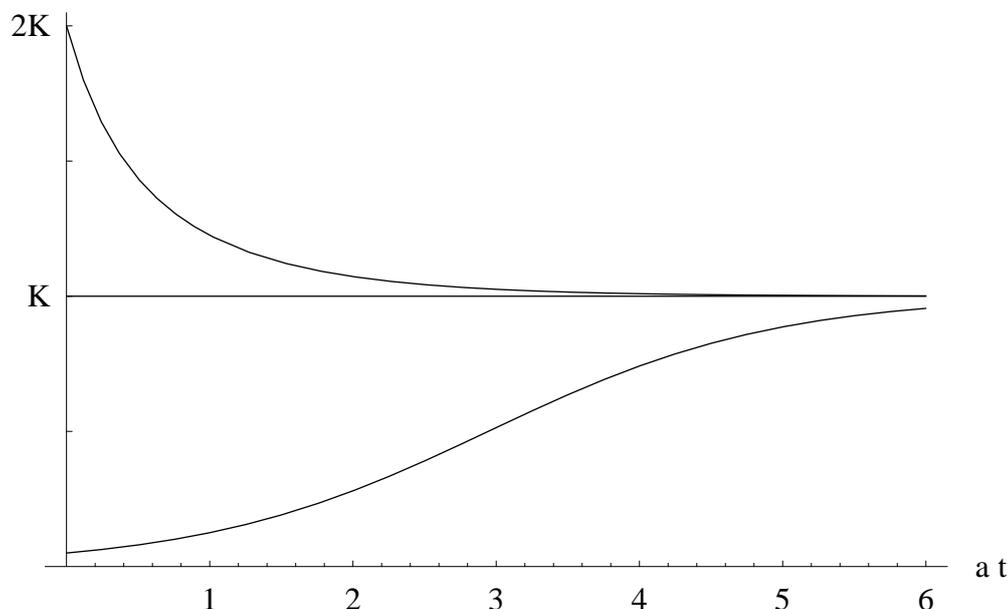


Abbildung 2.7: Lösungen (2.24) für $x_0 = 0.05K$, $x_0 = K$ und $x_0 = 2K$

Für $0 < x_0 < K$ hat $x(t)$ einen "S-förmigen" Verlauf. Dieser Fall liegt vor, wenn eine Population sich in einer Umwelt befindet, in der sie noch wachsen kann. Während die exponentielle Wachstumsrate mit wachsendem x linear fällt, erreicht die absolute Wachstumsrate für $x = K/2$ ihr Maximum, weshalb $x(t)$ dort einen Wendepunkt hat. Mit weiter wachsendem Bestand nähert sich dieser dann asymptotisch der Kapazität K .

Im Falle $x_0 > K$ sinkt dagegen der Bestand mit ständig fallender Geschwindigkeit und nähert sich von oben der Kapazität K . Dieser Fall entspricht der Situation, dass sich eine Population einer plötzlich gefallenen Kapazität ihres Habitats gegenüber sieht. Interessant ist, dass dieser Zweig der Lösungen (2.24) nicht für beliebig kleine t definiert ist: Die Lösung hat bei

$$t^- = \frac{1}{a} \log \frac{x_0 - K}{x_0}$$

einen Pol, sie ist daher nur für $t \in (t^-, \infty)$ definiert. Das lässt sich so interpretieren, dass vor dem Zeitpunkt t^- das Wachstum nicht der Gleichung (2.23) gefolgt sein kann (etwa weil davor das Habitat eine höhere Kapazität hatte).

Allen Lösungen (2.24) mit $x_0 > 0$ gemeinsam ist, dass sie für alle Zeiten $t \geq 0$ definiert sind und

$$\lim_{t \rightarrow \infty} x(t) = K,$$

wobei dieser Grenzwert streng monoton (fallend oder wachsend) erreicht wird.

Zusammenfassung : Eine Population, deren Wachstum der logistischen Differentialgleichung (2.23) genügt, nimmt langfristig einen konstanten Wert in Höhe der Kapazität K des Habitats an.

Aufgabe 2.13

Für das Epidemiemodell (c) bedeutet das, dass langfristig ein konstanter Anteil der Bevölkerung erkrankt ist, wobei die betroffenen Individuen wechseln. Untersuchen Sie, wovon (im Modell) dieser Anteil abhängt und durch welche Maßnahmen er sich verringern lässt.

2.3.5 Logistisches Wachstum in diskreter Zeit

Das einfachste numerische Verfahren zur Lösung gewöhnlicher Differentialgleichungen (genauer: Anfangswertaufgaben), ist das *explizite Eulerverfahren*, welches für die logistische Differentialgleichung (2.23) die Iterationsvorschrift

$$x(t + \Delta t) = x(t) + \Delta t a x(t) \left(1 - \frac{x(t)}{K}\right) \quad (2.25)$$

liefert mit einem fest gewählten $\Delta t > 0$. Das Verfahren besteht darin, dass die Geschwindigkeit des Wachstums (die absolute Wachstumsrate) während des Zeitintervalls $[t, t + \Delta t]$ konstant gleich dem Wert gesetzt wird, der sich zum Zeitpunkt t aus dem Bestand $x(t)$ errechnen lässt. Geometrisch bedeutet das, dass die Lösungskurve durch ein Geradenstück ersetzt wird, nämlich ihre Tangente im aktuellen Punkt, fortgeschrieben für ein Zeitintervall der Länge Δt . Die S-förmige Lösungskurve zwischen 0 und K (s. Abbildung 2.7) wird damit in ihrem unteren Teil unter-, in ihrem oberen Teil überschätzt.

Simulationsprogramme wie STELLA arbeiten mit diesem Verfahren, ohne dass der Benutzer das explizit sagen muss. Er muss selbst nur dafür sorgen, dass Δt "hinreichend" klein gewählt wird, um die Lösung der Differentialgleichung gut zu approximieren. In dieser Sicht ist (2.23) das "eigentliche" Modell und (2.25) seine mehr oder weniger mit Fehlern behaftete Approximation.

Man kann (2.25) aber auch als ein Modell "eigenen Rechts" ansehen, insbesondere nämlich bei Vermehrungsvorgängen, in denen es problematisch erscheint, die Zeit kontinuierlich zu wählen, wie das bei *Population mit nicht überlappenden Generationen* der Fall ist.

(f) Viele Arthropoden (Gliederfüßer, zu denen u. a. Spinnen und Insekten gehören), bilden eine kurzlebige Erwachsenenpopulation in der wärmeren Jahreszeit, wogegen nur die Eier überwintern. Zur Beschreibung des Wachstums solcher Populationen wäre eine Differentialgleichung etwa der Form (2.25) unangemessen, da sich damit die Übergänge zwischen den verschiedenen Stadien nicht beschreiben lassen. Wählt man dagegen einen festen Zeitpunkt im Jahr (z. B. Sommeranfang), so kann man versuchen, den aktuellen Bestand als Funktion des Bestandes im Vorjahr zu beschreiben. Damit liegt aber Δt fest (1 Jahr) und kann nicht beliebig gewählt und vor allem nicht beliebig verkleinert werden.

Bei fester Wahl von Δt lässt sich immer, durch entsprechende Festlegung der Zeiteinheit, $\Delta t = 1$ wählen. Wählt man außerdem noch K als Maßeinheit des Bestandes oder betrachtet man, was auf dasselbe hinausläuft,

$$y = \frac{x}{K},$$

so erhält (2.25) die Form

$$y(t+1) = y(t) + a y(t) (1 - y(t)) . \quad (2.26)$$

Wie schon die logistische Differentialgleichung hat auch (2.26) die beiden Gleichgewichtslösungen

$$y(t) = 0 \text{ bzw. } y(t) = 1 \text{ für alle } t \in \mathbb{N}_0 ,$$

und auch hier wächst der Bestand bei kleinen, aber positiven Anfangswerten zunächst einmal. Ob und in welcher Weise er für $t \rightarrow \infty$ gegen 1 konvergiert, hängt allerdings vom Parameter a ab: Nur für $a \leq 1$ ergibt sich wie bei der logistischen Differentialgleichung eine monotone Konvergenz, für größere Werte von a zeigt sich dagegen ein qualitativ anderes Verhalten.

Aufgabe 2.14

Führen Sie die Iteration (2.26) mit $y(0) = 0.01$ für verschiedene Werte von a durch und skizzieren Sie jeweils die Entwicklung des Bestandes. Beachten Sie besonders die qualitativen Veränderungen im Übergang von $a < 1$ zu $a > 1$, von $a < 2$ zu $a > 2$ und für wachsendes a zwischen 2 und 3. **Achtung:** Für $a > 3$ ergeben sich nach wenigen Schritten *negative* Bestände.

Es handelt sich bei (2.26) um das in der einschlägigen Literatur inzwischen Standard gewordene Beispiel für deterministisches Chaos in einer Dimension (vgl. PEITGEN / JÜRGENS / SAUPE 1994, 37 ff.), dessen komplizierte Eigenschaften dort ausführlich diskutiert werden.

Für eine Gleichung, die das Wachstum einer Bestandsgröße beschreiben soll, hat (2.26) den Nachteil, dass sich negative Werte ergeben können. Bei kleinen Anfangswerten treten sie zwar erst für $a > 3$ auf, doch offenbar kann die Gleichung für große Bestände nicht stimmen, unabhängig davon, wie groß a gewählt wurde. Die rechte Seite ist vielmehr so zu wählen, dass sie für beliebige $y \geq 0$ einen nichtnegativen Wert ergibt. MAY (1980, 11) macht hierfür unter anderem den Vorschlag

$$F(y) := y e^{a(1-y)} .$$

Der zugehörige Iterationsprozess

$$y(t+1) = F(y(t)) = y(t) e^{a(1-y(t))} \quad (2.27)$$

lässt sich ebenfalls als aus einem numerischen Verfahren für die logistische Differentialgleichung (2.23) entstanden verstehen. Formt man diese nämlich mit $u = \log x$ äquivalent um zu

$$\dot{u} = \frac{\dot{x}}{x} = a \left(1 - \frac{x}{K}\right) = a \left(1 - \frac{e^u}{K}\right)$$

und wendet darauf das Eulerverfahren an, so ergibt sich

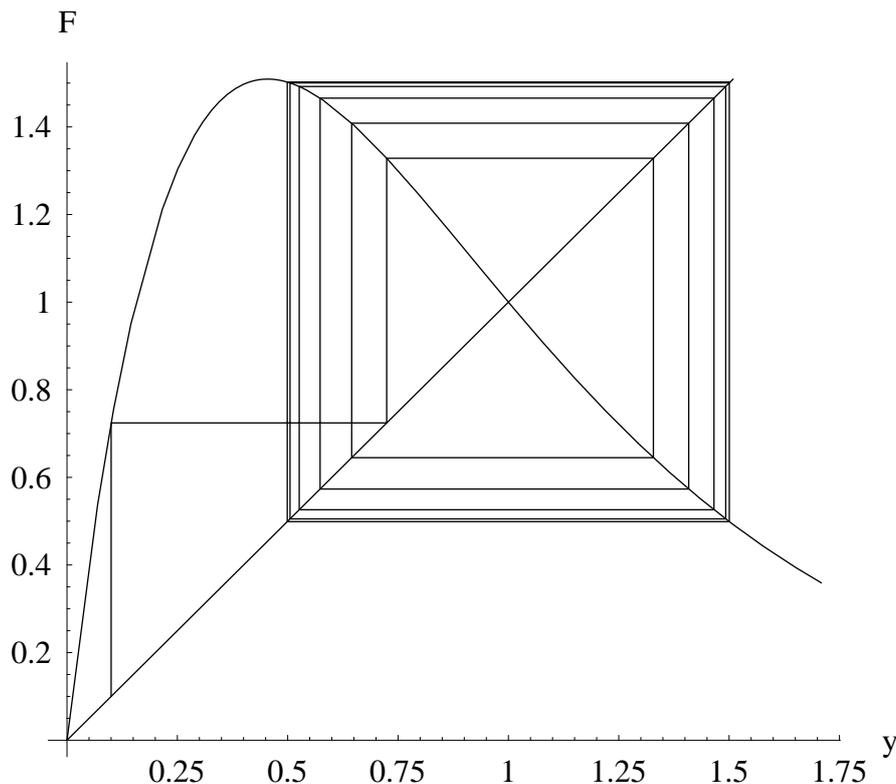
$$u(t + \Delta t) = u(t) + \Delta t a \left(1 - \frac{e^u}{K}\right)$$

und daher

$$x(t + \Delta t) = x(t) e^{\Delta t a \left(1 - \frac{x(t)}{K}\right)} ,$$

woraus sich mit der speziellen Wahl $\Delta t = 1$ und der Einführung von $y = x/K$ wie oben die Gleichung (2.27) ergibt.

Die Funktion F , mit der sich der Bestand des Folgejahres aus dem aktuellen Bestand errechnet, wächst zunächst von $F(0) = 0$ auf ihren maximalen Wert $F(1/a) = (1/a) e^{a-1}$ an und nähert sich anschließend asymptotisch dem Wert 0. Bei $y = 2/a$ hat sie einen Wendepunkt. Die Iteration (2.27) lässt sich geometrisch veranschaulichen, indem man zunächst den Graphen der Funktion F und die Winkelhalbierende im positiven Quadranten einzeichnet. Der jeweils aktuelle Bestand y wird als Punkt (y, y) auf der Winkelhalbierenden markiert. Den nächsten Punkt $(F(y), F(y))$ konstruiert man dadurch, dass man zum Punkt $(y, F(y))$ auf dem Graphen von F entlang einer vertikalen und von dort aus nach $(F(y), F(y))$ entlang einer horizontalen Strecke sich bewegt. So fortfahrend entsteht ein "Netz", das einen gewissen Eindruck vom Verhalten der Iterationsfolge vermittelt (vgl. Abbildung 2.8).

Abbildung 2.8: Iteration (2.27) mit $a = 2.2$ und Anfangswert $y(0) = 0.1$ **Aufgabe 2.15**

Machen Sie auf diesem Wege anschaulich klar, dass für $a \leq 1$ und Anfangswerte $y(0) < 1$ die durch (2.27) definierten Folgen für $t \rightarrow \infty$ monoton gegen den Grenzwert 1 streben.

Ob und in welcher Weise die Lösungen von (2.27) gegen den der "Kapazität" des Systems entsprechenden Gleichgewichtspunkt 1 konvergieren, hängt entscheidend von

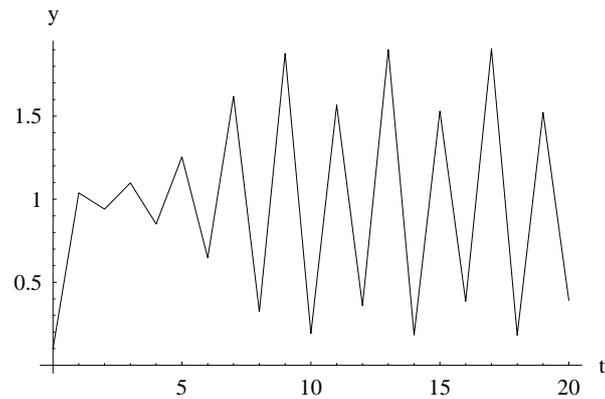
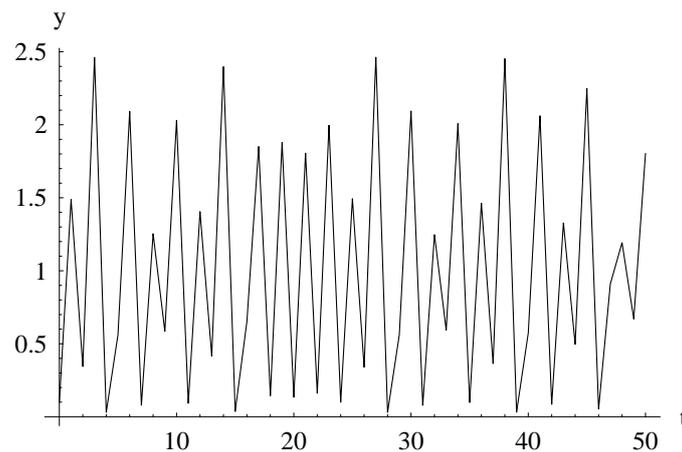
$$F'(1) = 1 - a$$

ab. Für $a > 1$ wird dieser Wert negativ, weshalb die Iterationsfolge in der Nähe von 1 alterniert. Solange $|F'(1)| \leq 1$, was für $a \leq 2$ der Fall ist, liegt Konvergenz vor, danach gehen die Folgen langfristig in einen Zweierzyklus über, wie in Abbildung 2.8 zu sehen. Mit weiter wachsendem a verdoppeln sich die Perioden der angestrebten Grenzyklen: 2, 4, 8, ... (vgl. Abbildung 2.9, in der ein Viererzyklus zu erkennen ist). Von einem bestimmten Wert für a an wird das Verhalten der Lösungen schließlich "irregulär", es sind keine regelmäßigen Muster mehr zu sehen (vgl. Abbildung 2.10). Unter gewissen hier weder genannten noch überprüften Voraussetzungen spricht man in solchen Fällen von *deterministischem Chaos*. Es ist mit bloßem Auge von Zufallsprozessen praktisch nicht zu unterscheiden.

Aufgabe 2.16

Führen Sie die hier angegebenen und weitere Iterationen auf dem Computer durch und achten Sie dabei besonders auf qualitative Veränderung im Verhalten der Lösungen bei wachsendem Parameterwert a .

Anders als bei der logistischen Differentialgleichung (2.23) lässt sich in der Iteration (2.27), wie übrigens schon in (2.26) der Parameter a und damit die Schnelligkeit des Wachstums nicht beliebig vergrößern, ohne das qualitative Verhalten der Lösungen zu verändern. Der entscheidende Grund dafür liegt darin,

Abbildung 2.9: Iteration (2.27) mit $a = 2.6$ und Anfangswert $y(0) = 0.1$ Abbildung 2.10: Iteration (2.27) mit $a = 3.0$ und Anfangswert $y(0) = 0.1$

dass die Funktion $F(y)$ auf der rechten Seite nicht monoton ist. Das genau aber wäre sicher zu stellen, um in diskreter Zeit dasselbe Lösungsverhalten zu garantieren wie bei der logistischen Differentialgleichung. Ein Beispiel für einen Iterationsprozess, der das leistet, ist

$$y(t+1) = \frac{(1+a)y(t)}{1+ay(t)}, \quad (2.28)$$

der sich aus (2.27) dadurch ergibt, dass man in

$$F(y) = y e^{a(1-y)} = y \frac{e^a}{e^{ay}}$$

Zähler und Nenner des Bruches jeweils durch ihre nach dem ersten Glied abgebrochene Taylor-Entwicklung ersetzt, also

$$e^a \approx 1+a, \quad e^{ay} \approx 1+ay$$

approximiert.

Aufgabe 2.17

Zeigen Sie mit Hilfe der oben an (2.27) durchgeführten geometrischen Methode, dass für jeden Anfangswert $y(0) = y_0 > 0$ die durch (2.28) definierte Folge $y(t)$ für $t \rightarrow \infty$ gegen 1 konvergiert, und

zwar monoton wachsend oder fallend, je nachdem ob $y_0 < 1$ oder $y_0 > 1$. Was verändert sich, wenn a wächst?

Hinsichtlich des *Modellierungs-Ansatzes* besteht ein wesentlicher Unterschied zwischen (2.27) und (2.28) darin, dass bei ersterem ein "Systemzusammenbruch" durch zu hohe Bestände möglich ist, während bei letzterem ein höherer Bestand stets zu einem höheren Folgebestand führt.

Aufgabe 2.18

Welches der beiden Modelle (2.27) und (2.28) erscheint Ihnen für welches der Beispiele (e) bis (f) geeignet?

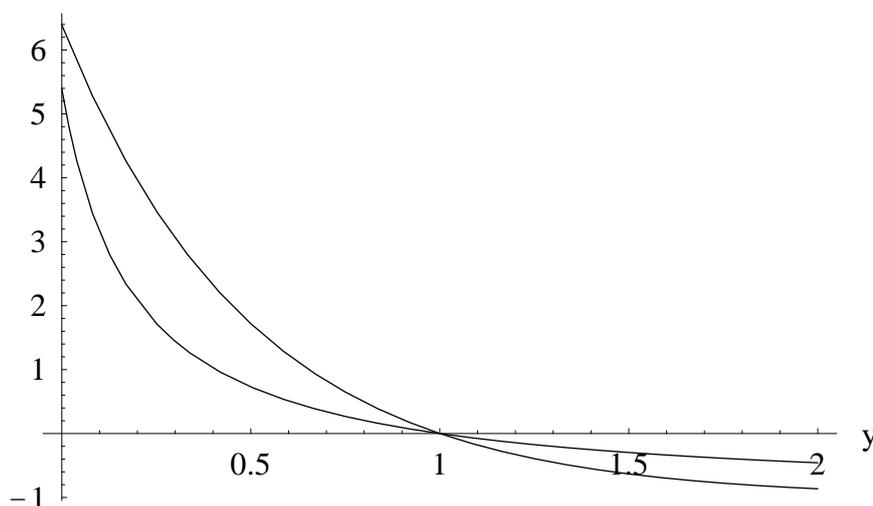


Abbildung 2.11: Relative Wachstumsraten von (2.27) und (2.28)

Eine Schwierigkeit bei der Wahl des "richtigen" Modells liegt darin, dass die relativen Wachstumsraten von (2.27) und (2.28) als Funktion des Bestandes y , also

$$e^{a(1-y)} - 1 \text{ bzw. } \frac{a(1-y)}{1+ay}$$

sich hinsichtlich des Funktionsverlaufs kaum unterscheiden (Abbildung 2.11): Beide sind streng monoton fallend und konvex und streben für $y \rightarrow \infty$ gegen -1 . Aus entsprechenden Daten allein lässt sich daher keine Entscheidung für den einen oder anderen Modellansatz treffen. Ganz anders sieht die Sache allerdings aus, wenn man statt der relativen die absoluten Wachstumsraten als Funktionen von y miteinander vergleicht.

2.3.6 Stochastische Modelle logistischen Wachstums

Soll eine Population durch die Anzahl der ihr angehörenden Individuen beschrieben werden, so ist man, wie schon beim exponentiellen Zerfall und Wachstum, auf stochastische Modelle verwiesen. Deren Formulierung hängt aber in hohem Maße davon ab, welche der Wahrscheinlichkeiten für mögliche Veränderungen (Geburten oder Todesfälle) von der vorhandenen Anzahl abhängen und welche konstant sind. Bereits die Frage, ob beliebig große Bestände möglich sind, wird davon tangiert.

Ausbreitung von Gerüchten und Krankheiten

In einer Bevölkerung, die aus n Personen besteht, verbreite sich eine Information oder auch eine Krankheit dadurch, dass Informierte / Infizierte mit Uninformierten / Uninfizierten in Kontakt treten (vgl. die obigen Beispiele (d) und (e)). Ich betrachte hier ein Modell in diskreter Zeit $t = 0, 1, 2, \dots$, in dem die Ausbreitung der Information bzw. Krankheit durch die Anzahl $N(t) = k$ der Informierten / Infizierten beschrieben wird, und mache in Analogie zu den bereits behandelten stochastischen Modellen die folgenden Annahmen:

- Jeder der $n - k$ Uninformierten / Uninfizierten treffe mit einer Wahrscheinlichkeit $p(k) = a \frac{k}{n}$ während des nächsten Zeitschritts auf einen Informierten / Infizierten und werde von ihm informiert / infiziert, unabhängig von allen anderen.
- Jeder der k Informierten / Infizierten verliere, ebenfalls unabhängig von allen anderen, während des nächsten Zeitschritts die Krankheit bzw. Information mit einer konstanten Wahrscheinlichkeit q

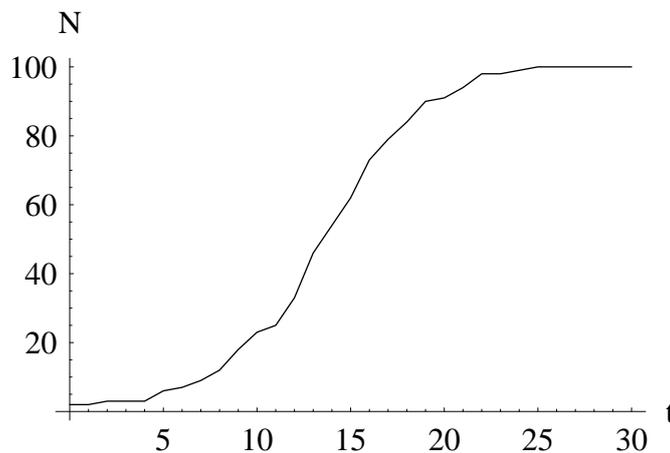


Abbildung 2.12: Simulationslauf mit $n = 100, a = 0.4, q = 0$ und $N(0) = 2$

Für jede der beiden Gruppen ist dann die Anzahl der Zustandswechsel binomial verteilt, und zwar ist

$$u_{jk} = \binom{n-k}{j} p(k)^j (1-p(k))^{n-k-j} \quad (j = 0, \dots, n-k)$$

die Wahrscheinlichkeit dafür, dass von den $n - k$ Uninformierten / Uninfizierten j informiert / infiziert werden, und

$$v_{ik} = \binom{k}{i} q^i (1-q)^{k-i} \quad (i = 0, \dots, k)$$

die Wahrscheinlichkeit dafür, dass von den k Informierten / Infizierten i die Information oder Krankheit wieder verlieren. Daraus lassen sich die Übergangswahrscheinlichkeiten

$$p_{jk} = P(N(t+1) = j \mid N(t) = k) = \sum_i v_{ik} u_{j-k+i,k}$$

leicht berechnen. Für Simulationen mit einem binomial verteilten Zufallsgenerator verwendet man die u_{jk}, v_{ik} lieber direkt.

Im Falle $q = 0$ (Verbreitung eines Gerüchts) kann $N(t)$ nicht fallen und muss daher mit Sicherheit irgendwann den Wert n erreichen. Abbildung 2.12 zeigt einen typischen Simulationslauf, der eine große Ähnlichkeit mit den Lösungen der logistischen Differentialgleichung erkennen lässt. Für $q > 0$ ist das

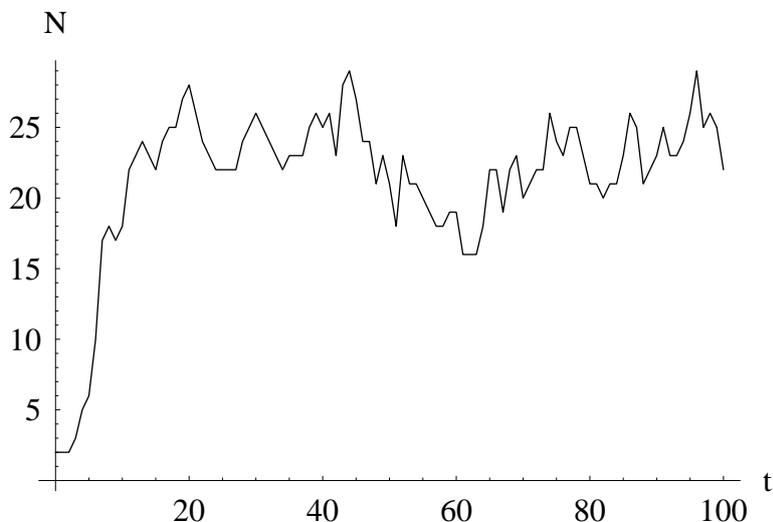


Abbildung 2.13: Simulationslauf mit $n = 30, a = 0.4, q = 0.1$ und $N(0) = 2$

nicht mehr der Fall, wie etwa Abbildung 2.13 zeigt. Die dort verwendeten Zahlenwerte könnte man als geeignet für die Simulation des Lausbefalls einer Kindergruppe halten: Sobald die Läuse entdeckt werden, werden sie entfernt, aber kein Kind ist immun gegen einen sofortigen neuen Befall.

Es macht den Eindruck, als würde $N(t)$ nach einer Wachstumsphase am Anfang um einen gewissen Mittelwert herumpendeln. Prädestiniert dafür wäre der Wert \bar{N} , bei dem die zu erwartende Anzahl der Neuinfizierten mit der zu erwartenden Anzahl der Geheilten übereinstimmt. Wie eine kurze Rechnung zeigt, ist das

$$\bar{N} = \left(1 - \frac{q}{a}\right) n,$$

bei den in Abbildung 2.13 gewählten Zahlenwerten also $\bar{N} = 22.5$, was mit der Grafik gut übereinstimmt. Doch der Eindruck trügt, denn es lässt sich zeigen, dass die Wahrscheinlichkeit $P(N(t) = 0)$ für $t \rightarrow \infty$ gegen 1 konvergiert, die Population der Infizierten also mit Sicherheit irgendwann ausstirbt.

Ein Zahlenbeispiel

Mitglieder einer vierköpfigen Familie seien von einer Krankheit des hier behandelten Typs befallen: Es gebe spontane Gesundungen, doch wer geheilt ist, kann sofort wieder angesteckt werden. Die Frage ist, was passiert, wenn die Familie gemeinsam, und nicht jeder Einzelne isoliert, in Quarantäne gesteckt wird. Zur Beantwortung wird das hier betrachtete Modell mit den (fiktiven) Zahlenwerten $n = 4, a = 4/5$

und $q = 1/4$ benutzt. Als Matrix der Übergangswahrscheinlichkeiten ergibt sich dann

$$P = (p_{jk})_{j,k=0,\dots,4} = \begin{pmatrix} 1 & \frac{16}{125} & \frac{9}{400} & \frac{1}{160} & \frac{1}{256} \\ 0 & \frac{12}{25} & \frac{33}{200} & \frac{21}{320} & \frac{3}{64} \\ 0 & \frac{39}{125} & \frac{157}{400} & \frac{81}{320} & \frac{27}{128} \\ 0 & \frac{37}{500} & \frac{33}{100} & \frac{27}{64} & \frac{27}{64} \\ 0 & \frac{3}{500} & \frac{9}{100} & \frac{81}{320} & \frac{81}{256} \end{pmatrix}$$

Für das Folgende entscheidend ist nun die Besonderheit des Zustands 0 (alle gesund), die nichts mit diesem speziellen Zahlenbeispiel zu tun hat:

$$p_{00} = 1, p_{j0} = 0 \text{ und } p_{0j} > 0 \text{ für } j = 1, \dots, 4.$$

Das heißt, dass der Zustand 0 von allen anderen Zuständen, vom Zustand 0 aus aber kein anderer Zustand erreicht werden kann. Es handelt sich hier um einen *absorbierenden Markov-Prozess* mit dem einen absorbierenden Zustand 0.

Bezeichne nun wieder $w_j(t) = P(N(t) = j)$ die Wahrscheinlichkeit dafür, dass zum Zeitpunkt t genau j Personen krank sind, und $w(t)$ den Vektor der $w_j(t)$. Es ist dann für alle $t \geq 0$

$$w(t+1) = P w(t),$$

und $w(0)$ ist einer der Einheitsvektoren, abhängig davon, wieviele Personen am Anfang krank sind. Daher ist

$$\begin{aligned} w_0(t+1) &= w_0(t) + \sum_{k=1}^4 p_{0k} w_k(t) \leq \\ &\geq w_0(t) + \min_j p_{0j} \sum_{k=1}^4 w_k(t) = w_0(t) + \frac{1}{256} (1 - w_0(t)) \end{aligned}$$

$w_0(t)$ ist daher monoton wachsend, andererseits durch 1 nach oben beschränkt und daher konvergent. Für den Grenzwert \bar{w}_0 gilt dann aber

$$\bar{w}_0 \geq \bar{w}_0 + \frac{1}{256} (1 - \bar{w}_0),$$

woraus $\bar{w} = 1$ folgt. Es gilt also unabhängig vom Anfangswert $w(0)$

$$\lim_{t \rightarrow \infty} w_0(t) = 1.$$

Insofern kann man die Familie gemeinsam in Quarantäne schicken, irgendwann sind alle wieder gesund. Es fragt sich nur, wie lange das dauert. Sei T_k die mittlere Anzahl von Zeitschritten, die nötig sind, um vom Zustand k zum Zustand 0 zu kommen. Offenbar ist $T_0 = 0$. Zwischen T_1, \dots, T_4 bestehen dagegen die Beziehungen

$$T_k = \sum_{j=1}^4 p_{jk} T_j + 1 \text{ für } k = 1, \dots, 4,$$

denn vom Zustand k gelangt man mit Wahrscheinlichkeit p_{jk} zum Zustand j und benötigt von dort aus im Mittel T_j Schritte zum Zustand 0. Das lineare Gleichungssystem für die T_k lässt sich eindeutig lösen und liefert im hier betrachteten Zahlenbeispiel näherungsweise

$$T_1 = 26.37, T_2 = 32.12, T_3 = 33.65, T_4 = 33.95.$$

Das ist natürlich erheblich länger als die mittlere Krankheitsdauer 4, die sich ergibt, wenn Wiederansteckung ausgeschlossen ist.

An dieser Stelle sollte noch einmal darauf hingewiesen werden, wie problematisch es ist, einen Krankheitsverlauf durch einen Markov-Prozess zu modellieren, der die Möglichkeit einer spontanen Genesung vorsieht (hier mit Wahrscheinlichkeit $q = 1/4$ an jedem Tag), aber keine Veränderung des Krankheitsbildes: Jemand ist hier entweder krank oder gesund, eine Abstufung gibt es nicht. Das hat zur Folge, dass jemand, der nach vier Tagen noch nicht gesund geworden ist, mit einer weiteren Dauer der Krankheit von weiteren vier Tagen rechnen muss, da die Vorgeschichte im Modell nicht erfasst ist, sondern einzig und allein die Tatsache zählt, immer noch krank zu sein. Vielleicht ist die einzige Krankheit, die damit angemessen beschrieben wird, der oben angeführte Lausbefall, bei dem die Genesung mit der zufälligen Entdeckung der Läuse einher geht.

Innerspezifische Konkurrenz

Innerspezifische Konkurrenz war in Beispiel (c) charakterisiert durch eine konstante relative Geburtenrate und eine mit dem Bestand wachsende relative Sterberate. Um diese Situation stochastisch zu modellieren und das Modell von möglichst wenig Parametern abhängig zu machen⁷, gehe ich so vor wie bereits in dem stochastischen Modell exponentiellen Wachstums, das in 2.3.3 genauer erörtert wurde:

Jedes Mitglied einer Population möge eine jede Periode $[t, t + 1]$ (t ganzzahlig) unabhängig von den anderen mit einer gewissen Wahrscheinlichkeit q überleben und sich in diesem Falle am Ende der Periode verdoppeln.

Im Unterschied zum exponentiellen Wachstum ist aber q keine Konstante, sondern eine monoton fallende Funktion des aktuellen Bestandes. Diese kann allerdings nicht linear sein, da q als Wahrscheinlichkeit zwischen 0 und 1 liegen muss. Der hier gewählte Ansatz

$$q = q(N) = b p^{N-1} \text{ mit den konstanten Wahrscheinlichkeiten } b, p \in (0, 1)$$

lässt sich so deuten: b ist die Wahrscheinlichkeit dafür, dass das Individuum ein Zeitintervall der Länge 1 überlebt, ohne Berücksichtigung der Effekte der innerspezifischen Konkurrenz. p ist die Wahrscheinlichkeit dafür, dass das Individuum nicht von einem ganz bestimmten anderen so empfindlich beeinträchtigt wird, dass es stirbt. p^{N-1} ist dann die Wahrscheinlichkeit dafür, dass das Individuum von keinem der $N - 1$ anderen auf diese Weise beeinträchtigt wird. Hierin steckt wieder die Annahme, dass die Beeinträchtigungen durch andere unabhängig voneinander erfolgen.

Auch bei diesem Modell ist die Anzahl der Individuen spätestens nach einem Zeitschritt gerade, weshalb von vornherein nur solche Anzahlen betrachtet werden. Sei $N(0) = 2n$ die Größe der Anfangspopulation. Wie in 2.3.3 ist dann

$$\begin{aligned} p_{jk} &= P(N(t+1) = 2j \mid N(t) = 2k) = \\ &= \binom{2k}{j} q^j (1-q)^{2k-j} = \binom{2k}{j} b^j p^{(2k-1)j} (1-b p^{2k-1})^{2k-j} \end{aligned}$$

⁷Der Grund liegt in der leichteren Analysierbarkeit des Modells. Das geht vielleicht auf Kosten einer größeren "Realitätsnähe"

die Übergangswahrscheinlichkeit dafür, dass in einem Zeitschritt aus $2k$ vorhandenen Individuen $2j$ sich entwickeln.

Anders als bei den zuvor betrachteten stochastischen Modellen logistischen Wachstums ist hier die Anzahl möglicher Individuen unbeschränkt, doch wächst die Wahrscheinlichkeit der Verringerung des Bestandes mit dem Bestand. Im Mittel ist bei einem Bestand $N(t) = 2k$ im nächsten Zeitschritt mit dem Bestand

$$\bar{N}(t+1) = 4 k b p^{2k-1}$$

zu rechnen, der dem vorhandenen Bestand $2k$ genau dann gleicht, wenn

$$p^{2k-1} = \frac{1}{2b},$$

woraus sich

$$2k = \bar{N} = 1 + \frac{\log 2b}{\log 1/p}$$

als Wert ergibt, um den herum die Population bei einem Simulationslauf vermutlich schwanken wird. Für den in Abbildung 2.14 dargestellten Lauf wurde

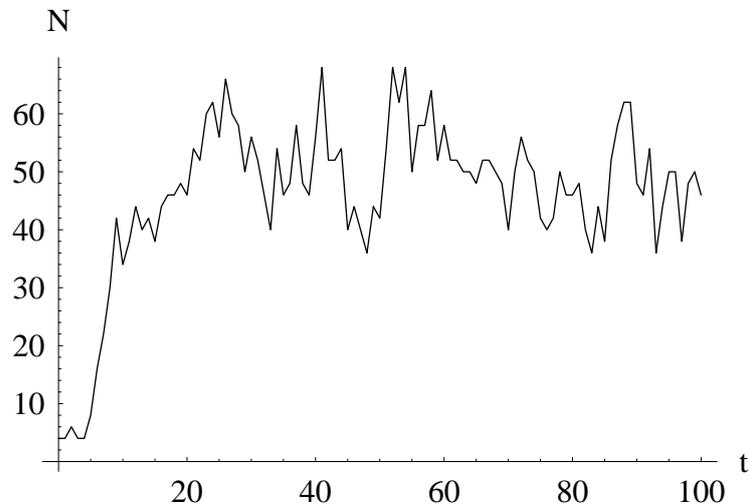


Abbildung 2.14: Simulationslauf mit $b = 0.818174$ und $p = 0.99$

$$p = 0.99 \text{ und } b = \frac{1}{2 p^{49}}$$

gewählt, so dass hier $\bar{N} = 50$ ist. Der Qualität nach ist kein großer Unterschied zum oben simulierten stochastischen Epidemiemodell zu erkennen.

Die Möglichkeit einer beliebig großen Individuenzahl ändert auch nichts daran, dass die Population langfristig mit Sicherheit ausstirbt: Es bezeichne wieder

$$w_k(t) = P(N(t) = 2k)$$

die Wahrscheinlichkeit dafür, dass die Population zum Zeitpunkt t aus genau $2k$ Individuen besteht. Die Wahrscheinlichkeit dafür, dass alle $2k$ vorhandenen Individuen in einem Zeitschritt sterben, ist

$$(1 - b p^{2k-1})^{2k},$$

sodass

$$w_0(t+1) = w_0(t) + \sum_{k=1}^{\infty} (1 - b p^{2k-1})^{2k} w_k(t). \quad (2.29)$$

Wegen

$$(1 - b p^{2k-1})^{2k} \geq 1 - 2k b p^{2k-1}$$

(Bernoulli'sche Ungleichung) und

$$\lim_{k \rightarrow \infty} k p^{2k-1} = 0$$

nähern sich die (positiven) Koeffizienten in der Summe der rechten Seite von (2.29) nicht beliebig der 0 (von einem bestimmten k an sind sie alle größer als 0.5), sie lassen sich daher nach unten durch ein $\varepsilon > 0$ abschätzen, sodass

$$w_0(t+1) \geq w_0(t) + \varepsilon \sum_{k=1}^{\infty} w_k(t) = w_0(t) + \varepsilon (1 - w_0(t)),$$

woraus wie oben

$$\lim_{t \rightarrow \infty} w_0(t) = 1$$

folgt.

2.3.7 Dichteabhängiges Wachstum: Modelle und Daten

Bei allen hier betrachteten Modellen handelt es sich um mathematische Konstruktionen, die einen "Idealfall" beschreiben. Das spricht keineswegs gegen sie, sondern liegt in der Natur der damit verbundenen Methodik, man denke an Galileis Einführung der gleichförmig beschleunigten Bewegung zur Beschreibung des freien Falls. Es bleibt dann allerdings in jedem konkreten Anwendungsfall die Frage, ob es sich hier tatsächlich um logistisches oder wenigstens dichteabhängiges Wachstum handelt.

Bevölkerungswachstum in den USA

Im Jahre 1940 versuchte Pearl, der Wiederentdecker der logistischen Differentialgleichung, zusammen mit zwei Koautoren⁸ das Bevölkerungswachstum in den USA durch die logistische Kurve, also die Lösungskurve der logistischen Differentialgleichung, zu beschreiben. In Tabelle 2.3 sind die Bevölkerungszahlen angegeben, die in den USA seit 1790 alle zehn Jahre erhoben werden, in der dritten Spalte steht der Zuwachs seit der jeweils letzten Volkszählung. Die Einwohnerzahlen sollen jetzt durch eine logistische Kurve der Form

$$x(t) = \frac{K C e^{a(t-1790)}}{1 + C e^{a(t-1790)}} \quad (2.30)$$

gefittet werden mit noch zu bestimmenden Parametern a , K und C . Da diese in (2.30) nichtlinear eingehen, ist es nicht so ohne Weiteres möglich, sie durch Lösung von

$$\sum_{i=0}^{15} \left(\frac{K C e^{a(t_i-1790)}}{1 + C e^{a(t_i-1790)}} - x_i \right)^2 = \min!$$

zu bestimmen, zumindest wäre für ein entsprechendes Verfahren der nichtlinearen Ausgleichsrechnung eine Näherungslösung erforderlich. Ich gehe daher anders vor und nutze aus, dass (2.30) eine Lösung der logistischen Differentialgleichung (2.23) ist. Dazu wird für jede Dekade $[t_{i-1}, t_i]$ die *mittlere exponentielle Wachstumsrate* α_i ermittelt, indem

$$x_i = x_{i-1} e^{\alpha_i(t_i-t_{i-1})} \text{ nach } \alpha_i = \frac{\log x_i - \log x_{i-1}}{t_i - t_{i-1}} = \frac{\log x_i - \log x_{i-1}}{10}$$

⁸PEARL, R. / REED, L. J. / KISH, J. F.: *The logistic curve and the census count of 1940*, Science 92, 486 - 488, 1940

Jahr t_i	Einwohner x_i	$x_i - x_{i-1}$	α_i	\bar{x}_i
1790	3929214			
1800	5308483	1379269	0.0300867	4584319
1810	7239881	1931398	0.0310299	6224319
1820	9638453	2398572	0.0286156	8382048
1830	12860702	3222249	0.0288416	11172239
1840	17063353	4202651	0.0282757	14863132
1850	23191876	6128523	0.0306869	19971139
1860	31443321	8251445	0.0304385	27108620
1870	38558371	7115050	0.0203987	34879982
1880	50189209	11630838	0.0263627	44118569
1890	62979766	12790557	0.0227013	56342726
1900	76212168	13232402	0.0190708	69385801
1910	92228496	16016328	0.0190748	83965896
1920	106021537	13793041	0.0139373	98964870
1930	123202624	17181087	0.0150188	114397128
1940	132164569	8961945	0.00702175	127631160

Tabelle 2.3: Bevölkerung der USA 1790 - 1940 Quelle: US Census Bureau www.census.gov

aufgelöst wird. Diese exponentielle Wachstumsrate unterstellt, ergibt sich als *mittlere Populationsgröße* während der Dekade $[t_{i-1}, t_i]$

$$\bar{x}_i = \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} x_{i-1} e^{\alpha_i(t-t_{i-1})} dt = \frac{x_i - x_{i-1}}{\alpha_i(t_i - t_{i-1})} = \frac{x_i - x_{i-1}}{\log x_i - \log x_{i-1}}.$$

Die Ergebnisse für α_i und \bar{x}_i sind in den letzten beiden Spalten von Tabelle 2.3 eingetragen. Logistisches Wachstum ist nun durch einen linear fallenden Zusammenhang zwischen der Populationsgröße und der exponentiellen Wachstumsrate charakterisiert. Die Ausgleichsgerade für die ermittelten Werte (α_i, \bar{x}_i) (s. Abbildung 2.15) liefert den funktionalen Zusammenhang

$$\alpha(x) = 0.031578 - 1.69127 \cdot 10^{-10} x = 0.031578 \left(1 - \frac{x}{186711059}\right),$$

woraus sich für zwei der drei Parameter in (2.30) die Schätzwerte

$$a = 0.031578 \text{ und } K = 186711059$$

ergeben. Den noch fehlenden Parameter C ermittle ich, indem ich (2.30) nach C auflöse und $t = t_i$ für $i = 0, \dots, 15$ einsetze, woraus sich

$$C = \frac{x_i e^{-a(t_i-1790)}}{K - x_i} \text{ für } i = 0, \dots, 15 \quad (2.31)$$

ergibt. Natürlich sind die Werte auf der rechten Seite von (2.31) nicht alle gleich. Das wären Sie nur, wenn die Daten tatsächlich auf einer durch a und K bestimmten logistischen Kurve liegen würde. Als Schätzung für C wähle ich daher den Mittelwert der rechten Seiten von (2.31) und erhalte

$$C = 0.0215101.$$

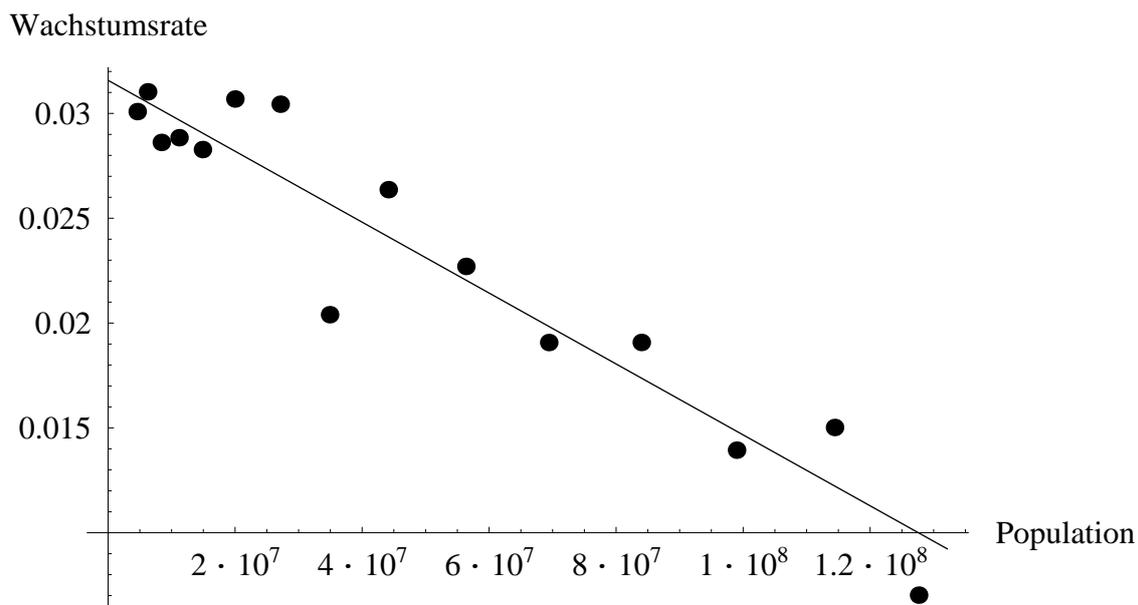


Abbildung 2.15: Zusammenhang zwischen Populationsgröße und exponentieller Wachstumsrate nach Tabelle 2.3

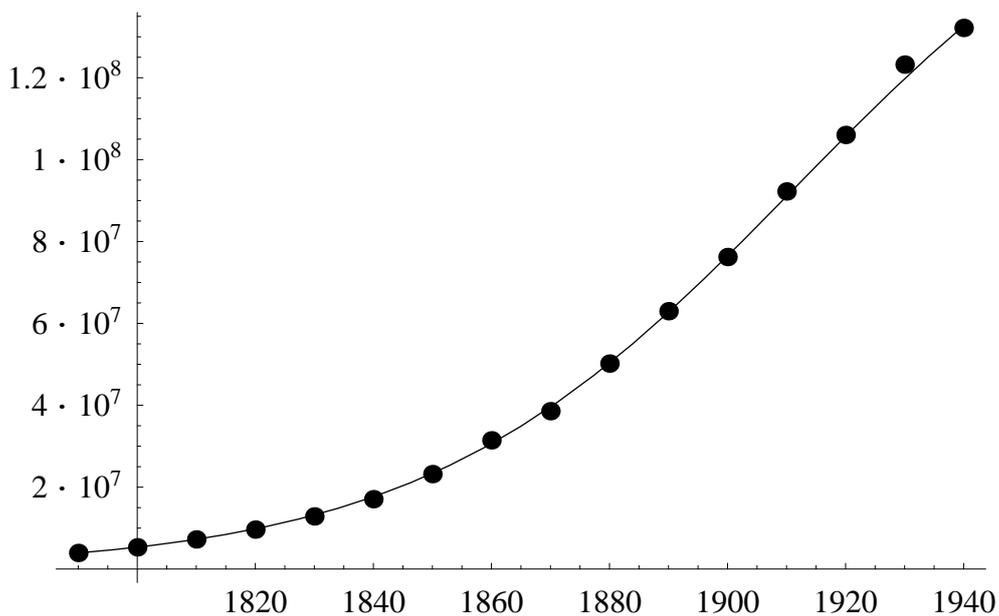


Abbildung 2.16: Daten nach Tabelle 2.3 und logistische Kurve

Vergleicht man das Ergebnis der so gewonnenen logistischen Kurve mit den Ausgangsdaten (Abbildung 2.16), so ist die Übereinstimmung so erstaunlich gut, dass eine weitere Nachbesserung der Parameterwerte wohl als unnötig erscheinen dürfte. Es lässt sich also feststellen, dass das Wachstum der US-Bevölkerung in den ersten 150 Jahren, in denen sie gezählt wurde, durch eine logistische Kurve sehr

gut beschrieben wird.

Doch was ist mit einer solchen Aussage eigentlich gewonnen? Was lässt sich mit ihr anfangen, und was nicht? Wenn der Zweck des hier an Bevölkerungsdaten angepassten mathematischen Modells (2.30) nur darin besteht, die Daten zu komprimieren, die Bevölkerungsentwicklung also möglichst knapp zu beschreiben, dann wurde dieser Zweck erreicht: Mit den drei Zahlenwerten für die Parameter a , K und C und der Formel (2.30) lässt sich die in den 16 Zahlenpaaren der Volkszählungen steckende Information praktisch reproduzieren. Das ist die wesentliche Funktion *deskriptiver Modelle*, eine andere haben sie nicht.

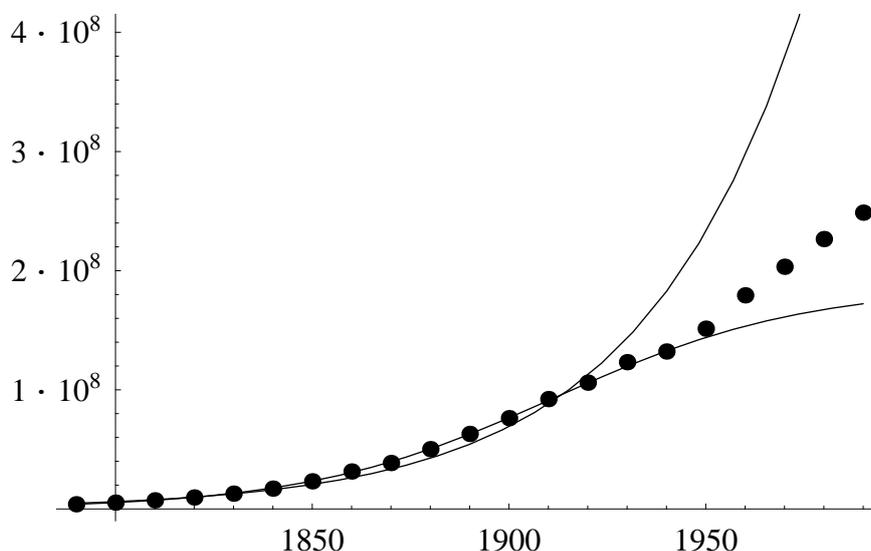


Abbildung 2.17: Exponentielles, logistisches und tatsächliches Wachstum der US-Bevölkerung 1790 - 1990

Sollte dagegen, wie es bei mathematischen Modellen ja oft der Fall ist, der Zweck in einer *Prognose* bestehen, so wurde dieser Zweck verfehlt, wie aus Abbildung 2.17 ersichtlich: Die berechnete Schranke $K = 186711059$ wurde bereits 1962 erreicht und 1990 um etwa ein Drittel überschritten. Der Grund dafür ist nicht darin zu suchen, dass "in Wirklichkeit" das Wachstum etwa exponentiell gewesen wäre⁹, sondern liegt schlicht darin, wie HEUSER (1989, 25) ironisch anmerkt, dass "ausgerechnet ab 1940", dem Jahr der Veröffentlichung der hier referierten Arbeit, "die US-Menschen dem druckfrischen Gesetz zum Hohn sich viel stärker vermehrten, als sie eigentlich durften."

Hängt die Wachstumsrate (nur) von der Populationsdichte ab?

Was ist hier schief gelaufen? Offenbar besagt die noch so perfekte Übereinstimmung eines Modells mit den Daten, die es erklären soll, nichts darüber aus, ob es die vorgefundene Empirie tatsächlich *erklärt*. Dazu müssten nämlich die *Wirkungszusammenhänge*, die für das beobachtete Phänomen verantwortlich sind, im Modell mit erfasst werden. Und das ist hier ganz offensichtlich nicht der Fall. Das Wachstum der Bevölkerung eines Staates hängt eben nicht von der Populationsdichte ab, sondern wohl eher von gesellschaftlichen Faktoren wie dem Gesundheitswesen, der Familienpolitik, der Organisation der Altersvorsorge, der Verbreitung und Effektivität empfängnisverhütender Mittel und besonders in den USA auch der Einwanderung.

⁹Unterstellt man von 1790 bis 1940 exponentielles Wachstum, so führt das Resultat zu noch schlechteren Prognosen, wie ebenfalls aus Abbildung 2.17 deutlich wird

Durch ein eindimensionales Wachstumsmodell wie dem logistischen lässt sich all dies nicht erfassen, denn der aktuelle Zustand wird hier immer nur durch eine einzige Zahl beschrieben, die Populationsgröße oder -dichte. Von den Ursachen, die auf die Wachstumsrate Einfluss nehmen, muss daher abstrahiert werden. Ein solches Modell kann deshalb das vorgefundene Wachstum nicht erklären. Das schließt natürlich nicht aus, dass das logistische Wachstum als *deskriptives Modell* zum Zwecke der Datenkompression geeignet ist wie etwa für die Bevölkerungsentwicklung der USA von 1790 bis 1940. Wenn das einmal so gut klappt wie hier, dann gibt es vielleicht Anlass, nach den Gründen dafür zu fragen. Das Modell selbst kann darauf aber keine Antwort geben.

Innerspezifische Konkurrenz

Der jetzt vielleicht nahe liegende Schluss, dass die logistische Differentialgleichung für die Beschreibung des Wachstums von Populationen generell ungeeignet sei, wäre allerdings etwas vorschnell. BEGON / HARPER / TOWNSEND: *Ökologie* (1991, 227) berichten u. a. von einer Population von Käfern (*Rizopertha dominica*), die im Labor in 10 g Weizenkörnern gehalten wurden, welche wöchentlich erneuert wurden; das Wachstum der Käferpopulation lässt sich durch eine logistische Kurve sehr gut beschreiben. Entscheidend ist hier natürlich, dass im Labor alle Faktoren, die das Wachstum beeinflussen könnten, konstant gehalten werden, außer eben die Populationsgröße selbst, sodass in der Tat die Wachstumsrate als Funktion nur der Populationsgröße bzw. -dichte angesehen werden kann.

Die nächste Frage ist dann aber, *wie* die Populationsdichte die Geburten- und Sterberate beeinflusst. In der obigen idealisierten Beschreibung der innerspezifischen Konkurrenz (Beispiel (c)) wurde davon ausgegangen, dass die (relative oder exponentielle) Geburtenrate konstant bleibt, während die Sterberate linear mit der Populationsdichte zunimmt. Eine konstante Sterberate bei linear fallender Geburtenrate würde zum selben deterministischen Modell führen. Die tatsächlichen Zusammenhänge sind in der Regel komplexer, auch im Labor, und durch Zwischenstufen vermittelt, die im Modell gar nicht erfasst sind, wie z. B. eine Verringerung der Eiablage, eine erhöhte Sterblichkeit der Juvenilen oder (bei Insekten) Larven oder (bei Pflanzen) der Einfluss der Dichte auf die Anzahl der Blüten je Pflanzen, die Anzahl der Samen je Schote und schließlich den Anteil der Samen, die neu keimen.¹⁰

Der Zusammenhang zwischen der Populationsdichte und der relativen oder exponentiellen Wachstumsrate dürfte nur selten linear sein. Tatsächlich ist dieser Zusammenhang im Allgemeinen aber nur in Gestalt einiger Datenpunkte bekannt, die aus Messungen gewonnen wurden. Und in diesem Falle bleibt einem oft nichts anderes übrig, als durch die wenigen und mit Fehlern behafteten Daten eine Gerade zu legen.

2.3.8 Ein allgemeiner Ansatz für dichteabhängiges Wachstum

Die logistische Differentialgleichung (2.23) ist ein Spezialfall der allgemeineren

$$\dot{x} = \alpha(x) x \quad (2.32)$$

für die Dichte bzw. den Bestand x einer Population. Dabei sei $\alpha : [0, \infty) \rightarrow \mathbb{R}$ eine Funktion von x , die der Einfachheit halber als stetig differenzierbar vorausgesetzt wird. Die Frage ist, welche qualitativen Eigenschaften der logistischen Kurven unter welchen (biologisch sinnvollen) Voraussetzungen auch den Lösungen von (2.32) erhalten bleiben.

¹⁰Eine ausführliche Diskussion der Effekte, die die Erhöhung der Populationsdichte für verschiedene Tier- und Pflanzenarten haben kann, führen BEGON / HARPER / TOWNSEND (1991, 215 - 232) und BEGON / MORTIMER / THOMPSON (1997, 50 - 79)

Satz 2.3.1

Es existiere ein $K > 0$, sodass

$$\alpha(x) \begin{cases} > 0 & \text{für } 0 < x < K \\ < 0 & \text{für } x > K \end{cases} .$$

Dann besitzt (2.32) zu jedem $x_0 \geq 0$ eine eindeutige Lösung $x : [0, \infty) \rightarrow \mathbb{R}$ mit $x(0) = x_0$. Ist $x_0 = 0$, so ist $x(t) = 0$ für alle $t \geq 0$. Ist $x_0 = K$, so ist $x(t) = K$ für alle $t \geq 0$. Für $0 < x_0 < K$ ist $x(t)$ streng monoton wachsend, für $x_0 > K$ streng monoton fallend, und in beiden Fällen ist

$$\lim_{t \rightarrow \infty} x(t) = K .$$

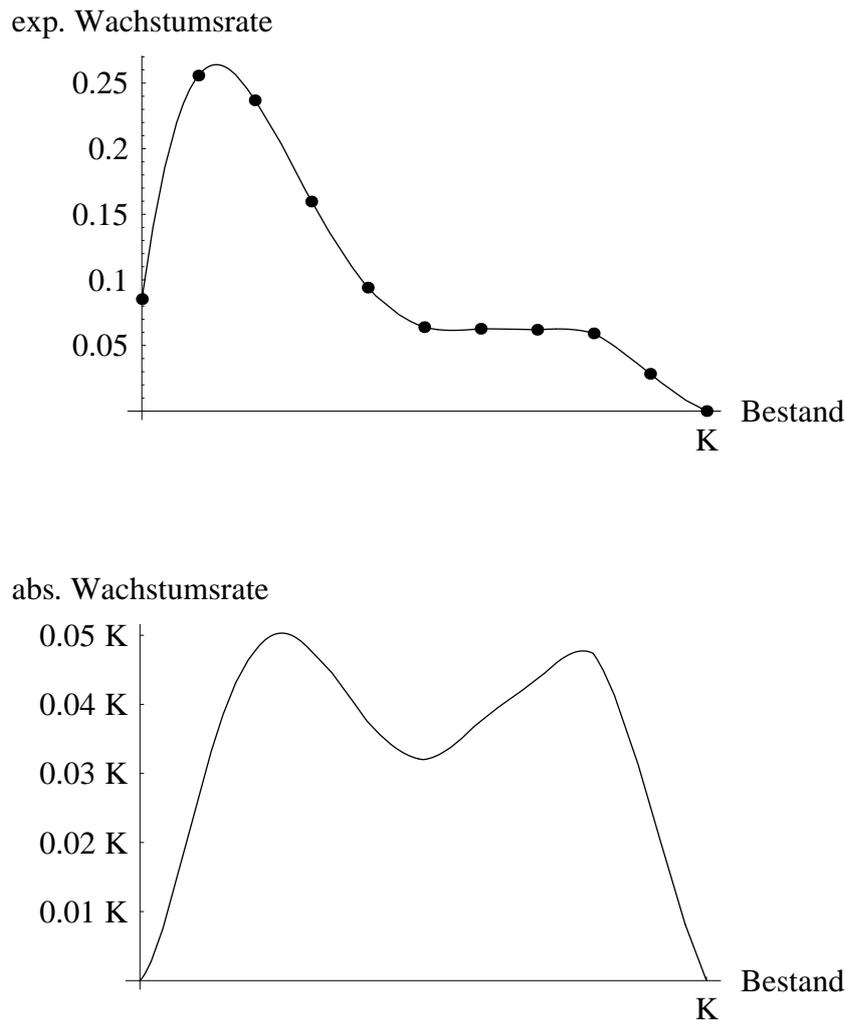


Abbildung 2.18: Exponentielle und absolute Wachstumsrate in Abhängigkeit vom Bestand für einen fiktiven Datensatz

In Abbildung 2.18 ist zur Veranschaulichung mit einem fiktiven Datensatz eine Situation konstruiert, auf die der Satz anwendbar ist, die aber aus dem Rahmen der logistischen Differentialgleichung fällt. Die durch eine Spline-Interpolation der Daten gewonnene exponentielle Wachstumsrate $\alpha(x)$ ist für kleine

Bestände nicht mehr monoton fallend. Hier liegt nicht Konkurrenz, sondern Kooperation vor. Das lässt sich etwa bei bestimmten Seevögeln beobachten, bei denen der relative Bruterfolg mit der Dichte einer Vogelkolonie zunächst wächst, weil größere Kolonien sich gegen Räuber besser wehren können.¹¹ Erst bei hohen Dichten treten gegenläufige Effekte ein, z. B. wegen eines Mangels an geeigneten Nistplätzen.

Von Fragen nach der Existenz und Eindeutigkeit der Lösungen einmal abgesehen, die in den Bereich der allgemeinen Theorie gewöhnlicher Differentialgleichungen fallen, lässt sich Satz 2.3.1 geometrisch plausibel machen, indem man beachtet, dass für positive Lösungen von (2.32)

$$\dot{x}(t) \begin{cases} > \\ = \\ < \end{cases} 0, \text{ falls } x(t) \begin{cases} < \\ = \\ > \end{cases} K.$$

In Abbildung 2.19 ist eine zu diesen Wachstumsraten gehörige, numerisch berechnete Lösungskurve wiedergegeben. Von einer logistischen Kurve unterscheidet sie sich qualitativ nur durch ihre weniger

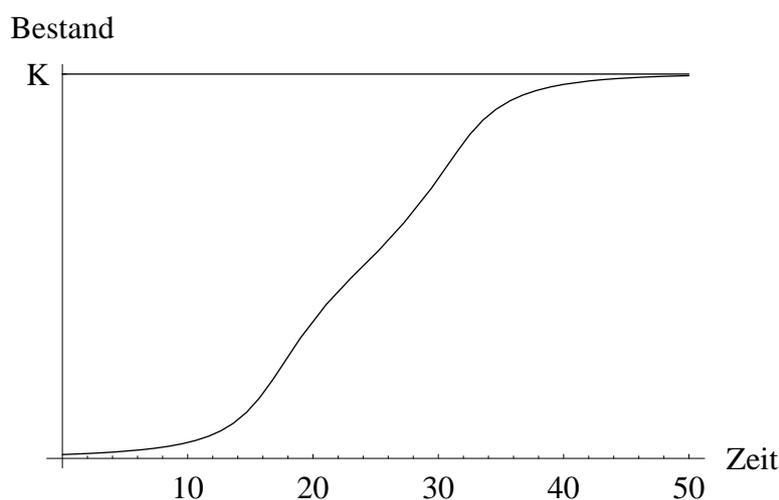


Abbildung 2.19: Lösungskurve von (2.32) für die Wachstumsraten aus Abbildung 2.18

präzise S-Förmigkeit. Der Grund dafür lässt sich an der Gestalt der absoluten Wachstumsrate als Funktion von x erkennen (Abbildung 2.18). Anders als bei der logistischen Gleichung hat sie nicht nur ein (dort bei $K/2$ liegendes) lokales Maximum, sondern deren zwei. Das hat zur Folge, dass die Steigung der Lösung nicht einfach anwächst und dann wieder fällt, sondern vielmehr hin und her schwankt, weshalb eine Schlangenlinie entsteht. Zugleich ist festzustellen, dass die qualitativen Unterschiede zur logistischen Differentialgleichungen an den Lösungen viel weniger deutlich werden als an den Wachstumsraten. Hieraus ergibt sich ein *Hinweis auf die Datenerhebung*: Es sollten besser die Wachstumsraten als die Entwicklung von Beständen gemessen werden, will man artspezifischen Unterschieden auf die Spur kommen.

Aufgabe 2.19

Durch menschliche Eingriffe sei die in Abbildung 2.18 beschriebene Population durch eine zusätzliche exponentielle Sterberate von 0.16 je Zeiteinheit belastet, d. h. $\alpha(x)$ verringere sich für alle x um diesen Betrag und werde also insbesondere für sehr kleine x negativ, weswegen Satz 2.3.1 nicht mehr anwendbar ist. Welche Auswirkungen ergeben sich für die langfristige Entwicklung des Bestandes? Wie hängt die Antwort auf diese Frage vom Anfangswert x_0 ab?

¹¹BEGON / MORTIMER / THOMPSEN (1997, 79 - 80) berichten von entsprechenden Untersuchungen bei Trottellummen (*Uria aalge*) in Südwales

2.4 Ein allgemeiner Rahmen zur Modellierung dynamischer Prozesse

In den letzten Abschnitten wurden Modelle für die zeitliche Entwicklung des Bestandes einer Substanz oder Population untersucht. Diese Problemstellung ist Spezialfall einer allgemeineren, welche nach der *zeitlichen Entwicklung* des Zustands irgendeines Systems fragt. Die Besonderheit des betrachteten Spezialfalls liegt darin, dass der Zustand durch eine einzige Zahl, den Bestand, gekennzeichnet ist.

Unter einem *dynamischen Prozess* soll ein System verstanden werden, dessen zeitliche Entwicklung durch bestimmte, theoretisch mehr oder weniger fundierte *Entwicklungsgesetze* festgelegt ist. Ein *dynamisches Modell* ist ein mathematisches Modell eines solchen Prozesses, in dem also die Entwicklungsgesetze in mathematischer Form gefasst sind. Das setzt voraus, dass auch der *Zustand* des Systems durch mathematische Größen definiert ist. In der Regel wird dazu eine einzige Zahl nicht ausreichen, selbst dann nicht, wenn der Modellzweck nur auf eine einzige Zahl, wie etwa den Bestand einer Population, gerichtet ist. So dürfte es etwa für menschliche Bevölkerungen klar sein, dass ihr Wachstum nicht nur von ihrer aktuellen Größe abhängt, sondern auch noch von anderen Faktoren, die daher in das Modell mit einbezogen werden müssen.

Die im Folgenden skizzierten Modellklassen unterscheiden sich hinsichtlich der mathematischen Struktur der Zeit, der Zustände und (damit zusammenhängend) der Entwicklungsgesetze.

2.4.1 Der Zustandsraum

Die möglichen Zustände eines Systems werden durch eine Menge X beschrieben, den *Zustandsraum*. Zu jedem Zeitpunkt t befindet sich das System in einem Zustand $x(t) \in X$. Die übliche mathematische Aufgabenstellung besteht darin, zu gegebenem *Anfangszustand* $x(0) = x_0 \in X$ auf der Basis der definierten Entwicklungsgesetze die Zustände $x(t)$ für alle Zeiten $t \geq 0$ oder auch nur für ein beschränktes Zeitintervall zu bestimmen.

$x \in X$ soll den Zustand des Systems im Sinne des Modells *vollständig* beschreiben, muss also die voneinander unabhängigen Größen, die auf die Entwicklung Einfluss nehmen und von ihr beeinflusst werden, enthalten. Welche das sind, ist eine Entscheidung, die in der Modellbildung zu treffen ist. Jede Modellbildung besteht darin, gewisse Größen als zum Modell gehörig und andere als externe, als konstant angenommen zu deklarieren.

Beispiele

- Der Zustand eines Systems interagierender Populationen lässt sich durch den Vektor der Populationsgrößen beschreiben. Abstrahiert wird dann u. a. von der Fitness, der Altersstruktur und der räumlichen Verteilung der Populationen, die dann auch in den Entwicklungsgesetzen keine Rolle spielen dürfen.
- Der Zustand einer Volkswirtschaft wird definiert als Vektor bestimmter wirtschaftlicher Indikatoren, von denen angenommen wird, dass die weitere Entwicklung durch sie und *nur* durch sie bestimmt ist (Bruttoinlandsprodukt, Beschäftigungsquote, Kapitalstock usw.).
- Soll die Entwicklung der Wassertemperatur eines Sees im Laufe des Jahres modelliert werden, so genügt es im Allgemeinen nicht, die Durchschnittstemperatur zu betrachten, da es auch auf die Wärmeverteilung ankommt (das Wasser in tieferen Schichten ist im Sommer kälter als in höheren. Der Zustand wäre in diesem Falle als Funktion entweder der Wassertiefe oder aber jedes Orts im See zu wählen.
- Hängen die untersuchten Prozesse explizit von der Zeit (bzw. der Jahres- oder Tageszeit) ab, so sollte man die Zeit (bzw. Jahres- oder Tageszeit) als Komponente in den Systemzustand mit auf-

nehmen. Damit lässt sich erreichen, dass die Entwicklungsgesetze durch Funktionen des Zustands beschrieben werden können.

Zur mathematischen Struktur des Zustandsraums

Die mathematische Struktur des Zustandsraums liegt irgendwo zwischen zwei Extremen: Endlichen Mengen auf der einen Seite, wenn das System nur endlich viele Zustände annehmen kann, und Teilmengen unendlichdimensionaler Räume (z. B. von Funktionen), wenn bereits der einzelne Zustand durch unendlich viele Parameter bestimmt ist.

- Betrachtet man eine Population und beschreibt ihren Zustand als Anzahl der ihr angehörenden Individuen, so ist der Zustandsraum entweder endlich, sofern die Populationsgröße von vornherein beschränkt ist, oder aber abzählbar unendlich, wenn jede Populationsgröße im Prinzip denkbar ist.
- Wird jeder einzelne Zustand durch endlich viele Zahlen beschrieben, so hängt die Struktur des Lösungsraumes davon ab, um welche Art von Zahlen es sich handelt. Sind nur ganze Zahlen erlaubt (z. B. Anzahlen, Schalterstellungen usw.), so ist der Zustandsraum endlich oder abzählbar unendlich. Handelt es sich dagegen um kontinuierliche Größen, so ist der Zustandsraum Teilmenge eines endlichdimensionalen, in der Regel reellen Vektorraums \mathbb{R}^n . Natürlich sind auch Mischformen möglich, in denen gewisse Komponenten des Zustands ganzzahlig und andere reell sind.
- Unendlichdimensionale Zustandsräume treten auf, wenn der einzelne Zustand durch eine unendliche Folge oder durch eine auf einem Kontinuum definierte Funktion beschrieben wird, wie etwa der Temperaturverteilung in einem See.

2.4.2 Diskrete und kontinuierliche Zeit

Für die Modellierung der Zeit gibt es im Wesentlichen nur zwei Möglichkeiten. Entweder wird sie als kontinuierlich fließend aufgefasst, wofür die reellen Zahlen ein Modell bilden, oder es werden feste Zeiträume gleicher Größe bzw. äquidistante Zeitpunkte betrachtet, was sich bei entsprechender Wahl der Zeiteinheit durch die ganzen Zahlen modellieren lässt.

Eine Zwischenform bilden Modelle mit *semidiskreter* Zeit, in denen äquidistante Zeitpunkte

$$m \Delta t, m \text{ ganzzahlig}$$

betrachtet werden, die Länge des Zeitintervalls Δt aber nicht festgelegt ist, sondern kontinuierlich variiert wird. Solche Modelle treten insbesondere bei der numerischen Approximation von Modellen mit kontinuierlicher Zeit auf.

2.4.3 Deterministische und stochastische Entwicklungsgesetze

Entwicklungsgesetze definieren im Rahmen dynamischer Modelle in der Regel die *zeitlich lokalen, kurzfristigen* Zustandsänderungen des Systems: Bei gegebenem aktuellen Zustand beschreiben sie, welchen Zustand das System in der allernächsten Zukunft hat. Die mathematischen Untersuchungen im engeren Sinne bestehen oft darin, hieraus Schlüsse auf das *länger- oder langfristige* Systemverhalten zu ziehen.

Die Gestalt der Entwicklungsgesetze hängt in starkem Maße von der Struktur der Zeit und des Zustandsraums ab. So lässt sich etwa bei diskreter Zeit der *Folgezustand* in Abhängigkeit vom aktuellen Zustand definieren, während bei kontinuierlicher Zeit der Begriff des Folgezustands gar keinen Sinn hat, weshalb hier zeitliche Ableitungen und damit Differentialgleichungen ins Spiel kommen.

Davon unabhängig lassen sich aber *deterministische* und *stochastische* Modelle voneinander unterscheiden: Liegt mit dem aktuellen Zustand der "Folgezustand", damit aber die ganze Zukunft eindeutig

fest, so spricht man von einem *deterministischen* Modell. Ist dagegen das Entwicklungsgesetz in Form einer Wahrscheinlichkeitsverteilung über die möglichen Folgezustände gefasst, so nennt man das Modell *stochastisch*. In diesem Falle betrachtet man oft die *stochastische Erweiterung* des ursprünglichen Modells, dessen Zustandsraum die Menge der Wahrscheinlichkeitsverteilungen über dem ursprünglichen Zustandsraum ist und das selbst wieder deterministisch ist. Diese Vereinfachung wird aber durch eine extreme Vergrößerung des Zustandsraums erkauft: Aus einem endlichen Zustandsraum beispielsweise wird so ein endlichdimensionaler, aber unendlicher.

Je nach Wahl der Struktur der Zeit und des Zustandsraums sowie der Art des Entwicklungsgesetzes kommt man zu verschiedenen Modelltypen, von denen hier nur einige gebräuchliche genannt werden.

2.4.4 Automaten

Mit diskreter Zeit $t \in \mathbb{N}_0$ und einem endlichen oder abzählbar unendlichen Zustandsraum X gelangt man zu *Automaten*, wie sie vor allem in der theoretischen Informatik untersucht werden, wobei je nach Art des Entwicklungsgesetzes zwischen deterministischen und stochastischen Automaten unterschieden wird. Ein *deterministischer Automat* lässt sich charakterisieren durch eine Funktion $F : X \rightarrow X$ des abzählbaren Zustandsraums in sich, die zu jedem Zustand den Folgezustand angibt:

$$x(t+1) = F(x(t)).$$

Diese Kennzeichnung ist allerdings sehr abstrakt und für Zwecke der Informatik alles andere als ausreichend. Insbesondere der Funktionsbegriff ist zu allgemein, die Funktion F muss sich auch berechnen lassen. Ihre genauere Bestimmung und die der Struktur von X führen zu verschiedenen Automatentypen.¹²

Bei einem *stochastischen Automaten* oder *Markov-Prozess* ist an Stelle der (eindeutigen) Funktion f für jeden Zustand eine Wahrscheinlichkeitsverteilung auf der Menge X gegeben, durch die die möglichen Folgezustände und die Wahrscheinlichkeiten ihres Auftretens definiert sind.

2.4.5 Deterministische dynamische Systeme

Bei kontinuierlichem Zustandsraum (z. B. einer offene Teilmenge des \mathbb{R}^n) und deterministischem Entwicklungsgesetz spricht man von *deterministischen dynamischen Systemen*, die sich hinsichtlich der Struktur der Zeit und der Dimension und Art des Zustandsraums unterscheiden können.

Diskrete dynamische Systeme operieren mit diskreten Zeitpunkten und sind durch ein Entwicklungsgesetz in der Form einer Iteration

$$x(t+1) = F(x(t)) \text{ für } t = 0, 1, 2, \dots \quad (2.33)$$

bestimmt. Formal gesehen ähneln sie deterministischen Automaten. Der Unterschied liegt in der Topologie des Zustandsraums bzw. dem dort definierten Konvergenzbegriff, was zu anderen Fragestellungen führt: Man kann etwa nach dem Grenzwert

$$\lim_{t \rightarrow \infty} x(t)$$

fragen. Bei Automaten bedeutet die Existenz eines solchen Grenzwerts immer, dass die Zustandsfolge *nach endlich vielen Schritten* zur Ruhe kommt, die Zustände sich also nicht mehr ändern.

¹²Die hier eingeführte Funktion F hängt regelhaft vom inneren Zustand des Automaten und den Eingabedaten ab. Im Sinne der hier vorgenommenen formalen Klassifizierung setzt sich der Zustand des Automaten also aus dem inneren Zustand und den Eingabedaten zusammen.

Soll die in einem Zeitschritt erfolgte Änderung besonders betont werden, so schreibt man das Entwicklungsgesetz in der Form

$$x(t+1) = x(t) + f(x(t)) \text{ für } t = 0, 1, 2, \dots,$$

was voraussetzt, dass X Teilmenge eines Vektorraums ist.

Bei der Modellbildung mit diskreten dynamischen Systemen muss immer festgelegt werden, was unter einem Zeitschritt zu verstehen ist, ob also die Zeiteinheit ein Jahr, ein Tag, eine Stunde, ... sein soll. Oft ist diese Zeiteinheit durch die Datenlage oder aus anderen in der Sache liegenden Gründen vorgegeben (z. B. monatliche oder jährliche Rechnungsabschlüsse in der Ökonomie, jährlicher Generationswechsel in der Populationsdynamik), doch oft wäre jede Wahl eines festen Zeitschritts willkürlich. In diesem Fall wird das Modell mit einem variablen Zeitschritt formuliert und unterstellt, dass die stattgefundenene Änderung zur Länge des Zeitschritts ungefähr proportional ist:

$$x(t + \Delta t) = x(t) + \Delta t f(x(t)) \text{ für } t = 0, \Delta t, 2\Delta t, \dots$$

In diesen *semidiskreten* Modellen wird die Zeit im Grunde bereits als kontinuierlich angenommen, da $\Delta t > 0$ beliebig klein gewählt werden kann.

Lässt man im semidiskreten Modell den Zeitschritt Δt gegen 0 gehen, so erhält man im Falle eines endlichdimensionalen Zustandsraums mit

$$\dot{x}(t) = \lim_{\Delta t \rightarrow 0} \frac{x(t + \Delta t) - x(t)}{\Delta t} = f(x(t))$$

ein System *gewöhnlicher Differentialgleichungen* oder, im modernen Sprachgebrauch, ein *kontinuierliches dynamisches System*: Die Geschwindigkeit (Größe und Richtung), mit der sich der Zustand des Systems ändert, ist eine Funktion des aktuellen Zustands.

Insbesondere im Zusammenhang mit biologischen Fragestellungen haben *Differentialgleichungen mit verzögertem Argument* in den letzten Jahrzehnten an Bedeutung gewonnen:

$$\dot{x}(t) = f(x(t), x(t-s)) \text{ mit einem } s > 0.$$

Die Änderung des aktuellen Zustands hängt hier nicht nur vom aktuellen Zustand selbst ab, sondern auch vom Zustand, der einen gewissen Zeitabschnitt vorher herrschte, wie es etwa für viele Reproduktionsprozesse der Fall ist (zeitlicher Abstand zwischen Zeugung und Geburt). Formal gesehen handelt es sich hier nicht einfach um eine Mischform von diskretem und kontinuierlichem dynamischen System: Zu einer eindeutigen Lösung kommt man nämlich erst, wenn man den Zustand nicht nur zu einem festen Zeitpunkt vorgibt, z. B. $x(0) = x_0$, sondern gleich in einem ganzen Zeitintervall: $x(t) = x_0(t)$ für $t \in [-s, 0]$, was selbst bei einer einzelnen Gleichung einen unendlichdimensionalen Zustandsraum zur Folge hat.

Unendlichdimensionale Zustandsräume liegen auch bei allen Modellen vor, die *partielle Differentialgleichungen* verwenden. Der aktuelle Zustand ist zu jedem Zeitpunkt eine Funktion beispielsweise von Raumkoordinaten. Das Entwicklungsgesetz enthält deren Ableitung nicht nur nach der Zeit, sondern auch der anderen Variablen, von denen sie abhängt, z. B.

$$\frac{\partial x}{\partial t} = c \frac{\partial^2 x}{\partial s^2}$$

als Modell für die Entwicklung der Temperaturverteilung x in einem von einer einzigen Raumkoordinaten s abhängigen Medium. Die Frage nach der Existenz und Eindeutigkeit von Lösungen lässt sich bei partiellen Differentialgleichungen nicht so einheitlich beantworten wie bei gewöhnlichen, sondern ist für jede Problemklasse gesondert zu behandeln.

2.4.6 Existenz und Eindeutigkeit bei gewöhnlichen Differentialgleichungen

Gibt man einen Anfangszustand $x(0) = x_0 \in X$ vor, so ist bei einem Entwicklungsgesetz der Gestalt (2.33) klar, dass damit die zugehörige Lösung $(x(t), t \in \mathbb{N}_0)$ eindeutig bestimmt ist. Vorausgesetzt werden muss dazu nur, dass $F : X \rightarrow X$ den Zustandsraum in sich abbildet. Bei gewöhnlichen Differentialgleichungen ist das nicht so einfach, es gibt aber leicht überprüfbare Voraussetzungen, unter denen die Existenz und Eindeutigkeit gesichert ist.

Satz 2.4.1 (AMANN 1983, 110/111, KNOBLOCH/KAPPEL 1974, 116-118)

Seien $X \subseteq \mathbb{R}^n$ offen, $f : X \rightarrow \mathbb{R}^n$ stetig differenzierbar und $x_0 \in X$. Dann besitzt die Anfangswertaufgabe

$$\dot{x} = f(x), x(0) = x_0 \quad (2.34)$$

eine eindeutige Lösung in folgendem Sinne: Es existiert ein eindeutig bestimmtes offenes Intervall $I = (t^-, t^+)$ mit $0 \in I$ und folgenden Eigenschaften:

- (a) Es gibt eine eindeutig bestimmte Funktion $x : I \rightarrow \mathbb{R}^n$, die der Anfangswertaufgabe (2.34) genügt.
- (b) Ist $\tilde{I} \subseteq I$ irgendein Intervall mit $0 \in \tilde{I}$ und genügt $\tilde{x} : \tilde{I} \rightarrow \mathbb{R}^n$ der Anfangswertaufgabe (2.34), so ist $\tilde{I} \subseteq I$ und $\tilde{x}(t) = x(t)$ für alle $t \in \tilde{I}$.
- (c) Ist $t^+ < \infty$, $t_k \in I$ eine Folge mit $t_k \rightarrow t^+$ und $x(t_k) \rightarrow x^+$, so liegt x^+ auf dem Rand von X .
- (d) Ist $t^- > -\infty$, $t_k \in I$ eine Folge mit $t_k \rightarrow t^-$ und $x(t_k) \rightarrow x^-$, so liegt x^- auf dem Rand von X .

Das im Satz genannte Intervall I wird das *maximale Existenzintervall* der Anfangswertaufgabe (2.34) genannt. Dieses kann, anders als bei diskreten dynamischen Systemen, nach beiden Seiten beschränkt sein. Die Teile (c) und analog (d) des Satzes schränken diese Möglichkeit allerdings ein: Ist $t^+ < \infty$, die Lösung von (2.34) also nicht für alle positiven Zeiten definiert, so liegt das entweder daran, dass die Lösung den Rand von X erreicht, oder daran, dass keine konvergente Folge $x(t_k)$ mit $t_k \rightarrow t^+$ existiert, was aber nur möglich ist, wenn $x(t)$ unbeschränkt ist.

Im Falle der Anfangswertaufgabe

$$\dot{x} = x^2, x(0) = 1$$

ist $X = \mathbb{R}$ zu wählen, hat also keinen Rand. Dennoch ist das maximale Existenzintervall nach oben beschränkt, denn die Lösung

$$x(t) = \frac{1}{1-t}$$

ist nur auf $(-\infty, 1)$ definiert.

Um im Falle $X = \mathbb{R}^n$ die *globale* Existenz der Lösung nachzuweisen, ist also auszuschließen, dass eine Komponente der Lösung in endlicher Zeit gegen unendlich streben kann. Das aber ist nicht immer einfach und nur unter Berücksichtigung der konkreten Aufgabenstellung machbar.

Bemerkung

Die hier betrachteten Differentialgleichungssysteme

$$\dot{x} = f(x)$$

sind *autonom*, d. h. ihre rechte Seite hängt nicht explizit von der Zeit ab. Nicht autonome Systeme

$$\dot{x} = f(t, x)$$

lassen sich dadurch formal in autonome Systeme überführen, dass man die Zeit als zusätzliche Zustandsvariable deklariert, die der (zusätzlichen) Differentialgleichung

$$\dot{t} = 1$$

genügt. Auf das so erweiterte System ist dann wieder Satz 2.4.1 anwendbar.

2.4.7 Positive Invarianz

Bei der Modellierung etwa einer biologischen Fragestellung durch ein deterministisches dynamisches System oder durch ein System von Differentialgleichungen sind die entstehenden Gleichungen oft auch für solche Zustände definiert, die biologisch gar nicht interpretierbar sind, z. B. auch für negative Größen, obwohl es um die Modellierung von Beständen geht, die nicht negativ werden können. In diesem Falle könnte es passieren, dass die mathematische Lösung sich in Bereichen des Zustandsraums aufhält, die im Sinne des Modells unsinnig sind. Das hieße, dass das Modell für seinen Zweck nicht geeignet ist, bei der Modellierung also ein Fehler aufgetreten sein muss. Wie aber lässt sich überprüfen, dass das nicht passiert?

Definition 2.4.2

Unter den Voraussetzungen von Satz 2.4.1 heiße eine Menge $M \subseteq X$ "positiv invariant", wenn für alle $x_0 \in M$ die Lösung von (2.34) für alle positiven Zeiten, für die sie definiert ist, in M liegt:

$$x_0 \in M \Rightarrow x(t) \in M \text{ für alle } t \in [0, t^+).$$

Zu überprüfen ist also, ob die Menge der im Modellsinne interpretierbaren Zustände positiv invariant ist. Dieselbe Definition lässt sich auch auf deterministische dynamische Systeme (2.33) übertragen. Die positive Invarianz von $M \subseteq X$ ist dann äquivalent zu $F(M) \subseteq M$. Im Falle gewöhnlicher Differentialgleichungen gilt

Satz 2.4.3 (AMANN 1983, 237)

Sei unter den Voraussetzungen von Satz 2.4.1 $M \subseteq X$ abgeschlossen und ∂M der Rand von M . M ist genau dann positiv invariant, wenn für alle $x \in \partial M$ gilt:

$$\liminf_{t \rightarrow 0, t > 0} \frac{d(x + t f(x), M)}{t} = 0,$$

wobei

$$d(y, M) := \inf_{m \in M} \|y - m\|$$

der Minimalabstand von $y \in \mathbb{R}^n$ zu M .

Dass für die positive Invarianz der Menge M nur deren Randpunkte untersucht werden müssen, liegt daran, dass die Lösungen stetig differenzierbare Kurven in X bilden, die eine Menge immer nur über ihren Rand verlassen können (anders als bei diskreten dynamischen Systemen gibt es keine Sprünge). In jedem Punkt $x \in X$ ist die Richtung, in der die Lösungskurve sich von x aus weiterbewegt, durch $f(x)$ festgelegt. Die angegebene Bedingung besagt, dass diese Richtung nicht aus M herausweisen darf, eine tangentielle Richtung ist gerade noch erlaubt.

Der nichtnegative Oktant

$$\mathbb{R}_+^n := \{x \in \mathbb{R}^n : x \geq 0\}$$

ist der Bereich, in dem sich die Zustandsvariablen als Bestandsgrößen interpretieren lassen. Als Kriterium für seine positive Invarianz ergibt sich nun

Satz 2.4.4

Sei über die Voraussetzungen von Satz 2.4.1 hinaus $\mathbb{R}_+^n \subset X$. \mathbb{R}_+^n ist genau dann positiv invariant, wenn für $i = 1, \dots, n$

$$x \geq 0, x_i = 0 \Rightarrow f_i(x) \geq 0.$$

Aufgabe 2.20

Die Wahrscheinlichkeitsverteilung $w(t)$ eines kontinuierlichen Markovprozesses mit n Zuständen genügt einem System linearer Differentialgleichungen

$$\dot{w} = A w$$

mit einer (n, n) -Matrix $A = (a_{ij})$, für die

$$a_{ij} \geq 0 \text{ für } i \neq j \text{ und } \sum_{i=1}^n a_{ij} = 0 \text{ (} i, j = 1, \dots, n \text{)} .$$

Zeigen Sie, dass die Menge

$$\{w \in \mathbb{R}^n : w \geq 0, \sum_{k=1}^n w_k = 1\}$$

der Wahrscheinlichkeitsverteilungen für dieses System positiv invariant ist.

Gelingt es, die positive Invarianz einer *kompakten* Menge $M \subset X$ nachzuweisen, so ist damit auch die Existenz der in ihr verlaufenden Lösungen für alle $t \geq 0$ nachgewiesen, denn diese können weder dem Rand von X beliebig nahe kommen noch unbeschränkt sein. Es gilt also

Satz 2.4.5

Ist unter den Voraussetzungen von Satz 2.4.1 $M \subseteq X$ kompakt und positiv invariant, so sind für alle $x_0 \in M$ die Lösungen von (2.34) für alle Zeiten $t \geq 0$ definiert.

2.5 Literatur

- AMANN, H.: *Gewöhnliche Differentialgleichungen*, Berlin 1983
- BARTH, F. / HALLER, R.: *Stochastik Leistungskurs*, 5. Aufl., München 1996
- BEGON, M. / HARPER, J. L. / TOWNSEND, C. R.: *Ökologie*, Basel 1991
- BEGON, M. / MORTIMER, M. / THOMPSEN, D. J.: *Populationsökologie*, Heidelberg 1997
- HEUSER, H.: *Gewöhnliche Differentialgleichungen*, Stuttgart 1989
- HUPPERT, B.: *Angewandte Lineare Algebra*, Berlin 1990
- KNOBLOCH, H. W. / KAPPEL, F.: *Gewöhnliche Differentialgleichungen*, Stuttgart 1974
- MANKIV, N. G.: *Grundzüge der Volkswirtschaftslehre*, Stuttgart 1999
- MAY, R. M. (Hrsg.): *Theoretische Ökologie*, Weinheim 1980
- NÖBAUER, W. / TIMISCHL, W.: *Mathematische Modelle in der Biologie*, Braunschweig 1979
- PEITGEN, H.-O. / JÜRGENS, H. / SAUPE, D.: *Chaos - Bausteine der Ordnung*, Berlin 1994
- OPFER, G.: *Numerische Mathematik für Anfänger*, 2. Aufl., Braunschweig 1994
- ORTLIEB, C. P. / v. DRESKY, C. / GASSER, I. / GÜNZEL, S.: *Mathematische Modellierung. Eine Einführung in zwölf Fallstudien*, Wiesbaden 2009
- STRAUSS, W. A.: *Partielle Differentialgleichungen*, Braunschweig 1995
- WALDMANN, K.-H. / STOCKER, U.M.: *Stochastische Modelle*, Berlin - Heidelberg - New York 2004

3 Modellierung mit linearen Gleichungen

Insbesondere durch den täglichen Umgang mit Waren und Geld erscheint uns Linearität als eine Art „natürlicher“ funktionaler Zusammenhang: Drei Kilo Äpfel kosten dreimal soviel wie ein Kilo, und der für drei Kilo Äpfel und zwei Kilo Birnen zu zahlende Preis ergibt sich als entsprechende Linearkombination der Kilopreise von Äpfeln und Birnen. Abweichungen von diesem Schema gibt es zwar, sind aber nicht der Normalfall und werden als „Sonderangebot“ oder „Mengenrabatt“ gebührend herausgestellt.

Ein linearer Zusammenhang zwischen zwei oder mehr variablen Größen ist die *einfachste* Modellannahme, die man machen kann und daher auch macht, solange keine andersartigen Informationen vorliegen. Nicht nur für die Modellierung, sondern ebenso für die anschließende mathematische Behandlung sind lineare Modelle besonders einfach, weil das Verhalten ihrer Lösungen sich in der Regel auf analytischem Wege charakterisieren lässt, was natürlich numerische Berechnungen, besonders bei großen Systemen, keineswegs überflüssig macht.

Die in diesem Kapitel untersuchten Modelle führen auf drei verschiedene mathematische Problemklassen, die jeweils durch eine quadratische, reelle *Systemmatrix* $A \in \mathbb{R}^{n,n}$ gekennzeichnet sind: Zum einen Iterationsprozesse der Form

$$x(t+1) = A x(t) + u(t) \quad (t = 0, 1, 2, \dots),$$

in diskreter und Systeme linearer Differentialgleichungen

$$\dot{x}(t) = A x(t) + u(t) \quad (t \in \mathbb{R})$$

in kontinuierlicher Zeit t , zum anderen lineare Gleichungssysteme

$$A x = b$$

von n Gleichungen für ebenso viele Unbekannte, sogenannte *statische* Modelle, in denen von der Zeit abstrahiert wird.

3.1 Drei lineare Modelle

3.1.1 Populationswachstum mit altersabhängiger Fruchtbarkeit und Sterblichkeit

In den Wachstumsmodellen des letzten Kapitels wurden die betrachteten Populationen als homogen angesehen, die Individuen also alle gleichermaßen zur Reproduktion fähig und vom Tode bedroht. Für menschliche Populationen ebenso wie für Populationen langlebiger Tierarten kann diese Annahme, also die Abstraktion von der Altersstruktur, zu einer völligen Fehleinschätzung der Populationsdynamik führen. Denn natürlich hängen die Reproduktionsfähigkeit und die Sterbewahrscheinlichkeit der Individuen von ihrem Alter ab.

Es kommt hinzu, dass die eigentlich interessanten Informationen, die ein Modell liefern soll, häufig gerade in der Altersstruktur stecken, während die absolute Größe der Population eher zweitrangig ist:

- So hängt etwa die Frage, „ob unsere Renten im Jahre 2020 noch sicher sind“, nicht an der absoluten Bevölkerungszahl, sondern daran, wie sich dann das Verhältnis zwischen Jungen und Alten gestaltet.

- Die Größe von Fischen ist wesentlich eine Funktion ihres Lebensalters. Der Fischfang mit einer bestimmten Maschenweite des Netzes fischt die größeren und damit älteren Fische ab. Bei der Festlegung von Fangquoten und erlaubten Maschenweiten der Netze ist daher die Altersstruktur der Fischpopulation zu berücksichtigen.

Bei Berücksichtigung der Altersstruktur kommt die Zeit gleich doppelt ins Spiel, nämlich in Form des je aktuellen Zeitpunkts t und des Lebensalters s . Der Zustand der Population zu jedem Zeitpunkt t ist durch die Angabe bestimmt, wieviele ihrer Individuen das Lebensalter s haben, wobei s das Intervall der möglichen Lebensalter durchläuft. Der Zustand zu jedem Zeitpunkt t ist also eine reelle Funktion auf der Menge der möglichen Lebensalter. Je nachdem, ob die Zeitvariablen t und s diskret oder kontinuierlich gewählt werden, kommt man zu verschiedenen Modellen.

Dabei bietet es sich an, die beiden Variablen gleich zu behandeln, da es ansonsten Schwierigkeiten bereitet, den elementaren Sachverhalt mathematisch zu erfassen, dass jedes Individuum in einem bestimmten Zeitintervall um die Länge eben dieses Zeitintervalls älter wird. t und s werden also entweder beide kontinuierlich oder beide diskret gewählt, wobei im zweiten Fall der Zeitschritt gleich groß sein sollte. In der Versicherungswirtschaft (z. B. Lebens- und Rentenversicherung) ist es üblich, menschliche Populationen in Jahrgangsklassen aufzuteilen und ihre Entwicklung in Zeitschritten von einem Jahr zu modellieren. In statistischen Jahrbüchern findet man die folgenden Daten für ein bestimmtes Jahr (fester Zeitpunkt t):

- Aufgeschlüsselt nach Frauen und Männern, die Anzahl der Personen im i -ten Lebensjahr (also vom Alter $s \in [i - 1, i]$ für $i = 1, \dots, n$ z. B. mit $n = 90$, wobei die über 90-Jährigen zu einer einzigen Gruppe zusammengefasst sind,
- aufgeschlüsselt nach Frauen und Männern, den Anteil der Personen des Alters i , die das Alter $i + 1$ erreicht ($i = 1, \dots, n - 1$),
- die Anzahl der Lebendgeborenen, die während eines Jahres auf 1000 Mütter der Altersklasse $i = 1, \dots, n$ kommen,
- den Anteil der Mädchen und Jungen an den Lebendgeborenen.

Die Beteiligung der Männer am Reproduktionsprozess wird als so geringfügig eingeschätzt, dass er vernachlässigt werden kann.

Setzt man die erhobenen Daten als konstant voraus, so ergeben sich daraus lineare Modelle zur Prognose des weiteren Bevölkerungswachstums.

Die weibliche Population. Einfacher Leslie-Prozess

Zur Entwicklung der weiblichen Population braucht wegen der in die Daten eingehenden Annahmen die männliche nicht betrachtet zu werden. Es bezeichne

- $x_i(t)$ die Anzahl der Frauen der Altersklasse i zum Zeitpunkt t ,
- u_i den Anteil der Frauen vom Alter i , die das Alter $i + 1$ erreicht,
- a_i die Anzahl lebendgeborener Töchter je Frau der Altersklasse i .

Die Größen der weiblichen Altersklassen des nächsten Jahres ergeben sich dann aus denen des aktuellen Jahres gemäß

$$\begin{aligned} x_1(t+1) &= \sum_{j=1}^n a_j x_j(t) \\ x_i(t+1) &= u_{i-1} x_{i-1}(t) \text{ für } i = 2, \dots, n \end{aligned}$$

oder für $x = (x_1, \dots, x_n)$

$$x(t+1) = A x(t)$$

mit

$$A = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_{n-1} & a_n \\ u_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & u_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \cdots & u_{n-1} & 0 \end{pmatrix}. \quad (3.1)$$

In dieser Formulierung bleiben alle Personen, die älter als n Jahre sind, außerhalb der Betrachtung.

Population mit zwei Geschlechtern

Über die bereits eingeführten Größen hinaus bezeichne

- $y_i(t)$ die Anzahl der Männer der Altersklasse i zum Zeitpunkt t ,
- v_i den Anteil der Männer vom Alter i , die das Alter $i+1$ erreicht,
- b_i die Anzahl lebendgeborener Söhne je Frau der Altersklasse i .

Die Größen der männlichen Altersklassen des nächsten Jahres ergeben sich dann aus den Größen der männlichen und weiblichen Altersklassen des aktuellen Jahres gemäß

$$y_1(t+1) = \sum_{j=1}^n b_j x_j(t)$$

$$y_i(t+1) = v_{i-1} y_{i-1}(t) \text{ für } i = 2, \dots, n$$

oder für $y = (y_1, \dots, y_n)$

$$y(t+1) = B x(t) + C y(t)$$

mit

$$B = \begin{pmatrix} b_1 & b_2 & b_3 & \cdots & b_{n-1} & b_n \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, C = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ v_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & v_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \cdots & v_{n-1} & 0 \end{pmatrix}.$$

Eine Mäusepopulation. Reiner Geburtenprozess

Zu Testzwecken soll hier noch ein Modell mit nur drei Altersklassen betrachtet werden: Jemand kauft sich ein Mäusepaar. Es handelt sich um eine unsterbliche Rasse, die sich nach den folgenden Gesetzmäßigkeiten fortpflanzt:

- Jedes erwachsene Mäusepaar bringt alle drei Wochen vier neue Paare zur Welt,
- jedes neugeborene Paar wird nach sechs Wochen erwachsen.

Um die beiden Zeiträume, die für die Reproduktion eine Rolle spielen, erfassen zu können, betrachte ich drei verschiedene Altersklassen und einen Zeitschritt von drei Wochen.

- $x_1(t)$ bezeichne die Anzahl der Mäusepaare, die zum Zeitpunkt t zwischen 0 und 3 Wochen alt sind (Kinder).
- $x_2(t)$ bezeichne die Anzahl der Mäusepaare, die zum Zeitpunkt t zwischen 3 und 6 Wochen alt sind (Jugendliche).
- $x_3(t)$ bezeichne die Anzahl der Mäusepaare, die zum Zeitpunkt t älter als sechs Wochen sind (Erwachsene).

Dann gilt offenbar

$$\begin{aligned}x_1(t+1) &= 4x_3(t) \\x_2(t+1) &= x_1(t) \\x_3(t+1) &= x_2(t) + x_3(t)\end{aligned}\tag{3.2}$$

Hier wird, anders als beim einfachen Leslie-Prozess, die letzte Altersklasse nicht wieder verlassen. Mit der Matrix

$$A = \begin{pmatrix} 0 & 0 & 4 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}\tag{3.3}$$

ergibt sich wieder die Form

$$x(t+1) = A x(t).$$

Mit einiger Phantasie lässt sich bei Betrachtung von Tabelle 3.1 vermuten, dass die Population sich

t	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_1	1	0	0	4	4	4	20	36	52	132	276	484	1012	2116	4052	8100
x_2	0	1	0	0	4	4	4	20	36	52	132	276	484	1012	2116	4052
x_3	0	0	1	1	1	5	9	13	33	69	121	253	529	1013	2025	4141

Tabelle 3.1: Entwicklung der Mäusepopulation mit $x_1(0) = 1$, $x_2(0) = 0$, $x_3(0) = 0$

langfristig in jedem Zeitschritt von drei Wochen verdoppelt und sich einem Größenverhältnis der Altersklassen von 2 zu 1 zu 1 annähert.

Der Gehalt dieser Vermutung wird deutlicher, wenn man sich die Eigenwerte und zugehörigen Eigenvektoren der Matrix A aus (3.3) ansieht, deren Bedeutung für die Lösung linearer dynamischer Systeme noch in allgemeinem Rahmen behandelt werden soll: Es ist $A w_k = \lambda_k w_k$ für $k = 1, 2, 3$ mit

$$\lambda_1 = 2, w_1 = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \lambda_{2,3} = -\frac{1}{2} \pm \frac{i}{2} \sqrt{7}, w_{2,3} = \begin{pmatrix} -1 \mp i \sqrt{7} \\ -\frac{3}{2} \pm \frac{i}{2} \sqrt{7} \\ 1 \end{pmatrix}.$$

Jedem Eigenwert λ_k und zugehörigem Eigenvektor w_k entspricht nun aber eine (reelle oder komplexe) Lösung

$$x_k(t) = \lambda_k^t w_k$$

von (3.2), denn

$$x_k(t+1) = \lambda_k^{t+1} w_k = A \lambda_k^t w_k = A x_k(t).$$

Alle Lösungen von (3.2) lassen sich dann darstellen als

$$x(t) = a_1 x_1(t) + a_2 x_2(t) + a_3 x_3(t) \text{ mit } a_1, a_2, a_3 \in \mathbb{C}.$$

Die *spezielle Lösung* von (3.2) mit

$$\mathbf{x}(t) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

ergibt sich daraus (nach Lösen eines linearen Gleichungssystems für a_1, a_2, a_3) zu

$$\mathbf{x}(t) = 2^t \begin{pmatrix} \frac{1}{4} \\ \frac{1}{8} \\ \frac{1}{8} \end{pmatrix} + \left(-\frac{1}{2} - \frac{i\sqrt{7}}{2}\right)^t \begin{pmatrix} \frac{3}{8} - \frac{i}{8\sqrt{7}} \\ -\frac{1}{16} + \frac{11i}{16\sqrt{7}} \\ -\frac{1}{16} - \frac{5i}{16\sqrt{7}} \end{pmatrix} + \left(-\frac{1}{2} + \frac{i\sqrt{7}}{2}\right)^t \begin{pmatrix} \frac{3}{8} + \frac{i}{8\sqrt{7}} \\ -\frac{1}{16} - \frac{11i}{16\sqrt{7}} \\ -\frac{1}{16} + \frac{5i}{16\sqrt{7}} \end{pmatrix} \quad (3.4)$$

Die komplexen Größen in dieser Formel treten allesamt in Gestalt von Summen konjugiert komplexer Zahlen auf, deren Imaginärteile sich deswegen wegheben. Es handelt sich also um eine Formel für eine reelle Lösung, die sich ohne den Umweg über die komplexen Zahlen aber nicht so ohne Weiteres angeben ließe.

Mit wachsendem t erhält $\lambda_1 = 2$, also der größte der drei Eigenwerte, deren t -te Potenz gebildet wird, in (3.4) ein immer größeres Gewicht. Der Absolutbetrag der anderen beiden Eigenwerte ist $\sqrt{2}$. Daher ist

$$\lim_{t \rightarrow \infty} \frac{\mathbf{x}(t)}{2^t} = \begin{pmatrix} 1/4 \\ 1/8 \\ 1/8 \end{pmatrix}.$$

Das ist die präzise Fassung der oben etwas vage formulierten Vermutung, dass die Population sich langfristig in jedem Zeitschritt von drei Wochen verdoppelt und sich einem Größenverhältnis der Altersklassen von 2 zu 1 zu 1 annähert.

$$\mathbf{x}(t) \sim 2^t \begin{pmatrix} 1/4 \\ 1/8 \\ 1/8 \end{pmatrix}$$

ist eine andere Fassung derselben asymptotischen Aussage. Die *absolute Abweichung* von diesem langfristigen Verhalten wird mit wachsendem t übrigens immer größer und ist von der Größenordnung $\sqrt{2}^t$. Dagegen geht die *relative Abweichung* gegen 0.

Aufgabe 3.1

Jemand kauft sich ein Kaninchenpaar, das einer von drei verschiedenen Rassen angehört. Das Wachstum der Kaninchen-Population, gemessen in Anzahlen der Paare, soll berechnet werden. Allen drei Rassen gemeinsam ist:

- Jedes neugeborene Paar wird nach sechs Wochen erwachsen.

Die drei Rassen unterscheiden sich hinsichtlich ihrer Reproduktions- und Sterberaten:

- Jedes erwachsene Paar bringt alle sechs Wochen ein neues Paar zur Welt und stirbt nie. (Überlegungen unter dieser Annahme gehen bereits auf Leonardo von Pisa (1170 - 1240), bekannter und dem Namen Fibonacci, zurück.)
- Jedes erwachsene Paar bringt einmal, nämlich sechs Wochen nach Eintritt ins Erwachsenenalter, zwei neue Paare zur Welt, bekommt anschließend keine Jungen mehr, stirbt aber nie.
- Jedes erwachsene Paar bringt einmal, nämlich sechs Wochen nach Eintritt ins Erwachsenenalter, drei neue Paare zur Welt und stirbt anschließend.

Welche der drei Rassen wächst am schnellsten? Welche prozentuale Verteilung auf die Altersklassen stellt sich langfristig ein?

3.1.2 Ökonomische Verflechtung

Zur Produktion von Gütern und Dienstleistungen werden in einer modernen Wirtschaft andere Güter und Dienstleistungen benötigt: Zur Produktion von Weizen etwa braucht man Saatgut, Dünger, Maschinen usw., zur Produktion von Maschinen Stahl und andere Werkstoffe, zur Herstellung bzw. Gewinnung von Stahl wiederum Maschinen, für alles braucht man Energie und menschliche Arbeitskraft, für die Energiegewinnung wiederum Maschinen und für die Reproduktion der Arbeitskraft Nahrungsmittel. Irgendwie hängt alles mit allem zusammen. Dadurch wird es schwierig, Fragen wie die folgende zu beantworten: Wieviel durch Ölverbrennung gewonnene Energie und wieviel menschliche Arbeitszeit „stecken“ in einem Kilo Brot? Zu ihrer Beantwortung genügt es nicht, den Produktionsprozess des Brotes zurück zu verfolgen, denn die zu seiner Herstellung benötigten Maschinen verbrauchen Energie nicht nur im Betrieb, sondern bereits zu ihrer Herstellung, die Energiegewinnung benötigte ihrerseits Maschinen usw.

In diesem Abschnitt sollen lineare Modelle vorgestellt werden, mit deren Hilfe sich derartige Fragen angehen lassen. Die Darstellung bewegt sich allerdings auf einer eher abstrakten Ebene. Was nämlich zur Beantwortung der oben gestellten Fragen hinzu kommen muss, sind Daten, und die sind nur schwer und nur in Einzelfällen zu bekommen.¹

Volkswirtschaftliche Gesamtrechnungen: Geldflüsse

In den Veröffentlichungen des Statistischen Bundesamtes der Bundesrepublik Deutschland findet man unter der Überschrift *Volkswirtschaftliche Gesamtrechnungen* Input-Output-Tabellen, in denen nachträglich Geldflüsse zwischen verschiedenen Wirtschaftssektoren aufgeführt sind, denen Flüsse von Gütern und Dienstleistungen (in entgegengesetzter Richtung) gegenüberstehen. Tabelle 3.2 ist dafür ein Bei-

Tabelle 3.2: Input-Output-Tabelle 2005, Inländische Produktion (Herstellungspreise) in Mrd. Euro;
Quelle: Statistisches Bundesamt www-genesis.destatis.de/genesis/online/, 16.05.2009

Aufkommen	Verwendung				
	Primärer Bereich	Sekundärer Bereich	Tertiärer Bereich	Letzte Verwendung	davon Exporte
Primärer Bereich	7.3	22.6	1.6	16.3	4.1
Sekundärer Bereich	8.0	571.6	114.5	1092.9	615.5
Tertiärer Bereich	10.5	314.9	640.6	1401.9	137.0
Importe	4.0	320.5	102.9		
Arbeitnehmerentgelte	7.6	362.5	760.8		
Gewinne	7.8	98.6	432.8		

spiel. Die Volkswirtschaft der Bundesrepublik Deutschland ist hier in drei Sektoren aufgeteilt, den primären Landwirtschaftssektor, den sekundären Sektor der Industrieproduktion und den tertiären Sektor der Dienstleistungen. Wirtschaftliche Aktivitäten sind hier gemessen durch das Geld, das für sie gezahlt wurde. Dabei fließt das Geld von den Spalten in die Zeilen, die realen wirtschaftlichen Leistungen von den Zeilen in die Spalten. Die Zahl 22.6 in der ersten Zeile und zweiten Spalte von Tabelle 3.2 bedeutet also, dass landwirtschaftliche Produkte zur Weiterverarbeitung in den Industriesektor geflossen sind, für die 22.6 Mrd. Euro gezahlt wurden. Die Zahl 760.8 in der fünften Zeile und dritten Spalte bedeutet, dass im Dienstleistungssektor 760.8 Mrd. Euro an Löhnen und Gehältern gezahlt worden sind zur Vergütung der dort von Arbeitnehmern geleisteten Arbeit.

¹Es ist schon bemerkenswert, dass wir über die von uns selbst geschaffenen gesellschaftlichen Verhältnisse oft sehr viel schlechter informiert sind als über „natürliche“ Sachverhalte, die wir gar nicht beeinflussen können.

Nicht erfasst bzw. nicht getrennt ist hier der staatliche Sektor (Steuern und Staatsausgaben). Wegen der Verschiedenheit der in den einzelnen Sektoren zusammengefassten Waren (Güter und Dienstleistungen) ist als einheitliches Maß nur das Geld möglich. Das hat zur Folge, dass sich Preise und Mengen aus der Tabelle nicht mehr getrennt herauslesen lassen, deren Zahlen immer schon aus der Multiplikation von Mengen und Preisen hervorgehen. Die Frage, wie sich Produktionspreise aus der Vernetzung verschiedener Sektoren ergeben, ist damit ebenso wenig zu beantworten wie die nach Materialflüssen.

Warenproduktion mittels Waren

In einer Volkswirtschaft oder auch einem Einzelbetrieb werden n Waren (Güter und Dienstleistungen) in n verschiedenen Produktionsprozessen hergestellt, für jede Ware einer.² Zur Produktion werden m weitere Waren benötigt, die nicht selbst erzeugt werden, wie z. B. Rohstoffe oder menschliche Arbeitskraft, und die hier als *Inputs* bezeichnet werden. Die Inputs gehen ebenso wie die selbst produzierten Waren in den Produktionsprozess ein, soweit das eben technisch erforderlich ist. Produziert wird für den Konsum, der hier nicht weiter ausdifferenziert wird.

Alle Waren einschließlich der Inputs werden in geeigneten *Mengeneinheiten*, also nicht in Geld gemessen. Es bezeichne im Folgenden $[ME_i]$ die Mengeneinheit der produzierten Ware i ($i = 1, \dots, n$) und $[MI_j]$ die Mengeneinheit von Input j ($j = 1, \dots, m$). Betrachtet wird der Produktionsprozess über einen

Tabelle 3.3: Allgemeine Input-Output-Tabelle (Mengeneinheiten)

	Verwendung			
	Produktion von			Konsum
	Ware 1	...	Ware n	
Ware 1	$A_{11} [ME_1]$...	$A_{1n} [ME_1]$	$B_1 [ME_1]$
⋮	⋮	⋮	⋮	⋮
Ware n	$A_{n1} [ME_n]$...	$A_{nn} [ME_n]$	$B_n [ME_n]$
Input 1	$C_{11} [MI_1]$...	$C_{1n} [MI_1]$	
⋮	⋮	⋮	⋮	
Input m	$C_{m1} [MI_m]$...	$C_{mn} [MI_m]$	

bestimmten Zeitraum, etwa ein Jahr. Es bezeichne (vgl. Tabelle 3.3) für $i, k = 1, \dots, n, j = 1, \dots, m$

- $A_{ik} [ME_i]$ die Menge der Ware i , die in Produktion der Ware k eingegangen ist,
- $B_i [ME_i]$ den Nettooutput der Ware i , also die für den Konsum produzierte Menge,
- $C_{jk} [MI_j]$ die Menge des Inputs j , die in die Produktion der Ware k eingegangen ist.

Die Summe der k -ten Zeile von Tabelle 3.3

$$X_k = A_{k1} + \dots + A_{kn} + B_k \quad (3.5)$$

ist dann die Gesamtproduktion der Ware k im betrachteten Zeitraum. Zu ihrer Herstellung wurden die in der k -ten Spalte von Tabelle 3.3 aufgeführten Mengen benötigt, also $A_{ik} [ME_i]$ der Ware i ($i = 1, \dots, n$)

²Nicht erfasst sind damit sogenannte *Kuppelproduktionen*, in denen ein einziger Prozess mehrere Waren gleichzeitig erzeugt, wie z. B. elektrische Energie und Fernwärme in einem Kraftwerk. Nicht erfasst ist auch die Möglichkeit, die Technik je nach anfallendem Bedarf zu wechseln, also das Problem der *Technikwahl*.

und C_{jk} [MI_j] des Inputs j ($j = 1, \dots, m$). Die Verhältniszahlen

$$a_{ik} = \frac{A_{ik}}{X_k} [\text{ME}_i/\text{ME}_k] \text{ und } c_{jk} = \frac{C_{jk}}{X_k} [\text{MI}_j/\text{ME}_k] \quad (3.6)$$

geben also an, wieviele Einheiten der Ware i bzw. des Inputs j zur Produktion einer Mengeneinheit der Ware k im Durchschnitt benötigt wurden. Die Linearität des hier entwickelten Modells besteht nun darin, diese empirisch festgestellten Durchschnittswerte *als Konstanten des Produktionsprozesses zu interpretieren*, also für $k = 1, \dots, n$

- a_{ik} [ME_i/ME_k] als Menge der Ware i , die zur Herstellung einer Mengeneinheit der Ware k benötigt wird ($i = 1, \dots, n$),
- c_{jk} [MI_j/ME_k] als Menge des Inputs j , die zur Herstellung einer Mengeneinheit der Ware k benötigt wird ($j = 1, \dots, m$).

Bruttoproduktion bei gegebener Nachfrage

Welche Bruttoproduktion $x = (x_1, \dots, x_n)$ und welcher Input $y = (y_1, \dots, y_m)$ ist erforderlich, um einen Nettooutput $b = (b_1, \dots, b_n)$ herzustellen, der durch einen gegebenen Konsumbedarf definiert ist? Zur Herstellung der Menge x_k [ME_k] der Ware k wird die Menge $a_{ik} x_k$ [ME_i] der Ware i benötigt. Diese Mengen sind über den Nettooutput hinaus zu produzieren, es gilt also:

$$x_i = \sum_{k=1}^n a_{ik} x_k + b_i \text{ für } i = 1, \dots, n$$

oder mit der *Produktionsmatrix* $A = (a_{ik})$

$$x = A x + b. \quad (3.7)$$

Die mathematische Frage in diesem Zusammenhang ist, ob und unter welchen Voraussetzungen zu jedem $b \geq 0$ eine eindeutig bestimmte Lösung $x \geq 0$ dieses linearen Gleichungssystems existiert, was damit gleichbedeutend ist, dass mit der (n, n) -Einheitsmatrix I die inverse Matrix $(I - A)^{-1}$ existiert und

$$(I - A)^{-1} \geq 0.$$

Eine Teilantwort auf diese Frage liefert die Ermittlung der Produktionskoeffizienten a_{ik} aus der Input-Output-Tabelle 3.3. Denn aus dem ersten Teil von (3.6) und (3.5) folgt

$$X = A X + B.$$

Zu dem in der Input-Output-Tabelle enthaltenen Konsumvektor $b = B$ besitzt also (3.7) die Lösung $x = X \geq 0$.

Definition 3.1.1

Eine nicht negative (n, n) -Matrix $A \in \mathbb{R}^{n,n}$ heie „produktiv“, wenn fur jedes $b \in \mathbb{R}^n$ mit $b \geq 0$ das lineare Gleichungssystem (3.7) eine eindeutige Losung $x \in \mathbb{R}^n$ mit $x \geq 0$ besitzt, wenn also $I - A$ invertierbar ist und

$$(I - A)^{-1} \geq 0.$$

Diese vor allem in der konomischen Literatur gebruchliche Wortwahl kommt natrlich aus dem Zusammenhang des hier vorgestellten Modells. Die Produktionsmatrix ist produktiv, wenn es mglich ist, beliebige Nettooutputs herzustellen, wenn also in allen Komponenten mehr Waren produziert als in der Produktion verbraucht werden. Im Falle $n = 1$, in dem die Matrix A nur aus einer einzigen Zahl $a \geq 0$

besteht, ist A genau dann produktiv, wenn $a < 1$. Für $n > 1$ ist die Situation komplizierter. Darauf wird noch eingegangen.

Sei $(I - A)^{-1} = (d_{ik})$. Offenbar ist dann

$$x_i = \sum_{k=1}^n d_{ik} b_k \text{ für } i = 1, \dots, n,$$

was die folgende Interpretation der Koeffizienten d_{ik} von $(I - A)^{-1}$ impliziert:

- d_{ik} $[ME_i/ME_k]$ ist die Bruttoproduktion der Ware i , die erforderlich ist, um die Nettoproduktion einer Mengeneinheit der Ware k zu gewährleisten.

Externe Inputs bei gegebener Nachfrage

Ist (im Falle einer produktiven Matrix A) bei gegebener Nachfrage b die Bruttoproduktion $x = (I - A)^{-1} b$ aus (3.7) errechnet, so lässt sich der Vektor $y \in \mathbb{R}^m$ der notwendigen externen Inputs sofort ermitteln: Zur Produktion der Menge x_k $[ME_k]$ der Ware k ist die Menge $c_{jk} x_k$ $[MI_j]$ des Inputs j erforderlich, für den gesamten Bruttoproduktionsvektor x also die Menge

$$y_j = \sum_{k=1}^n c_{jk} x_k \quad (j = 1, \dots, m).$$

Mit der Inputmatrix $C = (c_{jk}) \in \mathbb{R}^{m,n}$ ergibt sich also als Vektor der notwendigen externen Inputs

$$y = C x = C (I - A)^{-1} b. \quad (3.8)$$

Produktionspreise

Ein Betrieb, der die Waren $1, \dots, n$ herstellt, muss die zur Produktion notwendigen Inputs $1, \dots, m$ zu bestimmten Preisen einkaufen.

- q_j $[DM/MI_j]$ sei der Einkaufspreis des Inputs j ($j = 1, \dots, m$).

Die Produktion selbst verursache dagegen keine Kosten. Diese Vorstellung setzt voraus, dass alle anfallenden Kosten als externe Inputs (etwa Materialverbrauch, Arbeitseinsatz, Abnutzung von Maschinen) verbucht werden. Bei der Produktion des Nettooutputs b fallen dann wegen (3.8) die Kosten

$$K = \sum_{j=1}^m q_j y_j = q^T y = q^T C (I - A)^{-1} b = b^T (I - A^T)^{-1} C^T q$$

an (das hochgestellte T kennzeichne die transponierte Matrix). Der Vektor

$$(I - A^T)^{-1} C^T q$$

ist dann der Vektor der *Produktionspreise*: Die i -te Komponente dieses Vektors sind die Kosten in $[DM/ME_i]$, die bei der Produktion einer Mengeneinheit der Ware i anfallen. Würde der Betrieb alle von ihm produzierten Waren zu den Produktionspreisen verkaufen, so würde er bei beliebigem Nettooutput b weder Gewinn noch Verlust machen. Für einen sicheren Gewinn sind die Verkaufspreise also entsprechend höher anzusetzen.

Arbeitswerte

In einer geschlossenen Volkswirtschaft ohne Im- und Exporte ist die einzige Ware, die nicht aus Produktionsprozessen hervorgeht und daher als externer Input betrachtet werden muss, die menschliche Arbeitskraft. Für die ökonomische Klassik³ war daher der Wert einer Ware bestimmt durch die zu ihrer Herstellung im gesellschaftlichen Durchschnitt erforderliche Arbeitszeit. Zur Unterscheidungen von anderen Wert-Definitionen spricht man auch vom Arbeitswert. Unterstellt wird hier ein einheitliches Maß für jede Art von Arbeit, nämlich die Zeit. Es wird also nicht zwischen qualifizierter und weniger qualifizierter Arbeit unterschieden.

Dieser Ansatz lässt sich in das hier betrachtete Modell integrieren: Es ist $m = 1$, und $C = c^T = (c_1, \dots, c_n)$ ist ein Zeilenvektor.

- c_i [Arbeitsstunden / ME_{*i*}] ist die zur Herstellung einer Mengeneinheit der Ware i unmittelbar erforderliche Arbeitszeit.

c_i ist noch nicht der Arbeitswert der Ware i , dazu muss nämlich noch die Arbeitszeit hinzu gezählt werden, die in den anderen Waren steckt, die zur Produktion der Ware i benötigt werden. Die Arbeitswerte ergeben sich vielmehr als Spezialfall von (3.8): Um den Nettooutput b herzustellen, sind

$$c^T (I - A)^{-1} b$$

Arbeitsstunden erforderlich. Damit ist

$$(I - A^T)^{-1} c \tag{3.9}$$

der Vektor der Arbeitswerte. Die i -te Komponente dieses Vektors gibt an, welche Arbeitszeit erforderlich ist, um gerade eine Mengeneinheit der Ware i herzustellen.

Aufgabe 3.2

Führen Sie die Überlegungen dieses Abschnitts an Tabelle 3.2 durch, indem Sie die Geldeinheit (Mrd. DM) als Maß für die jeweiligen Mengeneinheiten nehmen. Welche Lohnsumme und welche Summe an Gewinnen und Abschreibungen „steckt“ jeweils in Waren für 1 Mrd. DM des primären, sekundären und tertiären Sektors?

3.1.3 Meinungsbildung in Gruppen. Das Abelson-Modell

Gegenstand der folgenden Modellierung ist eine Gruppe von n Personen, die sich zu einer bestimmten Frage eine Meinung bilden soll. Dabei werden bei den Betrachtungen alle möglichen inhaltlichen Aspekte von vornherein ausgeblendet. Es wird unterstellt, daß sich die Gruppenmitglieder in bestimmter Weise gegenseitig beeinflussen und daß die Meinungsbildung auf Grund dieses Einflusses und der Anfangsmeinungen zustande kommt. Ein mathematisches Modell auf dieser Basis wurde erstmals von ABELSON (1964) in die Literatur eingeführt. Es operiert mit einem System linearer Differentialgleichungen. Hinweise auf weitere Modelle, die zum Teil mit anderen mathematischen Strukturen, insbesondere diskreten stochastischen Prozessen arbeiten, gibt BODROŽIĆ (1998). Die folgende Darstellung bezieht sich ausschließlich auf das Abelson-Modell.

Die Meinung eines einzelnen Mitglieds wird durch eine einzelne Zahl beschrieben. Abgesehen von dem damit aufgeworfenen methodischen Problem, Meinungen in irgendeiner Weise zu quantifizieren, ist damit die Einschränkung verbunden, daß die Frage, um die es geht, „eindimensional“ ist, also mit „ja“ oder „nein“ oder dazwischen liegenden Abstufungen beantwortet werden kann. Die Meinung jedes Gruppenmitgliedes verändert sich kontinuierlich unter dem Eindruck der Meinung der anderen Mitglieder. Aus der Sicht eines Einzelnen läßt sich die Veränderung seiner Meinung so beschreiben: Als Gruppenmitglied erlebe ich die Abweichung der eigenen Meinung von der eines anderen als Konflikt, auf den ich mit Anpassung an die andere Meinung reagiere. Der Anpassungsdruck ist dabei umso größer,

³HEINE / HERR (2000, 221) nennen als ihre herausragenden Vertreter Adam Smith, David Ricardo und Karl Marx.

je stärker ich mich von dem anderen Mitglied beeinflussen lasse und je stärker dessen Meinung von der meinigen abweicht.

Es bezeichne für die Gruppenmitglieder $i = 1, \dots, n$

- $x_i(t) \in \mathbb{R}$ die Meinung von Person i zum Zeitpunkt t ,
- $a_{ij} \geq 0$ ein Maß für den Einfluss, den Person j ($j \neq i$) auf die Meinungsbildung von Person i ausübt.

Die Gruppenstruktur wird also durch die Zahlen a_{ij} beschrieben. $a_{ij} = 0$ bedeutet, dass Person j keinerlei Einfluss auf Person i ausübt. Ansonsten gilt: Je größer der Einfluss, desto größer a_{ij} .

Da sich in konkreten Gruppen derartige Einflussfaktoren schwerlich quantitativ fassen lassen, kann es bei mathematischen Modellen, die damit operieren, nicht um quantitative Prognosen gehen, sondern nur um qualitative, strukturelle Aussagen.

Mit den eingeführten Größen ist nun für $i \neq j$

$$a_{ij}(x_j - x_i)$$

der Anpassungsdruck, den Person i durch den Einfluss der Person j erlebt. Das Abelson-Modell geht nun davon aus, dass Person i diesem Druck einfach nachgibt, wobei sich die Einflüsse der anderen Gruppenmitglieder aufsummieren:

$$\dot{x}_i = \sum_{j \neq i} a_{ij}(x_j - x_i) \quad (i = 1, \dots, n). \quad (3.10)$$

Fasst man die aktuellen Meinungen der Mitglieder zu einem Vektor $x(t)$ zusammen, so erhält man die äquivalente Formulierung

$$\dot{x} = Ax$$

mit einer Matrix $A = (a_{ij})$, für die

$$a_{ij} \geq 0 \text{ für } i \neq j \text{ und } \sum_{j=1}^n a_{ij} = 0 \quad (i, j = 1, \dots, n).$$

Die Zeilensummen der Matrix A sind also alle gleich 0.

Die völlige Beliebigkeit der Wahl der x_i sieht zunächst etwas verwirrend aus, da nicht klar ist, wie man diese Zahlen interpretieren soll. Tatsächlich kann man sich die Meinungen auf einer Skala zwischen zwei Werten x_{\min} und x_{\max} angesiedelt denken, die dann im Verlaufe des Meinungsbildungsprozesses nicht verlassen wird. Es gilt nämlich

Satz 3.1.2

Sei $x(t)$ eine für alle $t \geq 0$ definierte Lösung von (3.10). Gilt dann für zwei reelle Zahlen x_{\min} und x_{\max}

$$x_{\min} \leq x_i(0) \leq x_{\max} \text{ für } i = 1, \dots, n,$$

so gilt für alle Zeiten $t \geq 0$

$$x_{\min} \leq x_i(t) \leq x_{\max} \text{ für } i = 1, \dots, n.$$

Beweis: Zu zeigen ist die positive Invarianz (AMANN (1983, 233 ff.)) des Würfels

$$X = [x_{\min}, x_{\max}]^n.$$

Diese ist damit äquivalent (AMANN (1983, 241/242)), dass für alle Randflächen von X der durch die rechten Seiten von (3.10) definierte Vektor nicht aus X hinausweist, daß also

$$\begin{aligned} x \in X, x_i = x_{\min} &\Rightarrow \dot{x}_i \geq 0, \\ x \in X, x_i = x_{\max} &\Rightarrow \dot{x}_i \leq 0. \end{aligned}$$

Diese Implikationen ergeben sich aber unmittelbar aus der Tatsache, daß im ersteren Falle $x_j \geq x_i$ und im letzteren $x_j \leq x_i$ für alle $j \neq i$. ■

Dieser Satz erleichtert nicht nur die Interpretation der Zahlen, die die Meinungen der Mitglieder darstellen sollen, sondern drückt auch eine spezifische Eigenschaft des Abelson-Modells aus: Das zu Beginn des Gruppenprozesses vorhandene Meinungsspektrum (beschrieben durch beliebige Zahlenwerte) wird im Verlaufe des Gruppenprozesses nicht verlassen. Meinungen können also nicht extremer werden, als sie zu Beginn bereits sind. Das liegt natürlich an den Modell-Voraussetzungen, nach denen die Gruppenmitglieder sich in ihrer Meinungsbildung aneinander anpassen.

Spezialfall eines Paares

In dem speziellen Fall $n = 2$ lautet das Abelson-Modell nach entsprechender Umbenennung der Variablen und Konstanten:

$$\begin{aligned} \dot{x} &= a(y - x) & , & & x(0) &= x_0 \\ \dot{y} &= b(x - y) & , & & y(0) &= y_0 \end{aligned}$$

mit $a, b \geq 0$. Damit sich die beiden Personen überhaupt in irgendeiner Weise beeinflussen, muss $a + b > 0$ sein, was im Folgenden vorausgesetzt werden soll. Für die Lösungen des Systems ergibt sich einerseits

$$\dot{x} - \dot{y} = -(a + b)(x - y) \text{ und daher } x(t) - y(t) = (x_0 - y_0) e^{-(a+b)t} ,$$

andererseits

$$b \dot{x} + a \dot{y} = 0 \text{ und daher } b x(t) + a y(t) = b x_0 + a y_0 .$$

Aus beiden Gleichungen zusammen folgt

$$\begin{aligned} x(t) &= \frac{b x_0 + a y_0}{a + b} + \frac{a}{a + b} (x_0 - y_0) e^{-(a+b)t} \\ y(t) &= \frac{b x_0 + a y_0}{a + b} + \frac{b}{a + b} (y_0 - x_0) e^{-(a+b)t} \end{aligned}$$

und damit langfristig

$$\lim_{t \rightarrow \infty} x(t) = \lim_{t \rightarrow \infty} y(t) = \frac{b x_0 + a y_0}{a + b} .$$

Langfristig tendiert das Paar also zu einer gemeinsamen Meinung, die ein gewichtetes Mittel der beiden Anfangsmeinungen ist. Die Gewichte sind dabei proportional zu den wechselseitigen Einflussfaktoren: b definiert den Einfluß des ersten auf den zweiten Partner und bestimmt das Gewicht von x_0 in der Endmeinung, a definiert den Einfluß des zweiten auf den ersten Partner und bestimmt das Gewicht von y_0 in der Endmeinung.

3.2 Die Zusammenhangsstruktur mehrdimensionaler Modelle

Besonders durch das zuletzt in 3.1.3 betrachtete Modell von Abelson zur Meinungsbildung in Gruppen sollte deutlich geworden sein, dass die quantitative Bestimmung von Modellparametern erhebliche Probleme aufwerfen kann. *In welchem Maße* etwa eine Person von einer anderen beeinflusst wird, lässt sich kaum feststellen. Ob das *überhaupt* passiert oder aber gar nicht sein kann, z. B. weil die beiden Personen sich nicht kennen, ist dagegen eine Information, an die schon eher heranzukommen ist. Was aber lässt sich allein mit solchen Informationen anfangen? Diese Frage soll jetzt untersucht werden.

In einem Modell mit n Zustandsvariablen wird (zunächst) festgestellt, durch welche anderen Zustandsvariablen sie beeinflusst werden. Für die Beispiele des letzten Abschnitts läuft das auf folgende Fragen hinaus:

- Durch welche (andere) Altersklasse des Vorjahrs ist die aktuelle Größe einer Altersklasse bestimmt?
- Welche (anderen) Waren fließen in die Produktion einer Ware ein?
- Welche anderen Personen beeinflussen die Meinungsbildung eines Gruppenmitglieds?

Stellt man die Zustandsvariablen jeweils durch einen Punkt in der Zeichenebene dar und kennzeichnet die wechselseitigen Beeinflussungen durch Pfeile von einem Punkt zum andern, so ergibt sich die mathematische Struktur eines gerichteten Graphen bzw. dessen grafische Veranschaulichung.

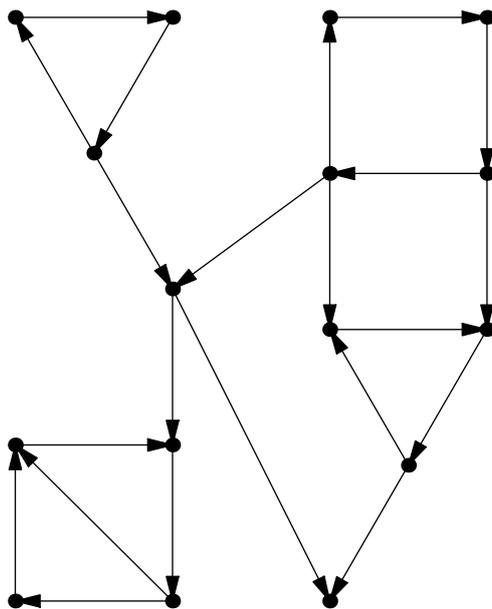


Abbildung 3.1: Beispiel eines Digraphen

Definition 3.2.1

Ein „gerichteter Graph“ oder „Digraph“ $G = (E, P)$ ist gekennzeichnet durch eine nichtleere, endliche Menge E und eine endliche Menge $P \subseteq E \times E$. Die Elemente von E werden als „Ecken“, die Elemente von P als „gerichtete Kanten“ oder „Pfeile“ bezeichnet. Anstelle von $(u, v) \in P$ schreiben wir auch $u \rightarrow v$; u heißt dann „Anfangs-“ und v „Endpunkt“ des Pfeils (u, v) .

Nach dieser Definition ist es erlaubt, dass ein Pfeil eine Ecke mit sich selbst verbindet. Für die im Folgenden vorgenommene Analyse (der Zusammenhangsstruktur von Digraphen) spielen derartige Pfeile aber keine Rolle.

Definition 3.2.2

Sei $G = (E, P)$ ein Digraph. Unter einem „Weg“ in G versteht man eine endliche Folge von Ecken $w = (u_0, u_1, \dots, u_{m-1}, u_m)$ mit

$$u_0 \rightarrow u_1 \rightarrow \dots \rightarrow u_{m-1} \rightarrow u_m .$$

m heißt dann die „Länge“ des Weges. u_0 heißt „Anfangs-“, u_m „Endpunkt“ des Weges w , w wird auch ein Weg „von u_0 nach u_m “ genannt. Ein Weg mindestens der Länge 1 mit gleichem Anfangs-

und Endpunkt heißt ein „Zyklus“. Ein Zyklus $(u_0, u_1, \dots, u_{m-1}, u_m = u_0)$ mit $m \geq 2$ und paarweise verschiedenen Ecken u_1, u_2, \dots, u_m heißt ein „Kreis“.

$v \in E$ heißt von $u \in E$ aus „erreichbar“, wenn ein Weg von u nach v existiert.

Nach dieser Definition ist für jede Ecke $u \in E$ die einelementige Folge (u) ein Weg (der Länge 0) von u nach u , weshalb in jedem Fall u von sich selbst aus erreichbar ist. (u) ist allerdings kein Zyklus, im Gegensatz zu (u, u) , falls $(u, u) \in P$. (u, u) ist aber kein Kreis. Kreise sind Zyklen durch mindestens zwei verschiedene Punkte, die keine echten Teilzyklen enthalten, in diesem Sinne also minimal sind.

Definition 3.2.3

Seien $G = (E, P)$ ein Digraph und $F \subseteq E$ nicht leer. Unter dem „von F erzeugten Unterdigraphen“ versteht man den Digraphen $H = (F, Q)$ mit

$$Q = \{(u, v) \in P : u, v \in F\}.$$

Die Ecken von H sind also die Elemente von F , die Pfeile von H sind *alle* Pfeile von G , die Ecken in F miteinander verbinden.

3.2.1 Zusammenhangskomponenten und Kondensations-Digraph

Definition 3.2.4

Ein Digraph $G = (E, P)$ heißt „kreisfrei“ oder „azyklisch“, wenn es in ihm keinen Kreis gibt.

In einem kreisfreien Digraphen kann es außer den Zyklen (u, u) der Länge 1 keine weiteren Zyklen geben.

Definition 3.2.5

Ein Digraph $G = (E, P)$ heißt *stark zusammenhängend*, wenn zwischen je zwei Ecken $u, v \in E$ ein Weg von u nach v in G existiert.

In einem stark zusammenhängenden Graphen mit mindestens zwei Ecken liegt *jede* Ecke auf einem Kreis. In diesem Sinne stehen die beiden zuletzt eingeführten Begriffe zueinander in scharfem Gegensatz. In einem noch zu präzisierenden Sinne sind sie aber auch zueinander komplementär: Jeder Digraph läßt sich auf azyklische Weise aus stark zusammenhängenden Unterdigraphen zusammensetzen.

Definition 3.2.6

In einem Digraphen $G = (E, P)$ heißt eine nichtleere Eckenmenge $F \subseteq E$ eine „starke Zusammenhangskomponente“ von G , wenn der von F erzeugte Unterdigraph stark zusammenhängend ist, dies aber für keine echte Obermenge von F gilt.

Satz 3.2.7 (ROBERTS (1976, 44/45))

Sei $G = (E, P)$ ein Digraph. Dann liegt jede Ecke $u \in E$ in genau einer starken Zusammenhangskomponente von G . Diese besteht aus allen Ecken $v \in E$, für die sowohl ein Weg von u nach v als auch ein Weg von v nach u in G existiert. Die Menge der starken Zusammenhangskomponenten bildet daher eine disjunkte Zerlegung der Eckenmenge E .

Definition 3.2.8

Der „Kondensations-Digraph“ $G^* = (E^*, P^*)$ des Digraphen $G = (E, P)$ ist definiert durch die Eckenmenge E^* der starken Zusammenhangskomponenten von E und die Pfeilmenge

$$P^* = \{(U, V) \in (E^*)^2 : U \neq V, \text{ es gibt } u \in U, v \in V \text{ mit } (u, v) \in P\}.$$

Zwei starke Zusammenhangskomponenten U, V werden also dann durch einen Pfeil miteinander verbunden, wenn in G entsprechendes für mindestens je eines ihrer Elemente gilt. Äquivalent dazu ist, daß von *jeder* Ecke in U zu *jeder* Ecke in V ein Weg in G existiert.

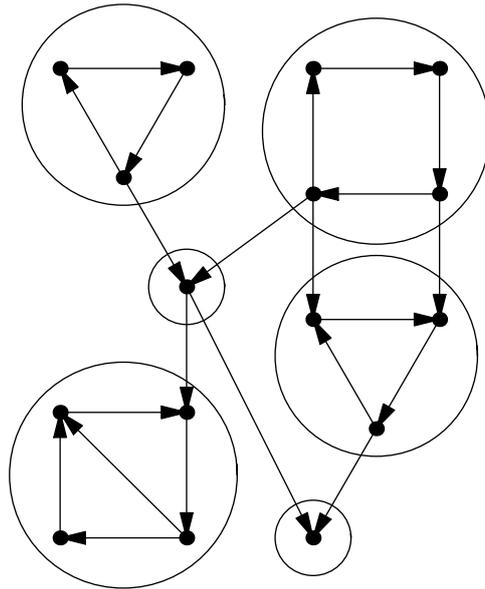


Abbildung 3.2: Starke Zusammenhangskomponenten des Digraphen aus Abbildung 3.1

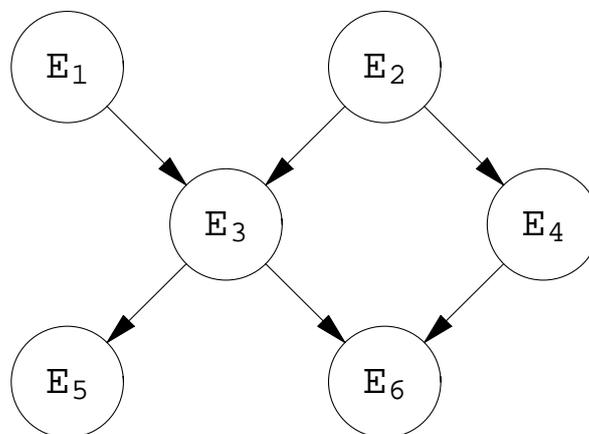


Abbildung 3.3: Kondensations-Digraph zu Abbildung 3.2

Satz 3.2.9 (ROBERTS (1976, 47))

Der Kondensationsdigraph eines beliebigen Digraphen ist azyklisch.

Definition 3.2.10

Sei $G = (E, P)$ ein Digraph. $B \subseteq E$ heißt „Eckenbasis“ von G , wenn zu jedem $v \in E$ ein $u \in B$ und ein Weg von u nach v in G existiert und wenn keine echte Teilmenge von B diese Eigenschaft hat.

In einem stark zusammenhängenden Digraphen bildet jede einzelne Ecke eine Eckenbasis. Ganz anders in einem azyklischen Digraphen:

Satz 3.2.11 (ROBERTS (1976, 47))

Ein azyklischer Digraph besitzt eine eindeutig bestimmte Eckenbasis. Sie besteht aus allen Ecken, die nicht Endpunkt eines Pfeils sind.

Definition 3.2.12

Die starken Zusammenhangskomponenten eines Digraphen G , die die eindeutig bestimmte Eckenbasis des Kondensations-Digraphen G^* bilden, werden als „Basiskomponenten“ von G bezeichnet.

3.2.2 Meinungsbildung in Gruppen

Für die in 3.1.3 betrachtete Situation der Meinungsbildung in Gruppen gibt der Kondensations-Digraph charakteristische Merkmale der Gruppenstruktur wieder. Innerhalb jeder Zusammenhangskomponente beeinflusst jedes Gruppenmitglied jedes andere direkt oder indirekt. Bei Anpassungsprozessen kann innerhalb einer Komponente langfristig mit einer einheitlichen Meinung gerechnet werden, was allerdings nicht bedeuten muss, dass jedes Mitglied innerhalb einer Komponente das selbe Gewicht hat.

Dagegen wird durch den Kondensations-Digraphen eine Hierarchie innerhalb der gesamten Gruppe aufgezeigt: Die Meinungsbildung innerhalb der Basiskomponenten wird autonom erfolgen, andere Mitglieder haben darauf keinen Einfluss. Die anderen Komponenten unterliegen dagegen dem Einfluss derjenigen Basiskomponenten, von denen aus sie erreichbar sind. Welche Stärke dieser Einfluss hat, ist damit noch nicht gesagt. Das hängt von dem quantitativen Modell und der Größe der Modellparameter ab.

3.2.3 Der Digraph einer nichtnegativen Matrix**Definition 3.2.13**

Sei $G = (E, P)$ ein Digraph, die Ecken seien durchnummeriert: $E = \{u_1, \dots, u_n\}$. Dann heißt die (n, n) -Matrix $A = (a_{ij})$ mit

$$a_{ij} = \begin{cases} 1 & \text{falls } u_j \rightarrow u_i \\ 0 & \text{sonst} \end{cases}$$

die „Adjazenzmatrix“ von G .

Beachten Sie die hier gewählte Reihenfolge der Indizes i und j . Nach Durchnummerierung der Indizes ist die Adjazenzmatrix eindeutig bestimmt, oder anders gesagt: Die Adjazenzmatrizen eines Digraphen gehen auseinander durch Vertauschung der Zeilen und entsprechende Vertauschung der Spalten hervor. Umgekehrt lässt sich zu einer gegebenen quadratischen Matrix mit Einträgen 0 und 1 der Digraph eindeutig rekonstruieren, dessen Adjazenzmatrix sie ist.

Definition 3.2.14

Sei $A = (a_{ij}) \in \mathbb{R}^{n,n}$ eine (n, n) -Matrix mit nicht negativen Einträgen $a_{ij} \geq 0$. Der Digraph $G = (E, P)$ mit $E = \{1, \dots, n\}$ und $P = \{(j, i) : a_{ij} > 0\}$ heiße der „zu A gehörige Digraph“ oder der „Digraph der Matrix A “.

Offenbar ist jeder Digraph der Digraph seiner Adjazenzmatrix.

Im folgenden soll es darum gehen, Eigenschaften von Matrizen durch die ihrer Digraphen bzw. die von Digraphen durch die ihrer Adjazenzmatrizen zu charakterisieren.

Definition 3.2.15

Eine reelle (n, n) -Matrix $A = (a_{ij}) \in \mathbb{R}^{n,n}$ heißt „irreduzibel“, wenn der Digraph der nichtnegativen Matrix $|A| = (|a_{ij}|)$ stark zusammenhängend ist.

Ist eine (n, n) -Matrix A nicht irreduzibel, so werden sie durch eine Umnummerierung der Zeilen und entsprechende Umnummerierung der Spalten auf eine übersichtlichere Form gebracht: Zunächst werden die Elemente jeder Zusammenhangskomponente des Digraphen von $|A|$ mit fortlaufenden Nummern versehen, sodann die Zusammenhangskomponenten in eine Reihenfolge gebracht, so dass

- die Basiskomponenten am Anfang stehen,

- die Komponente U vor der Komponente V steht, wenn im Kondensations-Digraphen ein Weg von U nach V führt.

Durch eine so erklärte (in der Regel nicht eindeutig bestimmte) Umnummerierung ergibt sich

Satz 3.2.16

Sei $A \in \mathbb{R}^{n,n}$. Besitzt der Digraph von $|A|$ m Zusammenhangskomponenten, davon r Basiskomponenten, so lässt sich A durch Umnummerierung der Zeilen und entsprechende Umnummerierung der Zeilen in die folgende Blockdreiecksform bringen:

$$A = \begin{pmatrix} A_{1,1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & A_{2,2} & & \vdots & \vdots & & \vdots \\ \vdots & & \ddots & 0 & \vdots & \cdots & \vdots \\ 0 & \cdots & 0 & A_{r,r} & 0 & \cdots & 0 \\ A_{r+1,1} & \cdots & \cdots & A_{r+1,r} & A_{r+1,r+1} & & \vdots \\ \vdots & & & & \ddots & \ddots & 0 \\ A_{m,1} & \cdots & \cdots & \cdots & \cdots & A_{m,m-1} & A_{m,m} \end{pmatrix}$$

mit irreduziblen Matrizen $A_{j,j}$ für $j = 1, \dots, m$.

Die Irreduzibilität nichtnegativer Matrizen lässt sich durch Untersuchung der Potenzen A^t feststellen, der starke Zusammenhang eines Digraphen also durch Betrachtung der Potenzen seiner Adjazenzmatrix. Das liegt daran, dass bei der Multiplikation von $A \geq 0$ mit sich selbst das Element

$$a_{ik}^{(2)} = a_{i1} a_{1k} + \dots + a_{in} a_{nk}$$

in der i -ten Zeile und k -ten Spalte von A^2 genau dann positiv ist, wenn für mindestens ein j sowohl $a_{ij} > 0$ als auch $a_{jk} > 0$, wenn also im zugehörigen Digraphen ein Weg von k nach i der Länge 2 existiert. Allgemeiner gilt (Beweis als **Aufgabe**):

Hilfssatz 3.2.17

Im Digraphen der nichtnegativen (n, n) -Matrix A existiert genau dann ein Weg von k nach i der Länge t , wenn das Element in der i -ten Zeile und k -ten Spalte von A^t positiv ist.

Wenn in einem Digraphen mit n Ecken ein Weg von der Ecke k zur Ecke i führt, dann immer auch ein Weg höchstens der Länge $n - 1$. Es genügt daher, die Potenzen A^t mit $t = 0, \dots, n - 1$ zu betrachten:

Satz 3.2.18

Für eine nichtnegative (n, n) -Matrix A sind die folgenden Bedingungen äquivalent:

- (a) A ist irreduzibel.
- (b) $I + A + A^2 + \dots + A^{n-1} > 0$.

Statt der hier gewählten hätte natürlich auch jede beliebige andere positive Linearkombination von $I, A, A^2, \dots, A^{n-1}$ gewählt werden können.

3.2.4 Positive Potenzen nichtnegativer Matrizen

Unter den Bedingungen von Satz 3.2.18 braucht keine Potenz A^t positiv zu sein. Ein Beispiel ist die Lesliematrix

$$A = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix}$$

für eine Population mit zwei Altersklassen. Für das diskrete dynamische System

$$x(t + 1) = A x(t), x(0) = x_0$$

mit der Lösung

$$x(t) = A^t x_0$$

hat das zur Folge, dass sich die Generationen in einem bestimmten Sinne nie durchmischen: Die Größe der ersten Komponente von x_0 etwa hat Einfluss auf die erste Komponente von $x(t)$ zu allen geraden Zeitpunkten t und auf die zweite Komponente von $x(t)$ zu allen ungeraden Zeitpunkten t , nicht aber auf die jeweils andere Komponente.

Das aber hat Einfluss auf das Langzeitverhalten, auf das später noch einzugehen sein wird. Als Charakterisierung ist der folgende Satz von Interesse:

Satz 3.2.19

Für eine nichtnegative (n, n) -Matrix A sind die folgenden Bedingungen äquivalent:

- (a) Es existiert ein $t \in \mathbb{N}$ mit $A^t > 0$.
- (b) Es existiert ein $t \in \mathbb{N}$, sodass $A^s > 0$ für alle $s \geq t$.
- (c) Der Digraph G von A ist stark zusammenhängend, und die Längen der Zyklen in G haben den größten gemeinsamen Teiler 1.

In dem speziellen Fall einer irreduziblen, nichtnegativen Matrix A mit mindestens einem positiven Diagonaleintrag $a_{ii} > 0$ sind die angegebenen Bedingungen erfüllt.

Beweis: Ist G stark zusammenhängend, so kann keine ganze Zeile oder Spalte von A nur aus Nullen bestehen. Mit $A^t > 0$ ist daher auch $A^{t+1} = A A^t > 0$ und damit $A^s > 0$ für alle $s \geq t$. Damit folgt die Äquivalenz von (a) und (b). Klar ist auch, dass (c) sowohl aus (b) als auch aus dem angegebenen speziellen Fall folgt. Zu zeigen bleibt, dass (b) aus (c) folgt:

Sei G stark zusammenhängend, und seien t_1, \dots, t_m Längen von Zyklen $\Gamma_1, \dots, \Gamma_m$ in G mit dem größten gemeinsamen Teiler 1. Seien $j_1 \in \Gamma_1, \dots, j_m \in \Gamma_m$ fest gewählt. Zu beliebigen $k, i \in \{1, \dots, n\}$ gibt es dann einen Weg von k über j_1, j_2, \dots, j_m nach i . Sei $s(k, i)$ dessen Länge. Dieser Weg lässt sich dadurch verlängern, dass man zusätzlich einen oder mehrere der Zyklen Γ_v einmal oder öfter durchläuft. Daher gibt es von k nach i Wege aller Längen

$$s(k, i) + \sum_{v=1}^m \alpha_v t_v \text{ mit } \alpha_1, \dots, \alpha_m \in \mathbb{N}_0.$$

Nun gibt es aber eine natürliche Zahl \tilde{t} , sodass sich alle natürlichen Zahlen $t \geq \tilde{t}$ in der Form $t = \sum_{v=1}^m \alpha_v t_v$ mit $\alpha_1, \dots, \alpha_m \in \mathbb{N}_0$ darstellen lassen (siehe nachstehender Hilfssatz). Für alle $s \geq s(k, i) + \tilde{t}$ gibt es daher einen Weg der Länge s von k nach i . Ist \tilde{s} das Maximum der $s(k, i)$, so ist daher $A^s > 0$ für alle $s \geq \tilde{s} + \tilde{t}$. ■

Zu zeigen bleibt noch

Hilfssatz 3.2.20

Seien t_1, \dots, t_m natürliche Zahlen mit dem größten gemeinsamen Teiler 1. Dann gibt es eine natürliche Zahl \tilde{t} , sodass sich alle natürlichen Zahlen $t \geq \tilde{t}$ in der Form

$$t = \sum_{v=1}^m \alpha_v t_v \text{ mit } \alpha_1, \dots, \alpha_m \in \mathbb{N}_0$$

darstellen lassen.

Beweis: Offenbar genügt es zu zeigen, dass sich t_1 aufeinander folgende natürliche Zahlen in der angegebenen Form darstellen lassen. Die Darstellung aller größeren ergibt sich dann daraus durch Addition entsprechender positiver Vielfacher von t_1 .

Nach einem bekannten Ergebnis der elementaren Zahlentheorie (vgl. LEUTBECHER (1996, 14)) lässt sich nun aber jede ganze Zahl s in der Form

$$s = \sum_{v=1}^m \beta_v t_v \text{ mit } \beta_1, \dots, \beta_m \in \mathbb{Z}$$

darstellen. Addiert man nun zu den Darstellungen

$$\begin{aligned} 1 &= \sum_{v=1}^m \beta_v^{(1)} t_v \\ 2 &= \sum_{v=1}^m \beta_v^{(2)} t_v \\ &\vdots \\ t_1 - 1 &= \sum_{v=1}^m \beta_v^{(t_1-1)} t_v \end{aligned}$$

eine Darstellung

$$\tilde{s} = \sum_{v=1}^m \alpha_v t_v$$

mit hinreichend großen natürlichen Zahlen $\alpha_1, \dots, \alpha_m$, so ergibt sich das gewünschte Ergebnis. ■

3.3 Der Lösungsraum linearer dynamischer Systeme

Für eine reelle (n, n) -Matrix A werden die mit ihr definierten dynamischen Systeme betrachtet, in diskreter Zeit also der lineare Iterationsprozess

$$x(t+1) = A x(t) \text{ für } t = 0, 1, 2, \dots \quad (3.11)$$

und in kontinuierlicher Zeit das System linearer Differentialgleichungen

$$\dot{x} = A x \text{ für } t \in \mathbb{R}, \quad (3.12)$$

ferner mit einer unendlichen Folge $u : \mathbb{N}_0 \rightarrow \mathbb{R}^n$ bzw. einer stetigen Funktion $u : \mathbb{R} \rightarrow \mathbb{R}^n$ die *inhomogenen* linearen dynamischen Systeme

$$x(t+1) = A x(t) + u(t) \text{ für } t = 0, 1, 2, \dots \quad (3.13)$$

bzw.

$$\dot{x} = A x + u(t) \text{ für } t \in \mathbb{R}, \quad (3.14)$$

(3.11) bzw. (3.12) wird auch das zu (3.13) bzw. (3.14) gehörige *homogene* System genannt.

3.3.1 Existenz und Eindeutigkeit bei Anfangswertaufgaben

Der Unterschied in den Definitionsbereichen bei den diskreten Systemen (3.11) und (3.13) einerseits und den kontinuierlichen Systemen (3.12) und (3.14) andererseits betrifft nicht nur die Frage, ob t diskret oder kontinuierlich zu wählen ist, sondern auch das Vorzeichen von t . Der Grund dafür liegt darin, dass bei vorgegebenem $x(0) = x_0$ im diskreten Fall und bei singulärer Systemmatrix A für negative t eine Lösung im Allgemeinen nicht definiert ist, im kontinuierlichen Fall dagegen wohl. Die Ergebnisse werden hier in zwei Sätzen zusammen gefasst. Der erste, die diskreten Systeme betreffend, folgt unmittelbar aus deren Definition. Der zweite bedarf dagegen eines keineswegs trivialen, hier nicht geführten Beweises und folgt im Übrigen auch nicht direkt aus dem allgemeinen Existenz- und Eindeutigkeitssatz für Systeme gewöhnlicher Differentialgleichungen.

Satz 3.3.1

Für jede Folge $u : \mathbb{N}_0 \rightarrow \mathbb{R}^n$ und jedes $x_0 \in \mathbb{R}^n$ existiert genau eine Lösung $x : \mathbb{N}_0 \rightarrow \mathbb{R}^n$ von (3.13) mit $x(0) = x_0$.

Satz 3.3.2

HEUSER (1989, 501)

Für jede auf einem reellen Intervall I mit $0 \in I$ definierte stetige Funktion $u : I \rightarrow \mathbb{R}^n$ und jedes $x_0 \in \mathbb{R}^n$ existiert genau eine Lösung $x : I \rightarrow \mathbb{R}^n$ von (3.14) mit $x(0) = x_0$.

Wählt man u als identisch Null, so sind mit diesen Sätzen auch die homogenen Systeme erfasst. Die gleichen Sätze gelten auch mit von t (stetig) abhängenden Systemmatrizen $A(t)$ (s. HEUSER (1989, 501)).

3.3.2 Die lineare Struktur des Lösungsraums

Mit x_1, x_2 ist auch jede Linearkombination $\alpha_1 x_1 + \alpha_2 x_2$ mit $\alpha_1, \alpha_2 \in \mathbb{R}$ eine Lösung des homogenen Systems (3.11) bzw. (3.12), wie sich durch Einsetzen sofort feststellen lässt. Die Abbildung, die jedem Anfangswert die zugehörige Lösung von (3.11) bzw. (3.12) zuordnet, ist daher linear. Sie ist außerdem injektiv, da die Lösungen zu verschiedenen Anfangswerten sich ja zumindest für $t = 0$ unterscheiden. Damit überträgt sich aber die lineare Struktur des Raums \mathbb{R}^n der Anfangswerte auf den Lösungsraum, es gilt also

Satz 3.3.3

Der Lösungsraum (die Menge aller Lösungen) von (3.11) bzw. (3.12) ist ein n -dimensionaler linearer Teilraum des Folgenraums $(\mathbb{R}^n)^{\mathbb{N}_0}$ bzw. des Raumes $C^1(\mathbb{R}, \mathbb{R}^n)$ der stetig differenzierbaren Funktionen von \mathbb{R} nach \mathbb{R}^n .

Definition 3.3.4

Unter einem „Fundamentalsystem“ von (3.11) bzw. (3.12) versteht man ein System von n linear unabhängigen Lösungen, also eine Basis des Lösungsraums.

Hat man ein Fundamentalsystem gefunden, so ergeben sich durch Bildung von Linearkombinationen alle Lösungen des homogenen Systems.

Für zwei Lösungen des inhomogenen Systems (3.13) bzw. (3.14) ist ihre Differenz stets eine Lösung des zugehörigen homogenen Systems (3.11) bzw. (3.12). Umgekehrt ist die Summe einer Lösung des inhomogenen mit der des homogenen Systems stets wieder eine Lösung des inhomogenen Systems. Daraus folgt

Satz 3.3.5

Ist z eine spezielle Lösung des inhomogenen Systems (3.13) bzw. (3.14), so ergeben sich alle Lösungen dieses Systems als Summe von z und den Lösungen des zugehörigen homogenen Systems.

Zur Bestimmung aller Lösungen eines homogenen Systems genügt es also, eine spezielle Lösung zu bestimmen, sofern ein Fundamentalsystem des zugehörigen homogenen Systems bekannt ist.

3.3.3 Die Rolle der Eigenwerte der Systemmatrix

Zur Bestimmung eines Fundamentalsystems liegt es nahe zu versuchen, möglichst viele linear unabhängige Lösungen zu bestimmen und zu hoffen, dass man n zusammen bekommt. Mit dem Ansatz

$$x(t) = \lambda^t w \text{ mit } \lambda \in \mathbb{R} \text{ und } w \in \mathbb{R}^n$$

für das diskrete System (3.11) bzw.

$$x(t) = e^{\lambda t} w \text{ mit } \lambda \in \mathbb{R} \text{ und } w \in \mathbb{R}^n$$

für das kontinuierliche System (3.12) ergibt sich durch Einsetzen:

Satz 3.3.6

$\lambda^t w$ ist genau dann eine von 0 verschiedene Lösung von (3.11), und $e^{\lambda t} w$ ist genau dann eine von 0 verschiedene Lösung von (3.12), wenn λ ein Eigenwert von A und $w \neq 0$ ein zugehöriger Eigenvektor ist, wenn also

$$A w = \lambda w .$$

Alle weiteren Überlegungen bestehen letztendlich darin, mit diesem Ansatz und gewissen Modifikationen ein Fundamentalsystem zu konstruieren. Eine erste Komplikation tritt dadurch auf, dass A auch komplexe Eigenwerte haben kann, sodass die Beschränkung auf reelle Eigenwerte nicht zu n linear unabhängigen Lösungen führen kann.

3.3.4 Komplexe Lösungen

Eine Folge $z : \mathbb{N}_0 \rightarrow \mathbb{C}^n$ bzw. Funktion $z : \mathbb{R} \rightarrow \mathbb{C}^n$, die der Gleichung (3.11) bzw. (3.12) genügt, heißt *komplexe Lösung* von (3.11) bzw. (3.12). Sie lässt sich darstellen in der Form

$$z(t) = x(t) + i y(t)$$

mit reellen Größen $x(t)$ und $y(t)$. Setzt man diese Darstellung in (3.11) bzw. (3.12) ein, so ergibt sich:

Satz 3.3.7

$z = x + i y$ ist genau dann eine komplexe Lösung von (3.11) bzw. (3.12), wenn x und y reelle Lösungen von (3.11) bzw. (3.12) sind.

Ist nun λ ein komplexer Eigenwert von A und $w \neq 0$ ein zugehöriger Eigenvektor, so ist wie oben für den reellen Fall $\lambda^t w$ eine komplexe Lösung von (3.11) und $e^{\lambda t} w$ eine komplexe Lösung von (3.12). Daraus lassen sich reelle Lösungen gewinnen: Mit

$$\lambda = \mu + i v = r e^{i \varphi} = r (\cos \varphi + i \sin \varphi) \quad (r \geq 0), \quad w = u + i v$$

ist

$$\begin{aligned} \lambda^t w = r^t e^{i \varphi t} w &= r^t (\cos \varphi t + i \sin \varphi t) (u + i v) \\ &= r^t [(u \cos \varphi t - v \sin \varphi t) + i (u \sin \varphi t + v \cos \varphi t)] \end{aligned}$$

und

$$\begin{aligned} e^{\lambda t} w = e^{\mu t} e^{i v t} w &= e^{\mu t} (\cos v t + i \sin v t) (u + i v) \\ &= e^{\mu t} [(u \cos v t - v \sin v t) + i (u \sin v t + v \cos v t)] . \end{aligned}$$

Es folgt

Satz 3.3.8

Ist $\lambda = \mu + i v = r e^{i \varphi}$ ein komplexer Eigenwert von A und $w = u + i v$ ein zugehöriger Eigenvektor, so sind

$$r^t (u \cos \varphi t - v \sin \varphi t) \quad \text{und} \quad r^t (u \sin \varphi t + v \cos \varphi t)$$

reelle Lösungen von (3.11) und

$$e^{\mu t} (u \cos v t - v \sin v t) \quad \text{und} \quad e^{\mu t} (u \sin v t + v \cos v t)$$

reelle Lösungen von (3.12).

Es sieht so aus, als bekäme man für jeden komplexen Eigenwert von A gleich zwei linear unabhängige, reelle Lösungen. Dabei ist jedoch zu beachten, dass mit λ auch die konjugiert komplexe Zahl $\bar{\lambda}$ ein Eigenwert der reellen Matrix A ist, der zu *denselben* reellen Lösungen führt.

Für die meisten Zwecke ist es einfacher, den Übergang von den komplexen zu den reellen Lösungen nicht zu machen, sondern gleich nach komplexen Fundamentalsysteme zu suchen.

3.3.5 Komplexe Fundamentalsysteme: Diagonalisierbare Systemmatrizen

m aus komplexen Eigenwerten und zugehörigen Eigenvektoren gewonnene Lösungen

$$\lambda_i^t w_i \text{ bzw. } e^{\lambda_i t} w_i \quad (i = 1, \dots, m)$$

sind genau dann linear unabhängig, wenn die Eigenvektoren w_1, \dots, w_m es sind (Beweis als **Aufgabe**). Mit dem hier durchgeführten Ansatz erhält man daher genau dann ein (komplexes) Fundamentalsystem, wenn A n linear unabhängige Eigenvektoren besitzt. Das wiederum ist damit gleichbedeutend, dass eine reguläre (komplexe) (n, n) -Matrix S existiert, sodass

$$S^{-1} A S = D$$

eine Diagonalmatrix ist. Die Diagonaleinträge von D sind gerade die Eigenwerte von A , und in den Spalten von S stehen entsprechende Eigenvektoren. (s. ARTMANN (1991, 197 ff.))

Ein sehr wichtiger Sonderfall ist der, dass A n *verschieden Eigenwerte* besitzt, denn dann sind die zugehörigen Eigenvektoren immer linear unabhängig. (s. ARTMANN (1991, 198))

3.3.6 Exkurs: Jordan'sche Normalform und Jordan-Sequenzen

Jede komplexe (n, n) -Matrix A ist bekanntlich (s. ARTMANN (1991, 324 ff.)) ähnlich zu einer Matrix in *Jordan'scher Normalform*, d. h. es existiert eine reguläre (n, n) -Matrix S , sodass

$$S^{-1} A S = B = \begin{pmatrix} J_1 & 0 & \cdots & \cdots & 0 \\ 0 & J_2 & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & J_{m-1} & 0 \\ 0 & \cdots & \cdots & 0 & J_m \end{pmatrix}$$

mit ober (r_i, r_i) -Matrizen

$$J_i = \begin{pmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \lambda_i & 1 \\ 0 & \cdots & \cdots & 0 & \lambda_i \end{pmatrix} \quad (i = 1, \dots, m),$$

den *Jordankästchen*, wobei $r_1 + \dots + r_m = n$. Echte Jordankästchen mit $r_i \geq 2$ treten für solche Eigenwerte auf, bei denen die algebraische Vielfachheit (Vielfachheit der Nullstelle des charakteristischen Polynoms von A) größer ist als die geometrische Vielfachheit (Dimension des Raums der Eigenvektoren).

Sind w_1, \dots, w_r diejenigen *Spalten* von S , die der Stellung eines bestimmten (r, r) -Jordankästchens (der Index i wird hier weggelassen)

$$J = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \lambda & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{pmatrix} \quad (i = 1, \dots, m)$$

in B entsprechen, so ist wegen $AS = SB$

$$Aw_1 = \lambda w_1, Aw_2 = \lambda w_2 + w_1, \dots, Aw_r = \lambda w_r + w_{r-1}. \quad (3.15)$$

w_1 ist also ein Eigenvektor zum Eigenwert λ von A, während w_2, \dots, w_r zugehörige *Hauptvektoren* verschiedener Stufe sind. r linear unabhängige Vektoren w_1, \dots, w_r mit (3.15) heißen auch eine *Jordansequenz* von A.

Die *Zeilen* von S^{-1} , die der Stellung des Jordankästchens in B entsprechen, spielen eine analoge Rolle: Sind f_1^T, \dots, f_r^T diese Zeilen, so ist wegen $A^T(S^{-1})^T = (S^{-1})^T B^T$

$$A^T f_r = \lambda f_r, A^T f_{r-1} = \lambda f_{r-1} + f_r, \dots, A^T f_1 = \lambda f_1 + f_2. \quad (3.16)$$

f_r ist also ein Eigenvektor von A^T (oder *Linkseigenvektor* von A) zum Eigenwert λ , während f_{r-1}, \dots, f_1 zugehörige Hauptvektoren verschiedener Stufe sind.

Es bezeichne (mit anderer Indizierung als eben!) w_1, \dots, w_n die Spalten von S und f_1^T, \dots, f_n^T die Zeilen von S^{-1} . Jedes $x \in \mathbb{R}^n$ ist darstellbar als Linearkombination der w_j :

$$x = \alpha_1 w_1 + \dots + \alpha_n w_n.$$

Multiplikation dieser Gleichung von links mit f_i^T liefert

$$f_i^T x = \alpha_i \text{ für } i = 1, \dots, n$$

(hier wurde ausgenutzt, dass $S^{-1}S = I$). Daraus folgt

$$x = (f_1^T x) w_1 + \dots + (f_n^T x) w_n \text{ für alle } x \in \mathbb{R}^n. \quad (3.17)$$

3.3.7 Komplexe Fundamentalsysteme: Allgemeiner Fall

Stimmen für mindestens einen (mehrfachen) Eigenwert der Systemmatrix A algebraische und geometrische Vielfachheit nicht überein, treten also in der Jordan'schen Normalform echte Jordankästchen auf, so lassen sich mit Lösungen der Form

$$\lambda^t w \text{ bzw. } e^{\lambda t} w$$

allein keine Fundamentalsysteme von (3.11) bzw. (3.12) bilden, da keine n linear unabhängigen Eigenvektoren w existieren. Hierzu kann man aber auf die Jordansequenzen zurückgreifen. Die beiden folgenden Sätze lassen sich durch Einsetzen in die jeweilig zu Grunde liegenden Gleichungen verifizieren:

Satz 3.3.9

Die Jordan'sche Normalform der reellen (n, n) -Matrix A besitze ein r -dimensionales Jordankästchen zum Eigenwert λ mit der zugehörigen Jordansequenz w_1, \dots, w_r entsprechend (3.15). Dann bilden im Falle $\lambda = 0$ die Folgen

$$(w_1, 0, 0, 0, \dots), (w_2, w_1, 0, 0, \dots), \dots, (w_r, \dots, w_1, 0, 0, \dots)$$

und im Falle $\lambda \neq 0$ die Folgen

$$\begin{aligned} & \begin{pmatrix} t \\ 0 \end{pmatrix} \lambda^t w_1 \\ & \begin{pmatrix} t \\ 1 \end{pmatrix} \lambda^{t-1} w_1 + \begin{pmatrix} t \\ 0 \end{pmatrix} \lambda^t w_2 \\ & \vdots \\ & \begin{pmatrix} t \\ r-1 \end{pmatrix} \lambda^{t-r+1} w_1 + \dots + \begin{pmatrix} t \\ 0 \end{pmatrix} \lambda^t w_r \end{aligned}$$

r linear unabhängige Lösungen von (3.11). Die Gesamtheit dieser Lösungen (über alle Jordankästchen von A) bilden ein Fundamentalsystem.

Satz 3.3.10

Die Jordan'sche Normalform der reellen (n, n) -Matrix A besitze ein r -dimensionales Jordankästchen zum Eigenwert λ mit der zugehörigen Jordansequenz w_1, \dots, w_r entsprechend (3.15). Dann bilden die Funktionen

$$e^{\lambda t} w_1, e^{\lambda t} (t w_1 + w_2), e^{\lambda t} \left(\frac{t^2}{2} w_1 + t w_2 + w_3 \right), \dots, e^{\lambda t} \left(\frac{t^{r-1}}{(r-1)!} w_1 + \dots + w_r \right)$$

r linear unabhängige Lösungen von (3.12). Die Gesamtheit dieser Lösungen (über alle Jordankästchen von A) bilden ein Fundamentalsystem.

3.3.8 Stabilität

Da alle Lösungen von (3.11) bzw. (3.12) als Linearkombinationen aus den Fundamentalsystemen ergeben, folgen aus den letzten beiden Sätzen für das Langzeitverhalten der Lösungen die folgenden Ergebnisse:

Satz 3.3.11

Sei A eine reelle (n, n) -Matrix.

(a) Die beiden folgenden Aussagen sind äquivalent:

- Für alle Lösungen x des homogenen, linearen, diskreten dynamischen Systems (3.11) mit der Systemmatrix A ist

$$\lim_{t \rightarrow \infty} x(t) = 0.$$

- Der Absolutbetrag aller Eigenwerte von A ist kleiner als 1.

(b) Die beiden folgenden Aussagen sind äquivalent:

- Alle Lösungen x des linearen diskreten dynamischen Systems (3.11) mit der Systemmatrix A sind beschränkt.
- Der Absolutbetrag aller Eigenwerte von A ist kleiner oder gleich 1, und für alle Eigenwerte mit dem Absolutbetrag 1 stimmen algebraische und geometrische Vielfachheit überein.

Satz 3.3.12

Sei A eine reelle (n, n) -Matrix.

(a) Die beiden folgenden Aussagen sind äquivalent:

- Für alle Lösungen x des homogenen Systems linearer Differentialgleichungen (3.12) mit der Systemmatrix A ist

$$\lim_{t \rightarrow \infty} x(t) = 0.$$

- Der Realteil aller Eigenwerte von A ist kleiner als 0.

(b) Die beiden folgenden Aussagen sind äquivalent:

- Alle Lösungen x des homogenen Systems linearer Differentialgleichungen (3.12) mit der Systemmatrix A sind auf $[0, \infty)$ beschränkt.
- Der Realteil aller Eigenwerte von A ist kleiner oder gleich 0, und für alle Eigenwerte mit dem Realteil 0 stimmen algebraische und geometrische Vielfachheit überein.

3.3.9 Inhomogene lineare dynamische Systeme mit konstanten Einträgen

Zur Bestimmung aller Lösungen der inhomogenen Systeme (3.13) bzw. (3.14) genügt es, sofern ein Fundamentalsystem des zugehörigen homogenen Systems bekannt ist, eine spezielle Lösung zu bestimmen. Auf allgemeine Lösungsformeln soll hier nicht eingegangen werden. Statt dessen wird hier nur ein Sonderfall betrachtet, dass nämlich die Einträge $u(t) = u$ konstant sind, also nicht von der Zeit t abhängen. Ich betrachte also inhomogene Systeme der Form

$$x(t+1) = A x(t) + u \tag{3.18}$$

in diskreter bzw.

$$\dot{x} = A x + u \quad (3.19)$$

in kontinuierlicher Zeit. Bei konstanten Einträgen u liegt es nahe, nach konstanten, also Gleichgewichtslösungen zu suchen, deren Bestimmung die Lösung eines linearen Gleichungssystems erfordert. im Falle des diskreten Systems ist das $x = A x + u$, im Falle des kontinuierlichen Systems $A x + u = 0$. Die eindeutige Lösbarkeit des jeweiligen Gleichungssystems vorausgesetzt, erhält man so einen eindeutig bestimmten Gleichgewichtspunkt \bar{x} , durch den eine konstante Lösung des jeweiligen Systems definiert ist.

Für eine *beliebige* Lösung x des jeweiligen inhomogenen Systems ist nun aber die Differenz $y(t) = x(t) - \bar{x}$ wiederum eine Lösung des zugehörigen homogenen Systems, auf die sich daher die zuletzt formulierten Sätze zur Stabilität anwenden lassen. Tut man das, so folgt:

Satz 3.3.13

Seien $A \in \mathbb{R}^{n,n}$ und $u \in \mathbb{R}^n$. Ist 1 kein Eigenwert von A , die Matrix $I - A$ also regulär, so besitzt das lineare Gleichungssystem

$$x = A x + u$$

eine eindeutig bestimmte Lösung \bar{x} , durch die eine Gleichgewichtslösung von (3.18) definiert ist. Ist darüber hinaus der Absolutbetrag aller Eigenwerte von A kleiner als 1, so gilt für alle Lösungen x von (3.18)

$$\lim_{t \rightarrow \infty} x(t) = \bar{x}.$$

Satz 3.3.14

Seien $A \in \mathbb{R}^{n,n}$ und $u \in \mathbb{R}^n$. Ist 0 kein Eigenwert von A , die Matrix A also regulär, so besitzt das lineare Gleichungssystem

$$A x + u = 0$$

eine eindeutig bestimmte Lösung \bar{x} , durch die eine Gleichgewichtslösung von (3.19) definiert ist. Ist darüber hinaus der Realteil aller Eigenwerte von A kleiner als 0, so gilt für alle Lösungen x von (3.19)

$$\lim_{t \rightarrow \infty} x(t) = \bar{x}.$$

Die beiden Sätze geben Bedingungen an, unter denen die Gleichgewichtslösung nicht nur irgendeine spezielle Lösung des inhomogenen Systems ist, sondern gerade diejenige, die langfristig „übrig bleibt“.

3.4 Nichtnegative Matrizen. Der Satz von Perron-Frobenius

Nichtnegative (n,n) -Matrizen A spielen für lineare Modelle insofern eine besondere Rolle, als sie den nichtnegativen Oktanten in sich abbilden: Mit $x \geq 0$ ist auch $A x \geq 0$. Ist x ein Vektor von Bestandsgrößen und soll auch $A x$ ein solcher sein, so ist das nur sichergestellt, wenn $A \geq 0$. Bei linearen, diskreten dynamischen Systemen

$$x(t+1) = A x(t)$$

mit einem Vektor $x(t)$ von Bestandsgrößen etwa ist das eine zwingende Forderung an ein sinnvoll formuliertes Modell.

Der Satz von Perron-Frobenius macht eine Aussage über den betragsmäßig größten Eigenvektor einer nichtnegativen Matrix : Dieser ist immer reell, und zu ihm gehört ein nichtnegativer Eigenvektor. Für irreduzible Matrizen lässt sich diese Aussage noch weiter verschärfen. Die besondere Rolle der Eigenwerte für die lineare Dynamik macht derartige Ergebnisse interessant.

Im Folgenden wird mit "Eigenwert-Ungleichungen" der Form

$$A w \geq \lambda w \text{ mit } \lambda \in \mathbb{R}, w \in \mathbb{R}^n, w \neq 0, w \geq 0$$

operiert. Mit $\lambda = 0$ ist eine solche Ungleichung immer erfüllbar. Ist sie erfüllt, so ist andererseits wegen

$$\lambda w_i \leq \sum_{j=1}^n a_{ij} w_j \leq \left(\sum_{j=1}^n a_{ij} \right) \max_k w_k$$

$$\lambda \leq \max_i \sum_{j=1}^n a_{ij}.$$

3.4.1 Der Perron-Frobenius-Eigenwert

Für eine nichtnegative (n, n) -Matrix A bezeichne

$$\lambda_0(A) := \sup \{ \lambda \in \mathbb{R} : \text{es gibt ein } w \in \mathbb{R}^n, w \neq 0, w \geq 0 \text{ mit } A w \geq \lambda w \}.$$

Im Folgenden soll gezeigt werden, dass $\lambda_0(A)$ ein Eigenwert von A mit den bereits angedeuteten Eigenschaften ist, der *Perron-Frobenius-Eigenwert* von A .

Hilfssatz 3.4.1

Für eine nichtnegative (n, n) -Matrix A ist $\lambda_0(A)$ eine nichtnegative reelle Zahl, und das Supremum in der Definition von $\lambda_0(A)$ wird angenommen.

Beweis: Nach den vorstehenden Überlegungen ist nur noch zu zeigen, dass für $\lambda_0 = \lambda_0(A)$ ein $w_0 \in \mathbb{R}^n$, $w_0 \neq 0$, $w_0 \geq 0$ mit $A w_0 \geq \lambda_0 w_0$ existiert.

Da die Ungleichung $A w \geq \lambda w$ stets auch für jedes positive Vielfache von w erfüllt ist, existiert zu jedem $\varepsilon > 0$ ein $w_\varepsilon \geq 0$ mit $\|w_\varepsilon\| = 1$, sodass

$$A w_\varepsilon \geq (\lambda_0 - \varepsilon) w_\varepsilon.$$

Wegen der Kompaktheit von $\{w \geq 0 : \|w\| = 1\}$ existiert dann aber eine Folge $\varepsilon_k \rightarrow 0$, für die w_{ε_k} konvergiert. Der Grenzwert w_0 dieser Folge hat die gewünschten Eigenschaften. ■

3.4.2 Positive Matrizen

Satz 3.4.2

Sei A eine (in allen Koeffizienten) positive (n, n) -Matrix. Dann ist die in 3.4.1 definierte Zahl $\lambda_0 = \lambda_0(A)$ ein positiver Eigenwert von A der algebraischen und geometrischen Vielfachheit 1 mit einem zugehörigen positiven Eigenvektor w_0 . Für jeden anderen Eigenwert λ von A gilt $|\lambda| < \lambda_0$.

Beweis: Es wurde bereits gezeigt, dass ein $w_0 \geq 0$, $w_0 \neq 0$ existiert, sodass $A w_0 \geq \lambda_0 w_0$. Wäre nun $A w_0 \neq \lambda_0 w_0$, so wäre $A(A w_0) > A(\lambda_0 w_0) = \lambda_0 (A w_0)$, da $A > 0$, und es gäbe ein $\lambda > \lambda_0$ mit $A(A w_0) \geq \lambda (A w_0)$, im Widerspruch zur Definition von λ_0 . Also ist $A w_0 = \lambda_0 w_0$, und wegen $A w_0 > 0$ ist $\lambda_0 > 0$ und $w_0 > 0$.

Sei $\lambda \neq \lambda_0$ ein anderer Eigenwert von A und $w \neq 0$ ein zugehöriger Eigenvektor. Sei $|w|$ der Vektor, dessen Komponenten die Absolutbeträge der Komponenten von w sind. Dann ist $A |w| \geq |A w| = |\lambda w| = |\lambda| |w|$, woraus nach Definition von λ_0 folgt, dass $|\lambda| \leq \lambda_0$. Zum Beweis der strikten Ungleichung $|\lambda| < \lambda_0$ betrachte man $A - \delta I$ mit einem so kleinen $\delta > 0$, dass $A - \delta I > 0$.⁴ Dann ist $|\lambda - \delta| \leq \lambda_0 - \delta$. Wäre nun $|\lambda| = \lambda_0$, so wäre mit $\lambda = \mu + i v$, da λ und λ_0 verschieden, $\mu^2 + v^2 = \lambda_0^2$, $\mu < \lambda_0$ und daher

$$|\lambda - \delta|^2 = \mu^2 - 2 \mu \delta + \delta^2 + v^2 > \lambda_0^2 - 2 \lambda_0 \delta + \delta^2 = (\lambda_0 - \delta)^2,$$

woraus sich ein Widerspruch ergibt. Also ist $|\lambda| < |\lambda_0|$.

⁴Man beachte, dass die Eigenwerte von $A - \delta I$ aus denen von A durch Subtraktion von δ hervorgehen.

Gäbe es zum Eigenwert λ_0 einen weiteren, von w_0 linear unabhängigen Eigenvektor w , so kann dieser (ggf. durch Übergang zum Real- oder Imaginärteil) reell gewählt werden. Wegen $w_0 > 0$ gibt es dann eine Linearkombination $y = \alpha w_0 + w$ mit $y \neq 0$, $y \geq 0$, aber nicht $y > 0$. Das aber ist wegen $\lambda_0 y = A y > 0$ nicht möglich. Der Eigenwert λ_0 von A hat also die geometrische Vielfachheit 1.

Wäre nun seine algebraische Vielfachheit größer als 1, so gäbe es zu λ_0 ein Jordankästchen von A mindestens der Dimension 2 und daher einen Hauptvektor $w \neq 0$ mit $(A - \lambda_0 I) w = w_0$. Mit dem positiven, zu λ_0 gehörigen Eigenvektor f_0 von A^T ergibt sich daraus der Widerspruch

$$0 = f_0^T (A - \lambda_0 I) w = f_0^T w_0 > 0.$$

Die algebraische Vielfachheit von λ_0 ist daher ebenfalls 1. ■

3.4.3 Allgemeiner Fall

Mit Stetigkeitsargumenten lässt sich der Satz von Perron-Frobenius auf beliebige nichtnegative Matrizen übertragen: Jede nichtnegative Matrix lässt sich durch positive Matrizen beliebig genau approximieren, und die Eigenwerte einer Matrix hängen in einem gewissen, hier nicht präzisierten Sinne von ihren Koeffizienten stetig ab. Daraus ergibt sich

Satz 3.4.3

Sei A eine nichtnegative (n, n) -Matrix. Dann ist die in 3.4.1 definierte Zahl $\lambda_0 = \lambda_0(A)$ ein nichtnegativer Eigenwert von A mit einem zugehörigen Eigenvektor $w_0 \geq 0$. Für jeden Eigenwert λ von A gilt $|\lambda| \leq \lambda_0$.

Dagegen muss bei nichtnegativen Matrizen der betragsmäßig größte Eigenwert nicht eindeutig und der Perron-Frobenius-Eigenwert muss nicht einfach sein. Gegenbeispiele sind die Matrizen

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

3.4.4 Irreduzible Matrizen

Das volle Ergebnis des Satzes von Perron-Frobenius für positive Matrizen lässt sich auf solche nichtnegativen (n, n) -Matrizen A übertragen, die eine positive Potenz A^t für ein $t > 0$ besitzen. Eine Charakterisierung dieser Eigenschaft ist in Satz 3.2.19 angegeben: A muss dazu irreduzibel sein, und die Längen der Zyklen des Digraphen von A müssen den größten gemeinsamen Teiler 1 haben. Mit dem Perron-Frobenius-Eigenwert λ_0 von A ist dann λ_0^t der Perron-Frobenius-Eigenwert von A^t . Nach Satz 3.4.2 hat er die algebraische Vielfachheit 1, was daher auch für den Eigenwert λ_0 von A gilt. Für alle anderen Eigenwerte λ von A ist ferner $|\lambda|^t < \lambda_0^t$ und daher $|\lambda| < \lambda_0$.

Satz 3.4.4

Sei A eine nichtnegative, irreduzible (n, n) -Matrix. Dann ist die in 3.4.1 definierte Zahl $\lambda_0 = \lambda_0(A)$ ein positiver Eigenwert von A der algebraischen und geometrischen Vielfachheit 1 mit einem zugehörigen positiven Eigenvektor w_0 . Für jeden anderen Eigenwert λ von A gilt $|\lambda| < \lambda_0$. Haben darüber hinaus die Längen der Zyklen im Digraphen von A den größten gemeinsamen Teiler 1, so gilt für jeden anderen Eigenwert λ von A sogar $|\lambda| < \lambda_0$.

Beweis: Ist die angegebene Zusatzvoraussetzung über die Zyklenlängen erfüllt, so existiert nach Satz 3.2.19 ein $t > 0$ mit $A^t > 0$, und die Behauptung folgt aus den oben bereits dargelegten Gründen.

In jedem Fall ist diese Zusatzvoraussetzung aber für die Matrix $I + A$ erfüllt, deren Perron-Frobenius-Eigenwert $\lambda_0 + 1$ daher algebraisch einfach ist mit einem zugehörigen positiven Eigenvektor w_0 . Daher ist auch λ_0 ein algebraisch einfacher Eigenwert von A mit dem selben positiven Eigenvektor w_0 . ■

Bemerkung

Die angegebene Zusatzvoraussetzung ist insbesondere dann erfüllt, wenn mindestens ein Diagonaleintrag von A positiv ist. Für beliebige irreduzible Matrizen braucht dagegen der Perron-Frobenius-Eigenwert nicht der einzige betragsgrößte Eigenwert zu sein, wie die Matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

zeigt. Umgekehrt kann aber natürlich auch eine nichtnegative Matrix, die nicht irreduzibel ist, wie etwa

$$\begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}$$

alle Aussagen des (starken) Satzes von Perron-Frobenius erfüllen.

3.4.5 Langzeitverhalten diskreter dynamischer Systeme**Satz 3.4.5**

Die nichtnegative (n, n) -Matrix A besitze einen positiven, algebraisch einfachen Eigenwert λ_0 , der größer sei als der Absolutbetrag jedes anderen Eigenwerts von A . w_0 bzw. f_0 seien zugehörige nicht-negative Eigenvektoren von A bzw. A^T mit $f_0^T w_0 = 1$. Dann gilt für die Lösung von

$$x(t+1) = A x(t), \quad x(0) = x_0$$

mit einem Anfangswert $x_0 \in \mathbb{R}^n$:

$$\lim_{t \rightarrow \infty} \frac{x(t)}{\lambda_0^t} = (f_0^T x_0) w_0.$$

Ist $x_0 > 0$ (oder $x_0 \geq 0$ und $f_0 > 0$), so ist $f_0^T x_0 > 0$.

Beweis: Nach (3.17) ist $f_0^T x_0$ der Anteil von w_0 in der Darstellung von x_0 als Linearkombination von Eigen- und Hauptvektoren von A . Für die Lösung der zugehörigen Anfangswertaufgabe gilt daher

$$x(t) = (f_0^T x_0) w_0 \lambda_0^t + \dots$$

mit zusätzlichen von t abhängigen Lösungsanteilen der Form

$$\begin{pmatrix} t \\ k \end{pmatrix} \lambda^{t-k}$$

mit anderen Eigenwerten λ von A . Nach Division durch λ_0^t streben diese Lösungsanteile für $t \rightarrow \infty$ gegen 0, woraus der erste Teil der Behauptung folgt. Der zweite folgt aus $f_0 \geq 0$, $f_0 \neq 0$. ■

3.4.6 Stabilität und Produktivität nichtnegativer Matrizen

Die Lösungen eines diskreten dynamischen Systems

$$x(t+1) = A x(t)$$

streben nach Satz 3.3.12 genau dann allesamt gegen 0, wenn der Absolutbetrag aller Eigenwerte von $A \in \mathbb{R}^{n,n}$ kleiner als 1. Im Falle einer nichtnegativen Matrix $A \geq 0$ ist das damit gleichbedeutend, dass

$\lambda_0(A) < 1$. Eine dritte hierzu äquivalente Bedingung ergibt sich aus Satz 3.3.13: $I - A$ ist invertierbar, und für alle $u \in \mathbb{R}^n$ und alle Lösungen x des inhomogenen Systems

$$x(t+1) = A x(t) + u$$

gilt

$$\lim_{t \rightarrow \infty} x(t) = (I - A)^{-1} u .$$

Ist hier speziell $u \geq 0$ und wählt man $x(0) \geq 0$, so ist $x(t) \geq 0$ für alle $t \geq 0$, was daher auch für den Grenzwert $(I - A)^{-1} u$ gelten muss, also für die Lösung \bar{x} von

$$x = A x + u .$$

Zu erkennen ist also eine Beziehung der angegebenen Stabilitätsbedingungen für das diskrete dynamische System zu dem in Definition 3.1.1 eingeführten Begriff der Produktivität der Matrix A . Genauer gilt:

Satz 3.4.6

Für eine nichtnegative Matrix $A \in \mathbb{R}^{n,n}$ sind die folgenden Bedingungen äquivalent.

(a) Für den Perron-Frobenius-Eigenwert $\lambda_0 = \lambda_0(A)$ gilt

$$\lambda_0 < 1 .$$

(b) Für jede Lösung x des diskreten dynamischen Systems $x(t+1) = A x(t)$ gilt

$$\lim_{t \rightarrow \infty} x(t) = 0 .$$

(c) $I - A$ ist invertierbar, und für alle $u \in \mathbb{R}^n$ und alle Lösungen x des inhomogenen Systems $x(t+1) = A x(t) + u$ gilt

$$\lim_{t \rightarrow \infty} x(t) = (I - A)^{-1} u .$$

(d) Für jedes $u \in \mathbb{R}^n$ mit $u \geq 0$ besitzt das lineare Gleichungssystem $x = A x + u$ eine eindeutige Lösung $\bar{x} \geq 0$.

(e) $I - A$ ist invertierbar, und $(I - A)^{-1} \geq 0$.

(f) Es gibt ein $\bar{x} \in \mathbb{R}^n$ mit $\bar{x} \geq 0$ und $A \bar{x} < \bar{x}$.

Beweis: Gezeigt wurde bereits die Äquivalenz der Aussagen (a), (b), (c) und ferner, dass jede von ihnen (d) impliziert. Dass (d) ebenso aus (e) folgt, ist klar. Umgekehrt hat (d) die Existenz der Inversen $(I - A)^{-1}$ zur Folge, und $(I - A)^{-1} \geq 0$ ergibt sich, wenn man in (d) für u speziell die Einheitsvektoren wählt. Schließlich ergibt sich auch (f) aus (d), wenn man dort $u > 0$ wählt.

Es gelte nun (f), dann ist $u := \bar{x} - A \bar{x} > 0$. Sei $f_0 \geq 0$ der zu λ_0 gehörige Eigenvektor von A^T . Dann ist

$$0 < f_0^T u = (1 - \lambda_0) f_0^T \bar{x} \text{ und } f_0^T \bar{x} \geq 0 ,$$

woraus (a) folgt. ■

3.4.7 Input-Output-Tabellen und Verflechtungsmatrix

Bei den in 3.1.2 betrachteten linearen Gleichungssystemen

$$x = A x + b ,$$

mit denen zu gegebenem Bedarfsvektor b die zu seiner Herstellung benötigte Bruttoproduktion x berechnet wird, ergab sich die Verflechtungsmatrix A aus der Erhebung einer *beobachteten* Bruttoproduktion X .

Aus Satz 3.4.6 folgt nun, dass A produktiv ist, wenn für diese eine, tatsächlich erfasste Bruttoproduktion X gilt

$$A X < X, \text{ wenn also } B = X - A X > 0.$$

Anders gesagt: Ist es möglich, *einen speziellen* positiven Bedarfsvektor mit einer nichtnegativen Bruttoproduktion zu bedienen, so ist das für alle nichtnegativen Bedarfsvektoren möglich.

Mit dieser Bedingung dafür, dass die Verflechtungsmatrix A produktiv ist, ist eine theoretische Lücke geschlossen, die in 3.1.2 zunächst noch offen geblieben war.

3.5 Analyse der Altersklassen-Modelle

Die in 3.1.1 eingeführten Modelle für ein Populationswachstum mit altersabhängiger Fruchtbarkeit und Sterblichkeit sollen jetzt mit den inzwischen entwickelten Hilfsmitteln analysiert werden. Die Tatsache, dass die hier zu Grunde liegenden Systemmatrizen im Allgemeinen nicht die Voraussetzungen des starken Satzes von Perron-Frobenius erfüllen, macht genauere Betrachtungen erforderlich.

3.5.1 Einfacher Leslie-Prozess. Langzeitverhalten

Betrachtet man nur die Entwicklung der weiblichen Population oder geht von einer Reproduktion fester Paare aus, so ergibt sich der einfache Leslie-Prozess

$$x(t+1) = A x(t),$$

wobei

$$A = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_{n-1} & a_n \\ u_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & u_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \cdots & u_{n-1} & 0 \end{pmatrix} \quad (3.20)$$

mit $a_1, \dots, a_n \geq 0$, $a_k > 0$ für mindestens ein $k \in \{1, \dots, n\}$ und $0 < u_i \leq 1$ für $i = 1, \dots, n-1$.

Die Zusammenhangsstruktur des zu A gehörigen Digraphen hängt wesentlich davon ab, welche der a_i positiv sind. Sei k die älteste reproduktive Altersklasse, k also der größte Index mit $a_k > 0$. Dann ist die zu den Altersklassen $1, \dots, k$ gehörige Teilmatrix

$$A_{1, \dots, k} = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_{k-1} & a_k \\ u_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & u_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \cdots & u_{k-1} & 0 \end{pmatrix}$$

von A irreduzibel, und $\{1, \dots, k\}$ ist die einzige Basiskomponente des Digraphen von A . Auf $A_{1, \dots, k}$ ist der Satz 3.4.4 von Perron-Frobenius für irreduzible, nichtnegative Matrizen anwendbar: $\lambda_0 = \lambda_0(A_{1, \dots, k})$ ist ein positiver, algebraisch einfacher Eigenwert von $A_{1, \dots, k}$, damit aber auch von A , denn aus der Gestalt von A geht sofort hervor, dass die Eigenwerte von A gerade die Eigenwerte von $A_{1, \dots, k}$ sind, ergänzt um

den $(n-k)$ -fachen Eigenwert 0. Der zu λ_0 gehörige Eigenvektor w_0 von A ist positiv: Aus der Eigenwertgleichung $A w_0 = \lambda_0 w_0$ folgt nämlich für seine Komponenten w_1, \dots, w_n

$$u_i w_i = \lambda_0 w_{i+1} \text{ für } i = 1, \dots, n-1.$$

Ungeklärt ist damit nur noch die Frage, ob für alle anderen Eigenwerte λ die strikte Ungleichung $|\lambda| < \lambda_0$ gilt. Da jedem Index i mit $a_i > 0$ gerade ein Zyklus der Länge i im Digraphen von $A_{1, \dots, k}$ entspricht, lässt sich diese Frage positiv beantworten, wenn der größte gemeinsame Teiler dieser Indizes 1 ist. Das ist z. B. dann der Fall, wenn $a_1 > 0$ oder (für menschliche Population realistischer) $a_i > 0, a_{i+1} > 0$ für zwei aufeinander folgende Altersklassen.

Zur Klärung des Langzeitverhaltens ist noch festzuhalten, dass die ersten k Komponenten des Eigenvektors f_0 von A^T zum Eigenwert λ_0 positiv, die letzten $n-k$ Komponenten jedoch 0 sind. Insgesamt ergibt sich daraus.

Satz 3.5.1

Für die in (3.20) definierte Leslie-Matrix A seien $a_1, \dots, a_n \geq 0$, und für ein $k \in \{1, \dots, n\}$ sei $a_k > 0$ und $a_{k+1} = \dots = a_n = 0$, ferner $0 < u_i \leq 1$ für $i = 1, \dots, n-1$. Die Indizes i mit $a_i > 0$ mögen den größten gemeinsamen Teiler 1 besitzen. Dann ist der Perron-Frobenius-Eigenwert λ_0 von A positiv, algebraisch einfach, und für alle anderen Eigenwerte λ von A ist $|\lambda| < \lambda_0$. Seien $x_0 \in \mathbb{R}^n$, $x_0 \geq 0$ und x die Lösung der Anfangswertaufgabe

$$x(t+1) = A x(t), \quad x(0) = x_0.$$

Sind dann die ersten k Komponenten von x_0 gleich 0, so ist

$$\lim_{t \rightarrow \infty} x(t) = 0.$$

Ist dagegen mindestens eine der ersten k Komponenten von x_0 positiv, so existiert der Grenzwert

$$\lim_{t \rightarrow \infty} \frac{x(t)}{\lambda_0^t} =: w > 0,$$

und für die Komponenten w_1, \dots, w_n des Eigenvektors w von A zum Eigenwert λ_0 gilt

$$\frac{w_i}{w_{i+1}} = \frac{\lambda_0}{u_i} \text{ für } i = 1, \dots, n-1.$$

Der Eigenvektor w gibt die Altersverteilung wieder, die sich langfristig einstellt. Aus der letzten Gleichung ist zu sehen: Je größer λ_0 ist, desto größer sind die jüngeren Altersklassen im Verhältnis zu den älteren. Ein schnell wachsende Population wie in vielen Ländern der dritten Welt bringt überproportional viele Junge, eine schrumpfende Population wie etwa in Deutschland dagegen ein Übergewicht der Alten mit sich.

3.5.2 Populationen mit zwei Geschlechtern. Langzeitverhalten

Aufgabe 3.3

Formulieren und beweisen Sie einen entsprechenden Satz für das in 3.1.1 eingeführte Modell für altersabhängiges Populationswachstum mit zwei Geschlechtern. Zeigen Sie insbesondere, dass der oben bestimmte Perron-Frobenius-Eigenwert λ_0 , der nur von den reproduktiven weiblichen Altersklassen abhängt, auch im zweigeschlechtlichen Modell gültig bleibt.

3.5.3 Wachstum oder Zerfall?

Eine auf Grund der Modellannahmen nahe liegende Vermutung hinsichtlich der Frage, ob die Population langfristig wächst oder ausstirbt, bezieht sich auf die mittlere Anzahl μ der Töchter, die eine Frau im

Laufe ihres Lebens zur Welt bringt: Ist $\mu < 1$, so wird die Bevölkerung schrumpfen, ist $\mu > 1$, so wird sie wachsen. Bevor diese Vermutung am Modell überprüft werden kann, ist sie zunächst präzise zu fassen, d. h. es ist festzustellen, wie sich μ aus den Modellparametern ergibt. Wegen der Bedeutung der u_i als Überlebenswahrscheinlichkeiten ist für alle $k = 2, \dots, n$ das Produkt

$$u_1 u_2 \dots u_{k-1}$$

die Wahrscheinlichkeit dafür, dass ein neu geborenes Mädchen die Altersklasse k erreicht. Ferner ist a_k die mittlere Anzahl von Töchtern, die eine Frau während ihrer Zugehörigkeit zur Altersklasse k zur Welt bringt. Damit ist

$$c_k := u_1 u_2 \dots u_{k-1} a_k$$

die mittlere Anzahl von Töchtern, die eine Neugeborene im Alter k zur Welt bringen wird, was übrigens auch für $k = 1$ richtig ist. Als mittlere Anzahl von Töchtern über die gesamte Lebenszeit einer Frau ergibt sich damit die Summe

$$\mu = c_1 + c_2 + c_3 + \dots + c_n = a_1 + u_1 a_2 + u_1 u_2 a_3 + \dots + u_1 u_2 \dots u_{n-1} a_n. \quad (3.21)$$

Die oben formulierte Hypothese lautet nun unter Verwendung des Perron-Frobenius-Eigenwerts λ_0 :

$$\lambda_0 \begin{cases} < \\ = \\ > \end{cases} 1 \iff \mu \begin{cases} < \\ = \\ > \end{cases} 1.$$

Damit ist natürlich nicht gemeint, dass $\lambda_0 = \mu$, denn λ_0 ist der Faktor, um den sich die Bevölkerung langfristig in einem Zeitschritt vermehrt, während μ sich auf die mittlere Länge einer Generation bezieht.

Eine Verbindung zwischen λ_0 und μ lässt sich über das charakteristische Polynom von A herstellen. Dieses lautet nämlich, wie sich durch Entwicklung der Determinante leicht feststellen lässt:

$$\begin{aligned} p(\lambda) &= \lambda^n - (a_1 \lambda^{n-1} + u_1 a_2 \lambda^{n-2} + \dots + u_1 u_2 \dots u_{n-2} a_{n-1} \lambda + u_1 u_2 \dots u_{n-1} a_n) \\ &= \lambda^n - (c_1 \lambda^{n-1} + c_2 \lambda^{n-2} + \dots + c_{n-1} \lambda + c_n). \end{aligned}$$

Offenbar ist $p(1) = 1 - \mu$. Nun ist aber für $\lambda > 0$ die Gleichung $p(\lambda) = 0$ äquivalent zur Gleichung

$$1 = \frac{c_1}{\lambda^1} + \frac{c_2}{\lambda^2} + \frac{c_3}{\lambda^3} + \dots + \frac{c_n}{\lambda^n},$$

deren rechte Seite in λ streng monoton fällt, weshalb $\lambda = \lambda_0$ der einzige positive (und einfache) Eigenwert ist. Da andererseits $p(\lambda) \rightarrow \infty$ für $\lambda \rightarrow \infty$, folgt

$$\lambda_0 \begin{cases} < \\ = \\ > \end{cases} 1 \iff p(1) \begin{cases} > \\ = \\ < \end{cases} 0 \iff \mu \begin{cases} < \\ = \\ > \end{cases} 1.$$

Damit ist gezeigt:

Satz 3.5.2

Unter den Voraussetzungen von Satz 3.5.1 liegen die in (3.21) definierte mittlere Anzahl von Töchtern eines weiblichen Individuums und der Perron-Frobenius-Eigenwert λ_0 von A auf derselben Seite von 1.

In ORTLIEB / V. DRESKY / GASSER / GÜNZEL (2009: 107 -110) findet sich eine Analyse des Leslie-Modells, die ohne Verwendung des Satzes von Perron-Frobenius auskommt.

3.6 Kooperative Matrizen

Im Falle linearer Systeme

$$\dot{x} = A x$$

gewöhnlicher Differentialgleichungen ist der nichtnegative Oktant genau dann positiv invariant, wenn für die Systemmatrix $A = (a_{ij})$ die Außerdiagonalelemente a_{ij} mit $i \neq j$ nichtnegativ sind. Für inhomogene Systeme

$$\dot{x} = A x + u$$

muss noch hinzu kommen, dass $u \geq 0$.

Definition 3.6.1

Eine reelle (n, n) -Matrix $A = (a_{ij})$ heißt „kooperativ“, wenn $a_{ij} \geq 0$ für alle $i, j = 1, \dots, n$ mit $i \neq j$.

Der Name rührt daher, dass sich in diesem Falle die Komponenten von x in gewissem Sinne „gegenseitig stützen“: Ist $a_{ij} > 0$, so wächst x_i desto stärker, je größer x_j ist. Ein Beispiel für ein System mit kooperativen Matrizen ist das oben eingeführte Modell von Abelson zur Meinungsbildung in Gruppen, bei dem der nichtnegative Oktant übrigens keine besondere Rolle spielt. Eine andere Beispielklasse, die genau mit den zum Abelson-Modell transponierten Matrizen operiert, sind die bereits früher eingeführten kontinuierlichen Markov-Prozesse.

3.6.1 Der Satz von Perron-Frobenius für kooperative Matrizen

Anders als bei diskreten dynamischen Systemen spielen für das Langzeitverhalten der Lösungen linearer Differentialgleichungen die Eigenwerte der Systemmatrix mit dem größten *Realteil* eine besondere Rolle. Im Falle kooperativer Matrizen handelt es sich dabei wieder um einen (eindeutig bestimmten) reellen Eigenwert. Um dieses Ergebnis aus dem Satz von Perron-Frobenius zu gewinnen, kann man diesen nicht direkt auf die kooperative Matrix A anwenden, die wegen der Möglichkeit negativer Diagonalelemente dessen Voraussetzungen nicht zu erfüllen braucht. Ein einfacher Trick hilft jedoch weiter: Durch Addition eines geeigneten Vielfachen der Einheitsmatrix zu A ergibt sich eine nichtnegative Matrix $\delta I + A \geq 0$, mit deren Perron-Frobenius-Eigenwert dann weiter gearbeitet werden kann, wobei die folgenden Sachverhalte zu beachten sind:

- Der Perron-Frobenius-Eigenwert einer beliebigen nichtnegativen Matrix ist zugleich deren eindeutig bestimmter Eigenwert mit dem größten Realteil.
- Die Eigenwerte von A ergeben sich aus denen von $\delta I + A$ durch Subtraktion der reellen Zahl δ , ihre Vielfachheiten und die zugehörigen Eigen- und Hauptvektoren bleiben dabei erhalten.
- Die Subtraktion von δ ändert nichts an der Anordnung der Realteile der verschiedenen Eigenwerte.

Damit ergibt sich aus den Sätzen des Abschnitts 3.4

Satz 3.6.2

Die reelle (n, n) -Matrix A sei kooperativ. Dann existiert ein reeller Eigenwert λ_0 von A mit einem zugehörigen nichtnegativen Eigenvektor $w_0 \geq 0$, sodass für den Realteil aller anderen Eigenwerte λ von A gilt: $Re(\lambda) < \lambda_0$. Ist A irreduzibel, so hat λ_0 die algebraische und geometrische Vielfachheit 1, und es ist $w_0 > 0$.

Definition 3.6.3

Der nach Satz 3.6.2 eindeutig bestimmte Eigenwert λ_0 der kooperativen Matrix A wird als „Perron-Frobenius-Eigenwert“ von A bezeichnet.

Beachten Sie, dass anders als bei nichtnegativen Matrizen der Perron-Frobenius-Eigenwert einer kooperativen Matrix auch negativ sein kann.

3.6.2 Langzeitverhalten linearer Differentialgleichungen

In völliger Analogie zu Satz 3.4.5 und mit denselben, hier nicht noch einmal vorgebrachten Argumenten ergibt sich für das Langzeitverhalten der Lösungen eines Systems linearer Differentialgleichungen mit kooperativer Systemmatrix:

Satz 3.6.4

Der Perron-Frobenius-Eigenwert der kooperativen (n, n) -Matrix A sei algebraisch einfach. w_0 bzw. f_0 seien zugehörige nichtnegative Eigenvektoren von A bzw. A^T mit $f_0^T w_0 = 1$. Dann gilt für die Lösung von

$$\dot{x} = A x, \quad x(0) = x_0$$

mit einem Anfangswert $x_0 \in \mathbb{R}^n$:

$$\lim_{t \rightarrow \infty} \frac{x(t)}{e^{\lambda_0 t}} = (f_0^T x_0) w_0.$$

Ist $x_0 > 0$ (oder $x_0 \geq 0$ und $f_0 > 0$), so ist $f_0^T x_0 > 0$.

3.6.3 Stabilität kooperativer Matrizen

Ebenfalls in völliger Analogie zu Satz 3.4.6 gilt (Beweis als **Aufgabe**)

Satz 3.6.5

Für eine kooperative Matrix $A \in \mathbb{R}^{n,n}$ sind die folgenden Bedingungen äquivalent.

- (a) Der Perron-Frobenius-Eigenwert λ_0 von A ist negativ.
- (b) Für jede Lösung x des homogenen Systems $\dot{x} = A x$ linearer Differentialgleichungen gilt

$$\lim_{t \rightarrow \infty} x(t) = 0.$$

- (c) A ist invertierbar, und für alle $u \in \mathbb{R}^n$ und alle Lösungen x des inhomogenen Systems $\dot{x} = A x + u$ gilt

$$\lim_{t \rightarrow \infty} x(t) = -A^{-1} u.$$

- (d) Für jedes $u \in \mathbb{R}^n$ mit $u \geq 0$ besitzt das lineare Gleichungssystem $A x + u = 0$ eine eindeutige Lösung $\bar{x} \geq 0$.
- (e) A ist invertierbar, und $A^{-1} \leq 0$.
- (f) Es gibt ein $\bar{x} \in \mathbb{R}^n$ mit $\bar{x} \geq 0$ und $A \bar{x} < 0$.

3.6.4 Strenge Monotonie des Perron-Frobenius-Eigenwerts

Der folgende Satz kann zur Abschätzung des Perron-Frobenius-Eigenwerts einer kooperativen Matrix nützlich sein, wenn man den einer anderen bereits kennt:

Satz 3.6.6

Seien A und B kooperative (n, n) -Matrizen, B irreduzibel, $A \leq B$, $A \neq B$. Dann gilt für die Perron-Frobenius-Eigenwerte λ von A und μ von B :

$$\lambda < \mu.$$

Beweis: Es werde zunächst angenommen, dass auch A irreduzibel ist. Nach Satz 3.6.2 existieren dann positive Vektoren w und f mit $A w = \lambda w$ und $B^T f = \mu f$. dann ist wegen $f > 0$

$$(\mu I - A^T) f \geq 0, \quad (\mu I - A^T) f \neq 0$$

und daher

$$0 < f^T (\mu I - A) w = (\mu - \lambda) f^T w,$$

und wegen $f^T w > 0$ folgt daraus $\lambda < \mu$.

Auch wenn A nicht irreduzibel ist, so ist es doch für jedes $r \in (0, 1]$ die Matrix

$$A(r) = (1 - r) A + r B .$$

Der zugehörige Perron-Frobenius-Eigenwert $\lambda(r)$ ist daher für $r \in (0, 1]$ streng monoton wachsend, und aus Stetigkeitsgründen folgt

$$\lambda = \lambda(0) < \lambda(1) = \mu .$$



3.7 Analyse des Abelson-Modells

Das in Abschnitt 3.1.3 eingeführte Modell zur Meinungsbildung in Gruppen von Abelson lautet

$$\dot{x}_i = \sum_j a_{ij} (x_j - x_i) \text{ für } i = 1, \dots, n$$

mit der durch eine Zahl x_i ausgedrückten Meinung des Gruppenmitglieds i und den Faktoren a_{ij} als Maß für den Einfluss, den Mitglied j auf Mitglied i ausübt.

Die Systemmatrix $A = (a_{ij})$ ist kooperativ mit der zusätzlichen Eigenschaft, dass alle Zeilensummen 0 sind:

$$\sum_{j=1}^n a_{ij} = 0 \text{ für } i = 1, \dots, n .$$

Umgekehrt kommen alle Matrizen mit diesen Eigenschaften für das Abelson-Modell prinzipiell in Frage. Sie sollen daher im Folgenden *Abelson-Matrizen* genannt werden.

Offenbar ist für jede Abelson-Matrix A 0 ein Eigenwert und der n -Vektor $\mathbf{1}$, dessen Komponenten sämtlich gleich 1 sind, ein zugehöriger Eigenvektor. Andererseits sind nach Satz 3.1.2 alle Lösungen von $\dot{x} = A x$ beschränkt, sodass A keine positiven Eigenwerte besitzen kann, ferner algebraische und geometrische Vielfachheit des Eigenwerts 0 übereinstimmen müssen. Es gilt also

Satz 3.7.1

Sei $A \in \mathbb{R}^{n,n}$ eine Abelson-Matrix. Dann ist 0 der Perron-Frobenius-Eigenwert von A . Seine algebraische und geometrische Vielfachheit stimmen überein. Für jede Lösung x von $\dot{x} = A x$ existiert ein $\bar{x} \in \mathbb{R}^n$ mit $A \bar{x} = 0$, sodass

$$\lim_{t \rightarrow \infty} x(t) = \bar{x} .$$

Dieser Grenzvektor \bar{x} muss kein Vielfaches von $\mathbf{1}$ sein, wenn 0 kein einfacher Eigenwert von A ist, wie etwa für die Nullmatrix, die sicher eine Abelson-Matrix ist.

3.7.1 Stark zusammenhängende Gruppen

Ist der Digraph von A stark zusammenhängend, A also irreduzibel, so ist der Perron-Frobenius-Eigenwert 0 einfach, und aus Satz 3.6.4 folgt

Satz 3.7.2

Die Abelson-Matrix $A \in \mathbb{R}^n$ sei irreduzibel. Dann gibt es einen eindeutig bestimmten Vektor $f = (f_1, \dots, f_n) > 0$ mit $A^T f = 0$ und $f^T \mathbf{1} = \sum_{j=1}^n f_j = 1$. Für die Lösung der Anfangswertaufgabe

$$\dot{x} = A x, x(0) = x_0$$

mit $x_0 \in \mathbb{R}^n$ gilt dann

$$\lim_{t \rightarrow \infty} x_i(t) = f^T x_0 = \sum_{j=1}^n f_j x_{0j} \text{ für alle } i = 1, \dots, n .$$

Die Gruppe tendiert also langfristig zu einer gemeinsamen Meinung aller ihrer Mitglieder, und diese ist ein gewichtetes Mittel der in den Meinungsbildungsprozess eingehenden Anfangsmeinungen. Für die Gewichte f_j gilt

$$0 = \sum_{i=1}^n f_i a_{ij} = \sum_{i \neq j} f_i a_{ij} - f_j \sum_{k \neq j} a_{jk}$$

und daher

$$f_j = \frac{\sum_{i \neq j} f_i a_{ij}}{\sum_{k \neq j} a_{jk}} \text{ für } j = 1, \dots, n.$$

Es handelt sich hier *nicht* um eine explizite Formel für das Gewicht f_j , mit dem die Anfangsmeinung des Gruppenmitglieds j in die Endmeinung der Gruppe eingeht. Zu erkennen ist aber, dass dieses Gewicht desto größer ausfallen wird, je größer die Einflussfaktoren a_{ij} auf die anderen Gruppenmitglieder und je kleiner deren Einflussfaktoren a_{jk} auf das Mitglied j sind.

3.7.2 Gruppen mit genau einer Führungsclique

Besitzt der Digraph von A genau eine Basiskomponente mit $r < n$ Ecken, so verläuft der Meinungsbildungsprozess innerhalb dieser „Führungsclique“ autonom, also unbeeinflusst von den anderen Mitgliedern, entsprechend Satz 3.7.2. Der Perron-Frobenius-Eigenwert 0 von A ist auch hier einfach, denn die sich aus den anderen Zusammenhangskomponenten ergebenden Eigenwerte haben nach Satz 3.6.6 alle einen negativen Realteil. Als Eigenvektor von A ergibt sich wieder $\mathbf{1}$, als Eigenvektor von A^T dagegen ein $\mathbf{f} \geq 0$, dessen Komponenten für die Mitglieder der Führungsclique positiv, für alle anderen aber 0 sind. Mit diesem Vektor gilt dann wieder Satz 3.7.2. Der Unterschied besteht also nur darin, dass die einheitliche Endmeinung der Gruppe ausschließlich durch die Anfangsmeinungen der Mitglieder der Führungsclique beeinflusst wird, während alle anderen sich vollständig anpassen.

3.7.3 Gruppen mit mindestens zwei autonomen Untergruppen

Besitzt der Digraph von A mindestens zwei Basiskomponenten, so ist es mit der einheitlichen Endmeinung der Gruppe vorbei, denn der Meinungsbildungsprozess verläuft in jeder dieser Untergruppen autonom, sodass sich dort einheitliche, von den anderen Basiskomponenten aber in der Regel verschiedene Meinungen herausbilden. Die algebraische und geometrische Vielfachheit des Perron-Frobenius-Eigenwerts 0 von A ist gleich der Anzahl der Basiskomponenten. Zu jeder Basiskomponente gehört ein Eigenvektor $\mathbf{w} = (w_1, \dots, w_n)$ von A mit $w_i = 1$ für alle i aus der Basiskomponente, $w_i > 0$ für alle i , die von der Basiskomponente aus erreichbar sind, und $w_i = 0$ für alle anderen i . Die Summe dieser Eigenvektoren über alle Basiskomponenten ist gerade $\mathbf{1}$. Ferner gehört zu jeder Basiskomponente ein Eigenvektor $\mathbf{f} = (f_1, \dots, f_n)$ von A^T mit $f_i > 0$ für alle i aus der Basiskomponente, $f_i = 0$ für alle anderen und $\sum f_i = 1$. Der Meinungsbildungsprozess innerhalb einer Zusammenhangskomponente, die keine Basiskomponente ist, braucht dort nicht mehr zu einer einheitlichen Meinung zu führen. Auf ihn nehmen alle Basiskomponenten Einfluss, von denen aus die betrachtete Nichtbasiskomponente erreichbar ist.

3.7.4 Beispiel

Die Abelson-Matrix

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & -2 & 1 \\ 0 & 3 & 2 & -5 \end{pmatrix}$$

beschreibt eine Gruppe mit zwei autonomen Einzelnen (1 und 2) und einer stark zusammenhängenden Untergruppe ($\{3,4\}$), dessen Mitglieder sich von jeweils einem anderen der autonomen Einzelnen beeinflussen lassen (3 von 1, 4 von 2). Eigenvektoren von A zum Eigenwert 0 sind

$$\left(1, 0, \frac{5}{8}, \frac{1}{4}\right) \text{ und } \left(0, 1, \frac{3}{8}, \frac{3}{4}\right),$$

Eigenvektoren von A^T die beiden ersten Einheitsvektoren. Für die Lösung x der Anfangswertaufgabe

$$\dot{x} = A x, x(0) = x_0$$

gilt daher

$$x_1(t) \equiv x_{01}, x_2(t) \equiv x_{02}, \lim_{t \rightarrow \infty} x_3(t) = \frac{5}{8} x_{01} + \frac{3}{8} x_{02}, \lim_{t \rightarrow \infty} x_4(t) = \frac{1}{4} x_{01} + \frac{3}{4} x_{02},$$

3.8 Markov-Prozesse

Ein System möge eine endliche Anzahl n von Zuständen einnehmen können. Der Übergang von einem Zustand in einen anderen erfolge zufällig, wobei die Wahrscheinlichkeit des Übergangs ausschließlich vom aktuellen Zustand abhängt und nicht von der Vorgeschichte, auch nicht davon, wie lange der aktuelle Zustand schon eingenommen wurde. Solche Systeme werden als *Markov-Prozesse* bezeichnet. Für eine mathematische Präzisierung dieses Begriffs muss unterschieden werden, ob die Zeit, in der sich die Zustandsänderungen vollziehen, als diskret oder kontinuierlich angenommen wird.

Definition 3.8.1

Ein „Markov-Prozess in diskreter Zeit“ oder einfach „Markov-Prozess“ wird beschrieben durch eine nichtnegative (n,n) -Matrix $P = (p_{ij}) \in \mathbb{R}^{n,n}$, deren Spaltensummen gleich 1 sind:

$$\sum_{i=1}^n p_{ij} = 1 \text{ für } j = 1, \dots, n.$$

p_{ij} wird hier interpretiert als Übergangswahrscheinlichkeit vom Zustand j zum Zustand i während eines Zeitschritts:

$$p_{ij} = P(N(t+1) = i \mid N(t) = j),$$

wobei $N(t)$ den vom System zum Zeitpunkt t eingenommenen Zustand bezeichne. Der Vektor $w(t)$ der Wahrscheinlichkeiten

$$w_i(t) = P(N(t) = i),$$

dass sich das System zum Zeitpunkt t im Zustand i befinde ($i = 1, \dots, n, t \in \mathbb{N}_0$), genügt dann der Iterationsvorschrift

$$w(t+1) = P w(t),$$

ist also Lösung eines linearen, diskreten dynamischen Systems mit nichtnegativer Systemmatrix, deren Spaltensummen sämtlich 1 sind.

Definition 3.8.2

Ein „Markov-Prozess in kontinuierlicher Zeit“ oder „kontinuierlicher Markov-Prozess“ wird beschrieben durch eine kooperative (n,n) -Matrix $A = (a_{ij}) \in \mathbb{R}^{n,n}$, deren Spaltensummen gleich 0 sind:

$$\sum_{i=1}^n a_{ij} = 0 \text{ für } j = 1, \dots, n.$$

Die a_{ij} sind hier für $i \neq j$ zu interpretieren als Wahrscheinlichkeiten pro Zeiteinheit. Für ein sehr kleines Zeitintervall Δt ist

$$a_{ij} \Delta t \approx P(N(t + \Delta t) = i \mid N(t) = j)$$

die Übergangswahrscheinlichkeit vom Zustand j zum Zustand i während eines Zeitintervalls der Länge Δt . Genauer:

$$a_{ij} = \lim_{\Delta t \rightarrow 0} \frac{P(N(t + \Delta t) = i \mid N(t) = j)}{\Delta t} \text{ für } i \neq j,$$

was natürlich voraussetzt, dass diese Grenzwerte existieren und vom aktuellen Zeitpunkt t nicht abhängen. Für die Wahrscheinlichkeit, während des Zeitintervalls Δt im Zustand j zu bleiben, ergibt sich

$$\begin{aligned} P(N(t + \Delta t) = j \mid N(t) = j) &= 1 - \sum_{i \neq j} P(N(t + \Delta t) = i \mid N(t) = j) \approx \\ &\approx 1 - \sum_{i \neq j} a_{ij} \Delta t = 1 + a_{jj} \Delta t \end{aligned}$$

und damit

$$a_{jj} = - \lim_{\Delta t \rightarrow 0} \frac{1 - P(N(t + \Delta t) = j \mid N(t) = j)}{\Delta t}.$$

Für die Wahrscheinlichkeiten

$$w_i(t) = P(N(t) = i)$$

gilt

$$w_i(t + \Delta t) = \sum_{j=1}^n P(N(t + \Delta t) = i \mid N(t) = j) w_j(t) \approx w_i(t) + \Delta t \sum_{j=1}^n a_{ij} w_j(t)$$

und daher

$$\dot{w}_i = \lim_{\Delta t \rightarrow 0} \frac{w_i(t + \Delta t) - w_i(t)}{\Delta t} = \sum_{j=1}^n a_{ij} w_j(t).$$

Der Vektor $w(t)$ dieser Wahrscheinlichkeiten ist also Lösung des Systems linearer Differentialgleichungen

$$\dot{w} = A w$$

mit kooperativer Systemmatrix, deren Spaltensummen sämtlich 0 sind.

Die Systemmatrizen kontinuierlicher Markov-Prozesse sind also gerade die Transponierten von Abelson-Matrizen, deren Eigenwerte und Eigenvektoren im letzten Abschnitt analysiert wurden. Die dortigen Ergebnisse lassen sich hier also sofort nutzbar machen. Auch die diskreten Markov-Prozesse machen in dieser Hinsicht keine Schwierigkeiten: Subtrahiert man nämlich von der Übergangsmatrix eines diskreten Markov-Prozesses die Einheitsmatrix, so erhält man die Systemmatrix eines kontinuierlichen Markov-Prozesses.

Ähnlich wie im Abelson-Modell hängt das Systemverhalten entscheidend von der Zusammenhangsstruktur des Digraphen der jeweiligen Matrizen ab.

3.8.1 Klassifikation von Markov-Prozessen und ihren Zuständen

Der Digraph der Matrix P bzw. A wird auch als *Übergangs-* oder *Transitionsdigraph* des zugehörigen Markov-Prozesses bezeichnet. Durch ihn werden die *möglichen Übergänge* gekennzeichnet. Für das Langzeitverhalten des Markov-Prozesses ist die Zusammenhangsstruktur des Übergangsdigraphen entscheidend. Bei diskreten Markov-Prozessen können die Längen der Zyklen des Übergangsdigraphen von Bedeutung sein. Auch Zyklen der Länge 1 (möglicher Übergang eines Zustands in sich) sollten daher gekennzeichnet werden. Im Falle kontinuierlicher Markov-Prozesse ist das nicht nötig.

Definition 3.8.3

Eine Zusammenhangskomponente des Transitionsdigraphen eines Markov-Prozesses heißt „ergodisch“, wenn kein Pfeil aus ihr herausführt, anderenfalls heißt sie „transient“. Ein Zustand eines Markov-Prozesses heißt „ergodisch“ bzw. „transient“, wenn seine Zusammenhangskomponente es ist. Ein Zustand eines Markov-Prozesses heißt „absorbierend“, wenn von ihm im Transitionsdigraphen kein Pfeil wegführt.

Jede ergodische Zusammenhangskomponenten bildet zusammen mit ihren inneren Übergangswahrscheinlichkeiten wiederum einen Markov-Prozess. Wie noch zu zeigen sein wird, sind die ergodischen Zustände diejenigen, die langfristig angenommen werden können, während die Aufenthaltswahrscheinlichkeit in transienten Zuständen langfristig gegen 0 tendiert. Absorbierende Zustände sind ergodische Zustände, die für sich allein bereits eine ergodische Zusammenhangskomponente bilden. Nimmt das System einen absorbierenden Zustand ein, so bleibt es dort für alle Zukunft.

Markov-Prozesse werden nun danach unterschieden, welche Art von Zuständen in ihnen auftreten. Dabei ist es nützlich, zwei Extremfälle besonders hervor zu heben, aus denen sich alle Markov-Prozesse in gewisser Weise zusammensetzen lassen.

Definition 3.8.4

Ein Markov-Prozess heißt „absorbierend“, wenn alle ergodischen Zustände absorbierend sind.

Absorbierende Markov-Prozesse „landen“ letztlich in einem ihrer absorbierenden Zustände.

Definition 3.8.5

Ein Markov-Prozess heißt „ergodisch“, wenn sein Transitionsdigraph stark zusammenhängend ist.

In einem ergodischen Markov-Prozess werden alle Zustände auch langfristig mit positiver Wahrscheinlichkeit angenommen. Zu fragen ist hier nach der asymptotischen Wahrscheinlichkeitsverteilung.

Fasst man in einem beliebigen Markov-Prozess die ergodischen Komponenten jeweils zu einem einzigen Zustand zusammen, so erhält man einen absorbierenden Markov-Prozess, dessen Analyse die Verteilung der langfristig möglichen Zustände auf die verschiedenen ergodischen Komponenten liefert. Die Verteilung innerhalb jeder ergodischen Komponente erfordert dann jeweils noch die Analyse eines ergodischen Markov-Prozesses.

3.8.2 Eine deterministische Interpretation von Markov-Prozessen

Interpretiert man die n Zustände eines Markov-Prozesses als Orte, so lassen sich die Zustandsänderungen als Ortswechsel eines Teilchens deuten, das zwischen den verschiedenen Orten zufällig umherirrt (Markov'sche Irrfahrten). Betrachtet man nun statt eines einzelnen Teilchens beliebig viele bzw. eine beliebig teilbare Substanz, die auf die n Orte verteilt ist und sich zwischen diesen bewegt, so ergibt sich eine deterministische Interpretation: In diskreter Zeit ist p_{ij} zu interpretieren als der Anteil der im Ort j befindlichen Substanz, der während eines Zeitschritts in den Ort i übergeht; in kontinuierlicher Zeit ist a_{ij} für $i \neq j$ der Anteil der im Ort j befindlichen Substanz, der *pro Zeiteinheit* in den Ort i abfließt, und $-a_{jj}$ ist der *pro Zeiteinheit* aus dem Ort j insgesamt herausfließende Anteil. Die zugehörigen linearen dynamischen Systeme

$$x(t+1) = P x(t)$$

bzw.

$$\dot{x} = A x$$

sind dann als Beschreibungen von Materialflüssen zu deuten. $x_i(t)$ ist die Masse der Substanz, die sich zur Zeit t im Ort i befindet. Dabei muss die Gesamtmasse (anders als bei den Wahrscheinlichkeiten) nicht 1 sein. Die Spaltensummen-Eigenschaften von P bzw. A sorgen für ein Gesetz der *Massenerhaltung*. Es ist nämlich

$$\sum_{i=1}^n x_i(t+1) = \sum_{i=1}^n \sum_{j=1}^n p_{ij} x_j(t) = \sum_{j=1}^n \left(\sum_{i=1}^n p_{ij} \right) x_j(t) = \sum_{j=1}^n x_j(t)$$

bzw.

$$\frac{d}{dt} \sum_{i=1}^n x_i = \sum_{i=1}^n \dot{x}_i = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_j = \sum_{j=1}^n \left(\sum_{i=1}^n a_{ij} \right) x_j = 0.$$

In dieser Interpretation stellen Markov-Prozesse einen allgemeinen Rahmen für lineare **Kompartiment-Modelle** dar bzw. sind der mathematischen Struktur nach mit ihnen identisch: Gefragt ist nach der Verteilung einer Substanz (Schadstoff, Nährstoff, Medikament) in einem bestimmten System (Ökosystem, Nahrungsnetz, menschlicher Körper). Dazu wird das System in verschiedene miteinander verbundene Abteilungen aufgeteilt, die Kompartimente, zwischen denen die Substanz transportiert wird, durch welche Mechanismen auch immer. Zu klären ist jetzt noch, welche Anteile der Substanz pro Zeitschritt oder Zeiteinheit von welchem Kompartiment in welches andere transportiert werden. Zu einem *linearen* Modell führt wie immer die Annahme, dass diese Anteile konstant sind.

3.8.3 Beispiele

- (a) ROBERTS (1976, 266) gibt ein einfaches Beispiel für ein **Kompartiment-Modell** an, in dem der Weg eines Phosphat-Moleküls verfolgt wird. Es kann sich in der Humusschicht des Bodens aufhal-

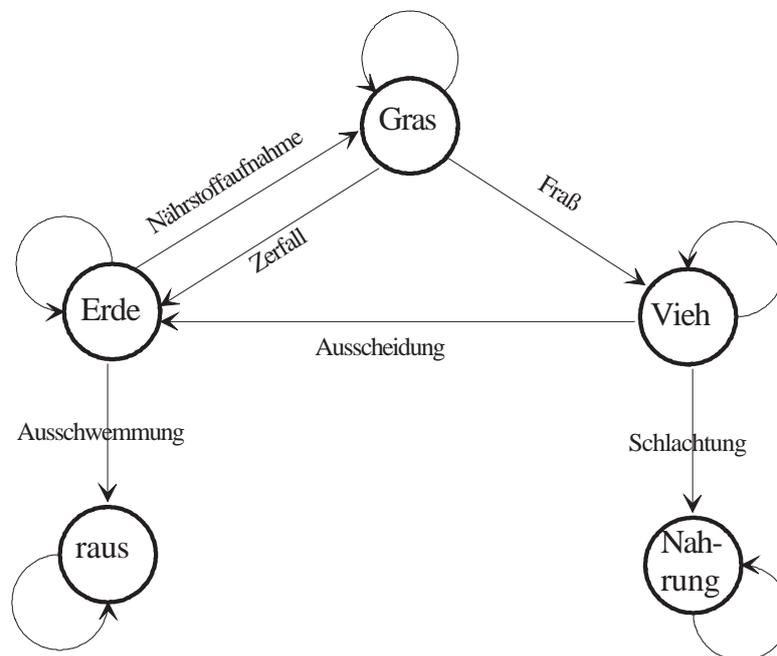


Abbildung 3.4: Wanderungen eines Phosphat-Moleküls

ten, vom Gras aufgenommen oder bei Zerfall des Grasses wieder abgegeben werden, vom Vieh gefressen und wieder ausgeschieden werden, bei Schlachtung des Viehs in die menschliche Nahrung gelangen oder auch ausgeschwemmt werden (s. Abbildung 3.4). Für die Modellierung als Markov-Prozess müssen konstante, nur vom aktuellen Aufenthaltsort des Moleküls abhängige Übergangswahrscheinlichkeiten angenommen werden. ROBERTS (1976) zitiert Zahlenwerte (leider ohne Angabe der Länge des unterstellten Zeitschritts), die mit der Numerierung der Kompartimente

$$\text{Erde} = 1, \text{Gras} = 2, \text{Vieh} = 3, \text{Nahrung} = 4, \text{raus} = 5$$

auf einen Markov-Prozess mit der Übergangsmatrix

$$P = \begin{pmatrix} 0.6 & 0.1 & 0.75 & 0 & 0 \\ 0.3 & 0.4 & 0 & 0 & 0 \\ 0 & 0.5 & 0.2 & 0 & 0 \\ 0 & 0 & 0.05 & 1 & 0 \\ 0.1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

führen. Geht man davon aus, dass die Übergänge zwischen den Kompartimenten kontinuierlich erfolgen, so ist ein kontinuierlicher Markov-Prozess zu formulieren, z. B. mit der Systemmatrix

$$A = \begin{pmatrix} -0.4 & 0.1 & 0.75 & 0 & 0 \\ 0.3 & -0.6 & 0 & 0 & 0 \\ 0 & 0.5 & -0.8 & 0 & 0 \\ 0 & 0 & 0.05 & 0 & 0 \\ 0.1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Es handelt sich hier offenbar um einen absorbierenden Markov-Prozess mit zwei absorbierenden Zuständen.

In der deterministischen Interpretation sagen diese Modelle etwas darüber aus, welche Anteile des im Boden befindlichen Phosphats wohin gelangen, etwa in die menschliche Nahrung. Um daraus Schlüsse auf die Phosphat-Konzentration im Rindfleisch schließen zu können sind allerdings weitere Informationen darüber erforderlich, welche Phosphatmengen in den Boden neu eingetragen werden, etwa über Dünger.

- (b) Ein Bootsverleih habe insgesamt n Boote. An jedem Freitagmorgen werden die defekten Boote in eine Werft gebracht, dort repariert und eine Woche später in intaktem Zustand wieder zurück gebracht. Die Wahrscheinlichkeit für ein intaktes Boot, im Laufe einer Woche durch unsachgemäßen Gebrauch defekt zu werden, sei $p \in (0, 1)$, und zwar unabhängig von den anderen Booten, vom Zeitpunkt der letzten Reparatur und von der aktuellen Woche. Mit welcher Anzahl von beim Bootsverleih stationierten Booten ist im Mittel zu rechnen?

Die Zeit ist hier offenbar diskret zu wählen mit der Zeiteinheit von einer Woche. Der Zustand des Systems werde beschrieben durch die Anzahl $N(t)$ der im Bootverleih stationierten, als *intakt* definierten Boote. Möglich ist jede ganze Zahl zwischen 0 und n . Sind in einer Woche j Boote intakt, also $n - j$ Boote defekt (in der Werft), so sind in der nächsten Woche mindestens $n - j$ und höchstens n Boote intakt. Jeder Übergang

$$j \rightarrow i \text{ mit } i \in \{n - j, \dots, n\}$$

ist also möglich. Auf Grund der Annahmen ist die Wahrscheinlichkeit für die möglichen Übergänge binomial verteilt, es ist mit $q = 1 - p$

$$P(N(t+1) = n - j + k \mid N(t) = j) = \binom{j}{k} p^{j-k} q^k$$

und damit

$$p_{ij} = P(N(t+1) = i \mid N(t) = j) = \binom{j}{i+j-n} p^{n-i} q^{i+j-n}$$

für $i, j = 0, \dots, n$. Im Falle $n = 3$ etwa führt das auf die Übergangsmatrix

$$P = \begin{pmatrix} 0 & 0 & 0 & p^3 \\ 0 & 0 & p^2 & 3 p^2 q \\ 0 & p & 2 p q & 3 p q^2 \\ 1 & q & q^2 & q^3 \end{pmatrix}.$$

Wegen der möglichen Übergänge

$$i \rightarrow n \text{ und } n \rightarrow i \text{ für alle } i = 0, \dots, n$$

handelt es sich hier um einen ergodischen Markov-Prozess, der wegen des möglichen Übergangs $n \rightarrow n$ sogar regulär ist.

HUPPERT (1990, 435) gibt dasselbe Modell für eine Situation an, in der es um den Ausfall und die Reparatur von Maschinen geht. Problematisch daran ist, dass der Ausfall einer Maschine in diesem Modell nicht davon abhängt, wie lange sie seit der letzten Reparatur läuft, Abnutzungserscheinungen also nicht berücksichtigt werden. Auch im Falle der Boote muss ausdrücklich angenommen werden, dass sie nicht durch Abnutzung, sondern nur durch unsachgemäßen Gebrauch defekt werden können. Bei Booten scheint mir das ein wenig plausibler zu klingen.

- (c) Für die Holzwirtschaft ist von Interesse, abschätzen zu können, wie schnell Bäume wachsen. Da nicht alle Bäume gleich schnell wachsen, die Gründe für unterschiedliches Wachstum aber nicht präzise bekannt sind, formuliert man das Wachstum als stochastischen Prozess. Die Bäume werden dazu in **Größenklassen** entsprechend dem Umfang ihrer Stämme eingeteilt, und dann wird erhoben, welche Übergänge von einer Größenklasse in die nächste während eines bestimmten Zeitintervalls stattfinden. MOONEY / SWIFT (1999, 130 ff) zitieren eine Untersuchung an einer Kiefernart (*Pinus sylvestris*) in Schottland. Die Bäume werden in 6 Größenklassen $1, \dots, 6$ eingeteilt, wobei 1 die dünnsten und 6 die dicksten Stämme bezeichnet. Im Abstand von 6 Jahren werden die

Klasse	Anzahl ₀	Anzahl ₁	geblieben	q_i
1	4 461	3 214	3 214	0.7205
2	2 926	3 276	2 029	0.6934
3	1 086	1 710	813	0.7486
4	222	444	171	0.7703
5	27	68	17	0.6296
6	1	11	1	1

Tabelle 3.4: Census-Daten zu *Pinus sylvestris* aus Corrou, Schottland

Bäume der verschiedenen Größenklassen gezählt (s. erste drei Spalten von Tabelle 3.4. Man beachte, dass die Gesamtzahl 8 723 der Bäume gleich geblieben, also kein Baum gestorben oder hinzu gekommen ist). Unter der *Annahme*, dass alle Bäume im Verlaufe der 6 Jahre entweder in ihrer Größenklasse geblieben oder in die nächsthöhere eingetreten sind, also keine Klasse übersprungen haben, lässt sich sofort ermitteln, wieviele in der jeweiligen Klasse verblieben sind (Spalte 4 von Tabelle 3.4). Hieraus wiederum lässt sich der Anteil q_i der in der Altersklasse i verbliebenen Bäume berechnen. Interpretiert man nun diesen Anteil als Wahrscheinlichkeit, dass ein Baum der Größenklasse i im Verlaufe von 6 Jahren in ihr verbleibt, so ergibt sich ein Markov-Prozess mit

der Übergangsmatrix

$$P = \begin{pmatrix} q_1 & 0 & 0 & 0 & 0 & 0 \\ 1 - q_1 & q_2 & 0 & 0 & 0 & 0 \\ 0 & 1 - q_2 & q_3 & 0 & 0 & 0 \\ 0 & 0 & 1 - q_3 & q_4 & 0 & 0 \\ 0 & 0 & 0 & 1 - q_4 & q_5 & 0 \\ 0 & 0 & 0 & 0 & 1 - q_5 & 1 \end{pmatrix}$$

und den q_i aus Tabelle 3.4. Untersucht werden kann nun die Verteilung auf die Größenklassen nach einer größeren Anzahl von Zeitschritten oder auch die mittlere Dauer bis zum Erreichen der größten Klasse von irgend einer anderen aus.

- (d) Als **soziale Mobilität** wird der Wechsel der gesellschaftlichen Klasse oder Schicht von einer Generation zur nächsten bezeichnet. Ein einfacher Indikator für Schichtzugehörigkeit ist das Familieneinkommen. ROBERTS (1976, 267) zitiert eine im Jahre 1949 in England und Wales durchgeführte Untersuchung, in der männliche Personen nach ihrer eigenen Schichtzugehörigkeit und der ihres Vaters befragt wurden. Unterschieden wurde dabei eine untere (1), eine mittlere (2) und eine obere (3) Schicht. Aus der Befragung lässt sich ermitteln, welche Anteile der jeweiligen Schichten in der nächsten Generation in welcher Schicht landen. Als Ergebnis entsteht die Übergangsmatrix eines Markov-Prozesses:

$$P = \begin{pmatrix} 0.448 & 0.054 & 0.011 \\ 0.484 & 0.699 & 0.503 \\ 0.068 & 0.247 & 0.486 \end{pmatrix}.$$

Ihr ist etwa zu entnehmen, dass 5.4 Prozent der Söhne von Mittelschicht-Vätern in der Unterschicht und 24.7 Prozent in der Oberschicht gelandet sind, während 50.3 Prozent der in die Oberschicht Geborenen in die Mittelschicht abgestiegen sind. Was lässt sich mit derartigen Daten anfangen?

Für sich genommen sind sie ein gewisser Indikator für die Festigkeit bzw. Verflüssigung von Klassenstrukturen: Die Einheitsmatrix würde eine extreme Klassengesellschaft kennzeichnen, in der jeder in der Klasse bleibt, in die er hineingeboren wurde. Umgekehrt wäre bei einer Matrix mit sämtlichen Einträgen $1/3$ die spätere Schichtzugehörigkeit unabhängig von der Schicht, in die jemand hineingeboren wurde. Für die Soziologie könnten daher vor allem Veränderungen der Koeffizienten im Laufe der Zeit von Interesse sein.

Mit einem Markov-Prozess $w(t+1) = P w(t)$ ließe sich die langfristige Entwicklung der Anteile der drei Schichten an der Gesamtbevölkerung untersuchen. Dazu müssen aber zwei Annahmen gemacht werden, die nicht unbedingt realistisch sind, dass nämlich die Koeffizienten im Laufe der Zeit konstant bleiben und dass die Anzahl der Kinder von der Schichtzugehörigkeit nicht abhängt.

3.8.4 Kontinuierliche Markov-Prozesse

Sei $A \in \mathbb{R}^{n,n}$ die Systemmatrix eines kontinuierlichen Markov-Prozesses. Betrachtet wird das System linearer Differentialgleichungen

$$\dot{x} = A x, \quad (3.22)$$

wobei x nicht unbedingt eine Wahrscheinlichkeitsverteilung zu sein braucht (Materialflüsse).

A^T ist eine Abelson-Matrix, deren Perron-Frobenius-Eigenwert 0 samt den zugehörigen Links- und Rechtseigenvektoren im letzten Abschnitt bereits analysiert wurde. Zu beachten ist dabei, dass die Baskomponenten (Führungscliquen) von A^T gerade die ergodischen Komponenten von A sind:

Hilfssatz 3.8.6

Der Perron-Frobenius-Eigenwert der Systemmatrix A eines kontinuierlichen Markov-Prozesses ist 0. Seine algebraische Vielfachheit stimmt mit der geometrischen überein und ist gleich der Anzahl der ergodischen Komponenten. Zu jeder ergodischen Komponente gehört

- genau ein Rechtseigenvektor von A zum Eigenwert 0 mit positiven Einträgen für die Mitglieder der Komponente und verschwindenden Einträgen sonst, wobei die Summe der positiven Einträge 1 ist,
- genau ein Linkseigenvektor von A zum Eigenwert 0 mit Einträgen 1 für die Mitglieder der Komponente und zwischen 0 und 1 liegenden Einträgen sonst.

Die Summe dieser Linkseigenvektoren ist 1, der n -Vektor, dessen sämtliche Komponenten 1 sind.

Hieraus ergibt sich als unmittelbare Folgerung:

Satz 3.8.7

Sei $A \in \mathbb{R}^{n,n}$ die Systemmatrix eines kontinuierlichen Markov-Prozesses. Für jede Lösung x von (3.22) ist

$$\sum_{i=1}^n x_i(t) = \sum_{i=1}^n x_i(0) \text{ für alle } t \geq 0 .$$

Mit $x(0)$ ist auch $x(t)$ für alle $t \geq 0$ eine Wahrscheinlichkeitsverteilung auf den Zuständen $1, \dots, n$.

Für jede Lösung x von (3.22) existiert ferner ein $w \in \mathbb{R}^n$ mit $A w = 0$, sodass

$$\lim_{t \rightarrow \infty} x(t) = w ,$$

wobei $w_i = 0$ für alle transienten Zustände i .

Im Falle nur einer ergodischen Komponente ist dieser Vektor w bis auf Vielfache eindeutig bestimmt:

Satz 3.8.8

Sei $A \in \mathbb{R}^{n,n}$ die Systemmatrix eines kontinuierlichen Markov-Prozesses mit genau einer ergodischen Komponente. Dann existiert eine eindeutig bestimmte Wahrscheinlichkeitsverteilung $w \in \mathbb{R}^n$ auf $\{1, \dots, n\}$ mit $A w = 0$. Es ist $w_i = 0$ für alle transienten und $w_i > 0$ für alle ergodischen Zustände $i \in \{1, \dots, n\}$, und für jede Lösung x von (3.22) ist

$$\lim_{t \rightarrow \infty} x(t) = \left(\sum_{i=1}^n x_i(0) \right) w .$$

Mit der „Gesamtmenge“ am Anfang steht also die Endverteilung eindeutig fest. Für Materialflüsse bedeutet

$$\lim_{t \rightarrow \infty} \dot{x}(t) = \left(\sum_{i=1}^n x_i(0) \right) A w = 0$$

nicht, dass die Flüsse zum Erliegen kommen, sondern nur, dass für jedes Kompartiment Zu- und Abflüsse in der Summe übereinstimmen, sodass sich an der Nettobilanz nichts mehr ändert.

Diffusion

Können sich die Teilchen einer Substanz zwischen je zwei Kompartimenten in beiden Richtungen in gleicher Weise bewegen, so ergibt sich für den zugehörigen kontinuierlichen Markov-Prozess eine symmetrische Systemmatrix A . Alle Spalten- und alle Zeilensummen von A sind Null, A ist also nicht nur Matrix eines kontinuierlichen Markov-Prozesses, sondern auch Abelson-Matrix, 1 ist sowohl Rechts-, als auch Linkseigenvektor von A zum Perron-Frobenius-Eigenwert 0, dessen algebraische Einfachheit in diesem Fall damit äquivalent ist, dass A irreduzibel. Für diesen speziellen Fall ergibt sich

Satz 3.8.9

$A \in \mathbb{R}^{n,n}$ sei kooperativ, symmetrisch und irreduzibel, alle Spalten- und Zeilensummen von A seien Null. Dann gilt für jede Lösung von (3.22)

$$\lim_{t \rightarrow \infty} x(t) = \frac{1}{n} \left(\sum_{i=1}^n x_i(0) \right) \mathbf{1}.$$

Wie bei Abelson-Matrizen üblich, lässt sich (3.22) hier auch in der Form

$$\dot{x}_i = \sum_{j=1}^n a_{ij} (x_j - x_i) \text{ für } i = 1, \dots, n$$

schreiben. Solche Systeme sind Modelle für Diffusionsprozesse, in den sich Teilchen in einem homogenen Medium zufällig bewegen, wobei keine Richtung bevorzugt ist. Der Raum ist dabei nicht kontinuierlich, sondern in endlich viele Teile zerlegt (Kompartimente), wobei die Teilchen sich von jedem Kompartiment in ein benachbartes Kompartiment bewegen können. Die Nachbarschaft zwischen zwei Kompartimenten i und j ist dabei durch $a_{ij} > 0$ gekennzeichnet. Der eben formulierte Satz besagt, dass diese Prozesse, wie generell bei Abelson-Matrizen, zum Ausgleich führen. Da andererseits Massenerhaltung gilt, ist der Endzustand durch den Anfangszustand festgelegt.

3.8.5 Reguläre Markov-Prozesse

Sei $P \in \mathbb{R}^{n,n}$ die Systemmatrix eines Markov-Prozesses in diskreter Zeit. Betrachtet wird das diskrete dynamische System

$$x(t+1) = P x(t), \quad (3.23)$$

wobei x auch hier nicht unbedingt eine Wahrscheinlichkeitsverteilung zu sein braucht (Materialflüsse).

Subtrahiert man von P die Einheitsmatrix, so erhält man mit $A = P - I$ die Systemmatrix eines kontinuierlichen Markov-Prozesses. Aus Hilfssatz 3.8.6 folgt

Hilfssatz 3.8.10

Der Perron-Frobenius-Eigenwert der Systemmatrix P eines Markov-Prozesses in diskreter Zeit ist 1. Seine algebraische Vielfachheit stimmt mit der geometrischen überein und ist gleich der Anzahl der ergodischen Komponenten. Zu jeder ergodischen Komponente gehört

- genau ein Rechtseigenvektor von P zum Eigenwert 1 mit positiven Einträgen für die Mitglieder der Komponente und verschwindenden Einträgen sonst, wobei die Summe der positiven Einträge 1 ist,
- genau ein Linkseigenvektor von P zum Eigenwert 1 mit Einträgen 1 für die Mitglieder der Komponente und zwischen 0 und 1 liegenden Einträgen sonst.

Die Summe dieser Linkseigenvektoren ist 1, der n -Vektor, dessen sämtliche Komponenten 1 sind.

Ohne Zusatzbedingungen lässt sich allerdings keine Aussage darüber machen, ob alle anderen Eigenwerte von P dem Betrage nach kleiner als 1 sind, was für das Langzeitverhalten der Lösungen von (3.23) große Bedeutung hat. Eine Bedingung dafür liefert Satz 3.2.19:

Definition 3.8.11

Ein Markov-Prozess in diskreter Zeit heißt „regulär“, wenn in jeder ergodischen Komponente die im Transitionsdigraphen auftretenden Zyklen den größten gemeinsamen Teiler 1 haben.

Hilfssatz 3.8.12

Bei einem regulären Markov-Prozess sind alle von 1 verschiedenen Eigenwerte der Transitionsmatrix P betragsmäßig kleiner als 1.

Für reguläre Markov-Prozesse lassen sich nun die Ergebnisse des kontinuierlichen Falls einfach übertragen:

Satz 3.8.13

Sei $P \in \mathbb{R}^{n,n}$ die Systemmatrix eines regulären Markov-Prozesses. Für jede Lösung x von (3.23) ist

$$\sum_{i=1}^n x_i(t) = \sum_{i=1}^n x_i(0) \text{ für alle } t \geq 0 .$$

Mit $x(0)$ ist auch $x(t)$ für alle $t \geq 0$ eine Wahrscheinlichkeitsverteilung auf den Zuständen $1, \dots, n$.

Für jede Lösung x von (3.23) existiert ferner ein $w \in \mathbb{R}^n$ mit $P w = w$, sodass

$$\lim_{t \rightarrow \infty} x(t) = w ,$$

wobei $w_i = 0$ für alle transienten Zustände i .

Satz 3.8.14

Sei $P \in \mathbb{R}^{n,n}$ die Systemmatrix eines regulären Markov-Prozesses mit genau einer ergodischen Komponente. Dann existiert eine eindeutig bestimmte Wahrscheinlichkeitsverteilung $w \in \mathbb{R}^n$ auf $\{1, \dots, n\}$ mit $P w = w$. Es ist $w_i = 0$ für alle transienten und $w_i > 0$ für alle ergodischen Zustände $i \in \{1, \dots, n\}$, und für jede Lösung x von (3.23) ist

$$\lim_{t \rightarrow \infty} x(t) = \left(\sum_{i=1}^n x_i(0) \right) w .$$

Aufgabe 3.4

Zeigen Sie in Beispiel (b) aus 3.8.3 (Bootsverleih) für die Wahrscheinlichkeit, dass in einer bestimmten Woche gerade i Boote intakt sind:

$$\lim_{t \rightarrow \infty} P(N(t) = i) = \frac{\binom{n}{i} p^{n-i}}{(1+p)^n} \text{ für } i = 0, \dots, n$$

Aufgabe 3.5

In Beispiel (d) aus 3.8.3 (soziale Mobilität) sei vorausgesetzt, dass die angegebenen Übergangskoeffizienten konstant und die durchschnittliche Anzahl der Nachkommen von der Schichtzugehörigkeit unabhängig sind. Welche Anteile der drei Schichten an der Gesamtbevölkerung ergeben sich dann langfristig?

3.8.6 Ein irregulärer Markov-Prozess

In Abbildung 3.5 ist ein stark zusammenhängender Digraph zu sehen, dessen Zyklenlängen allesamt ein Vielfaches von 3 sind. Ein Markov-Prozess mit diesem Digraphen ist ergodisch, aber nicht regulär. Auf der Menge der Ecken $\{1, \dots, 10\}$ ist durch

$$i \sim j :\Leftrightarrow \text{ es gibt von } j \text{ nach } i \text{ einen Weg, dessen Länge ein Vielfaches von 3 ist}$$

eine Äquivalenzrelation definiert. Die Ecken wurden so numeriert, dass

$$\{1, 2, 3, 4\} , \{5, 6, 7\} , \{8, 9, 10\}$$

die zugehörigen Äquivalenzklassen sind. Die zugehörige Transitionsmatrix hat die Gestalt

$$P = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & + & + & + \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & + \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & + & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & + & 0 \\ + & + & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ + & 0 & + & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ + & 0 & 0 & + & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & + & + & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & + & + & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & + & 0 & + & 0 & 0 & 0 \end{pmatrix} ,$$

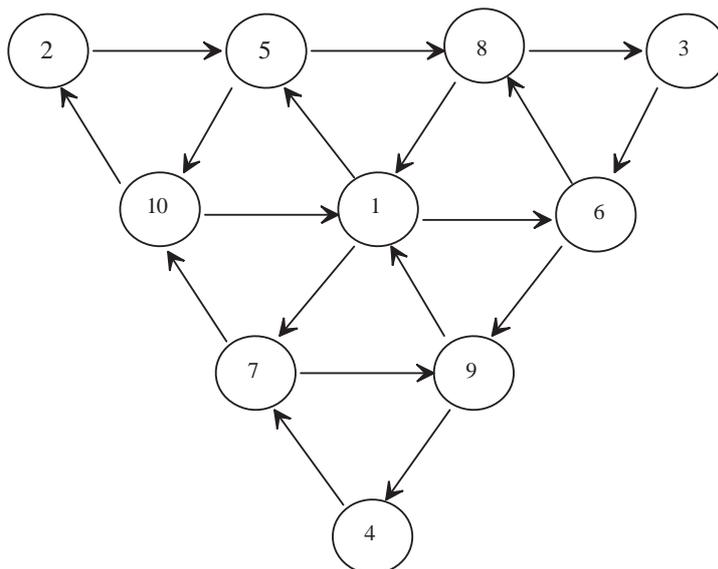


Abbildung 3.5: Digraph eines ergodischen, irregulären Markov-Prozesses

wobei „+“ für eine positive Zahl steht. Pfeile führen nur von der ersten zur zweiten, von der zweiten zur dritten und von der dritten zur ersten Äquivalenzklasse. Bildet man nun die dritte Potenz von P, so ergibt sich eine Blockdiagonalmatrix, da Wege der Länge 3 nur zwischen Ecken derselben Äquivalenzklasse möglich sind:

$$P^3 = \begin{pmatrix} + & + & + & + & 0 & 0 & 0 & 0 & 0 & 0 \\ + & + & 0 & + & 0 & 0 & 0 & 0 & 0 & 0 \\ + & + & + & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ + & 0 & + & + & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & + & + & + & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & + & + & + & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & + & + & + \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & + & + & + \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & + & + & + \end{pmatrix}.$$

Jede der drei Matrizen in der Blockdiagonalen definiert für sich einen Markov-Prozess und hat daher den Eigenwert 1, der damit ein dreifacher Eigenwert von P^3 ist. Andererseits ist 1 ein einfacher Eigenwert von P. Daher muss P auch die beiden anderen dritten Wurzeln von 1 als (einfache) Eigenwerte haben. P besitzt also die drei einfachen Eigenwerte

$$1, -\frac{1}{2} + \frac{i}{2}\sqrt{3}, -\frac{1}{2} - \frac{i}{2}\sqrt{3},$$

die allesamt den Absolutbetrag 1 haben. Deshalb konvergieren die Lösungen von (3.23) im Allgemeinen nicht gegen einen konstanten Vektor, sondern gegen eine periodische Lösung mit Periode 3.

Die hier am Beispiel durchgeführten Überlegungen gelten allgemeiner für ergodische, aber nicht reguläre Markov-Prozesse, vgl. HUPPERT (1990).

3.8.7 Absorbierende Markov-Prozesse

Aus den Sätzen 3.8.7 und 3.8.13 folgt unmittelbar, dass alle Lösungen von (3.22) bzw. (3.23) im Falle eines absorbierenden Markov-Prozesses langfristig gegen einen Vektor $w \in \mathbb{R}^n$ konvergieren mit $w_i = 0$ für alle transienten Zustände i . Die Frage ist, wie sich die übrigen Gewichte auf die absorbierenden Zustände verteilen, mit welcher Wahrscheinlichkeit also ein in irgend einem Zustand i startendes Teilchen in welchem absorbierenden Zustand landet. Die Antwort hängt vom Anfangszustand ab, sofern es, wie etwa in Beispiel (a) aus 3.8.3, mindestens zwei absorbierende Zustände gibt. Gibt es dagegen, wie in Beispiel (c) aus 3.8.3, nur einen absorbierenden Zustand, so ist die Antwort klar: Jedes Teilchen landet, unabhängig von seinem Anfangszustand, langfristig in dem absorbierenden Zustand. In beiden Fällen bleibt die Frage, wie lange das dauert, wie lange sich das Teilchen also in der Menge der transienten Zustände aufhält.

Zur Behandlung dieser Fragen ändere ich die bisherige Symbolik etwas ab, um die Unterscheidung zwischen transienten und absorbierenden Zuständen deutlicher werden zu lassen: Der Markov-Prozess möge n transiente und m absorbierende Zustände besitzen. Im Falle eines Markov-Prozesses in diskreter Zeit bezeichne p_{jk} die Übergangswahrscheinlichkeit vom transienten Zustand k in den transienten Zustand j und q_{ij} die Übergangswahrscheinlichkeit vom transienten Zustand j in den absorbierenden Zustand i für $i = 1, \dots, m$ und $j, k = 1, \dots, n$. Mit $P = (p_{jk}) \in \mathbb{R}^{n,n}$, $Q = (q_{ij}) \in \mathbb{R}^{m,n}$ und einer passenden Nullmatrix 0 und Einheitsmatrix I hat dann die Systemmatrix des absorbierenden Markov-Prozesses die Gestalt

$$\begin{pmatrix} P & 0 \\ Q & I \end{pmatrix},$$

und (3.23) erhält die Form

$$\begin{aligned} x(t+1) &= P x(t) \\ y(t+1) &= Q x(t) + y(t) \end{aligned} \quad (3.24)$$

für die Wahrscheinlichkeits- bzw. Materialverteilungen x auf den transienten und y auf den absorbierenden Zuständen. Die Dynamik auf den transienten Zuständen lässt sich ohne Betrachtung der absorbierenden Zustände untersuchen. Die nichtnegative Matrix P lässt sich durch die folgenden Eigenschaften charakterisieren: Ihre Spaltensummen sind alle kleiner oder gleich 1, und in ihrem Digraphen führt von jeder Ecke k ein Weg zu einer Ecke j , sodass die Summe der j -ten Spalte echt kleiner als 1 ist.

Entsprechend bezeichne im Falle eines kontinuierlichen Markov-Prozesses a_{jk} die Übergangswahrscheinlichkeit pro Zeiteinheit vom transienten Zustand k in den transienten Zustand $j \neq k$ und b_{ij} die Übergangswahrscheinlichkeit pro Zeiteinheit vom transienten Zustand j in den absorbierenden Zustand i für $i = 1, \dots, m$ und $j, k = 1, \dots, n$. Mit $A = (a_{jk}) \in \mathbb{R}^{n,n}$, $B = (b_{ij}) \in \mathbb{R}^{m,n}$ und passenden Nullmatrizen 0 hat dann die Systemmatrix des absorbierenden, kontinuierlichen Markov-Prozesses die Gestalt

$$\begin{pmatrix} A & 0 \\ B & 0 \end{pmatrix},$$

und (3.22) erhält die Form

$$\begin{aligned} \dot{x} &= A x \\ \dot{y} &= B x \end{aligned} \quad (3.25)$$

für die Wahrscheinlichkeits- bzw. Materialverteilungen x auf den transienten und y auf den absorbierenden Zuständen. Auch hier lässt sich die Dynamik auf den transienten Zuständen ohne Betrachtung der absorbierenden Zustände untersuchen. Die kooperative Matrix A lässt sich durch die folgenden Eigenschaften charakterisieren: Ihre Spaltensummen sind alle kleiner oder gleich 0, und in ihrem Digraphen führt von jeder Ecke k ein Weg zu einer Ecke j , sodass die Summe der j -ten Spalte negativ ist.

Langfristige Verteilung auf die absorbierenden Zustände

Zur Bestimmung des Langzeitverhaltens der Lösungen von (3.24) bzw. (3.25) werden die Recht- und Linkseigenvektoren zum Perron-Frobenius-Eigenwert 1 der Systemmatrix von (3.24) bzw. 0 der Systemmatrix von (3.25) benötigt. Diese lassen sich den absorbierenden Zuständen zuordnen: Zu jedem absorbierenden Zustand i gehört ein

$$\text{Rechtseigenvektor } w_i = \begin{pmatrix} 0 \\ e_i \end{pmatrix}$$

mit dem Nullvektor $0 \in \mathbb{R}^n$ und dem i -ten Einheitsvektor $e_i \in \mathbb{R}^m$ sowie ein

$$\text{Linkseigenvektor } f_i = \begin{pmatrix} v_i \\ e_i \end{pmatrix}$$

mit $v_i \in \mathbb{R}^n$ und dem i -ten Einheitsvektor $e_i \in \mathbb{R}^m$, wobei

$$v_i - P^T v_i = Q^T e_i \text{ bzw. } -A^T v_i = B^T e_i .$$

Es ist dann

$$f_i^T w_i = 1 \text{ und } f_i^T w_j = 0 \text{ für } i \neq j \text{ (} i, j = 1, \dots, m \text{)} .$$

Für jede Lösung (x, y) von (3.24) bzw. (3.25) mit Anfangswerten $x(0) = x_0, y(0) = y_0$ gilt

$$\lim_{t \rightarrow \infty} \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \sum_{i=1}^m (f_i^T \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}) w_i$$

und daher

$$\lim_{t \rightarrow \infty} x(t) = 0, \quad \lim_{t \rightarrow \infty} y(t) = \sum_{i=1}^m (v_i^T x_0 + y_{0i}) e_i .$$

Betrachtet man speziell die Anfangswerte $x_0 = e_k$ und $y_0 = 0$, so ist mit der k -ten Komponente v_{ik} von v_i :

$$\lim_{t \rightarrow \infty} y_i(t) = v_{ik} .$$

v_{ik} ist also zu interpretieren als Wahrscheinlichkeit dafür, dass ein im transienten Zustand k befindliches Teilchen langfristig im absorbierenden Zustand i landet. Es gilt also

Satz 3.8.15

Bei einem absorbierenden Markov-Prozess mit den oben eingeführten Bezeichnungen sei v_{ik} für einen absorbierenden Zustand i und einen transienten Zustand k die Wahrscheinlichkeit dafür, dass ein in k befindliches Teilchen irgendwann in i landet. Dann gilt für alle $i = 1, \dots, m$ und alle $k = 1, \dots, n$

- für einen Markov-Prozess in diskreter Zeit

$$v_{ik} = q_{ik} + \sum_{j=1}^n v_{ij} p_{jk} , \quad (3.26)$$

- für einen kontinuierlichen Markov-Prozess

$$b_{ik} + \sum_{j=1}^n v_{ij} a_{jk} = 0 . \quad (3.27)$$

Die bei einem Markov-Prozess in diskreter Zeit geltenden Gleichungen (3.26) lassen sich auch direkt durch wahrscheinlichkeitstheoretische Überlegungen ableiten: Befindet sich ein Teilchen im Zustand k , so geht es mit Wahrscheinlichkeit q_{ik} im nächsten Zeitschritt direkt in den Zustand i über; mit Wahrscheinlichkeit p_{jk} geht es in den transienten Zustand j über, von dem aus es mit Wahrscheinlichkeit v_{ij} irgendwann in Zustand i landet.

Mit der Matrix $V = (v_{ik}) \in \mathbb{R}^{m,n}$ lassen sich die Gleichungen (3.26) bzw. (3.27) auch schreiben als

$$V = Q + V P \text{ bzw. } B + V A = 0$$

mit der (wegen der Eigenschaften von P bzw. A) eindeutigen Lösung

$$V = Q (I - P)^{-1} \text{ bzw. } V = -B A^{-1} .$$

Beispiel: Wanderung eines Phosphatteilchens

In Beispiel (a) aus 3.8.3 ist ein absorbierender Markov-Prozess mit drei transienten und zwei absorbierenden Zuständen angegeben. In der diskreten Formulierung ist

$$P = \begin{pmatrix} 0.6 & 0.1 & 0.75 \\ 0.3 & 0.4 & 0 \\ 0 & 0.5 & 0.2 \end{pmatrix}, Q = \begin{pmatrix} 0 & 0 & 0.05 \\ 0.1 & 0 & 0 \end{pmatrix} .$$

Hier ist auf drei Dezimalstellen genau

$$V = Q (I - P)^{-1} = \begin{pmatrix} 0.135 & 0.180 & 0.189 \\ 0.865 & 0.820 & 0.811 \end{pmatrix} .$$

Daraus ist abzulesen, dass etwa ein im Gras befindliches Phosphatteilchen (Spalte 2) mit der Wahrscheinlichkeit 0.18 in die Nahrung gelangt und mit der Komplementärwahrscheinlichkeit 0.82 aus dem System ausgeschwemmt wird.

Die kontinuierliche Fassung mit

$$A = \begin{pmatrix} -0.4 & 0.1 & 0.75 \\ 0.3 & -0.6 & 0 \\ 0 & 0.5 & -0.8 \end{pmatrix}, B = \begin{pmatrix} 0 & 0 & 0.05 \\ 0.1 & 0 & 0 \end{pmatrix}$$

liefert das gleiche Ergebnis, was einfach daran liegt, dass hier $-A = I - P$ und $B = Q$.

Dauer des Aufenthalts in transienten Zuständen

Wie lange dauert es, bis ein absorbierender Zustand erreicht wird? Zur Beantwortung dieser Frage betrachte ich für eine Lösung x zunächst von (3.24) die Größe

$$u(t) := 1^T x(t) = \sum_{k=1}^n x_k(t),$$

wobei $u(0) = 1$ vorausgesetzt sei. $u(t)$ ist dann die Wahrscheinlichkeit, dass das Teilchen sich zum Zeitpunkt t noch in einem transienten Zustand befindet, oder in der deterministischen Interpretation, der Anteil der Substanz, die bis zum Zeitpunkt t noch nicht absorbiert wurde. Offenbar ist

$$u(t) - u(t+1)$$

dann gerade der Anteil, der zum Zeitpunkt $t + 1$ einen absorbierenden Zustand erreicht. Für die mittlere Dauer T bis zum Erreichen eines absorbierenden Zustands ergibt sich damit

$$T = \sum_{t=0}^{\infty} (t+1) (u(t) - u(t+1)) = \sum_{t=0}^{\infty} (t+1) u(t) - \sum_{t=1}^{\infty} t u(t) = \sum_{t=0}^{\infty} u(t) = 1^T \sum_{t=0}^{\infty} x(t).$$

Die Konvergenz dieser unendlichen Reihen ist gesichert, weil alle Komponenten von $x(t)$ sich aus geometrisch gegen Null konvergierenden Folgen zusammensetzen, da die Eigenwerte von P betragsmäßig alle kleiner als 1 sind. Nun ergibt sich aber aus

$$(I - P) \sum_{t=0}^s x(t) = \sum_{t=0}^s x(t) - \sum_{t=1}^{s+1} x(t) = x(0) - x(s+1)$$

im Grenzübergang $s \rightarrow \infty$

$$(I - P) \sum_{t=0}^{\infty} x(t) = x(0)$$

und daher

$$T = 1^T (I - P)^{-1} x(0).$$

Die k -te Komponente T_k des Vektors

$$T^T = (T_1, \dots, T_n) = 1^T (I - P)^{-1}$$

(es ist $T = T_k$, falls $x(0) = e_k$) ist gerade die mittlere Dauer, die ein im Zustand k befindliches Teilchen benötigt, um in einen absorbierenden Zustand zu gelangen. Schreibt man

$$T^T (I - P) = 1^T$$

komponentenweise aus, so ergibt sich

$$T_k = 1 + \sum_{j=1}^n T_j p_{jk} \text{ für } k = 1, \dots, n. \quad (3.28)$$

Diese linearen Gleichungen lassen sich wiederum direkt wahrscheinlichkeitstheoretisch begründen: Befindet sich das Teilchen im transienten Zustand k , so benötigt es mit Sicherheit mindestens einen Zeitschritt, um in einen absorbierenden Zustand zu gelangen; gelangt es im nächsten Zeitschritt in den transienten Zustand j , was mit Wahrscheinlichkeit p_{jk} eintritt, so benötigt es von dort aus im Mittel weitere T_j Zeitschritte bis in einen absorbierenden Zustand. Mit der Gegenwahrscheinlichkeit $1 - \sum_{j=1}^n p_{jk}$ gelangt es direkt in einen absorbierenden Zustand.

Im Fall des kontinuierlichen, absorbierenden Markov-Prozesses lässt sich eine analoge Rechnung durchführen. Hier ist

$$T = - \int_0^{\infty} t \dot{u}(t) dt = -[t u(t)]_{t=0}^{t=\infty} + \int_{t=0}^{\infty} u(t) dt = \int_{t=0}^{\infty} u(t) dt = 1^T \int_{t=0}^{\infty} x(t) dt.$$

Zu beachten ist dabei, dass in der deterministischen Interpretation des Markov-Prozesses $-\dot{u}$ der Anteil ist, der pro Zeiteinheit von einem transienten in einen absorbierenden Zustand übertritt. Aus

$$\int_0^{\infty} x(t) dt = \int_0^{\infty} A^{-1} \dot{x}(t) dt = A^{-1} \int_0^{\infty} \dot{x}(t) dt = -A^{-1} x(0)$$

ergibt sich

$$T = -1^T A^{-1} x(0)$$

und damit für die oben eingeführte Übergangszeit T_k vom transienten Zustand k in einen absorbierenden Zustand:

$$1 + \sum_{j=1}^n T_j a_{jk} = 0 \text{ für } k = 1, \dots, n. \quad (3.29)$$

Aufgabe 3.6

Bestimmen Sie die Verweildauer eines Phosphatteilchens im Erde-Gras-Vieh-System aus Beispiel (a) in 3.8.3 in Abhängigkeit von seinem aktuellen Aufenthaltsort.

Aufgabe 3.7

Wie lange braucht eine durchschnittliche Kiefer aus Beispiel (c) in 3.8.3, um von der ersten in die letzte Größenklasse zu wachsen?

3.8.8 Lineare Kompartiment-Modelle mit externen Einträgen

Das System der transienten Zustände eines absorbierenden, kontinuierlichen Markov-Prozesses

$$\dot{x} = A x$$

hat die Eigenschaft, dass für alle seine Lösungen gilt:

$$\lim_{t \rightarrow \infty} x(t) = 0 .$$

Die im System vorhandene Substanz fließt langfristig vollständig ab, wo immer sie sich befinden mag. Bei realen Systemen, die durch solche Modelle beschrieben werden, fließt die Substanz in der Regel an irgendwelchen Stellen auch in das System ein. So kann man etwa bei dem Phosphat-Modell des Beispiels (a) in 3.8.3 davon ausgehen, dass das Phosphat durch Düngung in die Erde eingebracht wird. Das Modell erhält dann die Form

$$\dot{x} = A x + u(t)$$

eines inhomogenen Systems linearer Differentialgleichungen. Dabei ist $u(t) \geq 0$, die j -te Komponente $u_j(t)$ ist der Eintrag der Substanz pro Zeiteinheit in das Kompartiment j . Empfängt das System nun dauerhaft solche externen Einträge, die andererseits aber auch ständig wieder abfließen, so stellt sich die Frage nach der langfristigen Massenverteilung.

Konstante Einträge

Es sei $u(t) = u$ konstant. Dann besitzt das inhomogene System eine eindeutige bestimmte konstante Lösung

$$x(t) = \bar{x} = -A^{-1} u .$$

Da alle Lösungen des inhomogenen Systems sich von dieser konstanten Lösung nur um die Lösungen des homogenen Systems unterscheiden, die jedoch langfristig verschwinden, streben alle Lösungen des inhomogenen System langfristig gegen \bar{x} , es handelt sich also um die Massenverteilung, die sich langfristig einstellt.

Beispiel: In dem Phosphat-System aus Beispiel (a) in 3.8.3 werde dauerhaft eine Mengeneinheit pro Zeiteinheit in die Erde eingetragen:

$$\dot{x} = \begin{pmatrix} -0.4 & 0.1 & 0.75 \\ 0.3 & -0.6 & 0 \\ 0 & 0.5 & -0.8 \end{pmatrix} x + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} .$$

Langfristig stellt sich dann (auf 4 Dezimalstellen genau) die Verteilung

$$\bar{x} = \begin{pmatrix} 8.649 \\ 4.324 \\ 2.703 \end{pmatrix}$$

des Phosphats auf die drei Kompartimente ein. Man beachte dabei die unterschiedlichen Maßeinheiten für x (Mengeinheiten) und u (Mengeinheiten pro Zeiteinheit). Die Unterschiede in der Größenordnung der Zahlenwerte kommen dadurch zu Stande, dass je Zeiteinheit nur 10 Prozent aus dem ersten Kompartiment (Erde) und 5 Prozent aus dem dritten Kompartiment (Vieh) abfließen, sodass am Ende mehr als 15 Mengeinheiten im System vorhanden sein müssen, damit Zu- und Abflüsse übereinstimmen.

Periodische Einträge

Gerade bei landwirtschaftlichen Systemen kann man davon ausgehen, dass externe Einträge von der Jahreszeit abhängen. In erster Näherung kann man sie vielleicht als jahresperiodisch ansetzen. Solche Funktionen lassen sich (mit der Zeiteinheit von $1/2\pi$ Jahren) durch trigonometrische Reihen

$$a_0 + \sum_{k=1}^r (a_k \cos kt + b_k \sin kt)$$

beliebig gut approximieren. Da sich spezielle Lösungen eines inhomogenen Systems

$$\dot{x} = A x + \sum_{k=0}^r u_k(t)$$

dadurch finden lassen, dass man spezielle Lösungen x_k der inhomogenen Systeme

$$\dot{x} = A x + u_k(t)$$

für $k = 0, \dots, r$ bestimmt und diese aufsummiert, ergibt sich das Problem, eine spezielle Lösung von

$$\dot{x} = A x + a \cos \omega t + b \sin \omega t$$

zu bestimmen mit konstanten Vektoren $a, b \in \mathbb{R}^n$ und einer Kreisfrequenz $\omega > 0$. Der Ansatz

$$x(t) = y \cos \omega t + z \sin \omega t$$

liefert nach Einsetzen in die Differentialgleichung

$$-\omega y \sin \omega t + \omega z \cos \omega t = (A y) \cos \omega t + (A z) \sin \omega t + a \cos \omega t + b \sin \omega t$$

und damit das lineare Gleichungssystem

$$\begin{aligned} -A y + \omega z &= a \\ -\omega y - A z &= b \end{aligned}$$

für die unbekanntenen Vektoren $y, z \in \mathbb{R}^n$, welches seinerseits äquivalent ist mit

$$(-A - i \omega I) (y + i z) = -A y + \omega z - i (\omega y + A z) = a + i b.$$

Da wegen der vorausgesetzten Eigenschaften die Realteile aller Eigenwerte von A negativ sind, A insbesondere also keine rein imaginären Eigenwerte hat, besitzt das Gleichungssystem bei jeder Wahl von ω eine eindeutige Lösung (vgl. HEUSER (1989, 172/478)).

Beispiel: In dem System aus Beispiel (a) in 3.8.3 werde periodisch Phosphat in die Erde eingetragen, etwa:

$$\dot{x} = \begin{pmatrix} -0.4 & 0.1 & 0.75 \\ 0.3 & -0.6 & 0 \\ 0 & 0.5 & -0.8 \end{pmatrix} x + \begin{pmatrix} 1 + \cos t \\ 0 \\ 0 \end{pmatrix}.$$

Für eine jahresperiodische Interpretation ist hier als Zeiteinheit $1/2\pi$ Jahre zu wählen und der Jahresanfang ist auf den Zeitpunkt des höchsten Phosphat-Eintrages zu setzen. Das oben beschriebene Vorgehen liefert die spezielle Lösung

$$\begin{pmatrix} 8.649 + 0.3027 \cos t + 0.8008 \sin t \\ 4.323 - 0.1366 \cos t + 0.1728 \sin t \\ 2.703 - 0.0860 \cos t + 0.0005 \sin t \end{pmatrix}.$$

Von dieser unterscheiden sich auch hier alle anderen nur um die langfristig verschwindenden Lösungen des homogenen Systems. Die hier angegebene Lösung ist also das, was sich langfristig einstellt.

3.9 Literatur

- ABELSON, R. P.: *Mathematical Models of the Distribution of Attitudes under Controversy*, in FREDERIKSON, N. / GULLIKSEN, H. (eds.): *Contributions to mathematical psychology*, New York 1964
- AMANN, H.: *Gewöhnliche Differentialgleichungen*, Berlin 1983
- ARTMANN, B.: *Lineare Algebra*, 3. Aufl., Basel 1991
- BODROŽIĆ, Z.: *Mathematische Modelle der "kognitiven Psychologie" für Entscheidungsprozesse in Gruppen*, Diplomarbeit, Fachbereich Mathematik der Universität Hamburg, 1998
- HEINE, M. / HERR, H.: *Volkswirtschaftslehre. Paradigmenorientierte Einführung in die Mikro- und Makroökonomie*, 2. Aufl., München 2000
- HEUSER, H.: *Gewöhnliche Differentialgleichungen*, Stuttgart 1989
- HUPPERT, B.: *Angewandte Lineare Algebra*, Berlin 1990
- KNOBLOCH, H. W. / KAPPEL, F.: *Gewöhnliche Differentialgleichungen*, Stuttgart 1974
- LEUTBECHER, A.: *Zahlentheorie*, Berlin 1996
- LUENBERGER, D. G.: *Introduction to dynamic systems. Theory, models, and applications*, New York 1979
- MOONEY, D. / SWIFT, R.: *A course in mathematical modeling*, Mathematical Association of America 1999
- ORTLIEB, C. P. / v. DRESKY, C. / GASSER, I. / GÜNZEL, S.: *Mathematische Modellierung. Eine Einführung in zwölf Fallstudien*, Wiesbaden 2009
- ROBERTS, F. F.: *Discrete Mathematical Models*, Englewood Cliffs 1976
- WALDMANN, K.-H. / STOCKER, U.M.: *Stochastische Modelle*, Berlin - Heidelberg - New York 2004